knowit CAPTO OPEN _FUTURE creative commons

OpenEPI

Open Earth Platform Initiative

Feasibility study and pre-project

September 2024

Report to Norwegian Agency for Development Cooperation (Norad)

Content

Executive summary

The Open Earth Platform Initiative (OpenEPI) aims to harness digital transformation to address the interconnected crises of climate change, biodiversity loss, and global inequality. By leveraging technology, OpenEPI seeks to improve access to high-quality climate data, thereby supporting informed decision-making and fostering sustainability. The initiative also seeks to actively support openness and democratized access to data, as a goal in itself. This feasibility study explores the potential of the concept of OpenEPI, focusing on its relevance particularly for sub-Saharan Africa, and outlines the necessary steps for its successful implementation.

We, the consortium behind this report, hereby gives our recommendations from our feasibility study and proof-of-concept, in response to the conditions under Norad's OpenEPI grant, which we are grateful to have been funded by.

The initiative aligns with the UN Secretary-General's call for utilizing technological advances to monitor and protect the environment. Digital transformation, through widespread dissemination and use of data, is critical in combating climate change and promoting global sustainability. Current climate adaptation efforts require reliable data, yet much of this data remains inaccessible or underutilized due to proprietary restrictions and vendor lock-ins. OpenEPI aims to rectify this by promoting open data and ensuring it is available and usable for developers addressing various stakeholders, including smallholder farmers who face significant climate-related challenges.

We suggest an OpenEPI platform, with data and services, and an organizational structure – dimensioned to about 22 FTEs as a starting point, increasing to about 60 FTEs at full-scale. The suggested architecture for the OpenEPI platform has been designed to ensure flexibility and scalability across multiple cloud environments, accommodating future growth and technological advancements. This approach emphasizes maintaining efficiency, cost-effectiveness, and alignment with core objectives while evolving to meet future challenges.

The primary goal for the OpenEPI concept is to enhance access to climate, nature, and environmental data, thus, supporting climate resilience in vulnerable regions. The initiative proposes a multi-faceted approach:

- Data Accessibility and Quality: Ensuring high-quality, relevant data is available through an open-source platform.
- User Engagement and Support: Providing services, capacity-building and community support to help users effectively utilize the data.
- Technological Development: Building a robust, user-friendly platform with a modern technological infrastructure.
- Collaborative Ecosystem: Fostering partnerships with relevant stakeholders and initiatives to secure access to data sources and promote collaborative efforts.
- Scalable and Sustainable Funding: Combining predictable funding from Norad with flexible partner and community contributions to ensure long-term viability and growth.

The feasibility study identifies several key risks and proposes mitigation strategies. This

includes operational and financial risks, data quality risks, user adoption risks, technological risks, privacy and security risks, and liability and reputational risks. It will be crucial that a permanent OpenEPI organization targets these risks thoroughly and conscientiously.

We think OpenEPI has the potential to significantly improve access to climate data, thereby supporting climate resilience and sustainable development. The initiative's success hinges on addressing identified risks, securing sustainable funding, and fostering a collaborative ecosystem. A detailed business plan and further plans for stakeholder engagement and demand side efforts on user uptake, are recommended to refine and implement the initiative effectively.

By enhancing data accessibility and usability of climate change related open data, we think OpenEPI can play a crucial role in mitigating the impacts of climate change, particularly in vulnerable regions and contribute to global sustainability goals.

1. The Open Earth Platform Initiative and this study

1.1 Background

Digital transformation is crucial to address the three distinct and inter-connected crises of climate change, nature and biodiversity loss and increasing inequality between countries and population groups. UN Secretary-General Guterres highlighted the importance of technology in protecting the environment in his "Roadmap for digital cooperation", submitted to the UN General Assembly in 2021:

> *"The recent advances in technology offer ground-breaking opportunities to monitor and protect the environment, as well as overall planetary health. By harnessing them appropriately, the digital revolution can be steered to combat climate change and advance global sustainability, environmental stewardship and human well-being."*[1]

In this context, digital transformation and use of technology means first of all a much more extensive dissemination and exploitation of data. The need for climate adaptation based on high quality climate data is established in global conventions, and the international community has started to respond. The Global Biodiversity Framework establishes a target to "*ensure that the best available data, information and knowledge, are accessible to decision makers, practitioners and the public (…)*".[2] In specific, the dissemination of climate data can make the difference for local smallholder farmers by getting timely information to tackle increasing climate change related weather-shocks and get access to relevant analysis and solutions related to for example soil and seeds that is crucial for their livelihoods.

Access to data for public interest purposes has never been more crucial. Global consultants, climate and agricultural technology (agtech) start-ups, insurance companies and many others, are all racing to meet the ballooning demand for information about climate dangers and how to prepare for them. However, today, large sums of scarce public and private international development finance are invested in closed climate data solutions. "Closed" in the meaning of not accessible and possible to exploit by others than those who are producing or collecting them in the first place. And it is not trivial whether data are open or proprietary. The UN Secretary-General's High-level Panel on Digital Cooperation, established that digital development efforts have been fragmented, overlapping and with weak scalability.[3] A key part of the response to address these shortcomings was the establishment of the Digital Public Goods Alliance and the subsequent development of a standard and definition of digital public goods.[4] Digital public goods are "open-source software, open data, open AI models, open standards, and open content that adhere to privacy and other applicable laws and best practices, do no harm by design, and help attain the Sustainable Development Goals

[1] United Nations (2021) Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation. Report of the Secretary-General. A/74/821 (https://documents.un.org/doc/undoc/gen/n20/102/51/pdf/n2010251.pdf)

[2] COP15: Nations Adopt Four Goals, 23 Targets for 2030 In Landmark UN Biodiversity Agreement | Convention on Biological Diversity (https://www.cbd.int/article/cop15-cbd-press-release-final-19dec2022)

[3] https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf

[4] Definition from Digital Public Goods » Digital Public Goods Alliance

(SDGs)". Applying this definition ensures scalability by design, prohibits vendor lock-in and enhances the rate of return for public investments in sustainable development.

While a lot of the data in principle is considered to be public, it is often voluminous, technically challenging and not particularly useful for people trying to evaluate their personal exposure. Even if we don't know exactly the magnitude of the problem, this applies most certainly to local farmers and other actors along the food chain in low- or middle-income countries, especially in sub-Saharan Africa - that has been our "trial domain" and primary scope for this assessment and feasibility study.

Together, we think the current state of affairs both reduces the cost-effectiveness of development aid and related initiatives, creates barriers to effective service delivery to poor and vulnerable populations, and limits competition and innovation. The expected exponential growth in data initiatives for nature exacerbates the problem and creates a sense of urgency to ensure that key data sets are made openly available for the future. Generally, our point of departure is that there is a need for localized, easy-to-use apps and other information services, supporting farmers in their climate change adaptation and mitigation efforts. Our observation is that solutions funded by many development initiatives are oriented to end-users and rather tailored to specific information products. More or less, they lack the focus on preparing for local innovation by providing open APIs and guidelines for local creativity and professional use of the available data for software development and localized digital content purposes. This lack of "fuel" for local innovation on relevant topics, applies not only for data, but also for reusable derived and more complex data products, as well as for training data for AI based solutions, algorithms and technology components appropriate for climate change related software solutions. Further, current open initiatives in the field of climate, nature and environmental data are often narrow, sector-specific or otherwise thematically or geographically limited.

Based on these worrying considerations and acknowledgements, Norad prepared a memo or concept note early in 2023, where they proposed an open data and tech access platform. The concept was centered around the idea of an open data platform and an initiative that addressed all the above-mentioned challenges. The concept was labeled "Open Earth Platform Initiative", or abbreviated "OpenEPI". The proposed platform was aiming at providing easy access to nature and climate change related open data and data products - and services related to those data and data products. It should be a robust infrastructure, service and resource (or digital public good) for developers and start-ups aiming at providing relevant software solutions for climate adaptation and nature management. This report examines the feasibility of the concept, as we as consultants and experts have dived into the rather broad range of issues, concerns and possible obstacles the concept raises.

One additional concern raised in Norad's original concept note is the present lack of large-scale, long-term solutions for data storage under public control - at a global level. This part of the overall "problem complex" is not something we treat or discuss in this feasibility study. However, facilitating the world to acquire better knowledge and data memory, and long time series of data, is one natural extension of what OpenEPI aims to meet in terms of needs.

1.2 About the initiative

The problem to be solved by the initiative is the apparent lack of decision support by available data, and the need for global investments in digital capabilities to support local innovation in digital solutions. Norad proposed the establishment of a data platform for sharing and innovative exploitation of open data from the climate, nature, environmental and geospatial domains - in any way of value in the efforts of combating climate change and uphold of food security, at first specifically in sub-Saharan Africa. The proposed platform aims to contribute to a more transparent, more complete and comprehensive view of and access to relevant data sources - and by metadata, descriptions, and easy-to-use instructions support local innovation. The intended target group for the platform and its offerings is local startups and app developers, who in turn can provide relevant and localized solutions for the users locally.

For Norad, the objective closely matched the ambition defined by the Norwegian Government's international strategy on climate adaption "Climate change, hunger and vulnerability"[5]. One of the key parts of this strategy is digital transformation and innovation. Norad has a strong track-record on digital public goods and innovation partnerships and a clear mandate to expand climate adaptation efforts and is thus well-placed to kick-off a project to develop such a platform with associated services. Norad also has a possibly very strong influence on the "data market players" through all their grants and funding of nature and climate change related, data intensive projects all over the world. Further, Norway is seen as a relatively neutral actor in the geopolitical landscape and can use its convening power to mobilize national and international actors around an initiative of this kind.

By June 2023, Norad published an open call for proposals and commissioned an extensive study to explore feasible solutions based on open source, open data and algorithms, and to investigate, assess and suggest necessary arrangements for the possible realization of an OpenEPI platform and organization. Our consortium, consisting of Knowit (Norway), Capto (Norway), Creative Commons (US) and Open Future (Poland/Holland) together applied for a grant, in competition with many other applicants. Our proposal that eventually received a grant of NOK 13 mill., has been designed as a pre-project and a feasibility study aiming to give clear recommendations to Norad, as support for Norad's potential decision on a full-scale realization of OpenEPI. This report presents the feasibility study and gives our conclusions and recommendations, to Norad and those who will be involved in any discussion on whether or not Norad should implement OpenEPI. We see our report as the main support for the decisions to be made.

In conducting the feasibility study, including building the proof-of-concept solutions at openepi.io supporting the study, we have had at least four goals in mind:

1. To show how to harness the potential of open data and open tech for the sake of building resilience against climate change, more specifically in the fields of agriculture and food production in sub-Saharan countries, but also for demonstrating the potential more in general.

[5] Government of Norway's strategy on "Climate change, hunger and vulnerability" (2021)

2. To demonstrate how we from a development agency perspective can spur local innovation by building digital infrastructures and public goods giving local students, tech startups and other innovators easier access to relevant global data, software components, advice and related services.
3. To demonstrate the feasibility in building an open source and vendor independent technology stack, supporting the sharing and open exploitation of data and tech components.
4. To assess the necessary governance and organizational and functional setup to run a data platform and the associated services, in a robust and sustainable manner.

In order to reach those goals, both Norad and we think an overarching loyalty to the principles of openness will be crucial. Therefore, open data and open source code are elements in the OpenEPI concept. The full-scale implementation of an OpenEPI will require a lot of work to enforce this openness and the FAIR principles. This is why we also introduce an OpenEPI specific policy on openness. More on what we mean by openness and open policy, in the definitions section in chapter 1.3.

Norad's 2023 call for proposals and initial idea of the OpenEPI initiative aims at delivering the following:

- A comprehensive list of open datasets across climate, nature and environment (building on existing collections), with initial focus on climate adaptation
- A curated collection of aggregate APIs and micro services to facilitate easy access to data and product development
- A comprehensive set of open AI models and open datasets relevant for nature and climate
- Reference implementations and pilot projects that showcase how the OpenEPI platform can be used for implementation
- Open data policy and guidelines for projects sharing data, aggregated products or components, or developing new platforms
- Technical working groups aiming to bridge the gap between data providers and startups and other users developing new products and services

This feasibility study is our response to Norad's call and the following grant agreement. During our assessment, technical exploration and prototyping we have discussed and learned a lot. In agreement with Norad, we have made modifications to the list above, excluded some but added some new features. Chapter 5 gives a description of the key functions and features we have focused on.

In conducting this study, we have adopted an open mind, a genuine curiosity and a fundamentally questioning and explorative attitude. At first, we really did not know whether we would be able to achieve the four goals above. For instance, we had an idea that setting up a fully open source based technical architecture would be doable, but we didn't know for sure. The same goes for access to climate change relevant data sources. Further, we thought we understood the stakeholder landscape, but not completely. As we see Norad's original idea - as manifested in their original concept note, there are many presuppositions and assumptions about both possibilities and causal relationships. We believe that a feasibility study is right to question most things - not because we think they are wrong, but to

ensure that we do not overlook conditions that will be important and decisive for the eventual realization of OpenEPI.

One basic question is what makes OpenEPI unique, and therefore worth the effort. Building a full scale OpenEPI will be a major investment, although that is one of the topics to be addressed by this study. We recognize the fact that there are already many initiatives on open data and on climate change adaptation. In a sense we are potentially competing with other initiatives aiming at some of the same goals.

Anyway, initially we think OpenEPI differs from other initiatives in at least two ways:

- OpenEPI targets developers, not end-users. That is, OpenEPI will potentially become a novel resource or building block (or blocks) for all developers aiming at building end-user IT solutions in the domains of climate change adaptation and resilience. This makes OpenEPI a true digital public good.
- OpenEPI will be truly open as a provider of data and components but will also be an effective guarantor of openness and availability more in general. By referring to OpenEPI, global actors, national development agencies and others will be able to set requirements for their different grants and support actions, fostering more reuse of data and thereby getting more value out of their efforts.

There will be some reference to the concept of “Digital public goods'' or DPGs in this study. The term is defined below. As a term, it appears as early as 2017, and has gained some attention with the growing recognition of the potential for new technologies to be implemented for the benefit of society. In specific, digital technologies have been identified by countries, NGOs and private sector entities as a means to achieve the UN sustainable development goals (SDGs). Several international agencies, including UNICEF and UNDP, are exploring DPGs as a possible solution for emerging economies to address the issue of digital inclusion. In the original concept note, Norad sees OpenEPI - or more precisely some of its data products and APIs - as a potential DPG.

The EU has traditionally had a strong focus on open data and the value of openness in an innovation context. In particular, the data that EU has listed as high value in the so-called High Value Data (HVD) list under the Open Data Directive are of interest, also as a global standard. HVD includes a great number of geospatial data themes and can at least serve as a “lense” for prioritizing and specifying relevant data also for initiatives like OpenEPI. In this feasibility study, we have therefore included some detailed study of EUs HVD initiative under the Open Data Directive - as a possible inspiration for how to specify requirements on data quality, data formats and metadata, in case of a full-scale implementation of OpenEPI. Experience from implementing HVD in European countries is especially interesting. The obstacles related to the introduction of any normative requirements can possibly raise awareness of some of the obstacles OpenEPI can face if adapting to the HVD approach.

Norad’s initial scan of ongoing initiatives also did indicate that no-one is currently working on a broad architecture to ensure long-term interoperability of open data for climate adaptation and nature. Instead, several organizations compile and use open data to solve climate adaptation and nature related issues, but only focusing on a very specific purpose, limited to one sector or otherwise too narrow to address the need for an open, broad data architecture.

To sum up, the long-term success criteria for OpenEPI are:

- actual provision of data and other offerings, relevant for the intended impact (on local innovation and on resilience to climate change)
- user adoption (number of developers, number of applications developed, share of developers aware of OpenEPI, number of data accesses, amount of advice given to developers etc.)
- compliance to open data policy
- an open source tech stack, securing vendor independence
- satisfactory handling of cross-cutting issues like data ethics, gender equality, privacy

In addition, OpenEPI as an initiative and an organization will depend on efficient mechanisms for scaling, funding, governance and trust among stakeholders.

1.3 Prerequisites and framework conditions for the study

In this feasibility study and pre-project, we demonstrate value from the OpenEPI initiative. Our point of departure is rather open, even though we are tied to the constraints under which our grant agreement. In this, we work under the following prerequisites and conditions:

- In designing solutions and establishing policies, we should be compliant to the Digital Public Goods Standard (a set of requirements or principles): Digital Public Goods Standard - Digital Public Goods Alliance
- We are always keeping in mind the global context. In this context, investments have to address the global south and to be targeted for global south's challenges.
- Norad is part of Norwegian governmental structures, and EUs priorities and initiatives on open data have relevance. The Open Data Directive and other legislative measures from the EU apply also for Norad and cooperating partners in the OpenEPI context - at least as a point of departure when it comes to governance, policies and considerations on cross-cutting topics like privacy and data security.
- Norad has a number of guidelines on different cross-cutting topics that have to be followed by OpenEPI.

The terms we use in this study are intentionally well defined and standardized. Some of the most important definitions are listed in the table below.

Open data	Open data is data licensed in a way that means it can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike
Closed data	Closed data is data published under a proprietary license, not allowing for re-use and adaptation.
Architecture	Norad's concept note deals with an architecture for open data comprising several "layers", often called a "stack" or a "platform". This architecture consists of data storage solutions, databases, artificial intelligence, application programming interfaces (APIs), data modeling

	tools and various end products in the form of applications, portals and other tools. As a shorthand, the concept note will collectively refer to all of these as “data solutions” or simply “data” throughout. Where needed, the concept note will specify which part of the platform it is discussing.
Nature data	Data solutions relevant for climate adaptation efforts, biodiversity and land-use are within the scope of this report. As a shorthand, the concept note will refer to this as “climate adaptation and nature data” or simply “nature data” throughout. Data solutions for climate change mitigation are generally considered out of scope, such as emissions data, however it must be noted that it is in some cases impossible to draw a clear line between what data are relevant for nature and climate change.
Climate data	Climate data includes weather, climate, hydrology, atmospheric composition, cryosphere, oceans, and space weather. These data represent the Earth system’s past, present, and future states, exchanged in real or near-real time, and from historical sources. Data types include observations, satellite data, GIS data, video/photo data, and derived products (e.g. weather forecast and hydrological model.)
Digital Public Goods	Digital public goods are openly licensed software, data, content, and standards that adhere to privacy and other applicable laws and best practices. They are designed to do no harm, are freely accessible, and help achieve the Sustainable Development Goals (SDGs). Digital public goods enable collaboration and innovation, allowing anyone to use, modify, and distribute them without restriction.
Data publisher or data providers	Data publishers are entities or individuals responsible for organizing, standardizing, and sharing data with the broader community. They play a critical role in ensuring the data's accuracy, relevance, and accessibility. Data publishers often collaborate with data hosts to facilitate online access.
FAIR principles	Refer to a set of guidelines that aim to make data Findable, Accessible, Interoperable, and Reusable. These principles guide the creation and management of data systems to ensure that data can be easily discovered and used by humans and machines alike, thereby fostering a culture of open science and data sharing. The application of the FAIR Principles[6] ensures that data is stored in accessible databases with appropriate metadata, uses standard formats for interoperability, and is accompanied by clear usage licenses to facilitate reuse.

[6] FAIR Principles - GO FAIR (go-fair.org)

1.4 Methods and approaches

The study has been running from August 2023 to June 2024. The dedicated assessment team has been led by Capto, the technical team exploring data and technology and developing the demonstrator - the developer platform with data offerings - has been led by Knowit, as well as the administrative functions of the project. At the beginning Open Future served as a discussion partner. Later Creative Commons has served as an internal advisory resource, especially contributing to licensing and openness related topics. Occasionally, Norad has taken part in discussions, workshops and meetings/interviews with relevant international stakeholders.

1.4.1 Assessments

The assessment team, led by Capto's Gjermund Lanestedt, has collected and analyzed data, gained insight and discussed a broad range of topics with the other consortium participants, with Norad and others. The modus operandi of the team has been interviewing, reading, discussing and revisioning report document versions. The report versions are written by using Google Docs and have been openly shared across all members of the consortium. For discussions among team members, and across the two teams, Discord has been used as a communication tool.

A major source of insight for the assessment team has been the interviews with several data platform owners and managers from different domains, giving input on crucial success factors, obstacles and governance issues. Hereby, we have also gained a better understanding of the demand side: the need for easy access, clear guidelines, and services and incentives for use. The list of interviewees and informants are included in the stakeholder overview shared in appendix 1.

The interviews have been complemented by extensive literature studies and desktop research, for instance use cases, policy documents and white papers on relevant topics. Many of the documents have been retrieved by us, but some of them have been recommended by our stakeholders or by those we have interviewed.

The initial phase was geared on exploring sources of relevant open data and deciding what open data sets we should use to demonstrate the possibilities of a developer portal. One point of departure for this was Creative Commons' overview of the state of openness and availability for climate change related data, globally - published as a part of their Climate Data Campaign[7], but also supplied by other sources identified through discussions with Norad, for instance an extensive scan of relevant datasets, conducted by Norad itself.

The feasibility study has included some delving with perspectives on how to organize and govern a data platform and related services, how to identify available high quality data, how to attract users and keep the platform services relevant, and different policy or governance related issues. We have had interviews with owners and managers of existing platforms and conducted an extensive study of open documentation on experiences and findings from a broad span of domains. In addition, we engaged an external consultant, in the Norwegian

[7] https://creativecommons.org/2023/08/08/surveying-the-open-climate-data-landscape/

consultant company Agenda Kaupang to assess the status of EU's implementation of the High Value Data (HVD) list - as a possible categorization of data types and normative specification of necessary metadata, and as possible learning points when it comes to implementation of policies. This specific study is included in this report, as an annex 3.

1.4.2 Technology and data exploration

As a starting point for this feasibility study and our work on the proof-of-concept, several existing data sources and data platforms were screened, giving a general impression of the difficulties and obstacles users of those data most certainly will experience. Primed with the notion that some data are more important than others for climate adaptation and food supply related issues, the data analyst team explored several data sources using criteria of openness and FAIR - but inspired by some of the use cases mentioned already in Norad's concept note.

The tech team, under the leadership of Knowit's Kenneth Stigen, has explored different technology configurations and setups throughout the entire pre-project period, focusing on providing a fully open-source technology stack combining different data sources, data handling mechanisms, integrating components, authentication mechanisms, user interfaces, metrics etc. A functional demonstrator has been built through iterations, serving potential users with a handful of relevant and quality assured datasets. The knowledge and insight gathered through this exploration has been most valuable.

In the initial phase of the project, our technical team was assembled to include a range of skills: one technical architect, two data scientists, one frontend developer, one UX designer, and one full-stack developer. This mix was crucial in the initial stages, to effectively handle the project's changing roadmap and landscape of the objectives as we gained more insight.

Since the details were not well-defined, we chose to use a flexible, Kanban-style method, supported by GitHub Projects. This tool was perfect for managing tasks that were still taking shape. It allowed us to set up a digital Kanban board, which made it easy to see who was doing what and track progress on tasks.

We also held daily stand-up meetings to ensure everyone was on the same page. These meetings were a chance for team members from different areas - like architecture, data science, UX design, and full-stack development - to coordinate and quickly address any issues. This approach, using straightforward digital tools and regular communication, created a productive environment that could adapt to changing project needs. Members of the assessment team also participated occasionally.

As part of the exploration phase, the project's tech team held a workshop in Kigali, Rwanda, with a selected group of students and professional developers - to gain insight into the users' needs, how the prototype platform and related services could meet those needs, and to identify the key factors for success and determine what core functionalities should be prioritized for the demonstrator platform. More on that in chapter 4.8.

1.5 Limitations of this study

This feasibility study has its obvious limitations, given the complexity and multi-objective nature of the OpenEPI idea. Not only is OpenEPI an advanced technical concept and ambition we have explored. As a concept OpenEPI also incorporates several independent objectives and impact goals to be reached. We have assessed and discussed possibilities and obstacles rather broadly in this preliminary project, for instance issues related to open policies, data relevance, stakeholder relations, and user adaptation. We have explored technical arrangements and studied the feasibility of an open source data platform, aimed at providing open data to innovators in the domains of food production, climate change adaptation and resilience in sub-Saharan Africa.

Such a comprehensive study necessarily brings with it a great deal of dependencies and uncertainties. For example, the selection of data sets depends on several factors including end-user information needs, the causal relationship between different natural phenomena and food security, the options and abilities to influence on practice, and much more. We do not know concretely enough what decision support farmers and other end-users will need now and in the future. Therefore, we will not be able to tell what kind of data and data products the developers will request. We think there is a need for a more thorough study and abstraction of the natural ecosystems in question, as well as of the causal relationships between data, decisions, change of behavior and a positive outcome (increased crops yields and/or improved food production). Such a study will require in-depth agriculture-scientific and ecological expertise. It will also be a question of how the relevant data could be collected in the most efficient manner. Data which are today collected on the ground, will in the future be available from satellite imagery - given an open access to and dissemination of satellite images, and the right technology at the end-user side. The data supply side will not be a stable value chain over time, but an object for immense innovation efforts.

Further, we cannot know for sure how data producers and data providers will react to the policy measures we suggest on open data requirements. If many of the requested data sets are still locked in by the data providers, it doesn't help much that OpenEPI has a policy to the contrary. Here the outcome will depend on a sufficiently strong voice - and the use of monetary measures on the part of the authorities' and the NGO's, to push for change. Still further, we must also consider the varying digital maturity in African countries, where the possibility for reaching the farmers by digital services can be very limited in some countries.

The possibly weakest part of the OpenEPI idea as a whole - and this feasibility study in specific - is the fundamental uncertainty about "the market" - or perhaps more precisely - the potential for digital innovation in regions like sub-Saharan Africa. We have not had access to an adequate overview of the general innovation challenges in the various countries, but we assume that they differ much. Factors like literacy and digital competency (especially among farmers), quality of network infrastructures, density of developers and startup businesses, and phone density among rural populations, vary substantially between countries and regions. The situation in Africa is presumably quite different from the situation in let's say South America - where the level of digital maturity among the populations is supposedly higher. Thus, we are a bit unsure or imprecise of what is likely the potential future user uptake of OpenEPI's offerings. Of course, this also affects the precision in our estimates of the size and scaling of OpenEPI and influences the discussions on the need for capabilities

like learning resources, guidelines, local support capabilities, language/translating support, local innovation advisors and consultants, etc. in this feasibility study. We assess some of what has been explored by others, and some of the learning by other projects and initiatives - but we have yet to understand the whole picture. The point of departure is therefore still a bit anecdotal when it comes to the demand side.

On the horizon of this initiative is also the recognition that there doesn't exist any publicly controlled, large-scale solutions for long-term storage of crucial data. In general, data storage is controlled by large American tech companies. Currently, a few dominant global commercial companies are responsible for storing some of the most crucial climate adaptation and nature data. Based on Norad's initial scan of the field, there are in fact no alternatives to using off-the-shelf solutions from those companies for data storage. There is a need for more research to be done on this, but some initial studies indicate that there is great risk that important datasets, over time, can be lost or subject to vendor lock-in. Even if it is a matter of importance for the sustainability of an initiative like OpenEPI, we are not dealing with this issue in this feasibility study.

1.6 Reading guide

This report intends to give a broad picture of the OpenEPI initiative, our pre-project assessments and our proof-of-concept. We are trying to lay a sound ground for further efforts in achieving the goals regarding climate change adaptation and resilience, regarding better supply of and availability to open data applicable for digital innovation in sub-Saharan Africa and other parts of the global South, and regarding reaping the benefits of open data, open source and open access to knowledge. The report therefore becomes rather comprehensive.

Chapter 1 describes the OpenEPI initiative and its scope and objectives, as it was designed and specified in the project proposal to the Norad 2023 call for proposals - named "Open Earth Platform Initiative", and the subsequent grant contract with Knowit and consortium partners Capto, Creative Commons and Open Future. Chapter 2 gives a descriptive review of the ecosystem of data providers, initiatives and open data in the climate, nature and environment domains. As a background, chapter 3 draws some general thoughts and the theory of change (ToC) for the conceptual idea of building a platform to achieve certain goals. Chapter 4 assesses the user needs (data, functions, services) we are addressing by the OpenEPI concept, including our empirical findings in this area.

Based on our practical efforts in assessing and building an open source technology demonstrator, with actual API services covering relevant data, chapter 5 describes the technological feasibility for OpenEPI as an infrastructure and concept for sharing of open, climate and nature related data and data products. A blueprint for the entire technology stack is provided, and different perspectives on the technology are thoroughly discussed. Chapter 6 describes our assessments on the operational feasibility for OpenEPI as an organizational entity, including dimensioning, competence profiles and cost estimates.

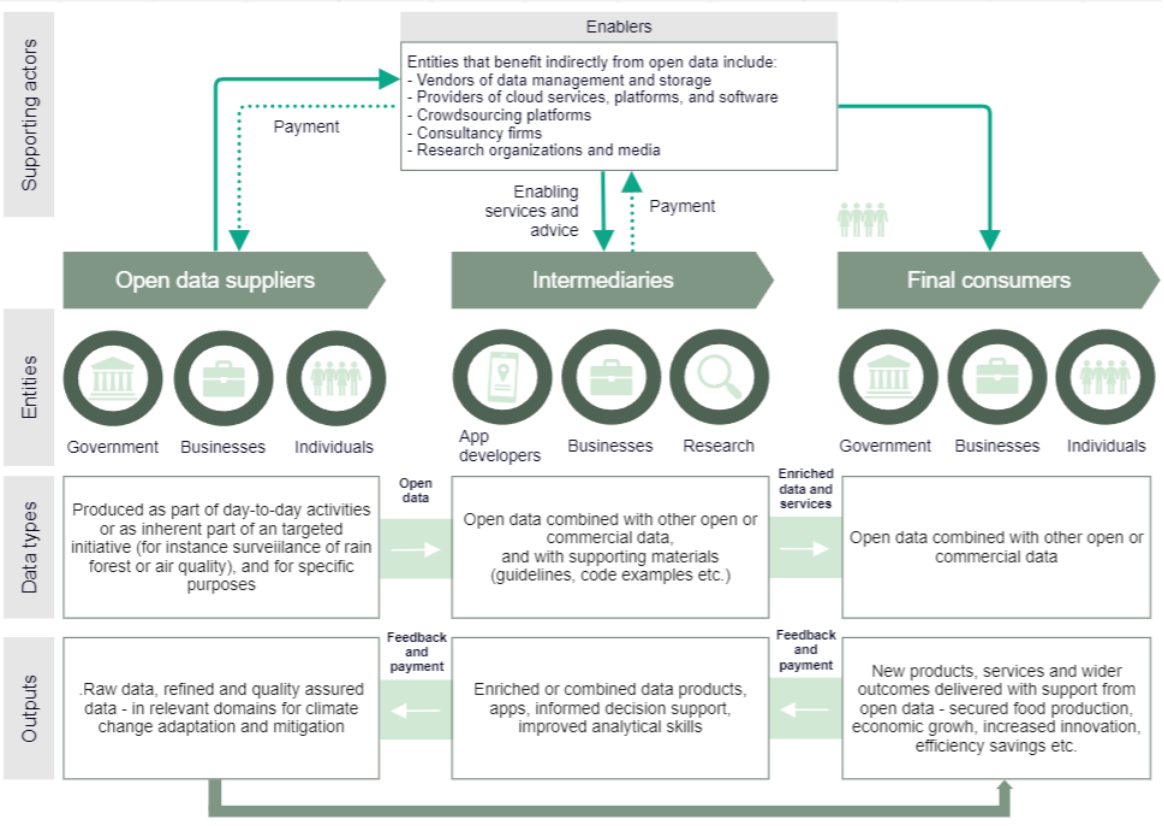
Chapter 7 describes all risk factors regarding the full scale realization of the Open Earth Platform Initiative. Chapter 8 draws our conclusions and advice to Norad. The last part (chapter 9) of the report contains appendices and annexes covering different specific issues.

2. The global climate, nature and environmental open data ecosystem

Geospatial data on climate, nature and environmental topics are collected and shared on a large scale, with many actors involved both on the dissemination and production, and on the user side. Presently, a multitude of global data platforms exists for distributing open data related to climate and nature. In the subsequent sections, we will delineate the stakeholders involved and their respective roles within the data ecosystem. Following that, we will give an account of to which extent data made available on these platforms are truly open and compliant with the FAIR-principles. Finally, we will investigate to what extent open data distributed on open data climate platforms are actually used to drive local innovation.

2.1 Roles and stakeholders in the data ecosystem

The global climate, nature, and environmental data ecosystem is quite simple, but still complex. It involves various roles and functions to ensure the production, collection, distribution, and utilization of data. We can largely distinguish between data suppliers, users or consumers, intermediaries and enablers.



Source: Inspired by Deloitte

Figure 1. The generalized open data ecosystem representing a combination of simplicity and complexity

The collaboration among these roles is crucial for creating a robust, interconnected and open global environmental data ecosystem. Some entities play multiple roles within the ecosystem. For example, a data user can also be a producer of climate data, and a sponsor can also be a data user.[8] Figure 1 illustrates in a stylized manner the open data ecosystem, inspired by some work from the consultancy firm Deloitte.

Digital ecosystems are very much influenced by the global nature of the digital economy. Actors are interconnected in many ways at the same time, giving very complex structures as a result. This also accounts for the domains of climate change resilience and food security in Africa. The web-like figure 2 below is not at all related to our specific context but can serve as an illustration of the complexity of the digital, global stakeholder landscape.

For our study, the complexity is further enhanced by the fact that local political and socio-economic conditions in the individual low-or-middle-income countries affect the innovation systems and provide variations in the opportunities for collaboration, interaction and exchange of data and information between the actors.

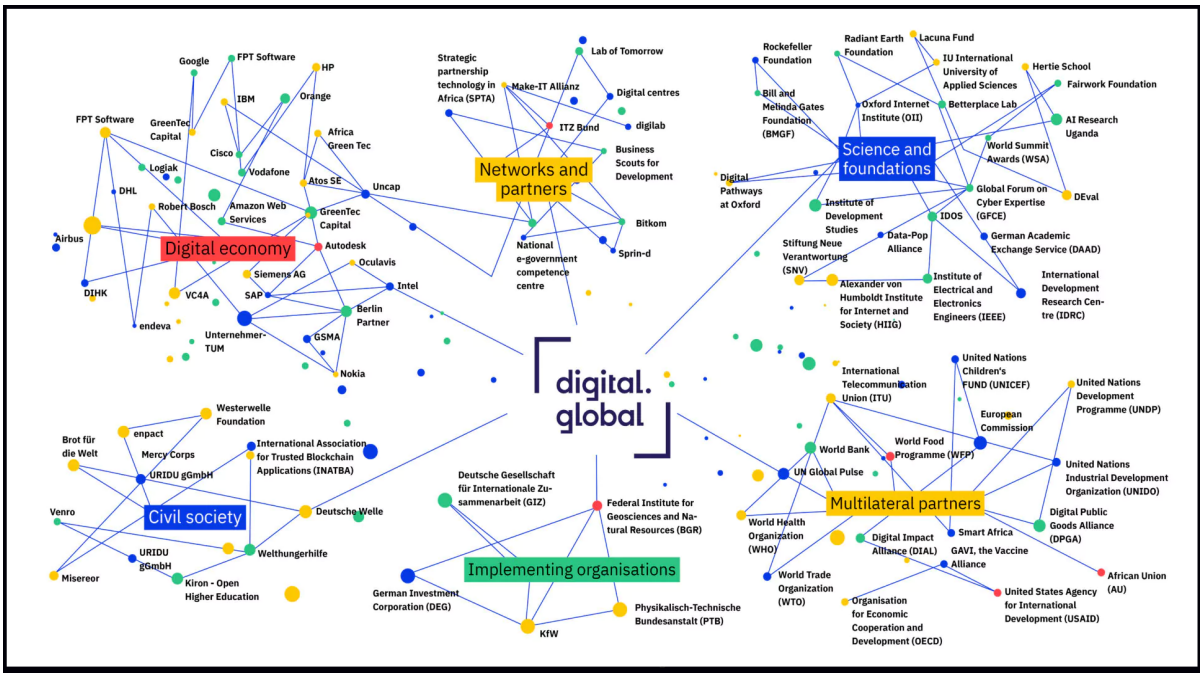


Figure 2. Illustrating the complexity and web-like relations between stakeholders, sectors and general and domain-specific networks in the digital landscape.

2.1.1 Data suppliers

Data suppliers or producers play a pivotal role in the ecosystem of nature and climate data. They are responsible for producing, collecting, monitoring or sharing raw data related to various aspects of the environment, like weather conditions and precipitation, water quality, hydrology, geology and soil chemistry, crops and natural vegetation covers, and so on. These entities are actively engaged in gathering raw data through a variety of methods,

[8] Recommended Best Practices for Better Sharing of Climate Data" by Taylor Campbell, Wanying Li, and Dr. Cable Green for Creative Commons is licensed CC BY 4.0., Creative Commons, 2024, https://docs.google.com/document/d/1PLRu-bLd91Zbndn4QbSfNQAn-vamPlblSq-7LFo7CP8/edit

including ground-based observations, analysis of satellite imagery, use of sensor networks, and through scientific research endeavors. In certain instances, data producers also create data via experiments, simulations, or modeling techniques to supplement or enhance the observational data. The way data producers manage, release and share their data significantly impacts its accessibility to the public.[9]

Three main sectors are generating most of the open nature and climate data: government agencies and international entities, the scientific community, and commercial private sector entities.

Government agencies and international organizations

Public bodies are significant data providers. Supporting their role and specific mission, they have established digital infrastructures for data monitoring and collecting, data storage and often also data sharing. They are managing for instance extensive networks of weather stations, environmental monitoring stations, and scientific research programs. Governments are to a growing extent also somewhat altruistic, aiming at value creation in the private sector and in other domains than their own. All over the world public sector entities are increasingly focusing on the importance of sharing their data to promote and stimulate business creation, innovation, research and knowledge sharing. The green transition has further underscored the need for governments to ensure access to timely, relevant, and high-quality data to strengthen resilience and encourage a unified societal response. While strides have been achieved in making climate and natural data more accessible, the OECD highlights the need for an expanded availability of open green data. Although geospatial and mobility data are widely available, there is a pressing need for increased efforts to ensure that earth observation, environmental, and meteorological data are accessible in high-quality formats suitable for further exploitation and user applications.[10]

In Europe, the concept of open data has been prioritized for years, guided by the EU's open data policy. This policy rests on the belief that data, especially when generated by public entities, ought to be freely accessible for both individuals and businesses to use and reuse. The overarching aim is to tap into the economic and societal benefits inherent in data, all the while safeguarding privacy and security measures. Sharing open public sector data has become an important part of the digital political agenda[11]. Central to the EU's approach is the Public Sector Information (PSI) Directive, which lays down a legal framework mandating EU member states to make certain public sector information available for reuse. It also mandates the European Commission to compile a list of high-value datasets (HVD) that are to be made available free of charge, intended to serve as foundational components for developing Artificial Intelligence solutions (see more on the high-value datasets in annex 3). Complementing the PSI Directive is the Data Governance Act (DGA), a key piece of the EU's wider data strategy that is designed to enhance data sharing across various sectors.

---

[9] Recommended "Best Practices for Better Sharing of Climate Data" by Taylor Campbell, Wanying Li, and Dr. Cable Green for Creative Commons is licensed CC BY 4.0., Creative Commons, 2024, https://docs.google.com/document/d/1PLRu-bLd91Zbndn4QbSfNQAn-vamPlblSq-7LFo7CP8/

[10] OECD (2023), "2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and key findings", *OECD Public Governance Policy Papers*, No. 43, OECD Publishing, Paris, https://doi.org/10.1787/a37f51c3-en.

[11] Open data | Shaping Europe's digital future: https://digital-strategy.ec.europa.eu/en/policies/open-data

The European Union has initiated several key programs aimed at promoting open access to data related to the “green deal”, with a special focus on geospatial, satellite, and meteorological data, primarily through the Copernicus program and through the Infrastructure for Spatial Information in the European Community (INSPIRE) Directive[12], which plays a pivotal role in the European Union's strategy for environmental sustainability and governance, acting as a foundational piece in the broader context of the PSI Directive and the Open Data Directive. INSPIRE is designed to enable the sharing of geospatial data across European borders. INSPIRE works by standardizing spatial data and services, making it easier for various stakeholders to access and use this data in an efficient and harmonized manner. The part of the Open Data Directive Addendum High Value Data list that covers geospatial data, builds on INSPIRE.

Further, following the INSPIRE Directive, the Copernicus program and EUMETSAT stand as significant initiatives enhancing the availability and accessibility of high-quality geospatial, satellite, and meteorological data. Copernicus is led by the European Commission in partnership with the European Space Agency (ESA) and is the largest earth observation data provider in the world. It serves a wide range of applications including environmental monitoring, climate change, and civil security, offering full, free, and open access to data. Copernicus is recognized for its Copernicus Data and Information Access Services (DIAS), which enable enhanced access to data and processing tools, promoting innovation and the creation of new business models based on earth observation data[13]. The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) further enriches the EU's capabilities in environmental monitoring and climate analysis. EUMETSAT provides critical data for weather forecasting and climate monitoring, operating a fleet of satellites that offer insights into global weather systems and environmental changes. This organization collaborates closely with the Copernicus program to deliver comprehensive meteorological data essential for a wide array of applications, from daily weather forecasting to tracking long-term climate change[14].

Norway has fully integrated the Public Sector Information (PSI) Directive, the Open Data Directive, and the INSPIRE Directive into national legislation. The Norwegian government's digitalization strategy aims to make open data available for reuse, facilitating the development of new services and value creation. In OECDs Open Government Data Index, Norway was in 2023 ranked 11th in open data maturity, performing above OECD average. Norway is placed in the “high performance group” together with Canada, Colombia, Finland, Czechia, Italy, Switzerland, Slovak Republic, Brazil, and Peru. These rankings highlight the progress these countries have made in implementing open data initiatives and setting standards for transparency and accessibility. It's important to note, however, that African nations, being non-members, are not included in the OECD's evaluations.

Several government agencies in Norway are collecting and distributing open geospatial and environmental data, including climate and nature data. Some of the most prominent agencies

[12] European Union, *Geospatial and Earth Observation data,* 04.03.2020, https://data.europa.eu/en/publications/datastories/geospatial-and-earth-observation-data
[13] European Union, *DIAS - Data and Information Access Services*, 19.01.2022, https://knowledge4policy.ec.europa.eu/projects-activities/dias-data-information-access-services_en
[14] https://www.eumetsat.int/

include the Norwegian Meteorological Institute, which focuses on weather data and forecasts; the Norwegian Mapping Authority (Kartverket), which deals with geospatial and property data; the Norwegian Coastal Administration (Kystverket), which provides maritime safety and infrastructure data and the Norwegian Water Resources and Energy Directorate (NVE) which shares hydrological and energy data, including water resources, flood and avalanche risks, and energy production statistics. BarentsWatch is also a significant agency involved in distributing climate and nature data, offering access to open data on maritime activities (like fisheries and aquaculture), sea and ice conditions in the Arctic.

Fact box

The Norwegian Meteorological Institute (MET Norway): METs official data sources and products are freely available to the public for use, dissemination and further processing, see https://www.met.no/en/free-meteorological-data
MET has for more than 10 years been involved in capacity building projects in several countries in Southeast Asia and Eastern Africa. From 2023 these projects are combined in a NORAD financed project named Sarepta (Institutional Support and Capacity Building for Weather and Climate Services). Capacity building activities include competence building related to climate services.

The government bodies within the geo- and environmental sector in Norway have over many years developed separate solutions for their collection and dissemination of environmental data. Many of these solutions are silo-based and designed or acquired for a specific purpose within a confined area. There isn't a shared data platform in Norway for environmental data, meaning that interfaces or integrations that exist are between individual systems, and they are costly and difficult to maintain.[15] In Europe, there are numerous well-developed portals for accessing large datasets. Examples of these portals include Reportnet[16], EMODNet[17], IPCheM[18], and Elixir[19], in addition to the already mentioned Copernicus and EUMETSAT databases. Norway must deal with some of these infrastructures and data sources without any influence over their future development, while in other cases, Norway could influence the design of the infrastructure. Anyway, European databases are continuously updated with environmental data from Norway.

In Africa, where OpenEPI will supposedly and intentionally have its main target users, open government data has been gaining momentum.[20] Several countries are making significant strides towards making public sector information more accessible and usable for innovation, transparency, and public engagement. However, the pace of adoption and the extent of data openness vary widely across the continent, reflecting differences in technological infrastructure, digital maturity, political will, and institutional capacity.[21] For instance, we have

[15] https://www.menon.no/wp-content/uploads/2021-153-Hovedrapport-KVU-Fremtidens-miljodata.pdf
[16] https://reportnet.europa.eu/
[17] https://emodnet.ec.europa.eu/en
[18] https://ipchem.jrc.ec.europa.eu/
[19] https://elixir-europe.org/platforms/data
[20] Data for development (D4D), *Sub-Saharan Africa and the State of Open Data 2022,* 2022, https://www.d4d.net/news/sub-saharan-africa-and-the-state-of-open-data
[21] Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre.

become aware of cases where local weather data sets are handed over to local private enterprises, as a part of national policies - thus becoming less available for open innovation.

Fact box

The OECD Open Government Index (OURdata Index) is based on three critical pillars[22]:

1. *Data availability: Measures the extent to which governments have adopted and implemented formal requirements to publish open government data. It also assesses stakeholder engagement for identifying data demand and the availability of high-value datasets as open data. For example, this pillar assesses if a country has an open data strategy.*
2. *Data accessibility: Measures the availability of requirements to provide open data in reusable formats, and the extent to which high-value government datasets are provided in open, timely and reusable formats, with good metadata quality, and through Application Programming Interfaces (APIs). It also assesses stakeholder engagement on the central open data portal and to improve data quality. For example, the pillar measures the percentage of high-value open datasets that are accessible through a central open data portal.*
3. *Government support to data reuse: Measures the extent to which governments play a proactive role in promoting the re-use of open government data inside and outside government. For example, it looks at events and partnerships with civil society and business actors to raise awareness about open government data and encourage re-use.*

Even if there are also examples of the opposite, the open data movement in sub-Saharan Africa has evolved substantially over the last five years.[23] Countries that are acknowledged to lead the way are Kenya, South Africa, Nigeria, and Ghana. Governments, backed by financial and technical assistance from international organizations, development partners and philanthropists, are developing and maintaining open data ecosystems on the continent. In recent years, governments have made a greater effort to support the production of data specifically aimed at monitoring the Sustainable Development Goals (SDGs), herunder data to adapt to climate change. This shift towards the production of data that can help monitor progress on the SDGs has also given prominence to central data repositories for African governments. 2022 marked a significant milestone with the United Nations Economic Commission for Africa (UNECA) launching its Open Data Portal, a platform aggregating statistical data across various levels - regional, national, and sub-national - showcasing progress towards SDGs. Similarly, the African Development Bank (AfDB) has been proactive in enhancing dataset accessibility for SDG monitoring across all 54 African Union member states through its Open Data platform. Tech driven transformations in other parts of the world have also raised many African political leader's awareness of the significance and influence of data. The COVID-19 pandemic marked a turning point for the open data movement, significantly influencing the progress of open data.[24]

Despite advancements, challenges remain, particularly in the realms of data collection and dissemination of open data. The State of Open Data report initially highlighted over 17 open

[22] OECD, *2023 OECD Open, Useful and Re-usable Data (OURdata) Index: Results and Key Findings, OECD Public Governance Policy Papers*, No. 43, OECD Publishing, Paris, 2023 https://doi.org/10.1787/a37f51c3-en.
[23] Data for Development, *Sub-Saharan Africa and the State of Open Data, https://www.d4d.net/news/sub-saharan-africa-and-the-state-of-open-data/*
[24] Data for Development, *Sub-Saharan Africa and the State of Open Data,* 2022, https://www.d4d.net/news/sub-saharan-africa-and-the-state-of-open-data/

data portals within sub-Saharan Africa in 2019. Yet, as of now, less than half of these portals remain active or have been updated in the past two years. Initiatives within francophone Africa have for instance suffered setbacks due to changing political climates. Political instability and the raised voices of youth facilitated by social media have led to a decline in public sector commitments to, and funding of, open data, with the primary focus of these nations shifting toward regaining control over territories and stabilizing their democratic processes.[25] Challenges related to data gaps and the availability of high quality open data in sub-Saharan Africa will be further discussed throughout this report.

Research Institutions

Research institutions play a pivotal role in the global effort to understand and address environmental challenges. Through rigorous scientific studies, experiments, and observations, these institutions generate a wealth of data on nature and climate, significantly contributing to our collective understanding of Earth's complex environmental processes and the changes it undergoes. They are at the forefront of employing innovative technologies for data collection, which not only enhances the efficiency and accuracy of the data collection and production processes but also ensures that the information gathered is of the highest quality and relevance.

Research institutions play a crucial role in the landscape of both data production and distribution. They are not merely centers for academic and scientific work on climate and nature related topics but also serve as foundational pillars for the dissemination of open data. The introduction of the FAIR principles marks a critical advancement in improving the management and utilization of research data. Originating within the research community, the FAIR principles encompass a detailed and comprehensive framework designed to ensure that data generated from research activities is Findable, Accessible, Interoperable, and Reusable. These guiding principles aim to elevate the quality and impact of research by ensuring that data can be easily located, accessed under clear and fair conditions, integrated with other datasets, and utilized for future research endeavors.

In Norway, national environmental institutes such as CICERO, NIBIO, NILU, NINA, and NIVA aim to contribute with high-quality research relevant to industry, governance, and society. These institutes engage in a wide range of environmental research activities, contributing to the body of knowledge necessary for informed decision-making in various sectors. Additionally, the Institute of Marine Research (IMR), recognized as one of Europe's largest marine research institutions, collects extensive data from Norwegian waters. Their data sets[26] are made freely available, with the condition of source acknowledgment, facilitating access for research and policy development purposes. The Nansen Center, an independent environmental research foundation, focuses on marine and Arctic conditions. It has established itself as a key entity in environmental research, particularly through the development of the Integrated Arctic Observation System[27]. This data portal, which consolidates open data from over 40 institutions, serves as a resource for the scientific community, enhancing the availability of research data on Arctic observations.

---

[25] Brookings, *Will rising insecurity erase West Africa's economic development gains?,* 03.02.2023, https://www.brookings.edu/articles/will-rising-insecurity-erase-west-africas-economic-development-gains/

[26] Research data | Institute of Marine Research (hi.no)

[27] Field work and other events | Integrated Arctic Observation System (nersc.no)

Internationally, a range of research institutions play crucial roles in the provision of essential climate and nature data. The Goddard Institute for Space Studies, operated by NASA, is noted for its extensive research into climate change, generating global climate data that supports various scientific and policy-oriented analyses. The European Space Agency contributes significantly to this field by facilitating access to satellite data crucial for environmental studies, covering observations of the Earth's lithosphere, hydrosphere, atmosphere, and biosphere. This data is instrumental in advancing our understanding of Earth's multifaceted environmental systems. The European Centre for Medium-Range Weather Forecasts also occupies a vital position in the landscape of environmental research by offering precise numerical weather predictions. These forecasts are fundamental for the operational planning of meteorological services within its member states, especially for managing weather-related risks.

In the realm of food security and agricultural policy, the International Food Policy Research Institute, a part of the CGIAR consortium[28], is recognized for its contributions to policy development aimed at poverty reduction and the enhancement of global food security. This work underscores the link between climate change impacts and food systems. CGIAR's Research Program on Climate Change, Agriculture and Food Security is specifically dedicated to the promotion of climate-resilient agricultural practices, with a focus on Africa. The program advocates for the utilization of open data to drive improvements in food security, highlighting the importance of climate-adapted agricultural methodologies.

In sub-Saharan Africa, a network of research institutions contributes to the generation and dissemination of important climate and nature data. Among these, the African Centre of Meteorological Applications for Development (ACMAD) provides climate services and data that are essential for weather-related disaster preparedness and climate change adaptation throughout Africa.[29] The African Climate Policy Centre (ACPC), under the auspices of the United Nations Economic Commission for Africa (UNECA), advises African governments on climate policy and resilience strategies, offering research and analysis on climate variability and its developmental impacts.[30] In South Africa, the Council for Scientific and Industrial Research (CSIR) conducts multidisciplinary research, including in environmental and climate science, contributing to the development of technologies and strategies for climate change adaptation and mitigation.[31] The Kenya Marine and Fisheries Research Institute (KMFRI) investigates marine and freshwater ecosystems, with a focus on the effects of climate change on aquatic environments, playing a vital role in the conservation and sustainable management of marine biodiversity and fisheries in Kenya.[32] Furthermore, the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL) offers climate services, education, and research facilities to several West African countries, aiming to improve resilience and adaptive capacity to climate change and variability.[33]

[28] https://www.cgiar.org/
[29] https://acmad.org/
[30] https://www.uneca.org/acpc
[31] https://www.csir.co.za/
[32] https://www.kmfri.go.ke/
[33] https://wascal.org/

Private sector entities

Corporations, particularly those operating in the data-driven technology sector, are increasingly recognizing the value of both utilizing open governmental and research data, and themselves contributing to the open data ecosystem. This participation not only serves public relations interests by demonstrating corporate social responsibility but also provides strategic advantages, including the enhancement of their own datasets through collaborative and crowdsourced improvements. Google for instance, runs Google Earth Engine, which is a public data catalog, offering an array of Earth science datasets - from geospatial to agricultural and oceanic data.

However, the reluctance among most private sector entities to share data openly remains a significant barrier, primarily due to competitive concerns. Actually, this is one of the major issues spurring the OpenEPI initiative. The fear that competitors might gain an edge by accessing their data keeps many companies from participating in the open data movement. Despite these challenges, innovative models of collaboration are emerging, particularly in regions like Africa, where data collaboratives represent a novel form of public-private partnership. These initiatives bring together diverse stakeholders across sectors to share data and analytical resources. This collaborative approach is making it easier for different companies to work together to benefit from each other's data and simultaneously to benefit the public.[34] One way they have been driving collaboration is by using competitions, where companies make data available to qualified applicants that compete to develop new apps or discover innovative uses for the data. One example known from the literature is the *Orange Telecom Data for Development Challenge* in the Ivory Coast and Senegal, where Orange Telecom hosted a global challenge that allowed researchers to use anonymized, aggregated Call Detail Record (CDR) data to help solve various development problems related to transportation, health, and agriculture. [35]

Companies like IBM, Google, and Meta have developed physical infrastructures within the African continent, motivated by the objective of customizing their offerings to meet the specific needs of African markets. The Data for Development Network notes that these firms are actively contributing to the enrichment of the open data ecosystem. They achieve this by financing open data projects that not only facilitate the adaptation of their technologies to local contexts but also support the broader agenda of open innovation. This strategy underscores a dual motive: advancing local development objectives while simultaneously creating opportunities for the utilization of their services and solutions, thereby generating profit.[36]

2.1.2 Data publishers/data platforms

The data publishers and data platforms (and complementary services) are a heterogeneous group of actors. They are responsible for making data available to the public. They may or may not be producing data, but they play a crucial role in facilitating accessibility. The goal of

[34] Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre.
[35] Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre.
[36] Data for Development, *Sub-Saharan Africa and the State of Open Data,* 2022, https://www.d4d.net/news/sub-saharan-africa-and-the-state-of-open-data/

open-data platforms is most often to enable not just data access but also enabling value creation. Data platforms have direct influence on how the data they publish is made available to the public[37].

There are many data platforms that distribute open climate and nature data. These platforms share some common traits. A notable characteristic of the platforms in this domain is their origin in specific research communities or groups. These origins reflect a collective desire to disseminate research data produced through various projects and spread them to a broader audience encompassing researchers, non-governmental organizations (NGOs), and policymakers. For instance, the Global Biodiversity Information Facility (GBIF), an international network and data infrastructure funded by some governments and originating from an OECD panel in 1999, has been pivotal in providing access to data about all types of life on Earth, supporting both scientific research and conservation efforts.

Advocacy for open data principles is another hallmark of these platforms. They champion the notion that environmental data should be openly accessible to all stakeholders, including the public, researchers, policymakers, and more. This advocacy is grounded in the belief that open access to data can drive informed decision-making and foster a more sustainable interaction with our environment.

Collaboration with a diverse range of stakeholders is also a defining feature. By partnering with governments, nonprofits, academia, and the private sector, these platforms enrich the quality and scope of the data they offer. An example of such collaboration is the Climate Watch platform, developed by the World Resources Institute (WRI)[38] in partnership with over 50 international institutions. This platform provides comprehensive data and visualizations to track and analyze national and global progress towards climate action.

There are numerous government and intragovernmental agencies across sub-Saharan Africa that are producing and disseminating climate and nature data. AGRA (Alliance for a Green Revolution in Africa) works towards a food secure and prosperous Africa. African Union Development Agency (AUDA-NEPAD) offers various datasets, including valuable geospatial data, to support development initiatives across Africa. This open data initiative enables better planning, analysis, and decision-making for projects aimed at sustainable development and economic growth. Many of the data platforms in these domains have researchers, authorities and environmental organizations as their main target group and their main goal is to provide data to support international policies and agreements related to environmental sustainability and climate change. For most platforms, developers are not specifically addressed. Instead of distributing raw data and providing APIs for development of innovative products, these platforms focus on end-user products - predominantly providing maps, charts, and other visualizations helping users understand complex environmental patterns and trends.

In recent years, there has been a significant increase in the overall number of global data platforms in the nature and climate domains. The World Resources Institute's creation of an

[37] Recommended “Best Practices for Better Sharing of Climate Data” by Taylor Campbell, Wanying Li, and Dr. Cable Green for Creative Commons is licensed CC BY 4.0., Creative Commons, 2024, https://docs.google.com/document/d/1PLRu-bLd91Zbndn4QbSfNQAn-vamPlblSq-7LFo7CP8/edit

[38] https://www.wri.org/initiatives/climate-watch

interactive visual map showcasing over 100 major climate data platforms exemplifies this growth.[39] This tool aids users in navigating the plethora of available platforms and selecting the most relevant ones for their needs. WRI's curation process revealed that many platforms were presenting similar datasets, suggesting a need for new platforms and datasets to build upon existing efforts, adhere to established standards, and provide actionable data. Before launching new initiatives, platform creators are advised to conduct thorough scoping to identify critical data gaps and assess how new datasets or platforms could address these gaps. WRI emphasizes the importance of integrating datasets into existing platforms where possible, to avoid redundancy and enhance the impact of open data efforts.

Some of the most impactful global platforms in the field of climate and nature data in sub-Saharan Africa are briefly presented below.

WORLD RESOURCES INSTITUTE	World Resources Institute (WRI): WRI is a non-profit organization that operates as a data publisher and provider. It focuses on producing and curating environmental data to promote sustainable development. WRI's data covers various aspects such as climate, ecosystems, water resources, and more. WRI develops and maintains platforms like the Global Forest Watch and Aqueduct, providing accessible tools for users to explore and analyze environmental data. WRI also uses its data to advocate for sustainable practices and policies, and it educates the public on environmental issues.
GEO GROUP ON EARTH OBSERVATIONS	Group on Earth Observations (GEO): GEO is an international organization that coordinates efforts to build a Global Earth Observation System of Systems (GEOSS). It facilitates the sharing of Earth observation data from various sources to support informed decision-making. Responsibilities. GEO brings together data from Earth observation satellites, ground-based sensors, and other sources, ensuring interoperability and accessibility. GEO works to enhance the capacity of countries and organizations to collect and use Earth observation data for environmental monitoring and management.
GODAN Global Open Data for Agriculture & Nutrition	Global Open Data for Agriculture and Nutrition (GODAN): An Africa-based collective initiative advocating for the open sharing of agricultural and nutritionally relevant data. While its primary focus is on food security and agriculture (but also other data domains of importance for economic growth), GODAN's work indirectly supports climate resilience and adaptation strategies by promoting the open availability of critical agricultural data.
AFRICA GEOPORTAL™ Powered by Esri	Africa GeoPortal: ESRI is a global vendor of GIS tools. They have established the Africa GeoPortal, including data, tools and learning materials. The portal was developed to remove the barriers to entry for working with geospatial data that currently exist throughout much of Africa and make it so that people can direct their energy toward building local solutions that help solve local problems.

[39] https://www.wri.org/data/overview-100-climate-data-platforms

2.1.4 Users

Users of open global nature and climate data span a diverse range of stakeholders, each with distinct purposes for harnessing the information. Typical users are researchers, policy activists, professional or amateur developers, civic hackers, data journalists and individual citizens. Balancing the diverse needs of all those stakeholders is of course a complex task. How data is shared and distributed by a platform or data provider often depends upon their business model and target market. Some user groups need access to specific raw data while others are more interested in access to processed data, or different combinations of data. Policy activists, for instance, look for accessible summaries, visualizations like charts and maps, and illustrative case studies. These resources help them communicate the status or the implications of actions to a broader audience, advocate for policy changes, and influence decision-makers by clearly demonstrating the real-world impacts of environmental challenges.

Conversely, developers and data analysts require access to downloadable, granular and flexible raw data that allow for creative and technical manipulation. The granularity of the data is crucial for them to perform sophisticated analyses and develop detailed visualizations, catering to specific analytical objectives or business needs. They further seek reliable API access to build applications or professional business services efficiently. They value stable, well-documented APIs that facilitate the integration of open data into their own tools and contexts. Researchers span both domains, needing both exhaustive, high-caliber raw datasets for in-depth study, and processed data for broader analyses. Like policy activists, they too rely on visual tools and case studies to underpin or supplement their research.

Below, in section 2.3, we will discuss further how open data platforms cater to different needs, and how the needs of developers, that are OpenEPI's main target group, are different to that of other user groups, like policy activists.

2.1.5 Consumers

Some second and third hand users could be labeled “consumers” of open data, more than solely “users”. Those are broadly speaking governments, businesses and individuals - that is, target groups for products, services, research and advocacy works produced by the firsthand users of open data.

Governments utilize open climate and environmental data for many purposes, including to make informed decisions on urban planning and the development of resilient infrastructure. For example, flood prediction models developed using open data can help in planning the construction of flood defenses and in the zoning of land use to minimize the impact from flooding. Agricultural businesses and farmers can benefit from applications developed using open climate data to optimize decisions on crops, harvesting and other farming practices. Apps that provide information on weather forecasts, soil moisture, and crop health can help farmers make decisions on planting, irrigation, and harvesting, leading to improved yields and reduced waste. Further, access to real-time data on natural disasters such as hurricanes, floods, or wildfires enables individuals to better prepare for emergencies. Apps and services that alert users of severe weather conditions can save lives and property by providing timely information for protective actions or evacuation.

2.1.6 Sponsors

Sponsors are organizations, private foundations or individuals that provide financial support or resources to facilitate the collection, processing, and distribution of climate, nature, and environmental data. They typically fund projects and specific, narrow research initiatives - or fund platforms that contribute to the broader climate ecosystem. They may have a direct influence on the accessibility of data resulting from sponsored research, even if the data are collected or produced by another organization.[40] Like other agencies for development cooperation, Norad is a substantial funder of data intensive initiatives in developing countries. In such, Norad has also potentially a very significant reach as a sponsor of open data and does actively also endorse open data initiatives.

The EU is a prominent actor in this landscape, significantly contributing through Horizon Europe (2021-2027), the program for research and innovation, which allocates funding to a wide range of open data research projects. These projects span various domains, including mobility, energy, pollution control, agriculture, and geospatial data, demonstrating the EU's commitment to advancing environmental and climate research through open data.[41] In parallel to Horizon Europe, the EU has also launched the Digital Europe program for the same period (2021-2027).[42] This program aims to bolster the digital transformation of Europe's societies and economies. It focuses on building the strategic digital capacities of the EU and on facilitating the wide deployment of digital technologies, to be used by Europe's citizens and businesses. One of the key components of the Digital Europe program is the development of common European data spaces[43], which are intended to create an environment in which data can be shared securely and efficiently across borders and sectors, fostering innovation and creating opportunities for new services and businesses.

The United Nations (UN) also plays a vital role in supporting and operating various regional and global open data initiatives. Through entities such as the United Nations Development Programme (UNDP), the United Nations Environment Programme (UNEP), the UN World Food Programme, the UN champions the cause of data openness. Additionally, several UN initiatives specifically focus on publishing open data, the United Nations Expert Group on Geospatial Information Management (UN-GGIM), the United Nations Office for Disaster Risk Reduction (UNDRR), the United Nations Satellite Centre (UNOSAT), and the United Nations Statistics Division (UNSD). These initiatives underscore the UN's commitment to leveraging open data for global benefits.

The World Bank, with its global reach and financial resources, supports numerous projects and initiatives aimed at improving access to climate and environmental data. This support often takes the form of funding for data collection, analysis, and dissemination efforts, contributing to the bank's broader goals of poverty reduction and sustainable development.

---

[40] Recommended "Best Practices for Better Sharing of Climate Data" by Taylor Campbell, Wanying Li, and Dr. Cable Green for Creative Commons is licensed CC BY 4.0., Creative Commons, 2024, https://docs.google.com/document/d/1PLRu-bLd91Zbndn4QbSfNQAn-vamPlblSq-7LFo7CP8/edit

[41] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

[42] https://digital-strategy.ec.europa.eu/en/activities/digital-programme

[43] https://digital-strategy.ec.europa.eu/en/policies/data-spaces

Also, other international organizations can act as sponsors for open data. For instance, there are regional international organizations across Africa who are supporting the distribution of open climate and nature data. For example, the Alliance for a Green Revolution in Africa (AGRA) works towards a food-secure and prosperous Africa, partly through the promotion of data openness in agricultural practices.[44] African Union Development Agency-NEPAD (AUDA-NEPAD) facilitates the exchange of best practices and offers technical assistance to African states in order to enhance governance, skills development, and effective project implementation. AUDA-NEPAD hosts The GeoHub that lets users explore, visualize, and download location-based development data.[45]

Finally, private foundations are also supporting the open data movement. One of the most prominent foundations in this space is the Bill & Melinda Gates Foundation. This Foundation has been a strong advocate for the open data movement, particularly in the areas of global health, education, and agricultural development in Africa. They invest in the development of open data platforms and tools that facilitate the easy sharing and accessibility of data. Occasionally, the foundation launches support initiatives and grants aimed at incentivizing the use of open data to solve specific problems. These initiatives encourage innovators and entrepreneurs to develop new solutions that leverage open data for social good.[46] The Mozilla Foundation is another example of a private foundation that is involved in several open data initiatives in Africa. They are focusing on leveraging technology for social good and enhancing internet accessibility and knowledge sharing. One notable initiative is the partnership with the German Development Cooperation initiative “FAIR Forward – Artificial Intelligence for All,” which aims to provide open, non-discriminatory, and inclusive training data, models, and open source AI applications. This initiative works with countries including Rwanda, Uganda, and Kenya to develop open AI training datasets in local languages like Kinyarwanda, Kiswahili, and Luganda, promoting local innovation and empowering marginalized groups through technology such as voice-based access to services and climate-smart agricultural advice.[47]

2.1.7 The open data community

The open data community consists of a wide range of organizations, networks, and platforms that advocate for the principles of openness, transparency, and free access to data. These entities play crucial roles in promoting open data across domains, including government, environment, science, and climate change. The focus of the open data movement may be seen as a bit instrumental. Normally, those attracted to the open data community are primarily concerned that data should be open, less about what the data are used for.

A major barrier to the open data movement is the commercial value of data. Traditionally, access to, or re-use of, data is often controlled by public or private organizations. Control may be through access restrictions, licenses, copyright or charges for access and re-use. Advocates of open data argue that these restrictions detract from the common good and that data basically should be available without any restrictions or fees.

[44] https://agra.org/
[45] https://www.nepad.org/
[46] https://www.gatesfoundation.org/
[47] https://www.bmz-digital.global/en/overview-of-initiatives/fair-forward/

Players in this field, such as Open Knowledge Foundation, Climate Action Tracker, Open Climate Knowledge, and Creative Commons - which is also standing behind this feasibility study - are particularly active in campaigning for open climate data. These organizations and initiatives advocate for the free and open sharing of climate-related data to enhance global response to climate change, support research and policymaking, and promote transparency and collaboration across borders.

2.2 How open are open data platforms?

As already mentioned, open data by definition should be freely available to anyone to use and reuse as they wish, without restrictions. However, *how* open data platforms are preparing for this, varies significantly. In a preliminary scan of open data platforms, carried out by Norad in December 2023, even self-proclaimed "open" data solutions for nature data are not truly open. A scan of more than 70 presumably relevant datasets showed that many were not licensed properly. Further, many datasets had an unclear provenance, and an unspecified quality. Almost all datasets were downloadable, but not available for direct consumption through APIs. Often, datasets were not described with appropriate metadata or failed in referring to defined standards, even if they were published under the correct license.

Research carried out by the Curtin Open Knowledge Initiative (COKI), revealed that only 47 percent of research articles on climate change published globally, are open.[48] Partly due to the research findings by COKI, Creative Commons, together with partnering institutions SPARC and EIFL, has run an "Open Climate Campaign" to promote open access to research and to accelerate progress towards solving the climate crisis and preserving global biodiversity[49]. The campaign's study from 2023[50] documents the openness of key large climate data providers from around the world, detailing how their climate data can be found, accessed and reused. Similarly to our scan of datasets, the picture is quite mixed. In general, though, many of the websites in question fail to provide open data.

To sum up, there are several reasons why open data fails to be really open:

- How it is possible to find the data, that is if they are identified with unique and persistent identifiers (like DOI), and by use of standardized metadata
- Licenses and crediting issues
- The degree of direct accessibility (that is, downloadable or possible to consume directly, to no costs and without registration procedures)
- Technical interoperability, for instance use of standardized file formats or data consolidation on the platform (contrary to distributed to third party, external sources)
- Reusability: data being properly equipped with describing metadata and provenance

OpenEPIs goal is to serve as a fully open data platform, where none of these or other factors contribute to creating obstacles for the use by developers.

[48] Open Climate Campaign, *Most climate change research is not accessible, 2023,* https://openclimatecampaign.org/
[49] https://openclimatecampaign.org/
[50] https://creativecommons.org/2023/08/08/surveying-the-open-climate-data-landscape/

2.3 Scarcity of platforms catering to developers

Our investigation into the open nature data ecosystem, informed by our study, including the stakeholder interviews we have conducted, has highlighted a critical oversight: the prevailing design of open data platforms in this domain does not cater to developers. Instead, they are chiefly aimed at facilitating research, policy building and decision-making. This focus on policy and decision support is also reflected in how data is made available—often as finished products like visualizations, graphs, and maps—rather than in raw, manipulable forms that can spark local innovation and be directly utilized by developers.

The scarcity of platforms that explicitly aim to stimulate innovation by addressing the specific needs of developers, highlights a significant gap between the potential of open data to drive local development and the current direction pursued by most data portals. Hence, there is a pressing need to realign open nature and climate change related data portals towards supporting local innovation. Prioritizing the needs of developers and technical users means to provide unmediated access to raw data. The design and functionality of the portals should enhance, rather than hinder, the accessibility of such data.

As Norad sees it, the inception of OpenEPI will be a strategic response to this identified need within the open data ecosystem. Our point of departure has also been that OpenEPI should be an open global data platform dedicated to serving developers, with the overarching aim of fostering local innovation. The purpose of this entire study is to assess the feasibility of this concept of a developer-centric portal or hub where developers eventually will find:

- Direct access to raw data: Unfiltered, comprehensive datasets available for download, enabling them to manipulate and analyze data according to their project needs.
- Robust APIs: Well-documented, reliable APIs that facilitate seamless integration of data into applications, simplifying the development process and encouraging the creation of innovative solutions.
- Developer resources and tools: A suite of tools and resources, libraries, and guides, tailored to streamline the development journey and empower them to effectively leverage open data.
- Community collaboration spaces: Forums and collaborative platforms that encourage dialogue, sharing of best practices, and co-creation among them, fostering a vibrant community centered around open data innovation.
- Support for data interoperability: Emphasis on data standards and interoperability to ensure that datasets can be easily combined and used across different platforms and projects, maximizing their utility and impact.

By focusing on these key areas, OpenEPI aims to bridge the current divide, transforming how developers interact with open data and, in turn, unlocking the full potential of open data to catalyze local innovation in the domains of nature management, agriculture and food production, and climate change mitigation. We recognize the challenge to meet the needs of all developers. Countries and regions are naturally different, both geographically and politically. At best we will be able to meet some of the existing needs for data. OpenEPI will have to be chasing needs and adapt both the selection of datasets and services to the varying needs of individual countries, regions and “markets”.

3. Designing a platform and services for impact

OpenEPI, like many other development initiatives, must rest on a set of assumptions about how and why the project should work to make a positive impact on society. In the realm of development projects, particularly those affiliated with UN agencies, it has become customary to employ a strategic framework known as the Theory of Change (ToC). [51],[52]

This approach meticulously articulates the project's requirements in a step-by-step manner, starting from its overarching objectives. It involves planning in reverse from these goals, specifying necessary actions, and identifying the resources required for realization. The ToC method elucidates how all the various moving parts must operate together to bring about a desired change or long term outcome. A ToC also involves identifying the potential risks that may hinder the achievement of the intended results and mapping appropriate mitigating measures. We think the ToC approach is crucial for OpenEPI success and will give it this chapter we discuss a ToC for OpenEPI, in a prescriptive approach.

3.1 Theory of change

Utilizing Theory of Change (ToC) in OpenEPI's future development will ensure that all strategic actions are purposefully aligned with its overarching mission of enhancing climate change adaptation and climate resilience in low-and-middle-income countries, in a fair and responsible manner. The methodology not only facilitates the identification of the most relevant and effective strategies and interventions but also fosters a deeper understanding of the contextual and operational dynamics that influence success in this domain. As part of this, ToC is inherently linked to the user perspective as it emphasizes the importance of stakeholder engagement, and the integration of the users' needs in the development process. This user-centric approach will be crucial for the success of OpenEPI, which aims to empower developers in developing countries. The ToC approach ensures that the platform's future development is guided by the real-world challenges and opportunities faced by this target group (and their "market"), enhancing the relevance and impact of OpenEPI.

In the following, we will explore the benefits of employing the Theory of Change methodology in the realization of OpenEPI, demonstrating how it can provide a structured path to achieving its objectives. We will outline and delve into OpenEPI's primary aim, demonstrating how the project's outcomes, outputs, and activities are interconnected to fulfill this objective. In chapter 4, we will explore the essential user needs that OpenEPI must meet to guarantee the achievement of its planned outputs, outcomes, and impacts. Chapter 5 outlines the technology building blocks and architecture blueprint for OpenEPI. In chapter 6 we assess the organizational setup and governance structures necessary to secure the planned outputs and outcomes. Chapter 7 offers an in-depth examination of possible risks that may prevent the realization of OpenEPI's chain of results, and outline strategies for their mitigation.

---

[51] Government of Canada, *Results-Based Management for International Assistance Programming at Global Affairs Canada: A How-to Guide,* 2022, https://www.international.gc.ca/world-monde/funding-financement/results_based_management-gestion_axee_resultats-guide.aspx?lang=eng&_ga=2.94634911.2043844598.1710503846-352763691.1710240499

[52] The United Nations Development Group (UNDG), *Theory of Change UNDAF Campanion Guidance,* https://unsdg.un.org/sites/default/files/UNDG-UNDAF-Companion-Pieces-7-Theory-of-Change.pdf

3.2 Building a model

Using the Theory of Change involves a structured approach that begins with the identification of the overarching impact of the initiative and then works backwards to map out the necessary outcomes, outputs, activities, and inputs needed to achieve these impacts. This groundwork ensures that all subsequent planning, from the selection of activities to the allocation of resources, is logically and strategically aligned with achieving its long-term impact. The figure below illustrates how we see the Theory of Change, with inputs, activities, outputs, outcomes and impact logically interconnected in the OpenEPI concept.

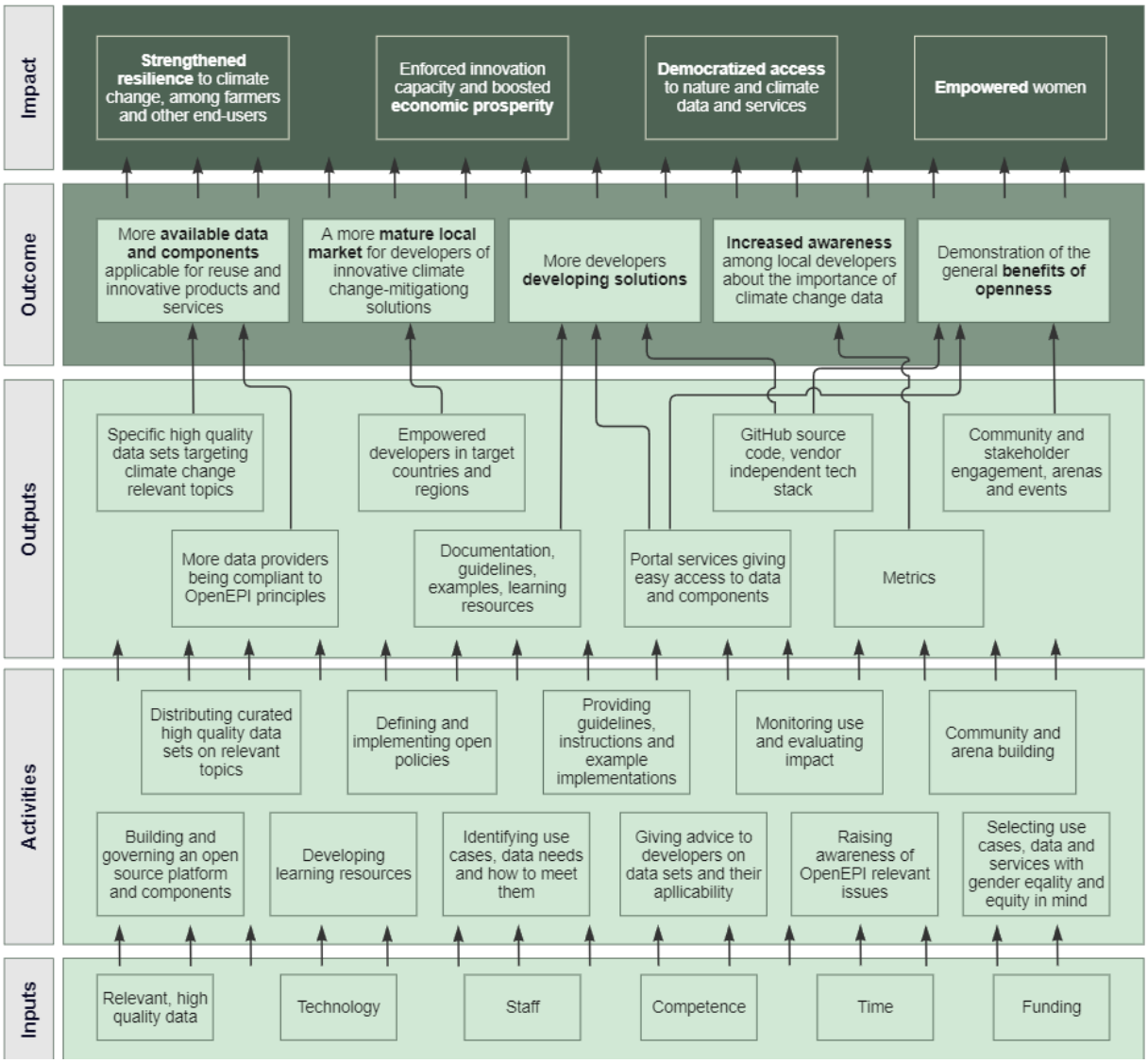


Figure 3. The Theory of Change for OpenEPI

3.2.1 Impact

The starting point for a ToC for OpenEPI is a clear, articulate long-term goal. This is the broad impact that the initiative aims to achieve in the future. The impact must be measurable and is dependent on the achievement of the outcomes.

We propose OpenEPI's overarching goal being *to significantly improve climate resilience among vulnerable populations in developing countries by enabling open and equitable access to actionable and relevant data.*

To be able to measure whether this has been achieved, the goal must be broken down into measurable metrics. We are proposing at least the following metrics, but those should be further elaborated:

- General acknowledgement of OpenEPI providing relevant data for improved decision support in sub-Saharan countries.
- Increased yield and economic prosperity in countries/regions that use OpenEPI-enforced digital solutions.
- A greater share of women is involved in local entrepreneurship and digital innovation activities, in countries where OpenEPI based services are widespread.

In deciding these metrics, we have investigated what specific challenges related to changes in climate conditions the populations of sub-Saharan countries face. Are there specific issues that are particularly acute? We have also discussed what challenges developers and startups face when they are trying to build digital solutions for this local "market" of farmers and agriculture workers combating climate change.

Though we have limited empirical support for it in this study, we must consider that different countries and regions represent rather different conditions for developers and for OpenEPI impact. We have some evidence from our visit to Rwanda, and some second-hand information from other African countries and regions around the world. This evidence is supplied by extensive document research and of course our interviews with existing data portals, but we really don't know if or how the logic from the ToC above applies to all countries relevant for climate change mitigation efforts. Further, to be able to measure the impact of OpenEPI we also had to consider if there are other stakeholders or platforms working towards the same goal and targeting the same groups, as mentioned in chapter 2.

3.2.2 Outcomes

Outcomes are the specific changes or benefits that result from the outputs from OpenEPI. The outcomes should together lead to the overarching goal of the project, proposed above. Outcomes can be categorized into immediate (or short-term) and long-term. For OpenEPI, an immediate outcome might be "*increased awareness among local developers about the importance of climate data*". A long-term outcome could be "*local farmers actively use solutions developed by using data and based on services from OpenEPI to mitigate local climate challenges.*"

Also, the outcomes should be measurable. Metrics for OpenEPI outcomes could be:

- The number of farmers using applications based partially on OpenEPI services.
- Changes in crop yields and income levels among users of OpenEPI based solutions.
- User satisfaction with the provided information on OpenEPI portal, measured by the nature of feedback or by satisfaction surveys among developers and startups.
- The extent of stakeholder engagement (for instance number of stakeholders engaged) and coordinated efforts in combating climate change.

Creating a visual representation, such as a flowchart or diagram, can help us to better understand the sequence of outcomes and how they contribute to achieving the goal. This visualization can also highlight dependencies and critical junctures where specific interventions are needed.

3.2.3 Outputs

Outputs are the direct product of OpenEPI's activities. They are tangible and measurable, and they contribute directly to achieving the desired outcomes. The most important OpenEPI outputs are:

- Data and data products relevant for developers in sub-Saharan countries, relevant to the efforts of climate mitigation and resilience, demonstrating the potential in data driven decision making.
- A blueprint for an open source, vendor independent technology stack building a functional platform targeting developers of solutions consuming the above mentioned data and data products.
- Services supporting developers in an adequate way.
- Policies, standards and routines securing long-term functionality and relevance (for instance updated data, covering new topics, exploiting new technology, etc.)
- Capacity, methods and arenas for stakeholder engagement and involvement.

Outputs will depend directly on the quality and relevance of OpenEPI’s activities as an organization (see chapter 6). For each identified outcome (above), we should determine what outputs are necessary to achieve it. For each output we should also define clear indicators that should allow us to measure whether and to what extent the output has been achieved. For the policies and standards example, indicators might include the perceived clarity (user survey) of the policies and standards described or claimed by OpenEPI, the number of queries or complaints from data providers or other stakeholders, or measurements of the general reputation of OpenEPI.

It will be of importance that planned outputs are feasible, given OpenEPI’s resources and implementation timeline, and that they are directly relevant to achieving the outcomes. Each output should be easy explainable as to how it contributes to the larger goals of the project.

3.2.4 Activities

Activities, together with inputs, are the building blocks of the OpenEPI, forming the foundation upon which outputs, outcomes, and ultimately, the long-term goal are achieved. Activities refer to the specific tasks or actions undertaken to produce the desired outputs, leading to the achievement of identified outcomes and contributing to the overarching goal of enhancing climate resilience among vulnerable populations.

For OpenEPI, activities could include:

- Developing and technically updating the platform, as new technology evolves: This involves the technical development, testing, and updating of the OpenEPI platform to ensure it is user-friendly, robust, and compliant to open source policy.

- Developing and updating the actual API’s and data products presented by the platform: Securing that available data provides accurate and timely climate, nature and environmental data of relevance for agriculture and climate change adaptation in sub-Saharan Africa.
- Capacity building, consultancy services and training workshops: Conducting workshops for and helping developers, community members, and other stakeholders to enhance their ability to effectively take advantage of OpenEPI. These workshops would cover how to access, interpret, and apply nature and environmental data for local climate resilience projects.
- Providing training materials: Development of high-quality, accessible training materials, including guidelines, video tutorials, and case studies, to support the capacity-building activities.
- Collaboration and networking: Establishing partnerships with environmental organizations, government agencies, and academic institutions to enrich the platform with diverse data sources and to foster a collaborative ecosystem around climate resilience.
- Outreach and advocacy: Implementing communication strategies to raise awareness about OpenEPI and its potential impact on climate resilience. This could involve social media campaigns, presentations at conferences, supporting development aid grant managers and receivers to be OpenEPI policy compliant in the case of relevant new data collection and production, and a broad range of other community engagement initiatives.
- Monitoring and evaluation: Setting up systems to track the usage of OpenEPI, gather feedback from developers, collect metrics, and in several ways assess the platform's impact on enhancing climate resilience, gender perspectives, and economic impact in the addressed countries and regions.

3.2.5 Inputs

Inputs are the resources invested in carrying out the activities. For OpenEPI, this includes:

- Financial resources: Funding is required for all aspects of the initiative, from platform development and governance, to conducting workshops and networking activities. The fundings might come over national budgets, or from grants, donations, or partnerships with organizations aligned with OpenEPI’s mission.
- Human resources: A team of skilled professionals is crucial, including backend platform architects, software developers, data scientists, climate (and other domains) experts, trainers, and project managers, to develop the platform, searching and gathering relevant data sources, conduct training, and manage the initiative.
- Data: The core of OpenEPI is the existence of and availability to relevant, high quality data - with metadata and descriptions. The data are provided either through quality assured APIs to the original data sources, or as derived or aggregated data products on the platform itself.
- Technological resources: Software, hardware, and internet services are essential for developing and maintaining the OpenEPI platform, as well as for facilitating virtual training sessions and online collaborations. The technology stack OpenEPI has demonstrated (see chapter 5) is open source and aims at vendor independence, thus contributing to the overarching goal of democratized access.

3.3 Summary: Theory of Change as a strategic tool

The Theory of Change (ToC) model presents a strategic framework for the initial as well as the future development and governance of OpenEPI. We believe the proposed model needs to be further developed and refined so that it has a well-thought-out content and logic. This should be done through work involving key stakeholders. Implementing a refined ToC will offer OpenEPI a clear roadmap for action, delineating the necessary steps to achieve its ambitious goals. Such a roadmap will aid in prioritizing activities, judiciously allocating resources, and adapting to unforeseen challenges or changes in the landscape. Furthermore, implementing a defined ToC will most certainly underscore the importance of partnerships and collaboration, identifying opportunities to enhance the platform's reach and impact by engaging with governments, universities, NGOs, and the private sector. Such collaborations will not only augment the platform’s data sources but also expand its dissemination channels, crucial for achieving widespread impact.

The ToC framework we are suggesting here will also enrich OpenEPI’s capabilities in monitoring, evaluation, and advocacy by establishing clear indicators for measuring progress and outcomes. This systematic monitoring also facilitates ongoing learning and adaptation, allowing OpenEPI to fine-tune its strategies based on empirical evidence of what works. By focusing on outcomes, the ToC will empower OpenEPI to make evidence-based improvements, scaling up successful initiatives while phasing out less effective ones. Moreover, a well-articulated ToC will serve as an efficient communication tool, enabling OpenEPI to effectively convey its strategy, logic, and impact to a broader audience. This not only aids in securing support and attracting funding but also enhances user engagement, laying a solid foundation for OpenEPI to achieve its mission of fostering local development of climate mitigation solutions.

4. User needs

In the previous chapter, we saw how important it was to build the logic around OpenEPI. Covering user needs is the central component of the rationale for OpenEPI. Before we dive into the technical assessments and description of what we have done in the pre-project, we therefore must discuss user needs, on the background of the perspectives drawn in the previous chapters.

User needs is both about the need for (access to) data and about the need for support for user uptake - i.e. support in building innovation capacity, referring to the ToC model. In other words, we see OpenEPI as both a supply-side measure and a demand-stimulating measure. The first to address the topics (climate adaptation, nature management, agriculture and food production resilience), the second to ensure that the data OpenEPI gives access to, provide value (i.e. has potential to cradle new solutions and creative applications, and thereby supporting local businesses). The supply side includes looking at the potential competition from other initiatives and innovative projects, although many of these "competitors" often have specific products and services also out in the final stages and are only partly concerned with providing data (although some of them also do this). In principle, the demand side has very little competition, and as far as we know, no such services as OpenEPI exist today.

4.1 Introduction

Open data platforms operate within a supply-demand ecosystem. Data are shared and distributed with the intention of solving real world challenges - like the challenges created by climate change or biodiversity losses. Hence, a platform's success lies in its usage; the users are the ones who generate its impact. And engaged users and their feedback and sharing of experiences attract new users. Without active users engaging with the data to solve real-world issues, the platform cannot achieve significant outcomes. Ultimately, addressing user needs ensures that open data not only exists but empowers and transforms.

For an open data platform to reach its set goals, it must be designed and promoted in a way that resonates with the needs of potential users. This involves not only making data easily accessible but also ensuring they are relevant, reliable, of satisfying quality (in a broad sense of the concept of quality) and presented in formats that users can easily apply to their contexts. Understanding the demand side normally requires thorough market analysis to identify the specific data needs of different user segments and different local contexts.

In addition, for an open data platform to be sustainable and impactful over the long haul, it's imperative not only to consider the primary users of the platform, but also the *secondary user*s (in this case, farmers or other secondary users across the agricultural industry of sub-Saharan Africa) who will interact with products or software solutions derived from the OpenEPI data. While direct engagement with secondary users might not always be feasible, overlooking the market dynamics and stakeholder interests surrounding the end-use of the platform's outputs can impede its societal impact. It is essential, therefore, that a platform caters not just to its direct users (the developers) but also ensures that the developed solutions and products based on the platform's data in their turn meet an existing demand,

and that end users are equipped (equally and fairly) to take advantage of those solutions and products.

The Theory of Change (ToC) approach emphasizes engaging with stakeholders to understand their needs, perspectives, and how they might interact with the platform. This can lead to the design of more user-friendly interfaces, relevant datasets, and effective outreach strategies. In the following we will discuss what user needs OpenEPI must address in order to have the desired outcome, both the needs of its direct users and that of its secondary users. We will then present our key learning points from the workshop we conducted with developers in Kigali, Rwanda, in November 2023.

OpenEPI will have locally residing developers as a main target group. They could be established professionals running their own startups or businesses, developers employed by governments or NGOs, or students still at universities. Although climate change adaptation and resilience, and agro-innovation and food security are indeed global challenges and areas of efforts, in this first phase the platform will mainly target developers located in sub-Saharan Africa - as climate change resilience in this specific geographical area has been the direct focus for Norad's initiative. For instance, we have already found that some of OpenEPI's potential services will be applicable and valuable for initiatives related to cadastral administration and deforestation mitigation in Brazil.

The typical user of the platform will most likely be a male developer, who has a university degree in data science or similar. The tech industry, both globally and within sub-Saharan Africa, tends to have much fewer female developers compared to male developers. This gender disparity is reflected in the anticipated user base of OpenEPI, where male developers are expected to outnumber female developers. Addressing this imbalance presents an opportunity for OpenEPI to implement initiatives aimed at encouraging greater participation and inclusion of women in technology and development, for instance at local universities and in agtech industry innovation projects. It is also relevant to address this through community engagement measures, like partnering with local women's groups and agricultural cooperatives to promote the concept and gather feedback from women - or more strategically by selecting data types and related services from a gender balance perspective, addressing needs faced specifically by women in the actual regions where we seek impact.

The user needs will differ based on their expertise and specific interests, including whether they intend to use OpenEPI for commercial projects, academic research, policy development, or societal improvement. To ensure that OpenEPI accurately understands and meets these varied needs, we here propose a differentiation of the category "users" into six distinct segments:

1. **Students and other emerging developers:** This segment encompasses university students and self-taught learners at the beginning of their technology careers, seeking foundational knowledge and skills.
2. **Commercially focused professional developers**: Experienced developers and technology professionals who operate within startups or established companies form this group. Their focus will be on leveraging OpenEPI in commercial ventures to drive innovation and business growth.

3. **Research-oriented professional developers**: Occupying positions at universities and research institutions, these users will be engaged in academic research, looking to OpenEPI for data that can support groundbreaking studies and findings.
4. **Policy- and community-oriented professional developers:** This group consists of developers working within government agencies, non-governmental organizations (NGOs), or international bodies focused on environmental or development aid. They will utilize OpenEPI to inform policies and initiatives for community and global benefit.
5. **University professors/teachers**: Professors and teachers in higher education will seek out OpenEPI as a resource their students can use to gain practical experience in how to use data to develop new and innovative solutions.
6. **Tech enthusiasts, hobbyists, and civic hackers**: A diverse group of technology aficionados, DIYers, and activists interested in using OpenEPI to explore new ideas or address societal issues, like climate change related issues, through innovation.

Section 4.2 below will delve deeper into how OpenEPI should plan to meet the requirements of these distinct user segments, ensuring that each group finds the platform valuable for their specific endeavors.

Fact box

Carnegie Mellon University

Carnegie Mellon University Africa (CMU-Africa), located in Kigali, Rwanda, serves as an extension of Carnegie Mellon University's efforts to expand its educational reach on the African continent. The institution focuses on providing advanced education in the fields of information technology and engineering, with a particular emphasis on information security, innovation and entrepreneurship, and mobile application development. The university places a significant focus on fostering innovation and entrepreneurship among its students. It encourages the development of startups by providing an environment where entrepreneurial ideas can be nurtured and developed into viable businesses.

Mobile technology, given its widespread use and impact in Africa, is another key area of emphasis at CMU-Africa. The institution supports students in developing mobile applications that address local challenges, including but not limited to healthcare, education, and financial services. This initiative is aligned with the broader goal of leveraging technology to solve societal issues.

CMU-Africa's student body is diverse, with over 300 enrolled students from 19 different countries. Among its student body, women represent 27 percent of the enrollment, underscoring the institution's ongoing efforts to enhance gender diversity in Science, Technology, Engineering, and Mathematics (STEM) fields. In addition to its current student population, CMU-Africa has produced over 550 alumni, many of whom have gone on to play significant roles in the technology sector. This network of former students contributes to the institution's impact on technological innovation and economic development within the continent.[53]

We have no overall figures on the number of African students in the STEM domain, which could give an indication of the potential uptake for OpenEPI data and services. Figures from the African Development Bank[54] show that less than 25 percent of African higher education

[53] https://www.africa.engineering.cmu.edu/about/index.html
[54] https://nexteinstein.org/promoting-stem-education-in-africa/

students pursue STEM-related career fields. This is substantially behind other parts of the world. Though there is an immense potential for STEM education to drive Africa's prosperity, concerted efforts are needed to overcome the challenges and encourage more students to pursue STEM fields. Anyhow, we assume there are great differences between different African countries, implying different measures to address this challenge, and different strategies for OpenEPI to secure user uptake. This must be further elaborated in the next phases of OpenEPI implementation.

4.2 Functionality - ease of access and use

OpenEPI's mission is to construct a developer-centric platform. Hence, the platform's development must be meticulously planned to address the specific needs of its various user groups. These groups surely have different requirements in terms of the functionality, such as data accessibility, platform usability, and data application capabilities. Understanding and addressing these different needs is crucial for maximizing the platform's utility and impact.

Based on the extraction from our research and discussions with stakeholders, the table below outlines the key needs for each user segment, shedding light on the specific requirements that OpenEPI must aim to satisfy. This segmentation allows for a targeted approach in the development of the platform and services, ensuring that features and resources are tailored to support the goals of each group effectively.

Fundamental for all user segments is the requirement for continuous open access to data provided by OpenEPI, available both through APIs and as complete, downloadable datasets.

User segment	User needs
Students and emerging developers	• **Learning resources and tutorials**: They will need accessible documentation, tutorials, and reference implementations to learn how to use OpenEPI effectively. • **User-friendly interface**: A straightforward and intuitive user interface will help lower the barrier to entry. • **Community support**: Forums or community platforms where they can seek help, share knowledge, and connect with peers and mentors.
Commercially focused professional developers	• **Advanced data analysis tools:** Tools and APIs that enable complex data analysis and integration into their commercial products. (*in its initial stage, providing tools will not be defined as an OpenEPI responsibility*) • **Scalability and reliability:** High-performance services that can scale for commercial applications and provide reliable data access, including real-time data. [crucial decision: can OpenEPI adopt this approach:] Some form of formal agreement on service level or availability to the services, or alternatively a disclaimer clearly forwarding (fully or partly) the responsibilities to the provider of the original data. • **Security and compliance:** Features that ensure data security and compliance with regulations, critical for commercial applications.

Research-oriented professional developers	● **Comprehensive data sets:** Access to high-quality, granular, and comprehensive data sets (including documentation of the quality and provenance of the data) to support rigorous academic research. ● **Collaboration tools:** Platforms for sharing data and findings with fellow researchers and participating in collaborative projects. (*in its initial stage, providing collaboration tools and facilities for researchers will not be defined as an OpenEPI responsibility*)
Policy and community-oriented professional developers	● **Access to multidisciplinary data:** Data that spans various fields relevant to policymaking and community development, including socio-economic, environmental, and health data. As this is not a focus for OpenEPI, there will possibly be a continuous pressure to develop the platform in this direction. ● **Engagement and feedback platforms**: Mechanisms for engaging with communities, gathering feedback, and iterating on projects based on community input.
University professors/teachers	● **Educational materials and datasets:** Curated datasets and materials that can be used for teaching purposes and to stimulate classroom discussions. ● **Assignment and project ideas:** Resources and suggestions for assignments and projects that students can undertake using the platform.
Tech enthusiasts, hobbyists and civic hackers	● **Open API Access:** Easy access to APIs for experimenting with data and developing personal or community projects. ● **Flexibility and creativity:** A platform that supports innovative uses of data and doesn't restrict the creativity of hobbyists and civic hackers. ● **Community showcase:** Opportunities to share projects with a wider community, receive feedback, and connect with like-minded individuals. The related functionalities may require some sort of moderating capacity.

4.3 Relevant and actionable data

For OpenEPI to effectively spur innovation and the launch of localized solutions for nature management and climate change adaptation and resilience, it is imperative that the platform provides data sets that are relevant and actionable for local developers. In our interviews, owners and managers of open nature data platforms emphasized how crucial it is to build capacity to supply users with high-quality, curated data sets that address their needs.

In order to have the desired impact - *to significantly improve climate resilience among vulnerable populations in developing countries by enabling open and equitable access to actionable nature and environmental data* - the entity operating OpenEPI must therefore have a comprehensive understanding of the specific climate-related challenges faced by farmers and others involved in food production, as well as the types of data that can empower solutions to mitigate these challenges.

4.3.1 The problem

Across sub-Saharan Africa, the population is highly dependent on agriculture for its livelihood. Agriculture is the key driver of economic growth. Consequently, this sector has been identified as one of the most critical for climate adaptation.[55] Farmers in sub-Saharan Africa are grappling with an array of challenges due to climate change, each significantly impacting agricultural productivity and sustainability.

One of the most immediate and visible impacts of climate change is the increased frequency and severity of droughts. These conditions drastically reduce water availability for irrigation - essential in arid and semi-arid areas. As water becomes scarce, the ability to irrigate crops diminishes, resulting in lower yields. Moreover, the scarcity of water impacts not only the current crop cycle but also the long-term sustainability of farming practices in the region. The dropping water tables, and declining soil moisture levels add further stress to the already vulnerable ecosystems.

Further, the unpredictable nature of rainfall patterns, characterized by irregular timing and intensity, poses a challenge to the agricultural schedule. Farmers, who rely on historical rainfall patterns to plan their planting and harvesting, face crop failures and significant income loss when rains are mistimed or absent. These inconsistent patterns complicate successful farming practices, as the risk of planting too early or too late could culminate in a total crop loss, endangering the food security and economic stability of communities.

Additionally, climate change has precipitated rising temperatures, which can stress crops beyond their tolerance levels, decrease fertility, and escalate the prevalence of pests and diseases. For instance, these days we see how climate change has dealt a harsh blow to West Africa, the heartland of the world's cocoa production. Droughts and rising temperatures have ravaged crops, leading to a substantial drop in global cocoa supply for this season. Generally, higher temperatures may inhibit plant growth, diminish seed germination rates, and lower yields. Moreover, warmer conditions favor the spread of pests and diseases, further reducing agricultural productivity. These factors not only diminish the quantity of produce but also its quality, challenging the maintenance of crop health and viability.

Furthermore, climate change exacerbates soil erosion, nutrient depletion, and land degradation, thus eroding the agricultural potential of the land over time. Increased heavy rains and flooding, direct consequences of shifting climate patterns, not only lead to the immediate waterlogging of crops but also erode fertile topsoil, rich in organic matter and essential nutrients. This erosion severely degrades soil quality, compounded by nutrient loss, making the land progressively infertile and less capable of sustaining agriculture. Such diminished productivity compels farmers to venture into new areas, risking further environmental degradation. Addressing the impacts of flooding becomes crucial in countering the challenges climate change presents to agriculture, highlighting the need for integrated water management and soil conservation strategies to bolster the resilience of agricultural landscapes.

[55] World Resources Institute, *Climate Change Open Data for Sustainable Development: Case Studies From Tanzania and Sierra Leone Prepared for the Global Partnership for Sustainable Development Data*, 2018,https://www.data4sdgs.org/sites/default/files/services_files/WRI%20Climate%20Data_FINAL2_optimized.pdf

Tackling these issues demands a comprehensive approach, including developing and introducing resilient crop varieties, adopting more knowledge based and sustainable farming practices, implementing efficient water management systems, and devising pest and disease control strategies. Collectively, these measures are vital for safeguarding the region's agricultural future.

4.3.2 Key data addressing the problem

Digital advancements have the potential to revolutionize agriculture worldwide by enhancing productivity, sustainability, and resilience against climate change. Despite the potential for advancements, agriculture remains among the least digitized industries in the global economy.[56] The management, sharing, and maintenance of data is crucial to succeed in digitizing the sector.

In regions like sub-Saharan Africa and other developing areas, even minimal progress in accessing and strategically using basic data sets, such as meteorological information, can drive major improvements in farming practices. Beyond meteorological data, other critical data types for addressing climate change include soil data, agricultural production forecasts, water resource management, biodiversity conservation, earth observation data, and other geospatial information. These datasets can support the development of solutions aimed at farmers as well as more general solutions aimed at food security, infrastructure risk assessment, and the understanding of local vulnerabilities to climate impacts.[57] In the sections that follow, we will explore some of these data sources in more detail, examining how we think they are relevant to local developers for innovative agricultural solutions.

Weather data

Meteorological data, often referred to as weather data, encompasses various types of information related to the atmosphere's state, such as temperature, humidity, wind speed, precipitation, and atmospheric pressure. This data is crucial for a wide range of applications, from daily weather forecasting to long-term climate research, agricultural planning, disaster preparedness, and the development of climate change mitigation strategies.

The main distributors of meteorological data globally are national meteorological and hydrological services, specialized UN agencies like the World Meteorological Organization (WMO), and various commercial players. The WMO plays a pivotal role in facilitating the international exchange of weather data, ensuring standardization, and supporting the development of meteorological services worldwide. The Norwegian Meteorological Institute (MET) is among the key players in the global distribution of meteorological data. This institution is renowned not only for its comprehensive collection and analysis of weather data within Norway but also for its significant contributions to international weather forecasting and climate research. The Norwegian Meteorological Institute operates Yr, a service that offers free weather data via its website and mobile app. Yr's accessibility and the accuracy of its

[56] Abbasi, Rabiya and Pablo Martinez and Rafiq Ahmad, *The digitization of agricultural industry – a systematic literature review on agriculture 4.0*, Smart Agricultural Technology, Volume 2, 2022, https://doi.org/10.1016/j.atech.2022.100042.

[57] World Resources Institute, *Climate Change Open Data for Sustainable Development: Case Studies From Tanzania and Sierra Leone Prepared for the Global Partnership for Sustainable Development Data*, 2018,https://www.data4sdgs.org/sites/default/files/services_files/WRI%20Climate%20Data_FINAL2_optimized.pdf

forecasts has made it a crucial tool for individuals, businesses, and researchers worldwide, aiding in everything from daily weather predictions to the planning of agricultural activities and the management of disaster risk.

Weather data is increasingly relevant for building climate mitigation solutions. In agriculture, for instance, accurate and timely weather information helps in making informed decisions regarding planting and harvesting times, irrigation scheduling, and the selection of crop varieties resilient to specific climate conditions. The demand for accurate meteorological data in sub-Saharan Africa is substantial, due to its critical role in agriculture, disaster management, and economic planning. In a case study conducted by the World Resources Institute (WRI) in Tanzania and Sierra Leone, respondents reported that access to up-to-date weather data was the most urgent need and that this was not available to them today. Obtaining high-quality meteorological data in this region presents significant challenges. There is a disparity in the quality of global weather data provided for Africa by global entities, including services like Yr, when compared to the data for Europe and the United States. This variance primarily stems from the sparse network of weather stations across the African continent, which yields data of lower granularity and spatial resolution. In addition, the access to local weather data is also notably poor across many sub-Saharan African countries, as the national meteorological institutes in the region are often hindered by infrastructural deficits and limited financial investment. These limitations have historically led to significant gaps in weather data collection and analysis, making weather forecasts less reliable and detailed than in regions with more developed meteorological infrastructure.

The repercussions of these data gaps are especially pronounced in the agricultural sector, where precise and timely weather information is indispensable for informed decision-making related to planting and harvesting. Enhancing the quality of meteorological data and leveraging it to create simple weather forecasting applications and services will empower farmers to make more informed decisions regarding planting and harvesting schedules, select crops suited to the anticipated weather conditions, and adopt water-saving measures in anticipation of droughts.

Soil data

Soil data encompasses information about the physical, chemical, and biological properties of soil, including its composition, nutrient content, moisture levels, pH levels, and organic matter content. Soil data is crucial for a wide range of applications, from agriculture and forestry to environmental management and climate change mitigation. The main distributors of soil data globally include national agricultural research organizations, international bodies such as the Food and Agriculture Organization (FAO) of the UN, the World Bank, and various non-governmental organizations (NGOs) specializing in environmental and agricultural development. Additionally, initiatives like the Global Soil Partnership and the International Soil Reference and Information Centre (ISRIC) play pivotal roles in collating and disseminating soil data at a global scale.

Soil data is instrumental in crafting climate mitigation solutions because it informs sustainable land management practices, soil conservation techniques, and the selection of appropriate crop types to enhance resilience against climate change. In a more long-term perspective, understanding soil carbon storage capabilities is essential for developing strategies to reduce greenhouse gas emissions and combat global warming. Moreover, accurate soil data aids in

assessing the potential of different regions for afforestation or reforestation projects, which are key components of climate change mitigation efforts.

In sub-Saharan Africa, the variability and availability of quality soil data significantly impact the region's agricultural productivity and food security, especially in the face of climate change and rapid population growth. Despite the critical importance of soil data for guiding sustainable farming and soil conservation techniques, many areas within sub-Saharan Africa suffer from sporadic coverage and often rely on outdated information. The challenges stem from the region's diverse ecosystems, limited research and development investments, and constraints in local data collection and analysis capabilities. These limitations are compounded by the vast geographical area of the region, making comprehensive soil sampling and analysis both logistically and financially challenging. However, there are concerted efforts and initiatives aimed at improving the availability and quality of soil data in sub-Saharan Africa. Collaboration is key to these improvements, with partnerships forged between African nations and international organizations like the FAO, the World Bank, and various NGOs. These collaborative efforts are geared towards strengthening soil research capabilities within the region. By enhancing local expertise, refining data collection methodologies, and promoting the sharing of crucial soil data, these initiatives aim to bolster sustainable agricultural practices, soil conservation efforts, and ultimately, food sustainability in the face of the ongoing climate challenges.

Earth observation data

Earth observation (EO) data, derived from satellite and aerial imaging, plays a crucial role in monitoring environmental changes, assessing natural resources, and supporting disaster management and climate change mitigation efforts globally. This data encompasses a wide range of information, including land use and land cover changes, vegetation health, water resources, atmospheric conditions, and more.

High-quality earth observation (EO) data are accessible via both proprietary and open sources. Private companies often distribute advanced, high-resolution datasets for a fee, catering to specialized needs across various industries like agriculture and urban planning. On the other hand, numerous space agencies and organizations globally offer significant amounts of EO data for free and under open licenses. Key providers include NASA’s EOSDIS, the ESA with its Copernicus Sentinel satellites, and the USGS’s Landsat program, all of which supply data crucial for environmental monitoring, scientific research, and policymaking.

Globally, the extensive availability of open-access EO data plays a vital role in supporting a broad spectrum of activities. EO data is integral to developing climate mitigation solutions as it provides detailed and accurate information on the Earth's surface and atmosphere. For instance, satellite data can track deforestation, monitor drought conditions, assess water quality, and observe agricultural productivity, enabling targeted and effective environmental and agricultural interventions. Furthermore, EO data supports the modeling of climate change impacts, helping policymakers and researchers develop strategies to mitigate these effects and adapt to changing environmental conditions. Despite its potential, the use of EO data faces hurdles in sub-Saharan Africa. The challenges encompass accessibility,

processing capabilities, and local validation efforts.[58] While many satellite datasets are freely accessible, acquiring the most sophisticated or detailed ones frequently incurs significant costs. This financial barrier poses a considerable challenge for researchers and policymakers operating within the confines of limited resources. Further, the specificity of sub-Saharan Africa's needs, such as in-depth agricultural or water resource data, might not always align with the scope of global satellite datasets. This mismatch, coupled with a general capacity constraint for analyzing and leveraging complex satellite data, further complicates the situation. Lastly, the effective deployment of earth observation data is contingent upon the presence of both a robust technological infrastructure and access to highly skilled personnel.

Geospatial data

Geospatial data encompasses all types of information linked to specific locations on the Earth's surface, proving indispensable for a myriad of applications across diverse sectors such as environmental monitoring, urban planning, agriculture, disaster response, and transportation. For instance, geospatial data is used to monitor climate change impacts, optimize agricultural practices, plan infrastructure projects, and manage natural resources. Geospatial data is a kind of collective term for data sets typically combining location information (usually coordinates on the earth) and attribute information (the characteristics of the object, event or phenomena concerned) with temporal information (the time or life span at which the location and attributes exist). In that way, when referring to a location, meteorological as well as soil and EO data are indeed geospatial data.

The distribution of geospatial data occurs through both open and closed channels. Open data sources, provided by governmental and international organizations, offer freely accessible geospatial datasets. Examples include:

- The United States Geological Survey (USGS), which supplies a variety of maps and data sets relevant to geography, natural resources, and hazards.
- OpenStreetMap, a collaborative project that creates a freely editable map of the world, providing data for mapping and navigation purposes.
- The European Space Agency (ESA) and NASA, which offer satellite imagery and other geospatial data that can be crucial for environmental monitoring and research.

The EU is mandating their member states to collect and govern geospatial data, defined by the Inspire Directive and further listed as “High Value Data” (HVD) under the Open Data Directive. Inspire specifications include administrative location data, cadastral data, EO and environment data, in addition to meteorological data and other data categories like statistical data. “Environment” in Inspire refers to data sets describing themes like “hydrology”, “land cover” and “soil” - themes typically of interest in the context of agriculture and climate change adaptation. By the legal instruments the EU lays the ground for open access to a great number of relevant geospatial datasets, covering European territories. More about the EU and their HVD list in annex 3.

[58] Thematic Research Network on Data and Statistics (TReNDS), *Addressing the challenges of using Earth Observation Data for SDG attainment: Evidence from the Global South and West Africa Region*, 2022, https://static1.squarespace.com/static/5b4f63e14eddec374f416232/t/624334541bf6aa586f04d166/1648571478139/Earth+Observation-3.29.pdf

Google provides global access to a vast array of geospatial data through various services and APIs, but it is important to distinguish between "open data" and "freely accessible" data. Google Maps, Google Earth, and the Google Maps Platform offer extensive geospatial data that includes detailed maps, satellite imagery, and location-based services. These tools are widely accessible to users and developers, enabling the creation of custom map-based applications and the analysis of geographic information. However, Google's geospatial data is not "open" in the strictest sense, as defined by open data principles that promote free use, redistribution, and modification. Some of Google's data are actually open, but still subject to terms of service and licensing agreements or are packed into products that restrict how the data can be used, shared, and modified.

The quality of geospatial data varies significantly depending on the source and the specific geographical coverage. In general, areas with high economic development, especially in the Western hemisphere, tend to more details and frequent updating, thanks to more resources for data collection and processing. This includes high-resolution satellite imagery, extensive ground-based sensor networks, and greater investments in geospatial technologies.

In contrast, the quality and availability of geospatial data in sub-Saharan Africa varies widely. While there is a growing availability of open geospatial data from international agencies, local data can sometimes be outdated or less detailed due to lack of resources. However, several initiatives are aiming to improve the situation, including:

- The Africa GeoPortal, which provides access to a wide range of geospatial data relevant to Africa's development.[59]
- Regional Centres for Mapping of Resources for Development (RCMRD), an African organization that supports geospatial data development and dissemination.[60]

Data type	**Data use**
Meteorological data (temperature, rainfall, storm systems)	• As an early warning system for agricultural communities • Changes in weather pattern for farmers • Changes in the national cropping calendar • Timing and duration of rainfall
Soil data	• Projected crop yields under different warming scenarios • Climate change impact on crop pest populations • Climate-resistant crop varieties and cropping pattern
Earth observation data	• Stream and river levels and quality • Condition of water point sources (wells and taps) • Types of irrigation facilities • Deforestation rates
Forests and land use	• Deforestation rates • Land use change emissions • Alternative livelihoods to reduce pressure on forest exploitation
Other geospatial data	• Demographic and geospatial data overlays to understand vulnerability to flooding, storm surge, and other climate risks

[59] https://www.africageoportal.com/
[60] https://rcmrd.org/en/

Access to high-quality geospatial data in sub-Saharan Africa is also being bolstered by international partnerships and technological advancements, including satellite technology, which are increasingly making detailed earth observation data available globally. Despite the challenges, these efforts are enhancing the availability and quality of geospatial data in the region, supporting its use in sustainable development, environmental conservation, and disaster management among other critical applications.

Although we have found few documented examples of data being used to develop climate mitigation solutions in developing countries, some successes have emerged. One success story from Uganda, a project nominated to the 2023 UN Global Climate Action Award, underscores the importance of high-quality weather data and demonstrates its transformative impact on farmers in these regions. Agriculture, which accounts for over 40 percent of Uganda's GDP and employs 80 percent of the labor force, is primarily rain-fed, making it vulnerable to drought. Yet, Ugandan farmers receive little or no relevant information to help them cope with drought and other climatic stresses. This project developed a climate information system that uses a set of ICT tools to collect, analyze and disseminate adaptation information to farmers. The system includes mobile-phone-based tools for gathering weekly crop and livestock market information from 46 local market outlets and daily weather data from 22 rural weather stations. This information is then disseminated via radio broadcasts, mobile phones, print media and community meetings with local authorities. This information helps more than 100,000 farmers decide what, when, where and how much to sell; gives guidance on low-cost rainwater harvesting techniques; and provides information on drought and flood coping mechanisms.

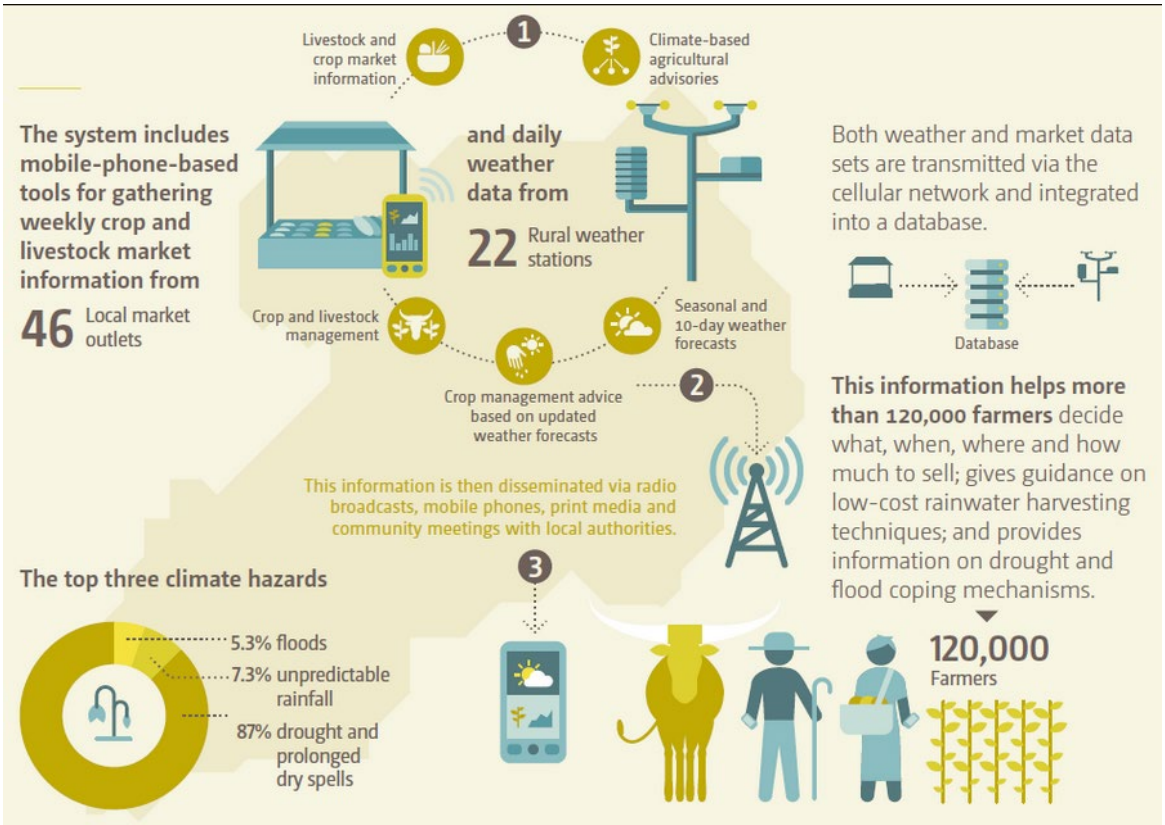


Figure 4: Enabling Farmers to Adapt to Climate Change, The UNFCCC secretariat (UN Climate Change)

As a direct result of the project, crop loss and damage has been reduced by up to 65 percent (USD 474 – 573 per household per year). The project is designed in such a way that it is generic enough to be replicated in other countries. All components of the system are based on widely used open source applications and can be used by any institution or country without the need to obtain licenses and subscription fees.[61]

Some other “success story” candidates are described in the fact boxes below.

Fact box

Predicting drought in Cuba

Participating in a pilot project by the Cuban Meteorological Institute, Cuban farmer Adalberto Martinez decided to take a risk by planting black beans in a region hit hard by droughts. He now receives weather forecasts aiding his risky planting of black beans, enhancing his crop management. This initiative is part of the SIN-Sequía project, funded by Canada's IDRC, aiming to provide accurate, long-term weather predictions using advanced data analysis and AI. It's a significant step towards enabling farmers to adapt to climate variability, showing promise for wider application across the region.[62]

Fact box

Climate forecasts models for small-scale farmers in Latin America

A partnership between the International Center for Tropical Agriculture, the CGIAR Research Program on Climate, Agriculture and Food Security and more than 10 partners including the Colombian Ministry of Agriculture and Rural Development, and the Honduran Secretariat of Agriculture have developed a suite of ICT tools and applications to help farmers in Colombia and Honduras make climate-smart decisions. The tools and applications enable technicians from farmer organizations to collect, analyze, and deliver information that allow farmers to understand variations in seasonal climate conditions, and thus adjust their management practices to cope with them. Guided by this information, an estimated 300 000 farmers now know whether to plant, when to plant and specifically which crops or crop varieties to plant. In addition, they have site-specific information on how much water and agrochemicals to use. This has increased agricultural productivity, food and income security, and allowed for more sustainable farming.[63]

4.3.3 The OpenEPI pilot - APIs for driving local innovation

Given the significant environmental and agricultural challenges confronting sub-Saharan Africa, including those directly impacting farmers, we have curated a collection of APIs tailored to meet the specific needs of local developers within the region. Recognizing the

[61] United Nations Climate Change, *Enabling Farmers to Adapt to Climate Change | Uganda,* https://unfccc.int/climate-action/momentum-for-change/ict-solutions/enabling-farmers-to-adapt-to-climate-change

[62] IDRC, *A project on predicting drought in Cuba*, 20.03.2024, https://idrc-crdi.ca/en/stories/project-predicting-drought-cuba

[63] United Nations Climate Change, *ICTs For Small-Scale Farmers: A Game Changing Approach to Climate Smart Agriculture in Latin America | Colombia and Honduras*, https://unfccc.int/climate-action/momentum-for-change/ict-solutions/icts-for-small-scale-farmers-a-game-changing-approach-to-climate-smart-agriculture-in-latin-america

Fact box

Indigenous peoples and local communities use satellite data to fight deforestation

The Saamaka tribal community, in the heart of Suriname's Amazonian rainforest, are using satellite data from Global Forest Watch, a platform developed by the World Resources Institute (WRI), to combat deforestation. This data provides near-real-time imagery and information that track deforestation and land degradation, which the communities use to detect unauthorized activities on their lands. In the case of the Saamaka, satellite images captured the unauthorized construction of a road through their territory, helping them gather tangible evidence of legal violations. Armed with this data, community representatives traveled to Washington D.C., to present their findings to the Inter-American Commission on Human Rights, aiming to enforce and strengthen their land rights.

Similarly, in other parts of the world like Indonesia and Peru, Indigenous and local communities are using satellite data to monitor illegal logging and deforestation. This technology allows them to document encroachments effectively and take legal action to protect their lands. The accessibility of this data and the ability to monitor large areas remotely reduce the need for costly and risky field patrols, enhancing community safety and the efficiency of conservation efforts. Overall, open satellite data has become an indispensable resource in the global fight for Indigenous and local land rights, enabling these communities to safeguard their territories against external threats.[64]

critical importance of accessible and relevant data in addressing these challenges, our initial release encompasses five distinct APIs: weather, flood, deforestation, soil, and geocoding. Each of the APIs is designed to provide vital, actionable information that can help mitigate the effects of environmental challenges and support sustainable agricultural practices.

The selection of these initial APIs for the OpenEPI proof-of-concept is tailored to address flood scenarios, as also suggested by Norad. Each API was specifically chosen for its direct relevance to flood monitoring and environmental analysis. The data providers for each of these APIs were selected based on their global reach and the absence of alternative APIs, reinforced by endorsements from Norad.

By providing local developers with access to these targeted data sets, we are aiming to empower them to create solutions that not only tackle immediate issues but also contribute to long-term sustainability and resilience in the agricultural sector of sub-Saharan Africa. The selected APIs have the potential to serve as foundational elements for developing applications that can significantly enhance decision-making, operational efficiency, and environmental stewardship among farmers and agricultural stakeholders in the region. In the pre-project and demonstration phase, our focus has been on providing API access to data, addressing the notable scarcity of such APIs among data providers - except for the weather services provided by met.no, which is commonly known and utilized in different digital solutions. For a full-scale operation of OpenEPI, the inclusion of complete, downloadable datasets will be prioritized where applicable, aligning with our commitment to ensure continuous and comprehensive access to data for all user segments.

[64] WRI, Indigenous Peoples and Local Communities Are Using Satellite Data to Fight Deforestation, 20.11.2023, https://www.wri.org/insights/indigenous-peoples-local-communities-use-satellite-data-deforestation?utm_medium=referral+&utm_source=GFWBlog&utm_campaign=GFWBlog%22target=%22_blank

The APIs defined in the pre-project version of OpenEPI are the following:

Weather API

The Weather API was developed to provide timely and accurate meteorological data essential for a wide range of applications, from agricultural planning to disaster response. Recognizing the critical need for up-to-date weather information, this API aims to support decision-making processes that rely heavily on weather conditions. Our decision to base the Weather API on data from met.no was influenced by met.no's reputation for comprehensive and reliable meteorological data. Met.no provides data that adheres to international standards of quality and accessibility, ensuring that users receive the most accurate and actionable weather information available.

The integration of data from met.no into OpenEPI was straightforward due to the high-quality APIs already provided by met.no. Consequently, there is minimal necessity to repackage this data for OpenEPI. The development of this API in the pre-project phase primarily serves as a demonstrative example, illustrating the process and potential of incorporating external data sources effectively in the platform.

For usage examples see our PoC developer portal (www.openepi.io), and the code examples provided there.

Geocoding API

The Geocoding API was developed to deliver precise and reliable geolocation data, vital for a variety of applications ranging from urban planning to logistics management. Recognizing the importance of accurate location information, this API is designed to support decision-making processes dependent on geographical data. The development of this API in the pre-project phase primarily serves as a foundational tool for using our other APIs effectively, illustrating the process and potential of incorporating external data sources.

Initially, we considered directly accessing data from OpenStreetMap datasets. However, to expedite development, we opted to utilize the existing API from photon.komoot.io. For a full-scale implementation of OpenEPI, it would be prudent to fetch data directly from OpenStreetMap datasets to alleviate undue strain on photon's services. The integration of data from photon.komoot.io into OpenEPI was accomplished with ease, leveraging the APIs already available from photon.komoot.io.

For detailed usage examples and code samples, please visit the developer portal.

Flood API

The Flood API was developed to deliver targeted flood warnings based on user-provided geographic coordinates. This capability is particularly important given the absence of such services in existing datasets like those provided by GloFAS

under the Copernicus Emergency Management Service (CEMS). This API aims to provide an efficient, localized flood risk assessment tool to support timely decision-making.

The Flood API utilizes a variety of data sources:

- 30-day Forecasted River Discharge Data: Retrieved daily from the Copernicus Climate Data Store (CDS), this data forms the core of our flood prediction model.
- Upstream Area Data: Accessed from the auxiliary data pages of CEMS, this data helps refine the flood forecasts.
- Return Period Threshold Data: Initially obtained directly from the GloFAS team, this crucial dataset will soon be accessible through the CDS. This data is used to assess the severity of potential flooding.

All the data used spans a global scale and is provided at a resolution of 5° by 5°. It's important to note that these datasets are governed by the CEMS-FLOODS datasets license, which is not a standard open license. In our pre-project phase, we use these data to fine-tune and validate our flood forecasting capabilities.

Before exposing data through our API, some processing is undertaken:

- Detailed Forecast: We calculate simple statistics daily from the forecasted river discharge data, utilizing GloFAS return period threshold data to provide a detailed discharge forecast.
- Summary Forecast: This forecast aggregates the detailed forecast over the 30-day period, similar to the GloFAS Reporting Point structure, which includes the intensity, tendency, and peak timing of a flood for each grid cell.

Additionally, we use upstream area data to exclude any grid cells with an upstream area smaller than 250 km². Our current focus is on a defined geographical part of Africa, spanning from -18.0° to 52.0° longitude and -6.0° to 17.0° latitude.

The Flood API can serve as a basis for applications across various sectors:

- Emergency management: Localized flood warnings assist in the rapid deployment of response measures and evacuation strategies.
- Urban planning and infrastructure development: The API's precise data helps in designing infrastructure resilient to flood risks.
- Research and environmental monitoring: Researchers can use the API to study flood patterns and their impacts on ecosystems.

For detailed usage examples and code samples, please visit the developer portal.

Deforestation API

The Deforestation API is designed to provide comprehensive and aggregated deforestation data covering the period from 2001 to 2022, based on the data supplied by the Global Land Analysis and Discovery (GLAD) laboratory at the University of Maryland in partnership with Global Forest Watch (GFW). This service aggregates vital environmental data over river basin polygons, sourced from HydroSHEDS,

making them crucial for stakeholders engaged in conservation efforts, policy-making, and environmental research.

The Deforestation API utilizes GLAD data, which is freely available under a Creative Commons Attribution 4.0 International License. The river basin polygons, integral for spatial analysis, are provided by HydroSHEDS and are available under a license that permits both non-commercial and commercial use.

This API allows users to query deforestation data either by specific geographic points or within a defined bounding box, ensuring flexibility in data retrieval:

- Point Queries: Users can input latitude and longitude to get deforestation data for that specific point.
- Bounding Box Queries: Users can define a geographical box to retrieve data for all river basins within that area.

Data can be accessed for any period between 2001 and 2022, allowing users to analyze trends over time. This service is documented in the OpenAPI specification, version 3.1.0, detailing endpoints, possible queries, and the structure of the responses, which are formatted in GeoJSON for geographical data representation.

Potential use-cases for this API:

- Environmental monitoring: Researchers can use the API to monitor deforestation rates within specific river basins, aiding in ecological impact assessments.
- Policy development: Policymakers can utilize the data to create informed, data-driven environmental regulations and conservation policies.
- Educational: Academics and educators can leverage detailed data for instructional purposes, helping students understand the dynamics of land use change.

For detailed usage examples and code samples, please visit the developer portal.

Soil API

The Soil API is created to deliver comprehensive and accessible soil data for a diverse range of applications, including agricultural planning, environmental monitoring, and land use management. This API leverages high-quality soil information from ISRIC – World Soil Information, which is celebrated for its extensive and detailed global soil databases.

ISRIC was selected as the preferred provider for our Soil API over other potential sources such as CGIAR, Varda, and FAO for several compelling reasons:

- Quality and Scope of Data: ISRIC provides both soil profiles and high-resolution SoilGrids, renowned for their precise and comprehensive representation of soil nutrients, composition, and texture.
- Programmatic Access: ISRIC stands out for offering programmatic access to its SoilGrids through WCS or WebDAV, facilitating smooth integration into various applications.

- Regular Updates and Open Licensing: ISRIC's data is not only regularly refreshed with the latest soil information but is also freely available under the CC BY 4.0 license, promoting broad usability and adherence to open data standards.

The Soil API exclusively sources its data from ISRIC via its WebDAV service, using SoilGrids data that provide detailed soil information at 250 meter resolution. The API organizes the data into two primary categories: soil type and soil properties.

Soil Type Data
This category captures the dominant soil type at a specified location, covering 30 distinct soil types such as Acrisols, Albeluvisols, and Andosols. Each type offers vital insights into the soil's characteristics and potential applications.

Soil Property Data
This segment provides continuous data on specific soil properties at various depths, including Bulk Density, Cation Exchange Capacity, and Clay content. Data is available across multiple depths, such as 0-5 cm, 5-15 cm, and up to 100-200 cm, with values presented in formats like mean, quantiles (0.05, median, 0.95), and uncertainty, thus delivering a thorough soil profile.

The API processes the data through various raster files, converting raw data into formats that are easy for users to employ. This process includes translating integer values into descriptive soil type names and appending relevant units to soil property data. For more extensive analyses, the API also compiles data to offer summaries of soil types within specified areas, indicating the prevalence of each type.

The methodical approach ensures that the API effectively supports a wide array of applications - from agricultural development and environmental research to urban planning - providing users with precise and actionable soil data.

Potential uses of the Soil API:

- Agricultural developers: Use the API for precision farming initiatives by evaluating soil health and orchestrating sustainable agricultural practices.
- Environmental scientists: Employ the API to track changes in soil quality and assess their ecological impacts, which is crucial for conservation efforts.
- Urban planners: Utilize the API to access soil data essential for sustainable infrastructure development and effective land management.

For detailed usage examples and code samples, please visit the developer portal.

Crop-health API

The Crop Health API is developed to enhance agricultural practices through early disease detection in crops using advanced image recognition technologies. This API utilizes a vast array of images from the Harvard Dataverse, which includes detailed datasets of crop leaves taken using mobile devices. These images cover a diverse

range of crops such as maize, cassava, beans, cocoa, and bananas, pivotal for agricultural activities in sub-Saharan Africa.

Comprehensive data collection
The Harvard Dataverse provides approximately 120,000 labeled images across different crop types, crucial for training our sophisticated machine learning models. These datasets were meticulously compiled by students and researchers from various universities across Africa, ensuring a rich and diverse data pool for accurate model training. All datasets use the following license: https://creativecommons.org/publicdomain/zero/1.0/

Advanced machine learning models
Our Crop Health API leverages a pre-trained ResNet model, optimized for early disease detection through further training on the aforementioned labeled crop images. We used the open-source PyTorch framework for model training due to its solid documentation and community support. We have developed three distinct models to cater to various needs:

- **Binary Model:** Determines the health of the crop, categorizing each image simply as healthy or diseased.
- **Single Disease Model**: Identifies if the crop is healthy or specifies one of twelve possible diseases, enhancing the granularity of our diagnostics.
- **Multi-Crop Health Model**: Alongside disease detection, this model classifies the type of crop when healthy, providing additional insights useful for agricultural applications.

Deployment and Accessibility
The models are deployed using TorchServe, a framework designed for efficient model serving. This setup facilitates easy access to our models via RESTful APIs, ensuring that users can effortlessly integrate our services into their applications. For more detailed usage and integration techniques, developers can refer to the code examples provided on the OpenEPI developer portal.

Potential uses of the Crop-health API:

- Agricultural advisory: Offering real-time advice to farmers on crop health, potentially preventing widespread disease outbreaks.
- Research: Enabling detailed studies on crop disease patterns and their impacts, which can inform future agricultural strategies.
- Policy making: Assisting in the formulation of targeted agricultural policies based on accurate, up-to-date field data.

Future Enhancements
Looking ahead, to expand the capabilities of the Crop Health API more datasets should be incorporated, and the models should be refined to include more crop types and diseases.

For detailed usage examples and code samples, please visit the developer portal.

4.3.4 Issues that need to be further addressed

For OpenEPI to achieve its intended outcomes and impact, it is crucial that there is a robust alignment between the data provided on the platform and the specific needs and demands

from its users. First, the data provided must target the environmental and agricultural challenges prevalent in sub-Saharan Africa and other similar developing regions. In the further work towards realizing the platform, it will be essential to maintain a vigilant focus on several key areas:

- **Bridging data gaps:** OpenEPI should identify and address data gaps and deficiencies in the datasets available for local innovation. Given the limitations of local data, especially in Africa, there's a growing recognition of the value of global datasets, which can sometimes offer superior quality and make developers less dependent on local data. OpenEPI should further explore the potential of utilizing earth observation data and other global data sources as alternatives or complements to local datasets, such as soil data, to better support environmental and agricultural decision-making processes.
- **Engaging with local experts and communities:** To ensure the relevance and effectiveness of the data and APIs provided, OpenEPI should actively engage with the local innovators, as well as potential users like local agricultural experts, research institutions, and farmer cooperatives. Gaining a deep understanding of the distinct data needs and challenges faced across various regions is essential. The data needs will most certainly differ between countries and regions, corresponding to varying socio-demographic conditions, political environment, digital capabilities and digital competence among end-users. This knowledge will guide the customization of APIs and data repositories to meet local requirements, thereby ensuring the utility and applicability of the platform's resources.
- **Push for more open data:** Eventually, OpenEPI should use its position to push producers and distributors of nature, environmental and climate data to release their data according to the open policy and the FAIR principles. Advocacy for open data policies that support the sharing of data will in the long term supposedly amplify the platform’s reach and effectiveness.
- **Monitoring and evaluation**: Implementing a robust monitoring and evaluation framework to track the usage and impact of using OpenEPI's APIs and other services will be vital. This includes efficient feedback mechanisms and the development of relevant metrics, and the capacity and methods to follow up on them. A monitoring and evaluation framework with robust mechanisms will lay the ground for continuous improvement, demonstrate value to stakeholders, and justify further investment in the platform. We suggest a specific program for developing metrics and measuring effects from locally developed solutions, addressing climate change related topics. As part of the implementation of such a program, OpenEPI could require developers to report on the perceived effects and impact for local end users, by their use of OpenEPI data and services in their digital solutions.

By improving data quality and availability, leveraging global data resources, and engaging with local expertise, OpenEPI can facilitate the creation of innovative solutions that address the pressing climate challenges facing the region.

4.4 Capability building services

The diverse user base of OpenEPI, encompassing local developers, startups, and various stakeholders across sub-Saharan Africa, will have varied needs for support and assistance in

leveraging the platform to its fullest potential and deriving tangible benefits from it. Recognizing and addressing these differential needs is key to fostering a productive and inclusive ecosystem around OpenEPI.

For local developers and startups, the assistance might include detailed documentation on how to access and use the platform's data, as well as technical support for integrating this data into their applications. A solid onboarding process for developers is crucial, hence our emphasis on creating a developer portal. The portal will serve as a central hub for essential resources, such as documentation, code samples, API key management, and community forums, facilitating easy access to the tools and information developers need. Incorporating feedback mechanisms within the portal allows for the continuous refinement of the onboarding process, ensuring it remains responsive to the evolving needs and experiences of the user community. Enhancing this foundational support, mentorship programs and workshops could further develop the users' capabilities in data analysis and application development.

Beyond technical assistance, many users will need guidance on how to commercialize their solutions effectively. This could involve training in business model development, marketing strategies, and navigating regulatory environments. Facilitating networking opportunities with potential investors, partners, and customers is also crucial for helping these innovators scale their solutions and achieve commercial success. Furthermore, some users might require help in understanding the broader implications of the data from OpenEPI. Educational resources on climate change, different subject matters like weather, soil and hydrology, environmental policies, and sustainable development practices can empower users to align their projects more closely with global sustainability goals and local needs.

Certain users of OpenEPI might find themselves in need of additional support and possibly financial assistance to secure the essential technical equipment required to fully utilize the platform. This encompasses both software, such as specialized data analysis tools and development environments, and hardware, which could range from basic computing devices to more advanced servers or environmental sensors. Access to these resources is crucial for users to effectively engage with OpenEPI and develop innovative solutions.

To cater to these diverse needs, users should have access to a multi-tiered support system that includes self-service resources, such as the developer portal and community-driven support mechanisms. OpenEPI is not intended to directly participate in community-driven support activities. Instead, we suggest OpenEPI to leverage a broad network of varied stakeholders to offer support indirectly.

By cooperating with educational institutions, like for instance Carnegie Mellon University in Africa, OpenEPI may build up and support a user base, leveraging the university's expertise to offer specialized training and development programs. Such efforts will prepare developers to use OpenEPI effectively and help to build an ecosystem around the platform that supports innovation of innovative solutions in the climate and nature domains.

Furthermore, collaboration with national development aid agencies like NORAD and different UN bodies with a strong local presence and deep understanding of the local context, can play a role in building capacity among potential users of OpenEPI. These bodies are uniquely

positioned to provide insights into the specific environmental challenges, technological needs, political conditions, and opportunities in various regions. Through their extensive networks and resources, these partners can facilitate access to funding, mentorship, and training programs for local developers and innovators. Importantly, collaboration with these organizations can also enhance efforts to recruit female developers and promote gender diversity in the tech sector. By leveraging specific initiatives and programs aimed at empowering women in STEM fields, OpenEPI can attract and support female talent, enriching the platform with diverse perspectives and skills. The implementation of joint capacity-building programs focused on equipping local developers, especially women, with the skills and knowledge to leverage OpenEPI for environmental projects can further foster innovation and capacity building in a manner that is sensitive to local contexts and committed to promoting diversity within the global tech community.

Finally, OpenEPI should consider forming partnerships with technology providers to guarantee that users have the necessary equipment and software tools to fully leverage the platform. Collaborating with these providers, OpenEPI could develop grant opportunities and subsidy programs, ensuring every user has access to the resources needed to utilize OpenEPI's capabilities to their fullest extent.

4.5 Community building

Throughout this study, we've observed that numerous open data platforms naturally evolved from established communities of researchers or domain experts, essentially having a built-in user base from the outset. Unlike these platforms, OpenEPI faces a unique challenge of having to cultivate its user community from the ground up. For OpenEPI to flourish and achieve long-term success, it is imperative to not only attract but also nurture a sense of belonging and cultivate a community spirit among its users. This necessitates that community-building initiatives be placed at the core of its strategy, aimed at encouraging user engagement and fostering active participation. Engaging users in this way is not just about increasing numbers; it's about creating a vibrant ecosystem where users feel motivated to participate and contribute, ensuring the platform's growth and sustainability.

For OpenEPI being able to scale as a service, the need for users to connect with one another is fundamental. Through mutual interaction, they can share insights, learn from each other's experiences, and draw inspiration. This could involve exchanging innovative ideas on data application, discussing best practices, or showcasing successful use cases.

OpenEPI may strengthen this community building digitally and by organizing real-life meetings. On the digital front, incorporating social media features such as forums, chat rooms and user groups can create vibrant spaces for users to engage and share. Beyond the digital realm, OpenEPI can organize real-world events like hackathons, which bring developers together to creatively use the platform, or annual conferences dedicated to highlighting the platform's activities, achievements, and inspiring further innovation and usage. OpenEPI can organize real world events together with local stakeholders, such as Carnegie Mellon University, local innovation hubs, government representatives and UN bodies. These community-building initiatives are pivotal in nurturing a dynamic, collaborative ecosystem around OpenEPI, ensuring its growth and relevance.

We suggest Norad to fund a demand side development program, as a separate initiative running alongside OpenEPI - to support efforts in community building in general, stimulating the demand for knowledge and data related to climate change in specific.

4.6 Market access and secondary user's needs

For OpenEPI to achieve its intended impact, it is essential that the outputs created through the platform lead to tangible outcomes in the real world. This means that products and solutions developed using OpenEPI, like flood prediction apps or soil condition analytics, must not only meet an existing market need but also be commercially viable. There must be a demand for these products in sub-Saharan countries, suggesting the presence of willing buyers, and the necessary technical infrastructure must be in place to support their use and dissemination.

This entails that OpenEPI must be preoccupied not just with the creation of innovative tools and applications from use of the platform, but also with ensuring they can be integrated and utilized effectively within the target markets. Bridging this gap between innovation and practical application will be crucial for OpenEPI's success, requiring a keen understanding of both the technological landscape and the market dynamics. To effectively distribute the outputs generated by OpenEPI, it is essential to develop a comprehensive dissemination strategy:

- Firstly, startups and developers often need support in terms of funding, business model development, and market access to bring their solutions to fruition and scale. Many users of the platform will need guidance on how to commercialize their solutions effectively. Facilitating networking opportunities with potential investors, partners, and customers is also crucial for helping these innovators scale their solutions and achieve commercial success. An essential strategy in this endeavor is to establish venues within local communities that facilitate direct interactions between developers and prospective buyers, promoting the exchange of ideas and fostering collaborations that could lead to the successful adoption of new data-driven solutions. In addition to leveraging local partnerships, OpenEPI should explore collaborations with Norad, and other international development aid organizations already present in Africa, in order to facilitate connections between developers and startups and potential user groups in the agricultural sector and beyond, such as urban planners, environmental agencies, and community developers. By aligning with programs that are already supported, OpenEPI can tap into established channels for technology dissemination and adoption, ensuring a wider reach and potentially greater impact of its solutions.
- Secondly, a key component of the strategy could be to develop specialized training programs for end-users, such as farmers, who stand to benefit significantly from the applications developed using OpenEPI data. These training programs could be designed and delivered in partnership with local agricultural organizations, Norad or other development aid organizations, which can play a crucial role in content validation and ensure the training is relevant and accessible to the local farming communities.
- Additionally, OpenEPI must invest in awareness and marketing activities, alone and in collaboration with stakeholders like UNDP and Digital Public Goods Alliance (DPGA),

showcasing the successful implementations and tangible benefits of applications built around data and services from the platform. Users of the platform must be encouraged to share best practice examples with other developers potentially using OpenEPI. By marketing successful applications, OpenEPI will help to build supply and demand for data-driven climate mitigation services.

4.7 Use cases

4.7.1 Flood warning

In the pre-project phase, we have developed a use-case demonstrator for a flood warning application. This initiative was undertaken in close collaboration with Norad and is intended to serve as one of our primary use cases. This demonstrator not only showcases the practical application of our APIs and client libraries but also provides a concrete example of how to effectively utilize the datasets available in the OpenEPI data catalog.

The primary objectives for creating this demonstrator are multifaceted:

- **Educational utility**: It serves as a practical example to illustrate how users can leverage the data within the OpenEPI data catalog, enhancing understanding and ease of use.
- **Reference implementation**: The demonstrator is built according to our style guide, providing a reference model that embodies best practices for software development within the OpenEPI framework.
- **Showcasing potential**: It highlights the value and potential applications of datasets hosted on OpenEPI, demonstrating what can be achieved with the resources available through the platform.

However, it is important to note that the functionality of the demonstrator is intentionally limited. Its primary role is to illustrate the potential applications of our platform; it is not designed as a fully functional application for end-user deployment.

The functionality included in this application encompasses:

- **User onboarding**: Includes user account creation and password setting, offering a straightforward entry point for new users.
- **Site addition**: Users can add specific sites using coordinates fetched from their phone’s GPS, facilitating location-specific interactions.
- **Data retrieval**: The application can retrieve and display flood warnings and weather information for the user-defined locations, providing relevant and timely data.

This demonstrator is an important step in demonstrating the practical application of OpenEPI's capabilities and serves as a foundation for further development and refinement of user-focused applications.

The following screens show some of the capabilities of the demonstrator.

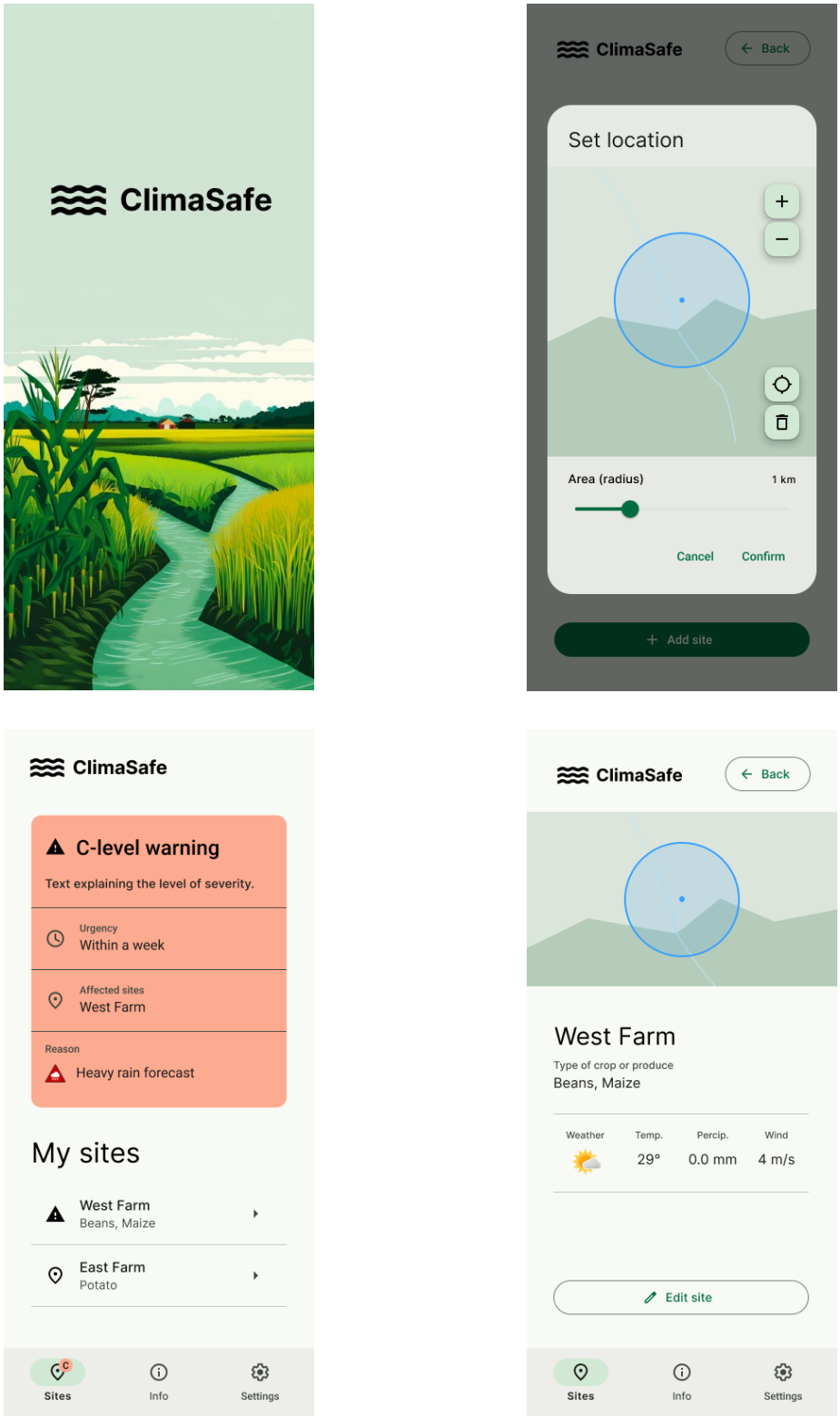


Figure 5. Screens from the OpenEPI based flood warning demonstrator.

4.7.2 Developer portal

In the early stages of the pre-project, the OpenEPI team recognized the need for an API hub as a central platform for presenting and accessing data. This initial concept was envisioned

as a straightforward way to facilitate user interaction with the data available through OpenEPI.

However, during the Rwanda workshops, it became evident that the needs were more complex than initially anticipated. Feedback indicated that a simple API hub would not suffice if users were to fully understand and utilize the APIs and the data's potential. This realization led to a pivotal decision, made in collaboration with Norad, to expand the concept into a more robust developer portal.

The developer portal was designed with several key objectives in mind:

- **Central access point**: It serves as the primary gateway for users to start engaging with data from OpenEPI.
- **Comprehensive data information**: The portal provides detailed descriptions of each dataset hosted on OpenEPI, including any processing that the OpenEPI team has applied.
- **Support resources**: It includes information about available client libraries, enhancing user ability to effectively utilize the datasets.

While we have not yet had the opportunity to conduct extensive testing of the developer portal with end users, we recognize the importance of such feedback. For a full implementation of OpenEPI, the developer portal's impact on user engagement and satisfaction needs thorough evaluation.

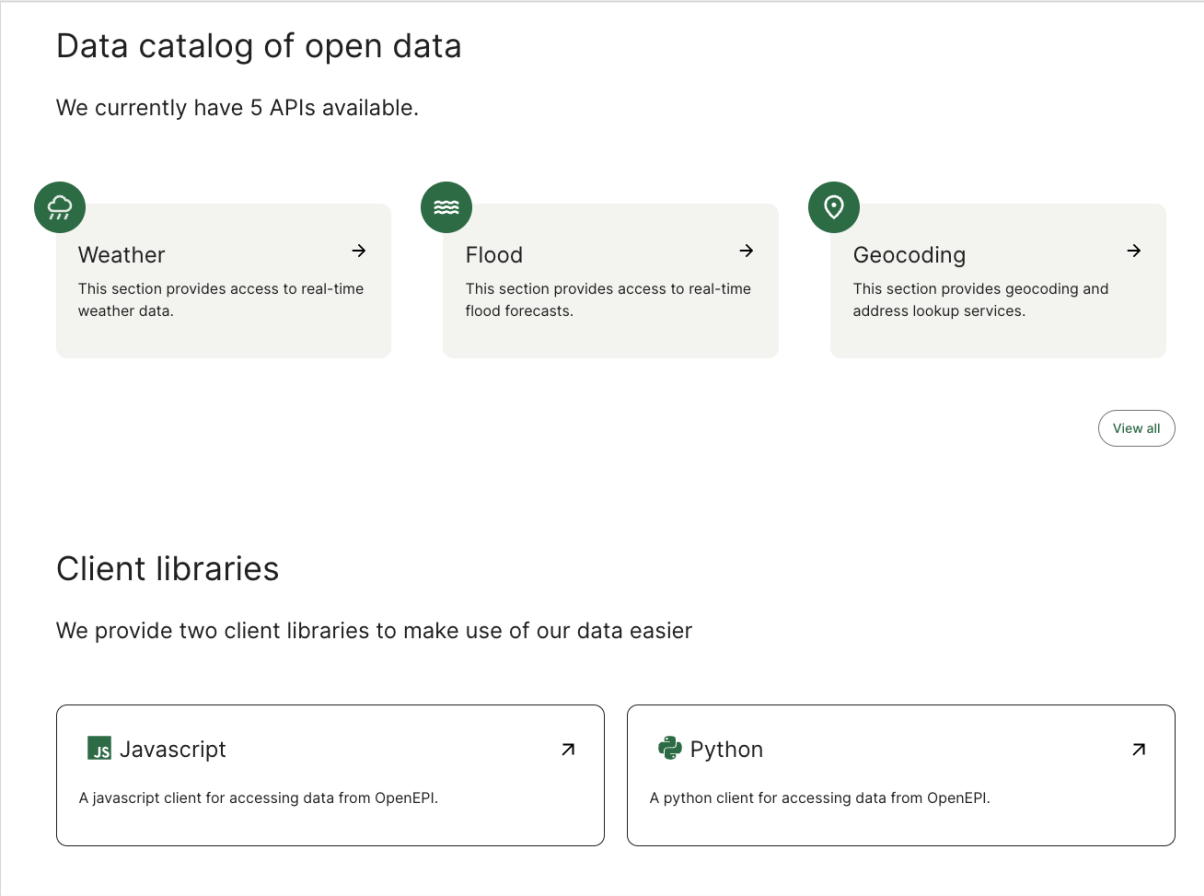


Figure 6. Key functions of the developer portal.

Looking forward, several enhancements to the developer portal are proposed for a comprehensive implementation of the project:

- **Enhanced user onboarding**: Implementing a system for user registration and management of client credentials to streamline access and usage.
- **Feedback mechanisms**: Integrating a feedback solution that would allow developers to report bugs and request new features, thereby fostering an interactive and responsive user community.

These proposed expansions are aimed at not just maintaining the functionality of the developer portal but enhancing its role as a critical resource for users to explore and harness the full capabilities of OpenEPI.

4.7.4 CAR - OpenEPI as a competency center for open architecture

In the spring of 2024, OpenEPI engaged in a series of meetings and workshops with representatives from Brazil's Rural Environmental Registry System (CAR). These interactions were designed to assess the feasibility of a collaboration between Norway and Brazil and to gain a better understanding of Brazil's forest management practices.

A technical review of CAR's current systems revealed that although they were established according to the highest standards available at their inception, there is now a significant need for technological updates and improvements. Specifically, the CAR system was not originally designed to handle the extensive usage it sees today. For instance, it relies on a single, centralized database that is crucial for the system's functionality.

With support from Norad, OpenEPI spearheaded the initial stages of modernization by hosting a preliminary workshop. This session aimed to establish a foundation for ongoing efforts to develop a contemporary and efficient architectural framework.

The significance of open principles is not widely recognized among developers and stakeholders. Naturally inclined to solve problems, developers often focus on addressing immediate challenges using whatever means are available. However, to create systems that are adaptable to future changes and avoid both vendor and organizational lock-ins, it is essential to continuously consider these principles. This consideration should inform design decisions, technology tool selections, and organizational strategies to ensure flexibility and sustainability in system development.

It is apparent that establishing an architectural advisory service will be of great value. Such a service would guide the architectural development with a focus on open principles, addressing the modern needs of software infrastructure. This would involve not only embracing open-source software to enhance flexibility and innovation but also designing systems to avoid vendor lock-ins and improve data and metadata accessibility.

Overall, the proposed advisory service should be designed to empower organizations to develop sustainable, adaptable, and forward-thinking technology ecosystems, leveraging the principles of openness to drive innovation. However, delivering such a service would require OpenEPI to cultivate specific expertise within its organization. It is crucial that the advisors

involved are not only experienced in architectural design but also deeply familiar with open principles. Further, given the demanding and complex nature of this work, it would necessitate the efforts of more than one person to ensure sustainability and prevent fatigue.

4.8 Insights from the Kigali Workshop

In November 2023, we organized a workshop in Kigali, Rwanda, with a selected group of students and professional, local developers. The workshop aimed to gain a deeper insight into typical users' needs, understand how the platform and its related services could meet those needs, identify the key factors for success, and determine what core functionalities should be prioritized for delivery.

4.8.1 Preparations

To investigate the technical and functional feasibility we developed a use case that we invited participants to experiment with. The use case involved four datasets - weather, flood, geocoding, and deforestation - accessible through a very simplified API hub. Additionally, we provided access to a more basic foundational layer such as improved quality and orchestration of satellite imagery. We also developed and presented in the workshop a demonstrator and preliminary sketches of a flood warning application, to set the stage for a constructive exchange of ideas and feedback. This use case was selected because we thought it would have high relevance across sectors and geographic areas.

Around 40 attendees were invited to the workshop, which spanned two days. Most of the participants were data science students from Carnegie Mellon University in Kigali. However, around 15 percent of the attendees were professional developers affiliated with local startups or more established firms. Just under 50 percent of the participants were women. The event involved an introduction to OpenEPI, a concept ideation, and the prototype development.

4.8.2 Stakeholder interactions

In conjunction with the workshop in Kigali, we arranged several meetings with local institutions, communities, and individuals, each offering valuable insights into the local tech ecosystem of Rwanda. We engaged in collaborative dialogues with Carnegie Mellon University Africa (CMU-Africa), who offer master's degrees in IT, Electrical and Computer Engineering, and Engineering Artificial Intelligence, focusing on creating a pool of high-quality engineering talent to address the continent's development needs.[65] Conversations with CMU centered on exploring the integration and practical uses of OpenEPI within their projects. These discussions, involving both faculty members and the director of CMU, sought to uncover potential paths for future collaboration and broaden the understanding of our initiative. A focal point of these engagements was the contemplation of a summer program. This program would enable CMU students to engage directly with the further development of OpenEPI, providing a hands-on experience that benefits both the students and the advancement of OpenEPI. These dialogues highlighted the mutual interest in leveraging our planned OpenEPI-capabilities to enrich CMU's projects, setting the stage for a collaborative

[65] CMU-Africa was established in 2011 through a partnership between Carnegie Mellon University and the Government of Rwanda. It represents a global extension of Carnegie Mellon University and is part of its College of Engineering, which is highly ranked, https://www.africa.engineering.cmu.edu/about/index.html

partnership that harnesses the strengths of both entities. However, due to budget constraints in the pre-project phase and after thorough discussions with Norad, the proposed summer program was temporarily shelved.

We also had interactions with the Rwanda Space Agency (RSA). RSA is tasked with coordinating the nation's space activities.[66] By focusing on critical areas such as agriculture, urbanization, and disaster management, RSA seeks to implement space technology solutions that can significantly contribute to the country's socio-economic development.[67] The talks with RSA were focused on understanding their specific data needs and how OpenEPI can contribute in any manner.

Last, we also paid a visit to Rwanda Mountain Tea, a privately owned holding company engaged in tea production. The company stands out for its innovative integration of Internet of Things (IoT) technologies, particularly in refining the fermentation process of black tea. One notable innovation is their development of a web-based interface that enables real-time monitoring of the tea fermentation. This technological advancement not only improves the accuracy and reliability of the tea fermentation process but also enables the factory to adapt more quickly to changes in environmental conditions and to maintain high-quality standards in their tea production.[68] Our exploration of Rwanda Mountain Tea provided an interesting look into the advanced technological applications within the agriculture and manufacturing sectors. It underscored the importance of adopting data-driven solutions tailored to the specific needs of the local environment.

4.8.3 Key learning points from the workshop

The workshop unveiled several insights, shaping our understanding and approach to technology deployment in diverse contexts. The visit illuminated the importance of developing and implementing technology solutions that are tailored to meet local needs, ensuring that they are both relevant and effective in addressing the unique challenges faced by the agricultural community.

Usage areas/application of data
The segment of the workshop focused on application areas for OpenEPI datasets proved very valuable. The discussion revealed a marked preference for innovative solutions aimed at mitigating flood impacts. Among the proposed solutions were advanced methods for optimal route planning to ensure efficient produce transportation and the development of a community-focused flood warning system. These ideas reflected a keen interest in leveraging technology to address critical challenges.

Functionality
Regarding the functionality of the OpenEPI platform, we garnered valuable feedback from the

[66] https://www.iafastro.org/membership/all-members/rwanda-space-agency.html
[67] https://www.africa-press.net/rwanda/policy/rwanda-space-agency-official-discusses-priorities-as-body-steps-up-awareness-effort
[68] Natukunda, Abraham and Adomako, Kwasi, Automation of black tea fermentation process leveraging the Internet of Things in Rwanda: Case of Rwanda Mountain Tea Factory Limited, 2017, Carnegie Mellon University, https://www.researchgate.net/publication/328531729_Automation_of_black_tea_fermentation_process_leveraging_the_Internet_of_Things_in_Rwanda_Case_of_Rwanda_Mountain_Tea_Factory_Limited

workshop participants, particularly from students who expressed a strong desire for access to complete datasets rather than solely relying on curated APIs. Given that a significant portion of these students are pursuing studies in data science, their interest in obtaining raw data is primarily driven by the need to utilize such information in machine learning (ML) projects, specifically as training data. This insight has been noted, and we are committed to consider the provision of unprocessed datasets as a crucial component of our ongoing project enhancements. More generally, this feedback emphasizes the necessity for OpenEPI to continually evolve and introduce innovative functionalities. As we move forward, the ability to cater to the sophisticated needs of our users by facilitating access to comprehensive and raw datasets will be a key factor in fostering innovation and supporting the development of cutting-edge ML services. This approach not only aligns with the professional aspirations of our users but also reinforces the significance of OpenEPI in contributing to the advancement of data-driven solutions in the years ahead.

Challenges and obstacles
One of the most significant findings from our Rwanda visit was the evident gap in the availability of current and accurate weather data. This deficiency poses a challenge to effective planning and decision-making in local agricultural practices, directly impacting productivity and sustainability. The availability of timely and precise weather information is critical for farmers to make informed decisions regarding crop management and disaster preparedness. Moreover, there was a noticeable reluctance to rely on international satellite data among local stakeholders, with a pronounced preference for data sourced from within Rwanda. This emphasizes the paramount importance of data quality and trustworthiness when developing technological solutions. The reliance on local data sources suggests a need for autonomy and relevance in the data utilized for local critical decision-making processes.

Additionally, the workshop highlighted substantial weaknesses in Rwanda's technical infrastructure, particularly in rural areas. The limited access to smartphones and the internet among the local farming community presents a significant barrier to the adoption of advanced technological solutions, such as innovative flood management systems. This digital divide not only restricts the reach of such solutions but also underscores the disparities in technological access between different population segments. The participants underscored the need for infrastructure improvements and the exploration of alternative, low-tech approaches to ensure the effectiveness and reach of these initiatives, thereby bridging the gap between technological advancements and real-world applicability.

To sum up, the workshop gave the following key learning point into the critical needs and opportunities that OpenEPI must address if it is to have the desired outputs, outcomes and impact:

- **The imperative of local context:** The realization that our Oslo-based team faced challenges in designing end-user solutions for Rwanda highlighted the importance of grounding technological solutions in the local context. This insight stresses the value of collaborative design processes that incorporate local perspectives and expertise.
- **Demand for local data:** A consistent theme across our interactions was the demand for accessible, high-quality local data. This is seen as crucial for the development of relevant and impactful solutions.

- **Trust and data quality:** A discernible apprehension towards international satellite data was observed, with a clear preference for reliable local data sources. This preference underscores the need for data quality and trustworthiness in developing technological solutions.
- **Developer-centric approach:** The feedback received placed a renewed emphasis on developers as the primary users of OpenEPI, acknowledging their pivotal role in translating available data into user solutions that resonate with local needs.
- **Strategic local partnerships:** The necessity for partnerships with local entities well-versed in the socio-political and cultural fabric of Rwanda was underscored. Such collaborations are essential for the effective dissemination of technology.
- **Varied understandings of open source:** The workshop highlighted diverse understandings of open source principles, underscoring the importance of clear communication and shared understanding in collaborative ventures. They also revealed a common misconception that open source projects cannot generate revenue, emphasizing the need to educate on the viable economic models that allow for profitability in open source software.

4.9 Concluding thoughts on building a collaborative ecosystem

During a three-day stop on our journey to Rwanda, project members also joined the annual meeting in the Digital Public Goods Alliance (DPGA), in Addis Ababa, Ethiopia. The Open Earth Platform Initiative was launched as part of the official meeting program, in an enthusiastic atmosphere. More importantly, the concept and our plans were discussed with members of the alliance, securing relevant input as well as support for the concept.

The workshop and interactions in Kigali, as well as our discussions in Addis Ababa, our visit to Brazil, and our interviews with established open data platform owners and managers, has provided pivotal insights into the critical role of understanding user needs at both the primary and secondary levels. The interviews with stakeholders of global open data platforms, emphasized unanimously the critical importance of engaging users in meeting the demand-side requirements. Our workshop sessions highlighted the need for a developer portal as an essential step toward fostering a collaborative ecosystem where technology meets local needs. This strategy is important to enhance the capacity for innovation, ensuring solutions are not just technologically advanced but also culturally attuned and impactful for the communities they are designed to serve. Moreover, prioritizing strategic partnerships and the dissemination of or interaction with high-quality local data is essential for OpenEPI to establish itself as an inclusive, effective, and trustworthy platform. These efforts, coupled with the importance of enhancing capability building and community engagement, will help in addressing not only the immediate technological and informational needs of its users but to also inspire and support the next generation of innovators.

In the paper titled "*Building open government data platform ecosystems: A dynamic development approach that engages users from the start*," by Andreas Hein and colleagues[69], an innovative model for crafting open data platform ecosystems is presented,

[69] Hein, Andreas and Martin Engert, Sunghan Ryu, Norman Schaffer, Sebastian Hermes, Helmut Krcmar, *Building open government data platform ecosystems: A dynamic development approach that engages users from the start*, Government Information Quarterly,

with a strong focus on fostering high levels of user engagement. This model diverges from the norm by involving ecosystem actors from the beginning, ensuring that the platform's development is highly relevant and practically useful for its users. The emphasis on **context specificity**, **continuous adaptation**, and **organic expansion** provides a comprehensive framework for building successful open data platforms, offering significant insights particularly applicable to our mission:

- The first learning, **context specificity**, emphasizes the importance of tailoring the platform to the unique needs and characteristics of the users and their local ecosystems. This involves close collaboration with local stakeholders to understand the unique environmental, socio-economic, and technological landscapes of the region. By focusing on the specific needs and data types that are most relevant to local developers and communities, OpenEPI can ensure high engagement and effectiveness of climate mitigation solutions. Consequently, OpenEPI will be a resource for developers building solutions for farmers in different countries, different natural environments, and different socio-demographic contexts. How OpenEPI will be able to support those obviously differentiated end-user needs, remains to be investigated in more detail.
- **Continuous adaptation** highlights the need for OpenEPI to evolve in response to the dynamic nature of its ecosystem. This means regularly updating and refining the platform's data offerings and technical capabilities based on feedback from local developers, emerging new technologies and data sources, changing climate patterns, and emerging environmental challenges. Adapting to these shifts ensures that OpenEPI remains a valuable and relevant resource for its users, facilitating the development of innovative and impactful climate solutions.
- **Organic expansion** suggests that OpenEPI can achieve growth and increased engagement by leveraging existing networks and relationships within the domains and geographical areas it operates. By adopting strategies that allow the platform to naturally extend its reach through the community and stakeholder connections, OpenEPI can foster a more integrated and collaborative approach to climate mitigation across the region.

Incorporating these strategies - context specificity, continuous adaptation, and organic expansion - into the development and implementation of OpenEPI will not only enhance user engagement but also maximize the platform's contribution to climate mitigation efforts in sub-Saharan Africa and other areas in the developing world. By tailoring its approach to the unique needs of the region, remaining flexible to changes, and growing organically through local ecosystems, we think OpenEPI can significantly impact the development of sustainable and effective climate mitigation solutions.

Volume 40, Issue 4, 2023, https://doi.org/10.1016/j.giq.2023.101878.

5. Technological feasibility

5.1 Embracing openness

In conducting this feasibility study, our consortium has embraced an approach defined by openness. We start with the foundational belief that openness significantly benefits citizens, businesses, and public administrations, fostering cooperation and collaboration. Echoing the views of organizations like Norad, we assert that open source, open data, open standards, and open access yield advantages across diverse sectors including health, food security, education, climate, intelligent transport systems, smart cities, and beyond. This chapter, "Embracing openness," explores how these principles are important in our proof-of-concept solutions and the broad impacts of such an approach.

5.1.1 Open data

The success of OpenEPI is closely tied to the availability to, and utilization of, open digital products and services. At the core of the initiative stands open data. To hit its target, OpenEPI relies heavily on the existence of open data, relevant for climate change adaptation and resilience. OpenEPI is therefore committed to promoting the use of open data and contributing positively to the development of openness in the actual domains. The following considerations are relevant for the effectiveness and impact of OpenEPI.

Transparency and accountability	Open data fosters greater transparency and can help hold governments, organizations, and corporations accountable. By making data publicly accessible, stakeholders, including the public, can scrutinize actions and decisions, leading to improved governance and business practices
Innovation and collaboration	The availability of open data encourages innovation by providing researchers, developers, and entrepreneurs with the resources needed to develop new technologies, applications, and services. It facilitates collaboration across different sectors and disciplines, potentially speeding up technological advancements and scientific breakthroughs.
Economic benefits	Open data can stimulate economic growth and competitiveness. Businesses can use open data to develop new products, improve services, and enter new markets.
Empowerment and engagement	Open data empowers citizens by providing them with the information needed to make informed decisions and to engage more effectively with their communities and governments. This can enhance public participation in decision-making processes and increase civic engagement.
Education and research	Open data is a valuable resource for educators and researchers, providing a rich basis for academic studies, papers, and teaching materials. Access to diverse datasets allows for more comprehensive research opportunities and can improve the quality of educational materials and methodologies.

Enhanced environmental monitoring	Open access to climate, nature, and earth data allows for more comprehensive monitoring of environmental changes and trends. This data can track deforestation rates, monitor air and water quality, and observe climate changes, providing essential information for environmental conservation and mitigation strategies.
Disaster preparedness and response	Open data related to weather patterns, geological surveys, and historical disaster tracking improves disaster response strategies. By making this data available, governments and organizations can plan more effectively for natural disasters, potentially reducing impacts and improving recovery times

These considerations and benefits demonstrate how open access to climate, nature, and earth data can serve as a catalyst for environmental protection, sustainable development, and an informed public discourse on global environmental challenges.

5.1.2 Open source software

Open source software (OSS) is pivotal in shaping the foundation and ongoing development of OpenEPI. This approach not only fosters innovation and collaboration but also ensures transparency and security across the evolving platform. By fully integrating open source principles, OpenEPI exemplifies the symbiosis between open data and open source software, enhancing both the utility and accessibility of data.

Encourages reuse and involvement
OpenEPI's commitment to using open source software enables the reuse of robust, community-vetted tools and frameworks, significantly accelerating development while reducing costs. This openness invites a broader community of developers to contribute to and expand upon the initiative in the years to come, enhancing the platform's capabilities and ensuring it meets a wide range of user needs. Except for the infrastructure code during the pre-project phase, all code developed by the OpenEPI team is released as open source. This practice actively fosters a culture of sharing and continuous improvement.

Security and vulnerability in OSS
While the transparent nature of open source software allows potential vulnerabilities to be visible to all, including potential attackers, it also means that these can be more swiftly identified and addressed by the community.

OSS complements open data
Utilizing open source software in conjunction with an open architecture and championing the open data movement ensures that both the software and the data are equally open and accessible. This cohesive approach not only enhances transparency and usability but also provides visibility throughout the entire value chain, from data collection to application development. Such openness facilitates a comprehensive view of the development process, where stakeholders can review changes, track progress, and contribute to the codebase. This level of transparency fosters a collaborative environment where innovations are shared and improved upon publicly, building trust among users and ensuring that the tools and data align with the needs and values of the community. This, in turn, simplifies the process for

researchers, developers, and analysts to employ these resources for a variety of applications.

Mitigating vendor lock-in with OSS
Utilizing open source software (OSS) is also a way to avoid vendor lock-in, which can constrain flexibility and increase dependency on specific suppliers. By leveraging OSS, OpenEPI can maintain control over its technological infrastructure, making it easier to adapt and scale systems without being tied to the proprietary technologies and pricing structures of single vendors. This independence allows for more agile decision-making and innovation, enhancing OpenEPI's ability to respond to new challenges and opportunities efficiently. Adopting OSS not only supports a sustainable technological ecosystem but also aligns with best practices for ensuring long-term operational flexibility and cost efficiency.

We suggest OpenEPI should fully embrace the principles of open source software in a full realization of the platform. By utilizing existing open source tools and libraries for building infrastructure and ensuring that any custom-developed software is also released under an open source license, OpenEPI can foster a culture of transparency and collaboration in the domains of climate change, nature management and environment studies. This strategy not only enables external developers to engage with and enhance the software, expanding its development potential and application scope, but also sets a benchmark for transparency in data platforms.

OpenEPI's adoption of OSS not only underpins its technological and operational framework but also drives a culture of innovation, collaboration, and transparency. By integrating OSS fully, OpenEPI ensures that its tools and data remain accessible, secure, and adaptable, supporting a sustainable and flexible platform that aligns with the evolving needs of its community.

5.1.3 Open standards and formats

Open standards are essential to the foundation and ongoing operation of platforms like OpenEPI, especially in fields like climate and nature data where consistency, reliability, and accuracy are crucial. Standards ensure that data collection, processing, and analysis can influence critical decision-making and policy formulation effectively.

Ensuring interoperability and integrity
For OpenEPI, which integrates diverse datasets across various environmental themes, adhering to open standards allows data sharing and processing. This ensures the integrity and interoperability of data, which is paramount in environmental research where collaboration across scientific and regulatory bodies is necessary.

Driving innovation and accessibility
OpenEPI's commitment to open standards extends to the use of universally recognized formats such as XML, JSON, Protobuf, and CSV for data exchange. These formats not only facilitate easy data manipulation and integration but also enhance the platform's accessibility and utility for a global community of researchers and developers. By using open formats, OpenEPI promotes reuse, innovation, and broad participation in environmental studies.

Metadata specification for enhanced data management
As part of its infrastructure, the pre-project has proposed a metadata specification that all data providers must follow. This specification is designed to make data easy to find and use, ensure consistent quality, and facilitate the creation of combined datasets with shared information, which is crucial for comprehensive environmental analysis.

By prioritizing open standards and formats, we not only uphold the principles of transparency and collaboration but also ensure that the OpenEPI platform remains adaptable, scalable, and aligned with the evolving needs of the environmental community. This approach not only simplifies the integration and analysis of complex datasets but also fosters a culture of innovation and shared knowledge, crucial for advancing global environmental initiatives.

5.1.4 Open APIs

The APIs of OpenEPI are foundational to the platform, facilitating access to the data hosted within. They are integral components and should be treated with utmost priority. It is crucial that these APIs are open, transparent, and accessible to everyone, thereby promoting inclusivity across all geographical locations and organizational affiliations. The following topics warrant careful consideration:

Comprehensive documentation: It is essential that OpenEPI utilizes the OpenAPI Specification to document its APIs. This standardized documentation enhances clarity and facilitates easier integration, helping to lower barriers for users across various domains and technical backgrounds.

Focus on ease of use: The APIs should be designed with simplicity and intuitiveness in mind. By making the APIs straightforward to use, OpenEPI can boost user engagement and simplify the process for users to obtain and utilize environmental data.

Ensure reliability: Prioritizing the reliability of the APIs is crucial. Reliable API services build trust and ensure consistent access to data, supporting the platform's users in their continuous research and operational activities.

Implement access management: While preserving the openness of the APIs, introducing access management through credentials is advised to facilitate the collection of metrics and anonymous monitoring of platform usage. This strategy ensures that all data remains accessible, while effectively managing resource use and maintaining data integrity.

Apply rate limiting: Introducing rate limiting on the APIs can prevent system overload and ensure equitable access for all users. This measure helps in maintaining the system’s performance and availability by managing traffic and resource allocation efficiently.

5.2 Key components of a modern infrastructure

In planning for the technology part of the OpenEPI initiative, we are aiming at a modern and flexible IT architecture, which allows efficient governance, management and further scaling, as more data, more solution developers and other users arrive, and more functionality is requested.

The pre-project has been designed to give our tech team the time and possibility to employ and evaluate the technologies we believe could be well suited for the purpose and for the realization of the overall aim of the OpenEPI project. In the following we describe and discuss this explorative work and our most important findings.

5.2.1 Cloud providers

To safeguard the sustainability, adaptability and long term goals of a platform like OpenEPI, it is crucial to avoid dependence on a single cloud provider. Sole reliance on one vendor for critical data provisioning, such as global climate information, can lead to significant risks, including possible service interruptions and a lack of competitive pricing options. Thus, we think it is vital to distribute the processing and storage of data across multiple providers.

Several factors must be considered when developing OpenEPI's cloud strategy:

- **Risk mitigation:** Utilizing multiple vendors helps guard against data loss or downtime resulting from failures at a single provider.
- **Geographical reach**: Different providers may deliver better performance in various global regions, thereby enhancing service access and quality for a worldwide user base.
- **Specialized services**: Cloud vendors often offer unique services and capabilities; selecting different providers allows the project to benefit from specific features tailored to its needs.
- **Cost**: Collaborating with multiple providers facilitates cost comparisons and enhances financial management due to varied pricing structures and potential discounts. However, adopting a multi-cloud strategy will likely increase the overall project expenses.

Choosing cloud vendors requires careful consideration to ensure that they possess robust infrastructures and advanced capabilities. For the full-scale implementation of OpenEPI, it is advisable to engage the three leading cloud vendors: Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure for the heavy lifting of processing in the platform.

Experiences from the pre-project

The following table outlines the cloud vendors engaged during the pre-project phase of OpenEPI:

Vendor	**Usage**
Google Cloud Platform	Application hosting
Amazon Web Services	Data processing and application hosting
Cloudflare	DNS Routing, Hosting of developer portal
Sanity	Headless CMS for openepi.io
Vercel	Hosting of openepi.io

To explore the benefits of a multi-cloud strategy and address the complexities of managing a platform distributed across multiple vendors, Google Cloud Platform (GCP) and Amazon Web Services (AWS) were selected as the primary providers for platform development.

While both GCP and AWS have capabilities for routing and DNS handling, the decision was made to assign these responsibilities to Cloudflare to diversify our service dependencies, which has proven effective. Vercel was initially chosen to host our project site, www.openepi.io. Although it has been a reliable platform, it is relatively expensive. Transitioning this service to Cloudflare could streamline operations and result in cost savings.

Key findings

Finding	Description
Complexity	• Utilizing multiple primary clouds significantly increases the complexity of the project. Maintaining two or more technology stacks requires expertise to manage effectively.
Cost	• There appears to be a higher cost associated with the AWS stack compared to GCP. This may be attributed to the more processing-intensive services currently running on AWS.
Distribution of services	• In the pre-project phase, we distributed the applications fairly equally between GCP and AWS, organizing them by theme. This thematic division has proven to be an effective strategy, simplifying the architecture and minimizing the need for cloud-to-cloud connectivity. API routing is managed by Cloudflare, which directs traffic to the appropriate cloud based on the service selected.

In conclusion, for OpenEPI to ensure sustainability, adaptability, and achieve the long-term objectives, integrating all three major cloud vendors—Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure—is crucial for the full-scale operation of the platform. These vendors will provide a robust infrastructure that leverages their diverse geographical reach and specialized services, crucial for enhancing the platform's performance and reliability. Additionally, including Cloudflare for DNS routing and hosting of the websites, will further solidify the system's efficiency.

From the lessons learned in the pre-project phase, it is also advisable to streamline web hosting services by transitioning from Vercel to Cloudflare for hosting openepi.io. This move, while not primarily for cost-saving purposes, will simplify operations and maintain consistency across service providers, which in turn supports better integration and management of technological resources. Adopting this comprehensive multi-cloud strategy will equip OpenEPI with a resilient and flexible infrastructure, well-suited to handle the dynamic demands of distributing global climate, nature and environmental data.

5.2.2 Version control, CI/CD and GitOps

All source code produced in the OpenEPI context must be securely stored to facilitate open sourcing and collaboration. Throughout the pre-project, GitHub has been utilized, aligning seamlessly with our goals and vision due to its status as the de facto standard for open source projects.

Continuous Integration (CI) and Continuous Deployment (CD) are crucial practices in modern software development, significantly enhancing the speed and quality of software delivery. CI involves developers frequently merging their code changes into a central repository, often multiple times a day. This practice helps detect and resolve conflicts early, incorporates automated testing to swiftly catch bugs, and thus improves software quality while reducing the time to validate and release updates. Continuous Deployment builds on CI by automatically deploying all code changes to a testing or production environment after the build and testing stages, ensuring that new features are quickly and reliably made available to users. Together, CI/CD automates and integrates the development and operations processes, streamlining the entire software release cycle.

Given its integration with GitHub and the benefits for open source projects, GitHub Actions is an ideal tool for facilitating CI and CD development within OpenEPI due to its vast ecosystem and cost-effectiveness.

A note about GitOps: While CI/CD has been a long-established practice, GitOps is a relatively newer concept rapidly becoming the gold standard for infrastructure and application management. GitOps employs Git as the single source of truth for declarative infrastructure and applications, enabling extensive use of Git for version control, collaboration, and change management throughout the deployment pipeline. This approach simplifies infrastructure management and enhances transparency, consistency, and reproducibility through pull requests.

Experiences from the pre-project

The following table outlines the products used for version control and ci/cd engaged during the pre-project phase of OpenEPI:

Product	**Usage**
GitHub	Version control and single source of truth
GitHub Actions	Tool used for continuous integration
ArgoCD	Tool used for GitOps-style continuous deployment

GitHub and GitHub Actions have proven to be highly effective and will be advantageous for the full-scale implementation of OpenEPI. The team did not explore other options during the pre-project phase, as GitHub's capabilities closely align with our vision for OpenEPI.

For continuous deployment using a GitOps flow, the team considered both FluxCD and ArgoCD. Ultimately, ArgoCD was chosen based on several compelling factors. ArgoCD not only facilitates a GitOps flow, enhancing version control and collaborative application management, but it also features a user-friendly graphical interface that provides clear visual representations of applications and their synchronization status. It consistently monitors application health and delivers detailed status reports directly within its interface, thus improving monitoring and alerting capabilities. Designed to efficiently handle large numbers of applications and clusters, ArgoCD is well-suited for the demands of OpenEPI.

Furthermore, ArgoCD boasts a large community and robust support channels, providing extensive resources and support for troubleshooting and implementing best practices.

While the pre-project phase has satisfactorily utilized ArgoCD, FluxCD would also likely serve OpenEPI well, considering its capabilities.

5.2.3 Infrastructure as Code

Infrastructure as Code (IaC) is an essential practice in modern software development, enabling teams to manage their IT infrastructure through configuration files or scripts. This approach is particularly significant for OpenEPI, given its multi-cloud approach, and aligns seamlessly with a GitOps methodology.

IaC enables precise and repeatable environment setups across various stages such as testing, staging, and production, and even development environments. This method counters the common "configuration drift" problem, where discrepancies emerge over time between different environments, often manifesting as the "works on my machine" issue. IaC also diminishes the reliance on specific team members' knowledge, allowing new members to quickly integrate and contribute without extensive training. This is enhanced by the capability to maintain and update infrastructure independently of any single individual's expertise or memory.

Moreover, IaC automates the deployment and configuration processes, significantly reducing the risk of human error and enhancing overall security. This automation ensures that deployments adhere to strict standards. However, it is important to acknowledge that the initial development time for setting up IaC can be greater than using traditional methods, such as manual configurations via web portals. Despite this, the long-term benefits of scalability, maintenance efficiency, and security justify the initial investment in setting up IaC.

Experiences from the pre-project

When selecting tools for the OpenEPI pre-project, the team considered various IaC solutions:

- Cloud-native IaC tools like Google Cloud Deployment Manager, AWS CloudFormation or AWS CDK, and Azure Resource Manager
- Terraform
- Pulumi

Given our commitment to a multi-cloud architecture, the decision was made against using cloud-specific tools to avoid the complexity of mastering multiple systems. The team ultimately preferred Pulumi over Terraform for managing infrastructure due to several key factors. In 2023, Terraform shifted their licensing from the open Mozilla Public License v2.0 to the more restrictive Business Source License (BSL) v1.1. This change influenced the decision as it potentially will limit the use of Terraform in certain environments.

Pulumi's support for mainstream programming languages such as Python and JavaScript greatly appealed to our OpenEPI team, whose expertise in these languages facilitates more sophisticated and adaptable infrastructure management scripts. This compatibility effectively reduces the learning curve and boosts productivity in managing cloud resources.

Additionally, Pulumi's commitment to open-source principles, evidenced by its Apache 2.0 licensing, ensures unrestricted usage in both private and non-commercial projects, aligning well with OpenEPI's open-source ethos.

A practical insight from the pre-project phase highlighted a challenge with Pulumi: performance issues arose when using a single stack for all infrastructure needs. A strategic revision in spring 2024 to implement micro-stacks significantly alleviated these problems, underscoring an important consideration for the ongoing development of OpenEPI.

5.2.4 Containerization and container orchestration

Containerization is a widely adopted virtualization strategy that encapsulates software code and all its dependencies into a single package or container. This method ensures that applications run consistently and efficiently across different computing environments, making containerization a de facto standard in software development. By isolating applications and their environments, containerization reduces conflicts between differing software versions and settings, highlighting its importance in modern software delivery.

Complementing containerization, container orchestration automates the management of these containers, especially in complex and large-scale environments. Container orchestration tools handle tasks such as deployment, scaling, and management of containerized applications, enhancing their robustness and agility. For projects like OpenEPI, these technologies enable seamless scaling and maintain high availability across diverse cloud platforms, proving essential for effective cloud infrastructure management.

Experiences from the pre-project

In the pre-project phase of OpenEPI, the technical team deliberated over two main deployment models for applications: **Docker** (containerization) and **serverless computing**. Both approaches offer distinct advantages and can be integral to any modern software project. Given our commitment to vendor neutrality, the team opted for Docker, which facilitates easy migration of applications across different cloud providers. While serverless computing also presents appealing features, transitioning between providers can entail a significantly higher workload.

Having chosen Docker for deployment, the next task was to determine the best method to manage, deploy, and scale applications. With AWS and GCP already selected as the cloud environments, the team explored various orchestration options offered by these platforms. Initially, AWS ECS and Google Cloud Run were considered due to their simplicity in container orchestration. However, the challenge of maintaining efficient deployment strategies across both clouds without resorting to cloud-specific code led the team to choose AWS EKS and GCP GKE instead.

By adopting **Kubernetes**, the team secured a consistent experience across both cloud environments, enabling the use of uniform tools for deployment, monitoring, and ingress handling, among other tasks. This approach not only streamlined operations but also aligned with our overarching strategy for OpenAPI, of maintaining flexibility and neutrality in its cloud infrastructure.

Kubernetes is a complex project that encompasses numerous components, demanding careful maintenance to ensure it operates smoothly and securely. During the pre-project phase, we took a cautious approach, implementing a few shortcuts, particularly in aspects of internal cluster security, which requires further refinement to reach production readiness.

For a full-scale implementation of OpenEPI, substantial efforts are necessary to develop a secure, scalable, and efficiently managed Kubernetes-based platform. This will involve enhancing the current setup to meet the rigorous demands of a production environment.

5.2.5 API management

The management of the APIs provided by OpenEPI is of the utmost importance. The APIs are the entry to data provided by OpenEPI and should be treated as first class citizens, with best-in-class documentation and management. When considering the management of the APIs, several important factors need to be discussed, as described in the table below.

API routing	Utilizing a common gateway for exposing APIs is crucial for streamlining the development process. This approach not only simplifies integration by offering a consistent entry point for all services but also reduces the risk of errors and the complexity associated with maintaining multiple, disparate systems. A standardized gateway ensures that all APIs adhere to unified protocols and patterns, enhancing reliability and easing the burden on developers. Many tools are available to help establish this unified system, each contributing to more efficient API management and better alignment with industry best practices.
Version management	Version management is important in maintaining the stability and predictability of APIs, ensuring that users can rely on consistent functionality and interface. This allows for new versions of the API to be introduced without disrupting existing implementations, as users can continue accessing previous versions. Additionally, version management facilitates backward compatibility, ensuring that updates or new features do not break applications built on earlier versions. It also simplifies maintenance and troubleshooting by clearly delineating changes between versions, which can help developers identify and address issues more efficiently. Moreover, structured versioning is key to meeting the evolving needs of users and adapting to feedback without compromising the user experience of current API consumers.
Rate limiting and throttling	Rate limiting and throttling are vital for API management to ensure service stability and prevent abuse. They cap and adjust the number of user requests based on server load, preventing resource monopolization and enhancing fair usage. These tools help maintain performance during peak traffic, reduce operational costs, and support security and compliance.
Authentication and authorization	While OpenEPI aims to keep APIs fully open, authentication can be important for rate limiting and usage monitoring. This process ensures that while access remains broad, the platform can still maintain control over API utilization to prevent abuse and manage system resources effectively.
API usage monitoring	To effectively understand and enhance the utilization of APIs, it's necessary to monitor their usage. Monitoring focuses on tracking the popularity of services and ensuring the ethical use of data. Importantly, this monitoring should be conducted anonymously, aimed solely at understanding the patterns and coherence in how the APIs are being used, without identifying individual users.

Experiences from the pre-project

The pre-project phase of OpenEPI was centered on prototyping and understanding the lifecycle management of APIs that provide data. To expedite development and concentrate on key technological elements, the project bypassed features like version management, rate limiting, and authentication. These components typically become essential when there are multiple versions of an API or a significant number of users, conditions not yet met.

Authentication and authorization were also deferred, given the early stage and scope of the user base. Instead, the team focused on establishing efficient routing and API usage monitoring. For routing within a Kubernetes environment, two tools were evaluated: Apache APISIX and Traefik. The project started with APISIX but later shifted to Traefik due to its simplicity and smoother integration with Kubernetes.

For API usage monitoring, metrics were implemented in Prometheus format to ensure detailed insight into API usage while maintaining user anonymity. Further information on these metrics can be found in chapter 5.2.7 Metrics and monitoring.

Moving forward, the team recommends continuing with Traefik as the API gateway, considering its proven ease of use and effectiveness in managing API traffic within the Kubernetes framework. This approach will support a streamlined and scalable infrastructure as the project moves into more advanced stages.

5.2.6 Authentication and authorization

In the OpenEPI project, the necessity for authentication and authorization emerges not just for public APIs but also for securing internal services like Grafana (for metrics and monitoring), ArgoCD (for continuous deployment), and Dagster (for pipeline management). These services are crucial for operational and infrastructure management, requiring controlled access to ensure security and integrity.

Experiences from the pre-project

For authentication management, OpenEPI chose to use Keycloak in the pre-project based on its strong industry reputation and previous positive experiences from other projects within the tech team. Keycloak is an open-source identity and access management solution known for its extensive support for modern authentication protocols, making it a reliable choice for projects requiring robust security frameworks.

Several open-source alternatives exist, including Gluu and FusionAuth, each offering comprehensive identity management capabilities. However, Keycloak was selected due to its familiarity to the team and proven effectiveness in similar project environments. If the project were to consider a closed-source option, Auth0 would be a viable candidate due to its advanced features and extensive support services.

At the time for this report writing, OpenEPI uses a basic realm setup in Keycloak without integration with external identity providers. This setup meets the immediate needs for the OpenEPI proof-of-concept while keeping the system manageable and secure. As OpenEPI progresses towards a full-scale implementation, our suggestion is to integrate with more

established identity providers such as GitHub to leverage its authentication mechanisms. This would not only enhance security but also streamline the login process for users across the platform.

5.2.7 Metrics and monitoring

As we already have argued in previous chapters, it will be important for OpenEPI to implement a thorough metrics and monitoring framework to ensure optimal functionality, continuous user feedback and platform improvement, and strategic foresight. This framework should be designed to capture detailed data on how APIs and datasets are utilized, which is pivotal for operational efficiency and the evolutionary path of the platform.

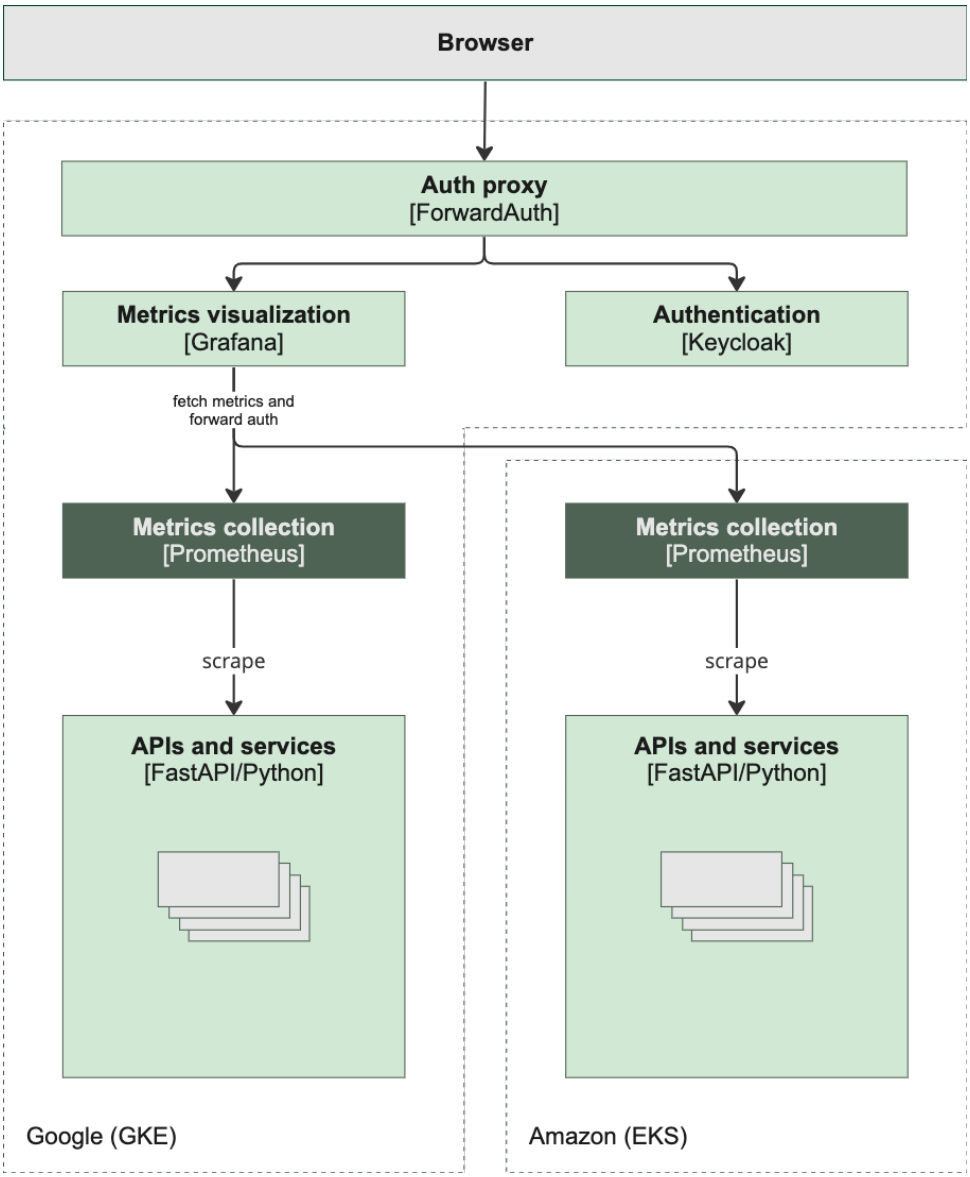


Figure 7. Collection and visualization of data, using Grafana.

By systematically tracking interactions with APIs and data sets, the OpenEPI team will be able to gain essential insights into usage frequencies, enhancing their ability to refine the

platform’s offerings. The continuous monitoring can help anticipate future demands and adapt to shifting usage trends, crucial for scaling the infrastructure effectively. Furthermore, analyzing the combined usage of data sets allows the team to understand complex data interdependencies, promoting the development of integrated data services while upholding ethical data usage standards.

Experiences from the pre-project

The team has in the pre-project implemented metric gathering by using Prometheus, and visualization of them using Grafana. To have valuable insights, the team has opted to only have one Grafana installation, which in the pre-project is running on GCP. However, to collect metrics across all cloud-providers, Prometheus has been set up at both AWS and GCP, allowing for collecting metrics from all APIs, and the Kubernetes cluster itself.

In order to not expose unnecessary information on the public internet, the Prometheus endpoints has been protected with an authentication layer, ensuring that only authorized users gain access to metrics. The sketch below illustrates how data is collected and visualized in Grafana, ensuring the confidentiality of metrics using open source components. This setup, incorporating Grafana, forward-auth, Keycloak, and Prometheus, serves as a solid foundation for our monitoring stack. However, the usage of the APIs on our OpenEPI pilot has been minimal, so the value of the monitoring data gathered so far is limited.

5.2.8 Programming languages and frameworks

In the pre-project we have chosen Python and TypeScript as the primary programming languages for the development of the OpenEPI platform, with a focus on leveraging frameworks that enhance the development of both backend and frontend functionalities.

Technology and frameworks used in the data handling of the platform is described in chapter 5.2.9 Data management.

Python for Backend development
Given its prominence in data-centric applications, Python is a good choice for OpenEPI, especially due to its extensive adoption within the scientific and data analysis communities, which aligns well with the data-heavy domain of OpenEPI. Python's extensive libraries and frameworks make it a top choice for developers working in fields that require robust data manipulation and analysis capabilities:

- **FastAPI**: This modern, fast (high-performance) web framework for building APIs with Python is based on standard Python type hints. The key features of FastAPI include speed and rapid development, and it's designed to be easy and intuitive, minimizing code duplication. It is licensed under the MIT license.
- **Pydantic**: This library is used primarily for data parsing and validation using Python type annotations. Pydantic enforces type hints at runtime and provides user-friendly errors when data is invalid. Like FastAPI, Pydantic also uses the MIT license.
- **Poetry**: A tool for dependency management and packaging in Python, Poetry allows developers to declare, manage and install dependencies, which ensures reproducible and consistent environments. It's also licensed under the MIT license.

TypeScript for Frontend Development
TypeScript is chosen for its ability to scale JavaScript by adding static types. Typescript's adoption is largely due to its powerful tools for large-scale applications and its ability to catch errors early through its static typing, making the codebase more maintainable and error-free. This aligns with the modern web development practices and is beneficial for projects like OpenEPI, where reliability and performance are important:

- **Next.js**: This React framework provides a lot of features that can enhance productivity in developing performant, server-side rendered React applications. It is especially known for its support for static site generation and server-side rendering. Next.js is open sourced under the MIT license.
- **React**: Known for its efficiency and flexibility, React is a JavaScript library for building user interfaces. It lets developers compose complex UIs from small and isolated pieces of code called components, and it uses a declarative paradigm that makes it easier to reason about your application and aims to be both efficient and flexible. React is also licensed under the MIT license.

Experiences from the pre-project

Python's popularity allowed our data scientists to contribute directly to API development, leveraging their expertise in the language. These choices have generally met our expectations, though continuous evaluation will ensure they align with OpenEPI's evolving needs as the project progresses. When developing a full-scale implementation of OpenEPI, it will be important to consider the efficiency of Python as a programming language. Due to its nature as an interpreted language, Python can be less efficient compared to compiled languages. This efficiency gap is important to consider when architecting a platform like OpenEPI.

The platform implementation in the pre-project has followed a microservices architecture style. This style allows for different components to use different technologies, meaning that programming language can be decided per service, and be different depending on the requirements of each service. However, it should be noted that the users of OpenEPI should be able to choose their programming language of choice, without being influenced by the core programming languages in the OpenEPI platform itself. Meaning, OpenEPI should provide client libraries in the most common languages used by developers in the target demographic.

5.2.9 Data platform

The primary function of the OpenEPI data platform is to ingest, process, and integrate data to be served by the APIs of the OpenEPI developer portal, subsequently aggregating it for usability. In essence, the data platform serves as a hub for collecting and preparing data from multiple sources, making it accessible and beneficial for developers and other prospective users of the OpenEPI platform.

A data platform consists of multiple parts, the following are the main ones, relevant for OpenEPI.

Data Ingestion

Geospatial data are stored and distributed in a wide variety of file format, which can be divided into three main categories:

Raster data	Raster data describes data that is organized in a two-dimensional spatial grid, potentially with multiple bands (layers) covering the same grid. This mainly includes satellite imagery and remote sensing data. Raster data is usually stored and distributed as GeoTIFF files, which are TIFF files that follow a metadata standard which embeds georeferencing information. Another option is Cloud Optimized GeoTIFF (COG), which is a standard based on GeoTIFF that allows for more efficient partial data reads from GeoTIFFs hosted on cloud systems. COG is entirely backwards compatible with GeoTIFF, and as there are virtually no downsides to using it, all newly created GeoTIFF files should follow the COG convention.
Multi-dimensional data	Outputs from meteorological, oceanographic and climate models usually have other dimensions in addition to the two spatial dimensions, i.e., time and elevation. The most common storage formats for this kind of data are NetCDF and GRIB. Multidimensional datasets are usually stored and distributed in one of three formats: NetCDF, GRIB and Zarr. Of these, the two first are oldest and the most widely used and supported, but Zarr seems to be gaining in popularity. While NetCDF and GRIB were originally mainly designed for use on local filesystems and HPC systems, Zarr is designed for efficient I/O on distributed systems like cloud object stores and can therefore often be much faster for cloud-based parallel computing applications.
Vector data	● GeoJSON ● Shapefile ● Parquet/Geoparquet ● CSV Geographic vector data describes geometries such as points, line-strings and polygons, and features associated with these geometries. For example, a dataset over flooding incidents could have the location of the incident as a point geometry, and the features might include the timestamp of the observation and the severity of the flooding incident. Vector data is the most common type of geographical data that is not produced by physics simulations or satellite imagery. GeoJSON is an open standard format for representing geographical vector features based on the JSON format. GeoJSON is the most widely used format for vector data. Shapefile is another format for vector data developed by Esri and is commonly used with GIS software such as ArcGIS

Data processing

Geospatial data differs from data in many other domains because it most often exists in the form of large raster files or multi-dimensional datasets. Processing this kind of data comes with a unique set of challenges and requirements compared to for example tabular data or smaller image files. The ability to process large raster files and multidimensional datasets in a scalable and reliable way largely determines the choice of data processing framework. As the files in many cases will be too large to be loaded into memory, this means that we need to use a data processing tool that allows us to process files incrementally (or in chunks). This makes using libraries like *NumPy* and *pandas* directly difficult, as they will try to load entire

files into memory regardless of size, which may cause out of memory errors. Another requirement is that it should be possible to dynamically scale the data processing capability to handle potentially terabytes of data. This largely leaves us to two alternatives: *Spark* and *Dask*. The details about both and why we chose to mainly go with the latter are explained in the Data Processing Framework section below.

Orchestration

As the data platform will contain many data pipelines belonging to different domains and categories, with pipeline dependencies both within and between groups, an orchestration system is required to have an overview over all the pipelines, keep track of pipeline runs, organize scheduled pipeline runs and to ensure that downstream pipelines incorporate upstream updates.

The pipeline overview should provide a unified view of all the data pipelines, showing dependencies between pipeline steps and run status. Run logs and metrics will be very useful for debugging and monitoring. The logs from previous runs should be stored and be easy to access so one can quickly find out why a pipeline fails.

Experiences from the pre-project

One of the main goals for us has been to provide an open source platform for geospatial data. The architectural decisions behind the OpenEPI data platform have in large part been influenced by this. For example, the restriction to using only open source components ruled out proprietary data platform solutions such as Databricks. Nevertheless, we allowed the use of proprietary solutions in the first phase of the project, with the reservation that we had a clear plan of how we would migrate to an open source solution. We chose to use Databricks in the initial phase because it would allow us to quickly set up some data pipelines and experience hands-on what functionalities we would need for the actual data platform. One of the strong points of using Databricks in this context is that it provided a configured data platform “out of the box”, which allowed us to quickly set up data pipelines and test out our data processing requirements without having to spend a lot of time setting up infrastructure in the beginning. While many of the core functionalities of Databricks are based on open source components like Spark and Delta Lake, it is not itself an open source platform. But as long as we limited our use of Databricks to mainly the open source components, migrating to a new data platform would be relatively easy. Building the first data pipelines on Databricks, we quickly learnt where it was suited to meet our requirements, and where it was not. The limitations we ran into were mainly related to processing multidimensional geospatial data. Some of these limitations made it unsuitable for our use case, setting aside the fact that it is not open source.

Apache Spark is the technology that allows Databricks to scale to huge workloads. Although Spark excels for processing tabular data, we found that it was not sufficiently suited for our demands, which in large part involves processing large raster files and multidimensional datasets. While Spark does have libraries that extend the functionality to geographic vector data and raster data, (e.g. Apache Sedona and GeoMesa for vector data, GeoMesa for raster data), there seems to be limited functionality for processing multidimensional data. Additionally, both GeoTrellis and GeoMesa use Scala, while we would prefer to use Python for the data processing as it has much more widespread use than Scala.

Based on our own experience and the knowledge derived from community geoscience projects like Pangeo, we have found that the most fitting framework for our needs is using Xarray in Python with Dask as the processing backend. Xarray and Dask are specifically designed for processing rasters and multidimensional data. This combination also makes it easy to perform local development and testing, but also scales up to very large workloads.

In conclusion, Databricks allowed us to quickly test and deploy data processing pipelines in the initial phase of the project. Since it is not open source we would eventually have to move away from the platform. Also, as we ran into limitations when processing raster and multidimensional data, the natural choice was to migrate away from Databricks sooner than we originally planned for.

Data Processing Framework

Framework	Description	Pros	Cons
Spark	Spark is an analytics engine for large scale data processing. Spark allows for distributed, parallel computing of large volumes of data, and can therefore scale to huge workloads. Spark has seen wide industry adoption since its release in 2014, especially for ETL workloads with tabular data. While there are extensions to Spark for vector data, frameworks for raster and multidimensional data are very limited	Well established and has seen wider adoption than Dask Well suited for tabular data Scales to huge workloads	Limited support for raster data Very limited support for processing multidimensional data (netCDF, Gribb, Zarr) Most geospatial libraries use Scala
Dask	Dask is a Python library for scalable parallel computing. Dask mimics the APIs of widely used frameworks like NumPy and pandas, making it in many cases a drop-in replacement. Xarray and geopandas are widely used libraries for processing raster and vector data respectively. However, as these libraries by default use Numpy and pandas as the compute backend, which loads all the data into memory once opened, you quickly run out of memory when working with large datasets. Dask fixes this problem as it reads, processes and writes data in chunks, and can distribute the workload over multiple processor cores and cluster nodes. We can specify Dask as the backend for xarray, which allows us to process large raster datasets in chunks and distribute the work over compute clusters	Python native framework Can be used interchangeably with xarray and pandas Well suited for raster and multidimensional data Can be deployed on Kubernetes Scales to huge workloads	Large overhead. Slow for small datasets Tabular data processing is limited compared to Spark

Orchestration Tools

Tool	Description
Apache Airflow	Apache Airflow is a well-established orchestration tool that has been around for several years. Among the alternatives, Apache Airflow stands out as the most well-established tool with a large and active community., which means that it has a wealth of features and integrations available. This active community also means that you can easily find help and resources online if you run into any issues while using Airflow. Its extensive feature set and wide range of integrations has for many years made it a popular choice for orchestrating workflows. However, some users may find Airflow to be less dynamic compared to newer alternatives, indicating a potential need for updates to keep pace with evolving requirements. It may not be the right choice for building a new data platform from the ground up. Deployment of Apache Airflow typically involves setting up a backend database and ensuring regular synchronization of Directed Acyclic Graphs (DAGs) and configurations across multiple nodes for efficient execution. While Airflow offers robust production deployment guidelines, users have several options such as Kubernetes with Helm, Docker, or Celery for orchestrating their workflows effectively.
Argo Workflows	Argo Workflows is another option that is particularly suited for Kubernetes-native deployments. Leveraging the tech stack already utilizing Kubernetes and Argo CD, Argo Workflows seamlessly integrates into existing infrastructures. However, its dependency on Kubernetes might pose a challenge for users unfamiliar with the platform.
Dagster	Dagster is an orchestration platform for data assets that was developed by Elementl. It has a declarative API that uses decorators to define data assets, rather than jobs. This makes it easy to use and understand, and it allows you to focus on defining the data assets that your pipelines produce, rather than the tasks that are required to produce them. One of the standout features of Dagster is its support for Python as a first-class citizen. Pipelines can be written in Python, which makes it easy to use for developers who are already familiar with this language (which is usually the case for data engineers and data scientists). Dagster seems to be well liked by developers, although it is a relatively new tool and does not have as large of a community as some of the other options on this list.
Prefect	Prefect is an open-source workflow management system for data engineers. It allows you to define pipelines and tasks using decorators, which makes it easy to create orchestrated pipelines with minimal changes to your code. Prefect also has built-in support for Dask, which means that you can easily use this library to parallelize and scale your pipelines.

After evaluating several different orchestration tools, including Apache Airflow, Argo Workflows, Dagster and Prefect, we decided to go with Dagster. One of the main reasons for our decision was Dagster's declarative and in Python-native API, which makes it easy to use and understand. We also appreciated the fact that pipelines can be written in Python, a language that many of our developers are already familiar with.

While Prefect and Dagster are in many ways very similar, one aspect where we preferred Dagster over Prefect was in the balance of features in the open source library vs hosted service. Both Dagster and Prefect offer a hosted service, however we perceived that wording in the documentation of the two projects was quite different, Dagster focusing more on the open source library, and Prefect focusing more on the hosted service. This made Prefect less appealing to us as an open-source solution. While this assessment is quite subjective, it was one of the multiple small differences that tilted the scales in favor of Dagster between two tools that are otherwise quite similar. Overall, we believe that Dagster is the best choice for our needs because of its modern API, support for Python, and full set of features in the open-source version.

5.3 Facilitating platform connectivity

OpenEPI stands as a critical intermediary platform that not only disseminates data but also facilitates the integration of these data to diverse external applications. Created to cater to developers needing effortless access to comprehensive datasets, OpenEPI can facilitate the effective use and integration of data sourced from various platforms. This dual role positions OpenEPI uniquely as both an expansive data platform and a powerful integration tool, making it an essential component within the data ecosystem.

Crucially, OpenEPI has the potential to also function as a reliable intermediary for data providers facing challenges with uptime. By serving as a resilient buffer or proxy, OpenEPI can have the capability to ensure uninterrupted access to data, which is vital for developers dependent on consistent data flows for their applications. OpenEPI needs to be implemented in a manner such that its infrastructure excels in caching and delivering data efficiently, minimizing interruptions that might result from source downtimes. This capability is crucial as it not only enhances the reliability of data access but also fortifies the trust that developers and end-users place on the platform for critical applications. As a result, OpenEPI can diminish the risks related to data unavailability and ensure a reliable data delivery service, cementing its position as an indispensable connector in the data supply chain.

To fully realize this potential, optimizing the architecture and functionality of OpenEPI involves meticulous planning and execution across several key areas.

Data ingestion

In the architecture of OpenEPI, the process of data ingestion is crucial for ensuring seamless integration and accessibility of datasets for developers. OpenEPI's strategy for data ingestion is aligned with the FAIR principles, which emphasize that data should be Findable, Accessible, Interoperable, and Reusable. These principles play a key role in enhancing the discovery, usability, and exchange of data across different systems and among various users.

An important component of this ingestion process is adhering to the OpenEPI Metadata Specification. This specification mandates that all data hosted on the platform is accompanied by standardized, high-quality metadata, ensuring the integrity and context of the data are maintained, making it easier for developers to understand and integrate with other datasets.

Open licensing also forms an integral part of the data ingestion framework. By mandating that datasets are shared under terms that facilitate broad use and distribution, OpenEPI nurtures a transparent and collaborative environment. This open exchange fosters innovation and extends the reach and impact of the data across different fields.

Additionally, OpenEPI recommends that each dataset ingested adheres to the data standards common to the data's domain. This ensures the datasets are not only relevant and accurate but also meet industry or academic benchmarks, enhancing their credibility and applicability. Compliance with these standards guarantees that the data can effectively meet the diverse needs of developers and researchers, facilitating data interoperability and integration within the platform.

Data transformation and adaptation

One of the key contributions of OpenEPI is its ability to simplify and adapt data, making it more accessible and easier to use for developers. This transformation process is crucial in ensuring that essential data fields are preserved without loss during the conversion steps. Simplifying data involves refining and reformatting data sets to enhance their compatibility and functionality within various applications.

During the pre-project phase, the OpenEPI team has specifically adapted flood and deforestation data to make it suitable for dissemination through an API. This adaptation process is meticulously managed to ensure transparency and maintain the accuracy of the data. It is critical that these transformations do not introduce any errors that could mislead users or distort the data's integrity.

An essential aspect of maintaining transparency and trust in the data transformation process is the open availability of all related source code. OpenEPI should commit to making all transformation code publicly accessible. By ensuring that all transformation steps are open and transparent, OpenEPI can reassure users that the data remains accurate and trustworthy after processing. This approach not only upholds the platform's integrity but also enhances its utility and reliability, making OpenEPI an indispensable resource in the data ecosystem.

Data delivery

OpenEPI's data delivery system should be designed to provide developers with seamless and reliable access to processed datasets, ensuring that every interaction with the platform is straightforward and productive. To achieve this, OpenEPI should leverage the OpenAPI Specification for all its API documentation, ensuring that developers have clear, comprehensive guidelines on how to utilize the APIs.

Moreover, OpenEPI needs to adhere to the licensing terms set by original data providers. This practice guarantees that users can confidently utilize the data within the legal frameworks established by the data sources, promoting ethical usage and distribution. To further enhance transparency, the developer portal should provide detailed descriptions of all data processing steps each dataset undergoes before being made available. This includes not only the transformations made but also examples of how to use the data, thereby helping

developers understand any modifications and rely on the accuracy and integrity of the data provided.

By combining comprehensive API documentation, strict adherence to licensing, maintained transparency in data processing, and ensuring reliable access, OpenEPI not only can fulfill its role as a critical intermediary but also strengthen the trust and reliance placed on it by developers and end-users.

5.4 A developer first approach

Chapter 4 discusses user needs and defines a set of core users for OpenEPI. Common for these users is that they are some types of software developer. In adopting a developer-first approach, OpenEPI prioritizes creating a seamless and intuitive experience for developers interacting with the platform. This chapter outlines the key elements to empower developers as they access and utilize resources in OpenEPI.

Developer Portal: The gateway to the data catalog

OpenEPI's developer portal serves as the essential access point for developers and should be designed to streamline their initial engagement and ongoing interaction with the platform. This portal is crucial in providing the tools and resources necessary for a comprehensive and intuitive development experience.

Continuing to maintain and enhance the data catalog within the portal is vital. It should include comprehensive details about each dataset's origin, licensing, and any processing conducted. Such transparency is fundamental for developers to fully understand the data's context and constraints, enabling effective utilization of available resources.

As OpenEPI evolves, adding a feature for API credential sign-up will be important to streamline access and manage developer interaction with the platform effectively. This functionality will ensure easier and more secure API access, supporting controlled use of the platform's capabilities.

User onboarding

A primary objective of the developer portal is to simplify user onboarding to the greatest extent possible. The process should be streamlined, allowing developers to quickly begin utilizing the platform's data with minimal preliminary steps.

Critical information should be prominently displayed to users as soon as they access the developer portal. This is why the data catalog is strategically placed in a highly visible location within the portal. Currently, in the pre-project phase, none of the APIs require credentials for access, facilitating immediate use. However, as OpenEPI moves towards full implementation, it will be essential to introduce a section for API credential sign-up. This addition aims to simplify the process of obtaining API access, enabling developers to easily secure and manage their interactions with the platform. Moreover, it will allow for controlled access, which is crucial for maintaining the integrity and security of both the data and the platform.

Client libraries and developer resources

The developer portal is equipped with client libraries for Python and JavaScript, which are essential for efficient data integration and developing applications with OpenEPI data. These libraries, currently in their basic form, are planned to be enhanced with more sophisticated functionalities beyond simple API call wrappers as OpenEPI moves towards full implementation. To cater to a wider developer base, the introduction of additional libraries for Java and Golang should be considered.

Additionally, the portal provides a variety of design resources currently used in developing the OpenEPI platform. With the project's progression, these offerings should be expanded to include frontend components, thereby enriching the suite of resources available. This expansion will further integrate and enhance the development environment, making it more comprehensive and accessible to developers.

Feedback and community engagement

All source code for OpenEPI is hosted on publicly accessible GitHub repositories, allowing users to easily report issues or suggest improvements. Additionally, we have established a Discord community to facilitate direct communication among developers, community members, and the OpenEPI technical team. This platform serves as a dynamic space for collaboration, discussion, and support.

To enhance this environment, the mechanisms through which users can formally propose changes or enhancements to the platform should be improved even further. Strengthening this aspect of community interaction will not only increase user engagement but also ensure that the platform continues to evolve in ways that meet the needs of its diverse user base.

5.5 Understanding system usage through metrics & monitoring

To effectively understand the usage and optimize the performance of OpenEPI, it is crucial to implement metrics and monitoring throughout the system. This chapter describes our approach to leveraging key open source tools like Prometheus and Grafana for system monitoring and analysis.

Prometheus

Prometheus, a monitoring tool, is installed on all clusters within the OpenEPI infrastructure. This setup ensures that we can collect detailed metrics from each component of our platform, regardless of where it resides. Every application and API within OpenEPI are configured to publish Prometheus metrics, which are then available for scraping. This consistent metric collection across all clusters is fundamental to obtaining a holistic view of the system’s health and performance. In addition to collecting metrics about the services and APIs, it will also collect metrics about the underlying cluster itself, allowing for detailed information about the system resources. This will allow for proactive management of resources needed to operate OpenEPI.

Grafana

For visualizing the metrics collected by Prometheus, we use Grafana, a popular analytics and monitoring solution known for its effectiveness in displaying complex data in an understandable format. Currently, Grafana is installed on a single cluster - presently on Google Cloud Platform (GCP), although this may change as our infrastructure evolves. Grafana provides the visualization capabilities needed to interpret the vast amounts of data collected by Prometheus, turning raw data into actionable insights. This enables the technical team to set up alerts when important events occur, and to actively monitor the usage of the platform.

Usage Metrics

Through Grafana, we can analyze usage metrics for each API and its specific endpoints. This granularity allows us to see not only the overall usage of the platform but also how individual components are performing. Metrics such as request counts, response times, and error rates for each endpoint help us identify popular features as well as areas that may require optimization. Additionally, we can observe trends in how APIs are used in combination with each other. Understanding these patterns is essential for optimizing API integrations and enhancing user workflows. This insight also aids in predicting future usage scenarios and planning capacity accordingly. Moreover, by analyzing how different datasets are accessed together, we can identify new opportunities for creating aggregate services that combine two or more datasets, offering more comprehensive solutions and insights to users.

By distinguishing between ad-hoc users and those who use API credentials for more extensive use of OpenEPI, we gain a deeper understanding of how and why the data within OpenEPI is being utilized. This differentiation not only helps us identify the diverse applications of our platform but also to some extent measures the broader impact of OpenEPI beyond our core developer community.

This insight into user behavior and platform engagement allows us to tailor our resources and support more effectively, ensuring that both casual and intensive users find value in OpenEPI. Additionally, by analyzing the usage patterns and feedback from these distinct user groups, we can innovate and expand our offerings to meet emerging needs and potential use cases.

Furthermore, metrics collected from different types of usage can help us understand the educational and research implications of OpenEPI. For instance, we can track which datasets are most frequently accessed and combined, providing valuable information about trending research areas or commercial interests. This data not only helps in forecasting future demands but also in assessing the socio-economic impact of the platform, as it is adopted across various industries and academic fields.

Overall, by leveraging detailed metrics, OpenEPI can continuously refine its services and extend its reach, enhancing its value as a pivotal resource in environmental data management and analysis.

5.6 Blueprint for a technical architecture

This section outlines the technical architecture of OpenEPI, emphasizing scalability, flexibility, and data management. We use the term "blueprint" cautiously, as OpenEPI's architecture is designed to evolve with technological advances and changing needs.

5.6.1 Architecture principles

A climate and nature data platform like the Open Earth Platform Initiative should adhere to architecture principles that ensure the platform's effectiveness, scalability, and sustainability.

Scalability and Flexibility: Design OpenEPI to be scalable, accommodating increasing data volumes and user numbers without compromising performance. The platform should be modular, allowing easy addition or removal of components to meet evolving needs.

Multi-Cloud Strategy: Ensure the platform functions across different cloud providers and on-premises environments to avoid vendor lock-in and enhance flexibility. Adopt a multi-cloud strategy, possibly designating a primary provider, to leverage unique offerings and increase system resilience.

Open Standards and Interoperability: Embrace open standards and ensure interoperability to facilitate integration with various systems and data formats. Utilize open-source components and libraries to foster innovation and enhance system compatibility and quality.

Data Management and Integration: Develop well-documented APIs and maintain standardized data formats and metadata to ensure seamless data exchange and integrity. Implement robust data quality control mechanisms and metadata management to support effective data discovery and utilization.

Innovative Technologies and Prototyping: Leverage best of breed technology, suited for the task at hand for insightful data-driven decisions. Utilize prototyping for quick validation, iterative refinement, and to ensure user-focused solutions.

Infrastructure and Development Practices: Rely on Infrastructure as Code (IaC) to simplify infrastructure management across different environments. Promote the use of suitable programming languages for data analysis and scalable web applications. Utilize containerization to enhance portability and management in cloud environments.

Data Portability: Ensure data remains portable and accessible across different platforms by using standardized protocols for data storage and exchange, maintaining control and ownership of data regardless of the cloud provider.

5.6.2 Standards and formats

Standards

Standards play a pivotal role in the field of climate and nature data by ensuring consistency, reliability, and accuracy in data collection, processing, and analysis. In environmental

research, where data influences critical decision-making and policy formulation, the integrity and interoperability of data are paramount. Standards facilitate the integration of diverse data sources, enabling comprehensive analysis and fostering collaboration across various scientific and regulatory entities.

OpenEPI itself does not set these standards; instead, the platform evaluates and adheres to existing standards used by data providers. OpenEPI will host data that span multiple themes and topics, each of them with its own set of data standards.

This approach ensures that the data integrated into OpenEPI meets high-quality benchmarks and is compatible with global scientific and technological communities. By evaluating the standards used by data providers when incorporating data into the platform, OpenEPI maintains a robust framework for data integrity and usability, enhancing the platform's reliability for users and stakeholders.

As a part of the pre-project OpenEPI has proposed a metadata specification to go along with data that are to be hosted through OpenEPI. This can be viewed at https://www.openepi.io/resources/metadata-specification. It is important to establish a clear set of rules for datasets, metadata specification. This is more than just a formality – it's a way to guide our partners, make data easy to find, ensure consistent quality, and create combined datasets with shared information.

Formats

Open standard formats promote accessibility, interoperability, reuse and innovation. Utilizing open formats allows data from OpenEPI to be easily shared, processed, and understood across different systems and applications without proprietary restrictions. This openness is essential for collaborative efforts, enabling researchers and analysts worldwide to utilize and extend the data for various environmental studies and initiatives.

OpenEPI commits to using open standard and common formats such as XML, JSON, Protobuf and CSV for data exchange. These formats are universally recognized and supported by a wide array of data processing tools and applications, facilitating seamless integration and manipulation of data. By prioritizing open formats, OpenEPI ensures that the data available through its platform is not only robust and scalable but also universally accessible and compatible with existing and emerging technologies in environmental research.

5.6.3 Scaling the platform

Scaling the OpenEPI platform involves not only expanding its technological infrastructure but also growing the team to support its development, deployment, and maintenance. The scalability of the platform is linked to the scalability of the team, ensuring that expertise covers various facets of the project's lifecycle.

For effectively scaling the OpenEPI platform, it's critical to architect services with scalability in mind. Stateless service design enhances scalability as it allows services to be duplicated across multiple instances without limits theoretically. By not maintaining any state within the services themselves, scalability becomes more manageable and efficient.

The initial implementation of OpenEPI utilizes Kubernetes, which frees application teams to focus solely on development by abstracting the underlying infrastructure complexities. However, managing a Kubernetes environment requires significant effort from the platform team, especially when scaling to accommodate heavy workloads. Kubernetes excels in handling extensive scaling through the addition of compute nodes, facilitating horizontal scalability.

To ensure the platform scales efficiently, it is vital to implement robust monitoring and auto-scaling systems that dynamically adjust resources based on real-time demand. Continuous optimization, which includes performance tuning and cost management, is crucial for maintaining operational efficiency and managing expenses.

Initially, OpenEPI operates with a single Kubernetes cluster per cloud provider, incorporating Dagster for data orchestration. For a full-scale deployment, it would be prudent to segregate data processing into a dedicated cluster. This separation allows for optimized environments tailored specifically for data processing and API serving, enhancing overall performance and resource utilization.

5.6.4 Security in the platform

Platform security is a critical aspect of developing OpenEPI. As highlighted in chapter 7.6 Security and safety risks, the core security challenges encompass confidentiality, integrity, and availability. Security discussions should primarily focus on infrastructure, clusters, and applications, ensuring that each layer is robustly protected to support the platform's goals and safeguard against potential threats.

Infrastructure Security

Security within OpenEPI's underlying infrastructure is an important focus during the lifetime of OpenEPI. As the platform will solely utilize cloud-based infrastructure initially, this aspect covers the security concerns associated with cloud services exclusively.

Secrets Management
Handling secret information in a secure manner is of utmost importance. This involves protecting sensitive information such as API key, passwords and other sensitive information that are needed for the operation of the platform. OpenEPI should use secure vaults and encrypted storage solutions to manage these secrets. Access to these secrets should be strictly controlled and audited to prevent unauthorized access.

In a multi-cloud setup, OpenEPI must consider that each of the cloud providers offer secret management solutions that are a bit different from each other and have different APIs.

OpenEPI has essentially three choices for handling secrets. Either have a centralized secret management system like Hashicorp Vault or Cyberark Conjur, rely on each cloud's own secrets management system, or let the IaC tool handle secrets and use Kubernetes secrets. There are pros and cons to these systems.

In the pre-project phase, we chose a straightforward approach, swiftly generating essential secrets using Pulumi and storing them as Kubernetes secrets. While this expedited

development, for the comprehensive implementation of OpenEPI, an alternative path is recommended. It's advisable to initially leverage each cloud provider's native secret management solution in the primary project phase.

	Centralized	**Cloud specific**	**Kubernetes secrets**
Pros	**Centralization**: All secrets are stored in a single location, providing a unified point of control and management. **Enhanced security**: Centralized systems often offer advanced security features such as encryption, access controls, and auditing capabilities. **Cross-cloud compatibility**: Works seamlessly across multiple cloud environments and on-premises infrastructure, promoting consistency and flexibility. **Automation**: Supports automation of key management tasks such as secrets rotation, reducing manual effort and minimizing the risk of human error.	**Native integration:** Seamlessly integrates with the respective cloud provider's ecosystem, simplifying deployment and management for organizations already utilizing these platforms. **Ease of use:** Built-in support for secrets management within the cloud platform's console or APIs may require less setup and configuration compared to external solutions. **Cost-efficiency:** Often included as part of the cloud provider's service offerings, potentially reducing costs associated with licensing and maintenance. **Automatic scaling:** Can automatically scale to accommodate changing workloads and demands without additional configuration or management overhead.	**Integration with Kubernetes**: Kubernetes Secrets provide a native solution for managing sensitive information within Kubernetes clusters, ensuring compatibility and seamless integration with containerized applications. **Visibility and Control**: Secrets managed through Kubernetes are visible within the Kubernetes ecosystem, providing transparency and control over access permissions and usage. **Scalability**: Kubernetes Secrets can scale alongside containerized applications and Kubernetes clusters, accommodating dynamic workloads and scaling requirements. **Ease of use**: Kubernetes secrets are an easy approach to delivering secret information to applications
Cons	**Dependency**: Organizations become reliant on a single vendor or solution, which can introduce risks such as vendor lock-in or disruptions in service. **Cost:** Centralized solutions may involve upfront costs for licensing, deployment, and ongoing	**Limited cross-cloud compatibility:** Solutions may not be easily portable across different cloud providers or on-premises environments, limiting flexibility and interoperability. **Reduced control:** Organizations may have less control over security policies, access controls, and audit	**Limited Security Features**: Kubernetes Secrets may lack advanced security features such as encryption at rest or fine-grained access controls compared to specialized secret management solutions like HashiCorp Vault or CyberArk Conjur. A truly secret setup with only Kubernetes secrets is a lot of work to do properly.

	maintenance, potentially making them less cost-effective than relying on native cloud solutions. **Integration challenges**: Integration with existing infrastructure and applications may require additional effort and customization, particularly in heterogeneous environments. **Potential single point of failure**: If the centralized system experiences downtime or security breaches, it can have widespread impacts on the organization's operations and security posture.	capabilities compared to using centralized systems. **Potential for fragmentation:** In multi-cloud environments, managing secrets across multiple native solutions can lead to fragmentation and increased complexity for security teams.	

5.6.5 Conceptual architecture

The diagram below provides a high-level overview of the suggested capabilities of the OpenEPI platform.

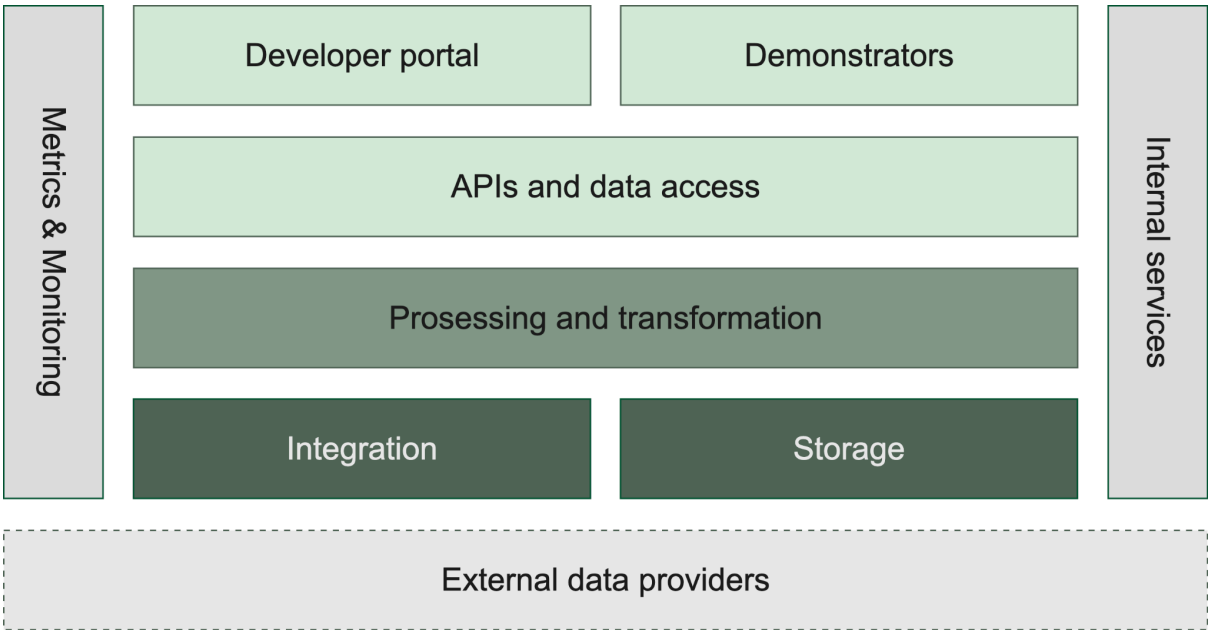


Figure 8. Overview of platform capabilities.

As discussed in the previous sections, the above conceptual architecture caters to the needs for a developer first approach, where all layers of the architecture support the easy access to data for application developers.

The platform can be divided into the following concepts:

Developer portal	The Developer Portal serves as the primary gateway for users to engage with the platform. It offers access to API documentation, enables users to explore demonstrators, and facilitates direct API interaction.
Demonstrators	Demonstrators provide practical examples of how to utilize APIs and client libraries effectively. These are not complete products but are designed as illustrative tools to showcase potential applications and encourage exploration.
APIs and data access	This component is crucial for end users as it grants them access to data and aggregated information. APIs serve as the conduit through which users can retrieve and interact with the platform's data.
Processing and transformation	This involves the internal refinement and adaptation of data received from providers to ensure it is optimized for API delivery. The process includes modifying the data to enhance its usability and integration within the platform.
Integration and storage	OpenEPI integrates with various data providers, employing both direct and processed forms of integration. For data that requires processing, the resulting products are stored directly on the platform, ensuring efficient data management and accessibility.
External data providers	Our partners play a vital role in supplying the platform with a diverse range of data, supporting the breadth and depth of services offered.
Metrics and monitoring	This function is essential for the technical team to monitor usage trends and overall system health. It enables proactive management of the platform's performance and scalability.
Internal services	This category encompasses all additional services required to operate a contemporary infrastructure platform. It includes continuous deployment, authentication, build pipelines, and infrastructure orchestration, ensuring the platform remains robust and agile in its operations.

5.6.6 Infrastructure architecture

The diagram below describes the OpenEPI infrastructure architecture.

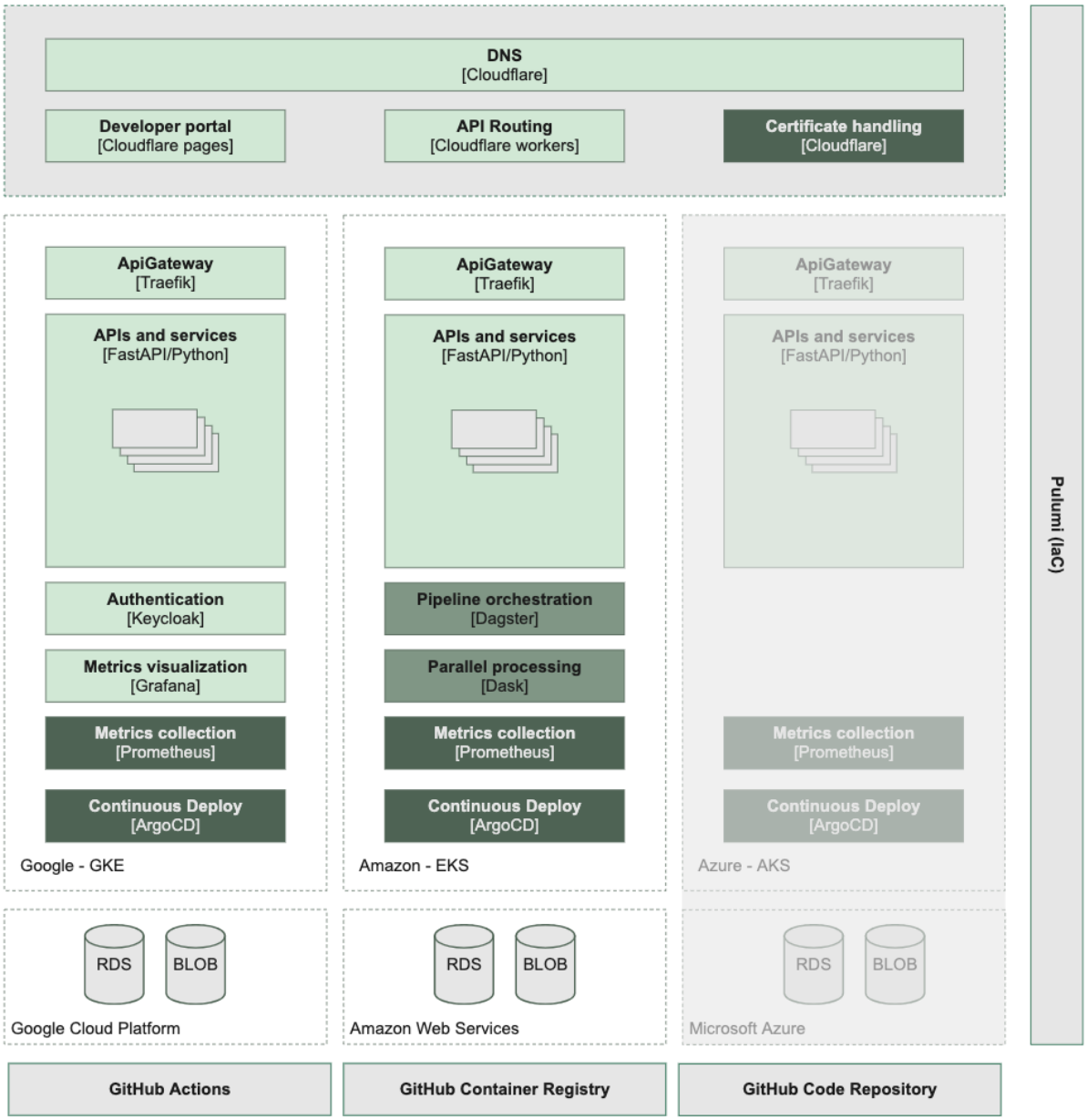


Figure 9. The OpenEPI infrastructure architecture.

The infrastructure consists of essentially four groups, as described in the following.

Cloudflare - DNS and routing

Cloudflare manages DNS and routing for the OpenEPI platform, including certificate handling. Its generous free tier and robust infrastructure make it an excellent choice for these functions. By outsourcing DNS and routing to Cloudflare, OpenEPI reduces its dependency on any single cloud provider like GCP or AWS, enhancing flexibility across different cloud environments. This approach is critical for OpenEPI's operations, supporting a versatile deployment strategy for services across various clouds. Additionally, Cloudflare Pages hosts the developer portal, offering a seamless method for deploying static websites. API routing is orchestrated by Cloudflare Workers, enabling OpenEPI to distribute its services across

multiple cloud vendors while centrally managing request routing to locate each service in the appropriate cloud.

Pulumi - Infrastructure as Code

In the development of the OpenEPI project, Pulumi is utilized as the Infrastructure as Code (IaC) tool of choice. It enables the team to deploy and manage infrastructure using Python, a language selected for its strong alignment with the project's data-centric focus. Through Pulumi Cloud, state management is handled, enhancing both security and reliability of configurations, particularly in automating the management of secrets. This setup allows for execution locally during the pre-project, with potential integration into broader CI/CD pipelines for continuous deployment across multiple cloud environments at a later stage. It is important to note that utilizing a single Infrastructure as Code (IaC) tool across multiple cloud vendors such as AWS, GCP, and Cloudflare simplifies the process of transitioning between these platforms.

GitHub - Source code and artifacts

All source code for the OpenEPI project, except for the infrastructure code, is hosted openly on GitHub, which aligns with the project's open-source ethos. The infrastructure repository remains closed source during the pre-project phase, with potential plans to open-source it later. GitHub was chosen as the primary platform due to its widespread adoption and support within the open-source community. It also provides GitHub Actions, which facilitates running tests and building artifacts at no cost for public repositories using standard GitHub Runners, sufficient for OpenEPI's current needs.

OpenEPI extensively utilizes Kubernetes to orchestrate container deployments. New Docker images are created for each version of an application or API, hosted on GitHub's Container Registry and made publicly available. This transparency allows developers worldwide to download and deploy instances of OpenEPI’s APIs with ease.

Application hosting - on Kubernetes

The core of the OpenEPI platform is built around applications and APIs that facilitate data access. The team has implemented Kubernetes for container orchestration to enhance the portability of APIs across different cloud providers. This flexibility theoretically allows the team to shift applications between providers by merely updating configurations.

During the pre-project phase, two Kubernetes clusters were set up: one on Google Cloud Platform (GCP) and another on Amazon Web Services (AWS). This dual-cloud strategy addresses the challenges associated with operating across multiple clouds. Each cluster hosts a suite of common services essential for maintaining the coherence of OpenEPI as a single platform, despite its multi-cloud deployment. These services include continuous deployment, metrics collection, and API gateway management, which are uniformly maintained across both clouds to facilitate seamless application development regardless of the hosting cloud.

For unique services like authentication, which require a single instance for functionality such as single sign-on across clouds, one cloud vendor is designated as the primary host. In the

pre-project, the authentication domain and Grafana are hosted on GCP, while data processing and pipeline orchestration reside on AWS.

Looking ahead to a full-scale deployment, integrating Microsoft Azure into OpenEPI's architecture is a natural next step to further enhance its multi-cloud capabilities.

5.6.7 Adaptive architecture: Evolving for future flexibility

The architecture of OpenEPI is conceived not as a static framework but as a dynamic, evolving structure poised to adapt alongside emerging technology trends. Recognizing that the initial architecture serves merely as a foundation, the platform is designed to flexibly respond to the changing needs and growth demands of its environment.

At the core of OpenEPI's architectural philosophy is the acknowledgment that today's technological decisions are based on the current landscape and the best information available. However, as technology advances and the team's knowledge deepen, it's anticipated that the architecture will naturally evolve. This iterative evolution is critical because it ensures that the platform remains at the cutting edge, effectively leveraging new tools, frameworks, and paradigms that enhance performance and scalability.

The initial design choices made during the pre-project phase were informed decisions based on the technologies and methodologies believed to be most effective at the time. These choices are considered starting points, each selected for its potential to provide a strong foundation while allowing for future adaptation and growth. For instance, the choice to utilize container orchestration through Kubernetes reflects a commitment to flexibility in deployment across multiple cloud environments, anticipating the need to scale and migrate services seamlessly as user demands and cloud technologies evolve.

Scaling the platform to accommodate growth is a foreseeable challenge that will likely require significant modifications to the technical implementation. As the user base expands and data throughput increases, the architecture must not only support scaling in terms of load handling but also maintain efficiency and cost-effectiveness. This might involve integrating more advanced data processing technologies, adopting more robust load balancing strategies, or even rearchitecting certain components to better handle increased operational demands.

Guiding all these adjustments are the architecture principles established in the project. These principles act as a compass for development, ensuring that while the architecture may change, it does so in a way that aligns with the core objectives and values of OpenEPI. Whether it's maintaining high availability, ensuring data integrity, or prioritizing security, these guiding tenets ensure that each architectural iteration not only addresses the immediate technical needs but also the broader mission of the platform.

In conclusion, the architecture of OpenEPI is designed with an inherent flexibility to adapt as new technologies emerge and as the platform scales. This approach ensures that the architecture will not only support the current needs of OpenEPI but will also evolve to meet future challenges, embodying a truly adaptive system that grows in capability and efficiency over time.

6. Operational feasibility

6.1 Introduction

Although we at the same time recognize the almost continuous stream of new initiatives in sharing open data and/or data for climate change related matters, we believe that OpenEPI has a potential role in the global open data ecosystem. In chapters 1-4 we tried to shed light over the entire landscape of initiatives, stakeholders and needs. There are indeed a lot of uncertainties and dependencies, especially in regard of the actual “market” for end-user solutions based on OpenEPI data. Still, we have concluded that OpenEPI will fill a gap, also in the long-term, contrary to other portals or services mostly offering specific “end-user products”.

In chapter 5 we have explained our proof-of-concept and assessment of the technological feasibility of OpenEPI, with compliance to our overarching ideals and policies on open data and open source. The chapter concludes with a proposed technical blueprint for OpenEPI. In chapter 6 we must now delve into the organizational and operational aspects of the establishment of OpenEPI and adjacent services. This is also a highly vital part of the entire feasibility study.

The original idea of OpenEPI as a concept includes the establishment of some form of business or organizational entity responsible for the platform and the services. The assessment team has profoundly evaluated various aspects of this building of OpenEPI as a permanent offering to developers and other users in LMIC countries. In addition to the resources needed to run and maintain the data platform and to provide relevant climate and nature related data sets, there is a need for capacity related to marketing activities, usage metrics and evaluation, support and advice, and other measures securing user adoption among developers and other data users in sub-Saharan Africa (and other regions combating climate change).

To conduct the ongoing pre-project and feasibility study, the OpenEPI consortium has built a temporary project organization - consisting of project management, assessment capacity and technology and organizational/systemic expertise. The team has mainly had Norwegian expertise (i.e. the private companies Knowit and Capto) but has been complemented by expertise from US-based Creative Commons and Polish-based Open Future - and sporadic contact with the grant giver Norad, the Norwegian development aid agency. The pre-project team has been recruited for development and assessment purposes, not for running a business operation. A permanent business unit will build and maintain the platform, based on the project’s blueprint and our recommendations and the guidance that this report constitutes. Furthermore, a permanent unit will operate the platform, manage (including efforts in improving data quality) data sets and data products, and offer a range of services aimed at developers - and manage the further scaling of operations.

Throughout this study, we have gathered insight from interviewing data platform owners and managers, assessed past, documented experiences, and learned from our workshops and discussions in the team. Based on this, we think it most certainly will be feasible to build a permanent OpenEPI organization. “Permanent” is to be understood as a sustainably funded

and robust organization, but not necessarily one single body. The OpenEPI organization could very well be a distributed organization, exploiting specific competencies and capabilities already established by existing initiatives globally. We think a core unit must handle and maintain the platform and APIs, have the responsibility for some basic services, for handling legal issues, maintain policies and take care of the administrative coordination of the entire OpenEPI initiative. However, the combined OpenEPI organization must eventually cover a multitude of topics and knowledge areas. OpenEPI will require a good understanding of relevant domain-related challenges (i.e. needs for specific data qualities, data governance routines, agtech innovations, the development of satellite imagery related techniques, etc.), know the data ecosystem and handle stakeholders with different needs all according to our discussions in the previous chapters – and, last but not least, manage the platform technology rig described in chapter 5.

We have assessed different organizational setups that are feasible, given the goals, user needs, technology stack, stakeholder landscape, and other aspects of the OpenEPI concept described in chapters 1-5. As mentioned, we have studied previous data platform initiatives, and interviewed several existing data platform owners, data HUBs and related initiatives. This has given us insight and input to functional needs, roles and competencies. We have supplied this by our general knowledge on organizational matters, especially regarding dimensioning, structuring and staffing of the type of organization OpenEPI will be.

6.2 Functions

We suggest several specific roles and functions for the OpenEPI organization, as described in the table below.

#	Role	Description
1	Team lead	Have responsibilities to facilitate daily standup, team backlog, prioritization, cross-team status meetings. Will also be a developer in any discipline
2	Tech lead	Technical responsible person, per team. Will function as team architect and responsible for following guidelines and technical roadmap as agreed with enterprise architect. Will have a technical speciality suitable for each team. Will also do hands-on development.
3	Platform engineer	Expertise in building cloud platforms.
4	Devops engineer	Expertise in streamlining developer experience
5	Data engineer	Data expert. Analyze and assess datasets. Do data and analytics development.
6	Backend developer	API developer, and general backend developer
7	Frontend developer	Expertise in JS/Typescript and proficient in web development
8	Product manager	

9	Technical architect	
10	Data architect	Responsible for data quality across teams.
11	UX Designer	Expertise in designing user interfaces
12	Head of unit / CEO	Management, decision making at business level
13	Administrative resource	Administration, supporting the overall business
14	HR consultant	Recruitment, competence planning, HR functions
15	CFO	Funding, financial issues, accounting
16	Legal expert / advisor	Competence on contractual arrangements with data providers and cloud services, SLA's (when relevant), IPR, privacy, relevant regulations, policy maintenance, standards, etc.
17	Communications	Stakeholder engagement, media contact, event management, SoMe & web presence, general branding of the initiative both globally and locally in LMICs
18	Marketing and user mobilizing	Responsible for marketing activities, hunting new startups, developers and other users, hunting relevant use cases, monitoring user uptake and support novice users
19	Customer handling	First line helpdesk and responding, handling of requests and queries, crowdsourcing and feedback mechanisms
20	Domain expertise	Specific support and subject matter consultancy and advisory services on data quality, relevant use cases and applications in different domains: agtech, biodiversity, soil, weather/flood etc.
21	Learning resource manager	Responsible for CMS and the governance of OpenEPI produced learning resources, guidelines, instructions, course activities etc.

In the descriptions of the organization in the next sections, we refer to these functions - where some functions may be populated by several employees.

In addition, it is crucial for OpenEPI to build relationships with existing data providers and portal owners in this "market", such as WRI, Copernicus, Yr.no and so on. Generally, OpenEPI must have the capacity to collaborate with others, to ensure access to relevant expertise, as well as to relevant data resources supporting the OpenEPI offerings.

Further, we suggest that the entity leans on an external strategic board, composed at least partly by members with academic credibility, from the domains that are covered by the platform offerings (i.e. soil, water, agriculture, forest, biodiversity). This board will be advisory, for strategic decisions, ethical discussions and other contextual and "political" issues.

6.3 Size and structure

We suggest an organization built up around technical teams dedicated to infrastructure and the specific data provisions, respectively. The technical teams are supported by management and staff, and experts covering cross-cutting topics. Furthermore, we suggest that OpenEPI as an organizational entity grows both in size and budget, from a gentle level to start with - to a maximum in 3-4 years, depending on the results from measuring and evaluating actual user uptake and success. Narrow follow-up by metrics on user uptake and on user feedback should be imperative.

6.3.1 Initial setup

Our assessment of possible organizational setups concludes with an initial size of the entity of approximately 22 FTEs, and a development pathway that ends at approximately 60 FTEs. Central for the OpenEPI organization will be the technical teams for governing the different data provisions, including data source identification and retrieval, data quality measures and routines, API management and description of data to the developers using the OpenEPI services. The number of teams will grow due to the expanding span of OpenEPI offerings across different thematic data of relevance for the developers. As a starting point for the organizational setup, we suggest an organization of 22 FTEs, structured in three technical or data management related teams, supported by one cross-technical team and managerial and other, cross-organization functions and non-technical expertise.

Platform team	
1	Tech lead/Developer
2	Platform engineer
3	Platform engineer
4	Devops engineer
Data/Product team 1	
5	Tech lead/Developer
6	Data engineer
7	Data engineer
8	Backend developer
Product team 3 - Demonstrators/Developer portal/Documentation/Client libraries	
9	Tech lead/Developer
10	Frontend developer
11	Backend developer
Cross-tech team functions	
12	Product manager
13	Technical architect
14	UX Designer

Management and admin

15	Head of unit
16	Administrative resource
17	CFO, financing/funding
Other cross-organization functions	
18	Lawyer, legal expert
19	Communications and media responsible
20	Marketing and general user mobilizing, analytics on metrics
21	User/customer handling
22	General domain expertise (climate change mitigation)

In our views, this setup could be seen as an absolute minimum viable configuration of OpenEPI as a provider of OpenEPI data and services.

6.3.2 Full scale operations

As an end point for the development of the OpenEPI organization, we suggest 60 FTEs, organized in seven technical or data management related teams, supported by one cross-technical team and managerial and other, cross-organization functions and non-technical expertise - as follows. Some of the teams could be located outside the core body, in a geographically distributed manner, but still under the same OpenEPI conceptual frame and interconnected with the necessary legal and administrative arrangements.

Platform team 1 - AWS	
1	Team lead
2	Tech lead - AWS Expert
3	Platform engineer
4	Platform engineer
5	Devops engineer
Platform team 2 - GCP	
6	Team lead
7	Tech lead - GCP Expert
8	Platform engineer
9	Platform engineer
10	Devops engineer
Platform team 3 - Azure	
11	Team lead
12	Tech lead - Azure Expert
13	Platform engineer
14	Platform engineer
15	Devops engineer
Data/Product team 1	
16	Team lead

17	Tech lead
18	Data engineer
19	Data engineer
20	Backend developer
Data/Product team 2	
21	Team lead
22	Tech lead
23	Data engineer
24	Data engineer
25	Backend developer
Data/Product team 3	
26	Team lead
27	Tech lead
28	Data engineer
29	Data engineer
30	Backend developer
Product team 4 - Demonstrators/Developer portal/Documentation/Client libraries	
31	Team lead
32	Tech lead
33	Frontend developer
34	Frontend developer
35	Backend developer
36	Backend developer
Cross-team functions	
37	Product manager
38	Technical architect
39	Data architect
40	UX Designer

Management and admin	
41	Head of unit
42	Assisting head of unit
43	Administrative resource 1
44	Administrative resource 2
45	HR consultant
46	CFO, financing/funding
Other cross-organization functions	
47	Lawyer, legal expert 1
48	Lawyer, legal expert 2
49	Communications and media responsible
50	Marketing and general user mobilizing, analytics on metrics 1

51	Marketing and general user mobilizing, analytics on metrics 2
52	User/customer handling 1
53	User/customer handling 2
54	Learning resources manager, responsible for demand stimulating measures 1
55	Learning resources manager, responsible for demand stimulating measures 2
56	Learning resources manager, responsible for demand stimulating measures 3
57	Domain expertise, differentiated data domain 1
58	Domain expertise, differentiated data domain 2
59	Domain expertise, differentiated data domain 3
60	Domain expertise, differentiated data domain 4

Although the dimensioning is different, both versions of the organization will cover the personnel resources and competence necessary for infrastructure operations; data management; statistics and data analysis; production of aggregated datasets; software development; licensing and other legal issues; information security; user contact, support and consultancy/advice; communication, external relations and marketing.

6.3.3 Recruiting and establishment

In order to build the new organization, the grant receiver will have to recruit the professional competencies for covering infrastructure operations; data management; statistics and data analysis; production of aggregated datasets; software development; licensing and other legal issues; information security; user contact, support and advice; communication, external relations and marketing. To start with, some of the roles may be temporarily sourced to ordinary consultancy and tech services providers in the market, but eventually the organization needs in-house resources.

The funding of the new organization must be further assessed. Norad could fund it directly over their budget, or more probably, under a multi-annual grant. We suggest that a small consultant-supported team get the responsibility to recruit the first key resources in the new permanent core body, transferring knowledge and arrangements from the present pre-project organization in Knowit, and arranging for practical matters like initial budget, office lease and so forth.

We assume that OpenEPI from the start will be a relatively small scale initiative, estimated to be about 22 FTEs and an annual budget of NOK 35-45 million. We suggest that the building of this first core unit is defined in a 2025 call, and funded by one single, but long-term grant. Regardless of how this business is organized, we expect it to need to scale up rather quickly to meet requests for more data sets and derived data products, more services, and not least, a more complex landscape of stakeholders and user needs. Flexibility and scaling possibilities for the OpenEPI organization will therefore be an important criterion for evaluating the proposals under the first call. We think it will be crucial to plan for some kind of multi-annual grant (like GRID-Arendal, also funded by Norad), creating a stable financial situation for the core OpenEPI operations.

OpenEPI will be governed under the laws and legislation of the specific country of the grant receiver – though there should be developed a specific legal framework regulating OpenEPIs

legal responsibilities. International regulations ensuring data security and privacy, like EU's GDPR regulations, should also be taken into consideration. The same counts for Open Data Directive, the definitions regarding High Value Data and eventually also future derived national regulations regarding mandatory sharing and reuse of open data.

Other platform initiatives have experienced a distributed organization to be feasible. We also suggest that OpenEPI is built as one core unit but thematically and geographically distributed. For instance, could existing initiatives (like WRI or ISRIC) be assigned to take care of some functions or domain-specific services, in accordance with the core unit. Anyway, the core unit should hold all necessary assets and capabilities and possess the technical expertise to manage and maintain the platform. If feasible, some technical services could be outsourced to market players. Further, it will be imperative for the OpenEPI organization to gradually - as it expands - build partnerships and to collaborate with important stakeholders among data providers, governments, NGOs, other platform owners, possible funders and so forth. OpenEPI could be seen as a platform-of-platforms, building partnerships and tight relationships to other platforms (and their stakeholders).

There will indeed be a need for dedicated efforts on building a common understanding of the OpenEPI open policies; the goals on local innovation and climate change adaptation; the importance of an open source, vendor independent technology stack; on how to develop the data ecosystem; and not at least how to strengthen the ability in the different countries to utilize open climate and nature data in their efforts on climate change adaptation. It will be a continuous task for the OpenEPI organization to explore potential partnerships with other organizations, data providers, or technology vendors to enhance the value proposition of OpenEPI, and thus expand the reach within the developer community.

6.4 Governance

In this context, governance refers to the framework, processes, and policies put in place to ensure that data and services are managed, controlled, and utilized effectively, securely, and in accordance with the overarching goals for, and the intentions behind, the initiative. This includes aspects such as securing data quality, information security, compliance with the openness requirements, regulations and standards, access control etc.

Effective governance will ensure that data is reliable, and the services from OpenEPI are trustworthy and available for developers while also mitigating relevant risks (see chapter 7). The recommendations below are based on our interviews with existing data platforms and are also inspired by experiences from previous data sharing projects in Norway.

6.4.1 Democratic values and OpenEPI approach

The way we suggest organizing ownership and governance of the platform, data management and services, is influenced by our Nordic point of departure and a true altruistic approach. The creation of a safe haven for climate and nature data is paramount in our mission to address the pressing challenges of climate change. Climate, nature and environmental data are invaluable resources for understanding the intricate relationships between environmental factors and human livelihoods. These data empower people everywhere to make informed decisions, develop effective policies, and implement strategies

for climate adaptation and conservation. Good governance based on democratic values is essential because it ensures that the actual data resources are managed transparently and equitably. It upholds principles of fairness, accountability, and inclusivity, enabling diverse voices and perspectives to be heard.

If applicable, decisions related to data collection, storage, and accessibility by OpenEPI should be made through processes where a wide range of stakeholders, including marginalized or vulnerable communities, are involved. Robust, democratic governance helps prevent data manipulation and misuse, safeguarding the integrity and authenticity of information.

Data diversity will be secured by the fact that OpenEPI will collect information from various sources and ecosystems, allowing us to build a comprehensive and nuanced understanding of the complexities in the local efforts on climate change adaptation and the securing of food production chains - and the data needs implied by these efforts. Data integrity will be utmost important for OpenEPI. It means that the data shared by or republished on the platform is trustworthy and not tampered with. This integrity ensures that data is reliable and can be used for innovation (and other use) with confidence. Satisfying data availability will also be crucial, as long as local business can become dependent on data streams - for instance in fine granular lookups or real time streaming. Data from OpenEPI should be accessible to those who need it, regardless of their geographical origin or technological infrastructure. OpenEPI should collaborate with relevant research communities to strengthen its professionalism and develop the ability to verify the quality and relevance of the data distributed to developers.

Furthermore, storing data in a cloud-agnostic and vendor-independent way is essential for the sustainability and long-term accessibility of OpenEPI. Cloud-agnostic storage means that OpenEPI data will not be tied to a specific cloud provider or vendor, ensuring that data remains accessible even if the cloud infrastructure at any specific time needs to be exchanged by another infrastructure. This approach prevents vendor lock-in and ensures that the data remains open and available for innovators globally, for the long term.

6.4.2 Prioritizing mechanisms

OpenEPI should implement a mechanism for deciding and prioritizing the backlog of requests for new services, data and datasets. Assuming that OpenEPI will be a success, the initiative will experience a never-ending stream of requests for new or better services and products. OpenEPI should of course listen to the developer community and follow their feedback and legitimate requests due to their efforts in supplying their local markets with innovative software products and digital services. As the resources get under pressure and OpenEPI doesn’t scale anymore, decisions on the expansion of offerings will be a priority task for the OpenEPI management. However, we suggest an extensive use of the previously suggested external strategic board - at least in discussing and anchoring the criteria for prioritizing (i.e. how to decide if high quality vegetation coverage data is more “important” than high quality soil data). Effects from using an external board to discuss criteria are decision robustness, anchoring and trust.

Among other perspectives, the criteria should reflect the “value” of different types of data in respect of climate change adaptation or innovative potential. To limit scope the definition of EUs High Value Datasets can be re-used as a criterion or prioritization mechanism. In any case, as far as possible criteria and prioritization processes should be transparent.

The governance of OpenEPI will include the continuous maintenance of an open policy for OpenEPI, as this will be crucial for OpenEPI’s position in the open data landscape. The governance also includes necessary gatekeeper mechanisms to ensure consideration of scalability, support capacity and a development of the platform that is in line with overall priorities and long-term goals.

The mechanism for governing the offerings from OpenEPI should be further assessed, exploiting other data platform owners’ experiences from their governing models.

6.4.3 Policy and regulatory framework

A key governance mechanism is the openness requirement, which is laid down in a policy for which the pilot project has prepared a beta version (www.openepi.io/resources/open-policy), also included in the report as annex 3. It is envisaged that our beta version of the policy will be further developed by a permanent OpenEPI, and then be an imperative norm for anyone who will use OpenEPI to share their data to developers, or others who want to utilize OpenEPI as a normative instrument or in some other way. For instance, development aid organizations and others who fund climate or nature data production can refer to OpenEPI and demand grant receivers to follow OpenEPI's principles, policies and standards - to secure findability, usability and fair use of the data. The OpenEPI policy is based on open data principles, FAIR principles and CARE principles. It is also inspired by the EU's Open Data Directive and the supplementary High Value Data list (describing requirements for especially valuable data sets from the public sector).

In general, it is often the case that copyrightable works produced for general and non-profit purposes, such as those for climate change adaptation and mitigation, are not openly licensed and disseminated widely to the public. We have found that many stakeholders and other members of the public are generally not aware of the resources created as a result of aid programmes, research projects and other international initiatives. Intentionally or unintentionally, the resources are often created and disseminated locally or disseminated to limited audiences. Even when the resources are known to exist, stakeholders and the public are not sure how to access them, what usage rights or permissions are necessary to use them, or how to obtain those rights or permissions.

The policy we suggest addresses these key problems. It explicitly gives permission to the public to access, reproduce, publicly perform, publicly display, and distribute the copyrightable work; prepare derivative works, and reproduce, publicly perform, publicly display and distribute those derivative works; and otherwise use the copyrightable work, provided that in all such instances attribution is given to the copyright holder. We believe that the implementation of this policy will result in significantly enhanced dissemination and use of deliverables created and provide stakeholders and members of the public with a simpler and more transparent framework to access, use, and modify these deliverables for the benefit of their communities.

The OpenEPI policy and its requirements apply to all data created or funded by OpenEPI and all data and other works distributed or made available from other sources, on the OpenEPI data portal, or in any way referred to or recommended by OpenEPI. As a digital innovation platform and digital public good, the OpenEPI policy will:

- enable the unrestricted access (except for an attribution requirement) and reuse of all collected or produced digital data created or funded by OpenEPI, and all data linked or referred to, or disclosed, reused, recombined, recommended or shared by OpenEPI, including any underlying data sets in projects that collect or produce data, or are in any way made available by OpenEPI.
- enable the unrestricted access and reuse of software, algorithms and models created or funded by OpenEPI, and all software, algorithms and models linked or referred to, or disclosed, reused, recombined, recommended or shared by OpenEPI.
- enable the unrestricted access and reuse of all peer-reviewed published research created or funded by OpenEPI, including any underlying data sets.
- enable the unrestricted access and reuse of educational resources created or funded by OpenEPI.

The policy includes licensing specifications, requirements regarding API implementation, open standards for metadata and file formats, and reference application. The policy will apply both to the deliverables themselves and any support materials necessary to the use of the deliverables. As part of the governance mechanisms and the routines in handling stakeholder relations, OpenEPI will not only review, but also openly endorse the data, software, algorithms, models and research publications linked or referred to, or disclosed, reused, recombined, recommended or shared by the platform.

More on metadata, metadata schemes, and other technical or documentation requirements on data in chapter 5.

When it comes to the regulatory landscape we must navigate in, we assume we must adapt to national legislation and regulatory measures - even if the target group for OpenEPI is situated in sub-Saharan Africa. Thus, OpenEPI must comply with existing and new regulations, and we have to consider more in depth how actual legislations will develop ahead. For instance, we still do not know exactly how EUs many directives relating to data and AI will be implemented in European countries, also affecting potential grant receivers in Europe. Implementing legislative work is still on-going, for instance the implementation of EU's Open Data Directive, or the NIS2 framework under security laws. Our point of departure is that OpenEPI - of course - must comply with the legislation in the country where data is stored, independently of where the organization pointing at those data is localized. It's relevant here to distinguish between data and services. The services from the organization must also comply with the legal and regulatory frameworks for data privacy, data security, intellectual property rights, and any other relevant regulations in the jurisdiction where it is legally situated. We suggest an initial legal assessment sorting out the relevant adaptations to different legal frameworks, aiming at giving advice on how to adjust the offerings from OpenEPI. We also draw the attention to the legal and ethical risks mentioned in chapter 7.

6.4.4 OpenEPI services

The OpenEPI organization will run a data platform and the necessary middleware to integrate different open source products and components. To run the platform includes optimizing performance, governing and performing information security measures, monitoring and reporting metrics, handling integration issues for new data sources and maintaining the front end solutions serving developers and others.

The OpenEPI data teams will deliver quality ensured data on climate, nature and environmental topics of relevance for farmers and others involved in food production in sub-Saharan countries. OpenEPI is not aiming at end-users like farmers or others involved in farming or food production. OpenEPI aims at local developers, startups and innovative businesses, addressing the mentioned end users by their commercial or other digital products.

OpenEPI will deliver data primarily through open and well documented APIs. The APIs and adjacent guiding materials enable easy access to data, fostering innovation and local development of new services.

The OpenEPI teams will govern aggregate services that are partly developed by partners but are included as part of the OpenEPI platform. The weather data from Yr.no is an example, as being offered by MET, but also a part of the OpenEPI offerings.

Together with the data providers, OpenEPI will design and run processes for data quality assurance, including work on assessing and documenting accuracy and integrity of the data shared by or produced on the portal, including activities on data quality improvement (correcting, adjusting, complementing, etc.), and including updating the portal with new algorithms. All partners and data providers delivering data to OpenEPI will have to follow OpenEPI’s open policy.

OpenEPI will have to develop and enforce robust data privacy and security policies to protect potential sensitive information. This includes data on endangered species or biotops. As part of this, OpenEPI must safeguard against data breaches and unauthorized access while respecting individual privacy rights.

Finally, as an organization OpenEPI must put effort in demand-side activities, securing the user uptake. This should be further strengthened by separate demand side programmes, including marketing, training and incentivizing measures. We think this is a crucial part of the work to be done, to secure impact on both climate change adaptation and local innovation. More on this in the next section.

6.5 Stakeholder relations

As already emphasized, the relation to and collaboration with important stakeholders will be crucial for OpenEPI’s legitimacy and efficiency. it will be necessary to establish and develop collaborative relationships and institutionalized joint arrangements with external partners, to increase either the data offerings, the data quality, the user uptake or the legitimacy of the entire concept.

We suggest marketing and promotion activities to be a substantial part of OpenEPI activity, especially the first few years. Our estimates of organization dimensioning, and suggested functions above also reflect this. Still, OpenEPI cannot be dimensioned to cover all there is to be done in this area. We suggest Norad address the need for “market development” by parallel capacity building programmes and initiatives targeting different markets and regions. Especially important is the raising of awareness of OpenEPI offerings and encouraging adoption by developers. Existing platform owners we have been in contact with, advise us not to underestimate the work that is needed to adopt new users, including social media engagement, marketing in local written media (in local languages), arrange developer outreach events, and establishing partnerships with relevant industry influencers (i.e. leading agtech vendors in Africa). This will truly not be an easy task, we think. We must consider African countries to have rather varying “innovation climates”, thus demanding differing measures and intensity of activities. Probably, OpenEPI must also depend on local partners to be “ambassadors”. We suggest Norad elaborate this more in detail before deciding on full scale implementation of OpenEPI.

Stakeholder relations also include how OpenEPI delivers its services. Although OpenEPI will not have any formal Service Level Agreements (SLA), the service level and OpenEPI organization’s handling of users should follow a certain predictable - and openly declared - standard. The OpenEPI core entity will have to decide the level of customized services, opening hours (considering an eventual global service), volume on consultancy services and so on.

A part of the image and reputation building related to service level, is the level of user testing of new data and services, before launching new offerings, and what kind of feedback mechanisms OpenEPI arranges for the different offerings.

Stakeholder relations include demand side activities like support and training, to help developers effectively use the data portal and troubleshoot any issues they encounter. It will be wise for OpenEPI or any complementary initiative to invest in the capacity building of individuals and organizations at the demand side to enhance their understanding of available data, best practices in use, relevant standards, and ethical considerations.

OpenEPI should also collaborate with governmental bodies (where and when applicable), international non-governmental organizations, research institutions present in Africa, and private sector partners holding interesting data, technologies or use cases. A widespread collaboration with others is time-consuming, but rewarding when it comes to reputation, user uptake and impact. We suggest dedicated programs for developers and start-ups (perhaps differentiated between countries or regions) to ensure re-use of OpenEPI data and services, where OpenEPI seeks specific funding to support the activity, and where one goal should be institution development in the partner countries. An important part of the stakeholder relations is about chasing donors and funding opportunities by all kinds of stakeholders, without ever compromising the open policy of OpenEPI.

A partnering program could be extended to reach the different stakeholder categories, more generally. Thus, partner programs for data providers, donors, countries, research institutes and community partners (startups, universities/students, NGOs) respectively, could be arranged for. To support this, it should also be considered to develop high quality learning

resources for all the target groups - including developers, startups, government agencies and UN organizations.

User adoption will of course be critical for the success of OpenEPI. High user adoption demonstrates the platform's impact and relevance. It reinforces the importance of the platform in addressing critical issues related to climate change adaptation and mitigation, thereby attracting more users and stakeholders. With more users engaging with the platform, there is also an increased capacity for monitoring and evaluating the effectiveness of implemented strategies. This feedback loop allows for continuous improvement and adaptation based on real-time data and user experiences.

Ultimately, high user adoption demonstrates the value of OpenEPI. It supports funding, sustainability, and future development by showcasing the platform's relevance and usefulness to governments, donors, data providers and other stakeholders.

By design, one major challenge for OpenEPI will be to measure its impact (on climate change adaptation and on the local innovation system). This raises the question on what kind of metrics and KPIs OpenEPI should implement. Inspired by other known open data initiatives, some suggestions on possible KPIs are:

- Number of visitors,
- Number of downloads,
- Demand and frequency of use of data,
- Number of data providers offering their open data to OpenEPI
- Number of applications that have arisen from the use of data from OpenEPI.

All those indicators are possible KPIs for OpenEPI, although some of them can be difficult to extract. Furthermore, one should bear in mind that metrics on the use of data does not measure impact. Arguing for impact from open data without looking at actual social, political, economic or environmental effects has not much meaning.

6.6 Cost estimates and funding model

Based on our assessments and estimates, OpenEPI will be a rather small, but still substantial business. As a most preliminary budget estimate, we suggest NOK 35-45 million as a starting point. This figure is based on an organization dimensioning of 22 FTEs (times an average of NOK 1.1 million per FTE, including overhead) or an annual personnel cost base of NOK 24 million. To this, we must add substantial ICT related costs, in addition to office expenses, traveling costs and so on. We suggest this “basic” organization to be financed by one single, multi-annual grant of NOK 35-45 million.

As calculated in previous sections, we suggest a flexible organization size, according to new challenges and user needs, and the following needs for data sets and new functionality supporting new topics and domains. We suggest a midpoint personnel cost of NOK 43 mill., up to NOK 65-70 mill. at full scale. Supplementary to these figures, there will be substantial ICT related costs, in addition to office expenses, marketing and traveling costs, and so on. A full scale operational OpenEPI will have an estimated budget of up to NOK 90-100 mill. annually.

We suggest that OpenEPI in the long term is funded partly by partners in the ecosystem, thus channeling efforts in new data sets to where the needs are prioritized and where someone is willing to support new activity. For instance, agreements with FAO, WRI or others, and targeted grants to initiatives in Africa related in some way to climate change mitigation, could bear costs also for OpenEPI operations. OpenEPI will in some degree have to chase these funds from international organizations, as well as from authorities in other countries or private donations. There should always be transparency to developers and other users who are at any time financing OpenEPI data and services, as transparency should indeed be a fundamental value for OpenEPI. Here, it will be wise to draw on the experience from other platform projects and how they have ensured enough funding and handling of external funders, also in the long term.

As a starting point, OpenEPI will focus on data and technology of relevance for regions in Africa. In a longer perspective, the datasets that become part of OpenEPI's offerings should not only be relevant for Africa, but for all severe climate change-exposed parts of the world. This further scaling will also entail a gradually broader foundation on the funding side as well.

We think that by combining predictable Norad funding and a more flexible partner- and community funding, we will balance robustness in the core business with the challenge- and demand-driven search for new data sets and data products. Although this could be a rather realistic funding strategy, we suggest that a more detailed and realistic business plan is provided before the final decisions on OpenEPI implementation

7. Key risks and mitigating strategies

Addressing the complexities and challenges associated with operating an open data platform like OpenEPI - and reaching the intended outcomes - requires a nuanced approach. That is why this chapter, concerning key risks, has a rather comprehensive form. Our concern entails not only recognizing and understanding the multitude of risks - ranging from operational and financial, to data quality, user adoption, privacy, security, technological infrastructure, ethics, gender equity, legal and regulatory, and reputational - but also implementing strategic and thoughtful mitigation measures, addressing those risks specifically. Addressing and mitigating risks is crucial to ensure the public trust and the successful running of the platform, to obtain lasting funding, and to make sure that the use of the data is responsible, respects human rights, and promotes social good.

In the following we outline the internal and external risks that we have assessed, and that may affect or result from the realization of the OpenEPI initiative. If Norad decides to implement a full scale OpenEPI, we claim that a responsible handling of these risks will be of utmost importance for the ultimate success of the initiative.

7.1 Summary of key risks

A summary of key risks and their mitigation strategies is given below, emphasizing the holistic management necessary for the platform's success and trustworthiness.

To effectively mitigate risks related to confidentiality, integrity, and availability, OpenEPI should employ a variety of strategies tailored to address specific vulnerabilities – as shown in the table below.

By addressing these risks through comprehensive strategies, OpenEPI can not only mitigate potential negative impacts but also strengthen its position as a trusted, valuable, and inclusive open data platform - as a digital infrastructure component to count for. The platform's success hinges on its ability to adapt to evolving challenges (in the nature management, food chain and climate change domains), maintain high standards of data quality and ethics, and foster a supportive, engaged user community locally in low and middle income countries. Achieving these objectives requires ongoing vigilance, collaboration, and a commitment to continuous improvement and innovation.

Risk type	Mitigating strategies
A: Operational and financial risks	• Develop a skilled workforce and continuously retain talent. • Secure long term funding (either national budget funding or funding by multi-annual grants). • Diversify funding sources to ensure financial stability. • Robust planning for scalability and future growth.
B: Data quality risks	• Implement robust data governance frameworks. • Engage domain expertise in the data curation process. • Establish continuous monitoring and feedback mechanisms from developers and end users.

C: User adoption risks	● Ensure data with high quality and relevance. ● Enhance platform usability and technical functionality. ● Provide support and training resources for users.
D: Privacy and security risks	● Adhere to European GDPR and other data protection regulations. ● Implement robust anonymization techniques and data governance.
E: Security and safety risks	● Encrypt stored data and data in transfer (SSL) ● Collect anonymized data on use ● Utilize checksums for datasets ● Collaborate with domain experts at data providers ● Prevent DDoS at edge ● Use of multiple cloud vendors
F: Data ethics risks	● Adopt the CARE Principles for data ethics. ● Conduct thorough risk assessments and audits. ● Educate users of the platform on responsible data use.
G: Technological infrastructure risks	● Keep abreast of technological advancements and updates. ● Reduce reliance on single vendors through open source, diversified services and local partnerships
H: Gender equity risks	● Ensure that OpenEPI is managed by a gender balanced team. ● Stimulate initiatives aimed at increasing data analytics and software development skills among female students and in local, female led start-ups. ● Select data types, data products, use cases and services which truly address needs of both genders in the actual regions.
I: Legal and liability risks	● Regularly consult with legal counsel to stay informed and compliant with regulations. ● Recruit legal competence to the OpenEPI staff. ● Consult with Creative Commons on the implementation of open licenses.
J: Reputational risks	● Develop a crisis response strategy and focus on equity and inclusion. ● Engage with local communities and ensure the right perspectives in data selection, collection and governance structures.

7.2 Operational and financial risks

The term “operational risks” in the context of managing an open data platform like OpenEPI, is multifaceted and spans various dimensions including operational, financial, and strategic risks.

Operational risks are associated with the day-to-day running of the platform and can encompass anything from technical failures, data breaches, and system downtime, to inefficiencies in data management and platform maintenance. These risks can directly impact the platform's reliability, performance, and user trust - and the entire initiative. Much of the

risks are about the ability to build and maintain a professional, capable and service oriented organization. To ensure this, the organization must be well managed, competent, equipped with sufficient resources and run professionally and efficiently as an organization.

As long as the entity is basically funded by stable mechanisms, under long term conditions (like multi-annual grants from development agencies), funding issues are marginal and a question of political an international prioritizing. It also counts for the strategic role of an OpenEPI. As long as stimulating local innovation and combating climate change mitigation are crucial components of the development aid policies by most development coordination agencies, OpenEPI will in many ways have an obvious mandate and role.

Some of the operational risks are of course also tightly related to the risks covered later in this chapter, for instance data quality risks, security and safety risks or reputational risks.

7.2.1 Financial risks

Securing sustainable funding will be pivotal for OpenEPI. An over-reliance on volatile funding sources or insufficient budgeting can compromise the platform's operational stability and growth. Our advice is that the long term existence of OpenEPI is secured by a basic funding by long-term grants and budgets. On the other hand, cost control will be crucial, especially concerning the operational costs associated with data transfer and using cloud services, which can escalate quickly as the platform scales. The tech teams will have architectural planning for optimal data transfer volumes as one of their roles, as this will affect the operational costs, as well as the overall carbon footprint of the entire OpenEPI initiative.

Financial risks also encompass the potential lack of funding for further scaling of the services by new datasets - as well as excluding datasets that are no longer relevant or demanded. Limited resources and capacity for quality and relevance evaluation, collection and “cleaning” of datasets related to specific new user needs, may result in dissatisfied users or in being drawn to less accurate and comprehensive datasets. This also counts for the necessary update, maintenance and perhaps expansion of the technological infrastructure, and demand for new data set or service offerings (i.e. AI models or training data). Long term sustainability for OpenEPI requires ongoing resources for both data and platform maintenance and updates. Without this, the data offerings and the OpenEPI services may become outdated or obsolete, and OpenEPI irrelevant for developers serving local needs.

Financial risks are also related to essential operations like cybersecurity measures, work on regulatory compliance (to data protection laws, IPR and other regulations), campaigns and other stakeholder engagement activities. The OpenEPI entity must have enough financial flexibility to handle the costs from scaling and “extended” activities, even if a major part of the total cost is related to personnel and IT costs for the running operations. We suggest that OpenEPI is given the opportunity to attract extended funding from partners and other external funders. How this could be arranged for in the context of the conditions given under the grant by Norad, is to be further assessed through discussions with the appropriate government bodies.

7.2.2 Capacity and scalability risks

The platform design and architecture must support the expansion of the OpenEPI offerings to new datasets, new data products (i.e. large datasets for AI training) and an increase in the traffic, related to growing user numbers or new user habits. As the platform eventually grows, OpenEPI must also efficiently manage and estimate operational and infrastructure costs. A sudden spike in usage or data volume can significantly strain resources. Ensuring that the infrastructure can support massive growth spikes and maintaining scalability across different operational domains are crucial. Those domains could be for instance cybersecurity issues or API management and maintenance, and most certainly stakeholder engagement activities.

Challenges also arise from the potential lack of knowledge and demand for innovation in open data, particularly in developer communities of the low and middle-income countries that OpenEPI addresses. Certainly, there are also possible challenges from the lack of policy implementation at the supply side. One concern is the efficiency of the incentives on the data dissemination side: are data producers in the climate, nature and environment domains really providing open data according to the OpenEPI policy – easily reusable and valuable for software developers in sub-Saharan countries? We think it will be necessary to be strict on the sanction side in this. Norad and other actors involved in development aid and coordination should use their economic grants as sanctional instruments, demanding compliance to open and data quality principles, laying the ground for enhanced data sharing and for more data driven innovation. The question is to what degree OpenEPI will succeed in maintaining the necessary awareness on this from Norad and other development coordination agencies, in the long term.

Furthermore, mitigation measures include planning for scalable growth, engaging in capacity-building efforts, and fostering partnerships to enhance knowledge and both supply of and demand for open data. Here, one must keep in mind that partnership and knowledge building activities are costly and time consuming - as the experiences from some of our sources clearly show.

7.2.3 Competence

The success of an open data platform hinges on the acquisition and retention of skilled personnel, including technical, security, and domain-specific expertise - as described in chapter 6, and implicit in chapter 5. High turnover or a scarcity of talent can surely lead to operational inefficiencies.

As a mitigation strategy for the competence challenge, the entity should be localized to a major city, near to educational institutions able to serve as a supplier of relevant expertise. This will also make it possible for the organization to hire external capacity from tech partners and consultants, as a supplement to those who are permanently employed.

Further, to some degree the mitigation to the competence challenges must also include playing on values. The entity must be marketed as a unique and outstanding organization with a very meaningful, global mission, thereby playing on the employees' value-based assessments. This is also about good and proper management, as well as how the

Norwegian national authorities contextualize the initiative in the preparation and political decision process for the initiative.

7.2.4 Stakeholder involvement and support

Engaging with stakeholders—from governments and international bodies to the open community and commercial entities—presents both opportunities and risks. Effective management of these relationships is essential but requires resources for coordination and negotiation. OpenEPI may face challenges in securing political support, navigating skepticism from the global south, or combating resistance from commercial business models.

Mitigating these risks involves proactive stakeholder engagement, transparent communication, and demonstrating the value and impact of the platform to garner broad support. One important mitigation measure will be to keep some sort of uniqueness to OpenEPI. The entity should avoid redundancies, by focusing on confirmed user needs and ensuring that new datasets and data products are complementary to and not redundant with existing efforts from other data portals and “competitors”. When possible, it will be strategically wise to build on other, existing initiatives and platforms, not competing but rather joining them in win-win collaborations.

7.2.5 Mitigating strategies

To sum up, by implementing the following mitigation measures OpenEPI can navigate operational, financial, and strategic risks, ensuring a resilient platform capable of driving innovation in the domains of nature data and climate change mitigation. The aim is a strong foundation for sustainable growth, stakeholder engagement, and continued relevance in the face of evolving global data needs:

- **Diversification of funding sources:** Even if we suggest the basic funding to be secured by long-term funding from Norad, OpenEPI should seek funding from a variety of sources including relevant government grants, private sector partnerships, philanthropic organizations, and user donations - all to reduce the long-term dependency on a single source.
- **Cost control and financial planning for compliance:** Implement strict budget management practices and regularly review expenditures to optimize resource allocation. Consider adopting cloud solutions that offer scalable and cost-effective infrastructure. Allocate funds for regulatory compliance and cybersecurity within the operational budget, this includes investing in data protection measures and adhering to international standards.
- **Scalable infrastructure:** Design the platform with scalable architecture from the start, allowing for easy expansion in response to increased demand. Utilize cloud services and establish an architecture that can dynamically adjust to traffic and data load.
- **Talent management:** Invest in talent acquisition and retention strategies, such as professional development opportunities and a positive work culture. Play on the value-based perspectives of OpenEPI as unique, having an important global mission, and established to make a difference.

- **Expanding expertise:** Continuously develop the platform team's expertise in climate, environmental, and nature domains through professional development, collaborations with research institutions, and hiring subject matter experts.
- **Stakeholder management:** Develop a framework for managing partnerships that includes clear communication channels, mutually agreed-upon goals, and regular review meetings to ensure alignment and address any issues promptly.
  Foster inclusive dialogues with stakeholders from the global south to understand their perspectives and needs, addressing skepticism towards external initiatives. Tailor the platform's features and data sets to meet these specific requirements.
  Identify potential sources of resistance early and engage directly with these stakeholders to understand their concerns and explore collaborative and win-win solutions. Highlight the mutual benefits of open data and seek common ground.
- **Community building:** Cultivate a strong open data community by hosting workshops, hackathons, and forums that encourage participation from various stakeholders, including governments, NGOs, private and public data providers, and the tech community.

7.3 Data quality risks

The information that can be extracted from data indeed depends on the quality of the data. Poor-quality data will almost always lead to poor data analysis and results. OpenEPI's ability to offer high-quality data to the developers is therefore crucial to the platform's success and ability to reach its objectives.

What constitutes good data quality is not clearly defined for nature or climate data. In our report, we have used OECD's definition of data quality where quality is defined along seven dimensions: *relevance, accuracy, credibility, timeliness, accessibility, interpretability, and coherence*[70]. The OECD's definition considers that data quality depends on the *intended use* of the data: data that are of good quality for certain applications can be of poor quality for other applications. In other words, even if data are of good general quality, their use can lead to wrong results if the data are irrelevant and do not fit the business or scientific questions they are supposed to answer. Data quality hence needs to be viewed as a multi-faceted concept, considering the *specific context* of data use.

In our interviews with open data platform owners and managers, they have emphasized the critical link between their success and their capacity to supply high-quality, curated data sets. Being a consistent provider of high quality data is, as highlighted by several of the data platforms, vital for the platform to obtain and preserve trust among both collaborators and users.

According to insights provided by the OECD, one significant factor behind the underutilization of open data platforms, compared to the expectations of many open data providers, is the uncertainty regarding the quality of the data they offer.[71] If OpenEPI does not ensure the provision of high-quality data, this shortfall could significantly deter users from interacting with

[70] Quality framework and guidelines for OECD statistical activities, version 2011/1

[71] OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, https://doi.org/10.1787/276aaca8-en.

the platform. The availability of reliable, accurate data is crucial for fostering trust and encouraging the active use of the platform.

Furthermore, the consequences of distributing poor-quality data are substantial and carry the risk of causing harm to individuals and communities. This risk stems from the fact that substandard data can lead to erroneous decisions and reinforce existing biases. For instance, if source data exhibits gender bias, it can exacerbate social inequalities and injustices, perpetuating discrimination and undermining efforts towards equity and inclusion. Thus, the integrity of the data provided by OpenEPI is paramount not only for maintaining user engagement and stakeholder collaboration but also for ensuring that the information disseminated does not inadvertently contribute to societal harm. Ensuring the accuracy, fairness, and reliability of data is essential to prevent the propagation of misinformation and to support informed decision-making that benefits all segments of society.

7.3.1 Reasons for poor quality data

There are various reasons for why the data can be of poor quality. An important factor that contributes to varying data quality, is the lack of *common metrics* for quality. There is no shared definition of what constitutes 'good' data quality. Research shows that there are many different interpretations and ways of measuring data quality. Since people use data for different purposes, certain data qualities matter more to a user group than others. Different methods for collecting and registering data further explain the difference in data quality. Data on the same topics can also be of different quality as a result of a lack of harmonization of collection and registering methods.[72]

From our interviews, we learned that other open data platforms invest a lot of time and resources into data cleaning and verification of data before they are made available on the platform. The fact that quality enhancing activities are resource-intensive and costly can however lead to poor quality data. Many data providers do not have neither the resources nor competence to implement robust data governance frameworks and routines in the organization[73]. Inadequate governance frameworks, lack of data quality metrics, insufficient data management policies and validation mechanisms to verify the accuracy, consistency, and reliability of the data can result in poor-quality datasets being distributed. Further, data that is not regularly updated becomes outdated and loses its relevance, potentially leading to inaccurate analyses or decisions.

Incomplete data collection, or data gaps, may also lead to poor quality data. Data gaps can compromise the relevance and utility of the datasets. Open data is often unequally distributed, as there is more capacity and a more pronounced culture for gathering and making data openly available within certain domains and geographies than others. This uneven distribution of open data is particularly pronounced between developed and developing countries, where the latter often grapple with limited resources, inadequate data collection infrastructure, and less established open data practices. As a result, datasets from developing regions might be not only sparse but also lack the necessary quality to meet

[72] Menon Economics og A2 (2021), *Fremtidens miljødata*, Menon-publikasjon nr. 153/2021, https://www.menon.no/wp-content/uploads/2021-153-Hovedrapport-KVU-Fremtidens-miljodata.pdf
[73] Climateworks Foundation, *How to find relevant climate data,* 2023, https://www.climateworks.org/blog/how-to-find-relevant-climate-data/

users' needs. Moreover, the challenges in collecting and distributing open data in developing countries may be exacerbated by factors such as political instability and government intervention. These elements may affect both the availability and reliability of data. For instance, governments may restrict access to certain types of data or fail to update existing datasets due to political agendas, resource constraints, or bureaucratic inefficiencies. In section 7.4, we will delve deeper into the specific challenges associated with the scarcity of local data and outline strategies to mitigate these issues.

7.3.2 Mitigating strategies

To address the risk of poor data quality and ensure high-quality data on OpenEPI, it's crucial to implement a comprehensive set of mitigation strategies. These measures not only help in improving the trustworthiness and usability of the platform but also in enhancing user adoption. Here are the most important measures:

- **Implement a robust data governance framework:** Robust data governance routines (and sufficient funding to back it) is vital to make sure the platform publishes high quality data. In addition to requiring the data providers to apply to open policy and the FAIR principles, OpenEPI should aim at requirements covering policies and standards for data management, including data collection, storage and processing. All to secure a robust value chain. This includes defining roles and responsibilities for data stewardship and establishing accountability mechanisms.
- **Implement data quality metrics and standards**: Create and enforce specific metrics and standards for data quality, tailored to the needs of the users and the objectives of the platform. These metrics should cover OECDs data quality dimensions, namely: relevance, accuracy, credibility, timeliness, accessibility, interpretability and consistency.
- **Transparency and documentation**: Provide detailed metadata and documentation for each dataset, including its source, methodology of collection, any limitations, and the context of its intended use. Transparency builds trust and helps users assess the relevance and reliability of data.
- **Access to domain expertise**: Involve experts with domain-specific knowledge in the data review process to validate the accuracy and applicability of the data. This can also help in interpreting complex datasets and ensuring they are meaningful for users.
- **Stakeholder engagement and collaboration:** Work closely with data providers, users, and other stakeholders to understand their needs and concerns. Engaging stakeholders in the data curation process can help in identifying high-value datasets and revealing data gaps. Offer guidance, resources, and tools to data providers to help them understand the importance of data quality and how to achieve it. This includes best practices for data collection, preparation, and submission. In one of our interviews, GEO highlighted the importance of educating their data providers on the importance of data quality. Investing in building data governance principles and educating their collaborative partners in these principles, had proved to have a very positive impact on the overall data quality they were able to offer on the platform.
- **Continuous monitoring and feedback loops**: Establish mechanisms for ongoing monitoring of data quality and user feedback. This allows for the early detection of quality issues and the continuous improvement of datasets based on user input and changing needs.

7.4 User adoption risks

Actual user adoption among the developers will of course be essential for the success of the portal. Perhaps this is the most severe risk for OpenEPI. Widespread user adoption is paramount in order for the platform to reach its objectives, and indeed for securing long term funding of the platform and its services.

Several factors, both within and beyond OpenEPI's control, may hinder widespread uptake of the platform. Below, we delineate the key internal factors, those that OpenEPI can influence to varying degrees, and external factors, which lie outside of OpenEPI's direct influence, that may obstruct the platform's widespread adoption.

7.4.1 Internal risk factors that may prevent high user adoption

There are several OpenEPI internal risk factors:

Limited availability of, and access to, local data
The quality, accuracy, and relevance of data are pivotal in attracting and retaining users on the platform. Outdated, incomplete, or unreliable data can deter potential users from depending on the platform for developing innovative solutions. Significant data gaps, such as the absence of reliable and pertinent local datasets, could gradually diminish the platform's usefulness and appeal. Users seeking to address local nature and climate-related challenges, such as those in agriculture, may find the platform inadequate for their needs if it fails to provide local data to support the development of effective services and solutions for a local context. Our interviews with managers of open data platforms revealed that while global data is valuable, local data is essential if the goal of the platform is to foster local innovation.

However, from our interviews with open data providers, it has emerged that securing access to and sourcing relevant local data can be challenging. We find this documented from other studies as well, for instance from Tanzania and Sierra Leone[74]. Distributing local datasets publicly presents numerous challenges, notably when these datasets contain personal data. The inclusion of personal information complicates the process of making these datasets publicly available due to privacy concerns and regulatory requirements. Additionally, governments in developing countries often exhibit greater reluctance to publish open data compared to their counterparts in regions like Europe. This hesitancy stems from concerns that open data, particularly from less developed areas, might be exploited by entities in more developed countries for commercial gain. Such exploitation could potentially widen existing disparities rather than contribute to equitable development. Furthermore, apprehensions regarding the dual-use nature of data—whereby information intended for beneficial purposes can also be used in ways that may cause harm or undermine privacy and security—make governments cautious about data sharing. These dual-use challenges, coupled with the fear of neo-colonial exploitation of local resources and data, contribute to the reluctance of these governments to participate fully in open data initiatives.

[74] Report from World Resources Institute. *Climate Change Open Data for Sustainable Development: Case Studies From Tanzania and Sierra*: Leonehttps://www.google.com/url?q=https://www.data4sdgs.org/sites/default/files/services_files/WRI%2520Climate%2520Data_FINAL2_optimized.pdf

To navigate these challenges, it's essential for OpenEPI to engage in diplomatic and collaborative efforts with local governments, research institutions, NGOs, data providers and data platforms. This involves demonstrating the mutual benefits of open data, implementing strict data governance policies to protect personal and sensitive information, and ensuring that data use respects local norms and contributes to local development. By fostering trust and showcasing the potential for open data to drive positive change, OpenEPI can encourage more open data contributions from developing regions, enriching the platform's dataset diversity and relevance.

Not sufficiently addressing user needs
Low user adoption can also follow from lack of investment in mapping out and analyzing user needs and necessary preconditions for using the service. From interviews with open data platforms, we have learned that open data initiatives tend to be more successful when they are clearly optimized for an intended audience or user base from the start. Particularly in developing economies, a clear, detailed understanding of the problem to be addressed by open data can help to ensure that efforts are targeted and optimized. According to a survey conducted by Verhulst and Young, some of the most effective open data projects in developing economies have had a strong focus on a specific user group or identified user needs.[75] Clearly defining the problem can also aid in the development of metrics of success and a strategy for monitoring progress.

Lack of stakeholder engagement
Engaging key stakeholders—including governments, supranational organizations, domain expertise, data providers, and advocates of the open data movement—is crucial for amplifying OpenEPI's impact and expanding its user base. Their involvement will help to ensure that the platform remains aligned with societal needs and user expectations, thereby enhancing its relevance and utility. Moreover, these stakeholders play a pivotal role in promoting OpenEPI, leveraging their networks and resources to attract and direct users to the platform. By actively participating in the development process, stakeholders can contribute their unique insights and expertise, enriching the platform's features and data quality. Stakeholder engagement must be seen not just as a strategic necessity for OpenEPI, but a catalyst for building a robust, relevant, and widely adopted platform.

Inadequate support, training and community building
Inadequate support and training mechanisms can significantly impede user adoption of platforms like OpenEPI. The absence of detailed tutorials, user guides, and responsive help desks can leave users, especially those less familiar with technical platforms, feeling overwhelmed and unsupported. Given that OpenEPI caters to a wide audience, including developers and startups with diverse skill levels, it is imperative to offer a broad spectrum of support resources. These should cater to both novice users and those with advanced technical knowledge, ensuring the platform is approachable and navigable for all. Tailored support resources, such as step-by-step guides, and FAQs, can greatly enhance user confidence and competence in utilizing the platform.

Moreover, the lack of interactive elements, forums for discussion, and opportunities for collaboration can lead to a reduction in user engagement and satisfaction. Interactive

[75] Verhulst, Stefaan G. & Andrew Young, *Open Data for Developing Economies*, 2017

features that encourage user participation, such as Q&A sections, live chats, and community forums, foster a sense of belonging and support among users. They serve as vital channels for peer-to-peer assistance, feedback, and the sharing of ideas and best practices to spur local innovation.

Lack of user-friendly interface and innovative functionality
An intuitive, user-friendly interface is crucial for any open data platform, particularly those in the nature and climate domain, where the complexity and volume of data can be overwhelming. If a platform is cumbersome to navigate, lacks streamlined interfaces, or demands a significant learning investment, users may struggle to engage effectively with the data, potentially leading to reduced adoption and engagement.

To enhance user experience and encourage broader utilization it is further important that OpenEPI offers high-quality and well-designed APIs. This not only reduces the barrier to entry for non-expert users but also enables developers to innovate and create more effective solutions to environmental challenges.

Further, users will increasingly anticipate that open data platforms not only provide access to vast amounts of data but also offer innovative, AI-driven functionalities. These can range from predictive analytics forecasting climate trends to automated image recognition for monitoring biodiversity. This means that OpenEPI must regularly update their offerings with the latest functionalities and ensure that the platform stays at the forefront of technological advancements, if they are to stay attractive for users in the years to come.

Unstable or unreliable access to data
It is not common for open data platforms to have formal Service Level Agreement (SLA) with their data providers[76]. Agreements with data providers are commonly based on Creative Commons licenses and mutual trust. However, the absence of an SLA may leave open data platforms in a precarious position, unable to guarantee their users consistent access to high-quality data. This uncertainty can potentially lead to a gradual erosion of trust and satisfaction among users, particularly if they frequently encounter issues such as data inaccessibility. Over time, this reliability gap can significantly hinder user engagement and growth.

For startups and businesses that will rely on OpenEPI to build their services, the stakes are even higher. These entities depend on stable, continuous access to data to develop and maintain their offerings. Without the assurance of data availability provided by SLAs, these companies face increased risks and uncertainties, making it challenging to plan and potentially jeopardizing their sustainability. This instability can deter startups from investing in innovative solutions that leverage open data, thereby stifling entrepreneurship and technological advancement.

7.4.2 External risk factors that may prevent high user adoption

There are several factors affecting the adoption of OpenEPI's data provisions and service offerings by developers. The most important are the following:

[76] SLAs outline the expected standards for data quality and availability, setting clear benchmarks for both data providers and users of the platform. These agreements also detail the repercussions should these standards not be met, ensuring accountability and a mechanism for addressing issues as they arise.

Lack of access to necessary technical infrastructure
OpenEPI's primary user base will be situated in some of the world's most economically disadvantaged regions. Challenges such as low internet penetration, the high expenses associated with smartphone usage, and a pronounced digital divide within these communities could significantly hinder uptake of the platform.

There is a steady increase in unique mobile subscribers within sub-Saharan Africa. However, as of 2022, the mobile penetration rate stood at 43 percent, with projections suggesting a potential rise to 50 percent by 2030. Still, there are great variations between countries, making it difficult to apply a general approach for the region. Further, 50 percent is considerably lower than the global average of 73 percent expected in the same period. Moreover, the region faces a substantial mobile internet usage gap, with only 25 percent of mobile subscribers enjoying internet access. This gap is primarily attributed to issues of affordability and a lack of digital literacy. Additionally, the gender divide in mobile phone ownership further exacerbates this issue, with women in low- and middle-income countries being 17 percent less likely than men to own a smartphone. In sub-Saharan Africa women are 36 percent less likely to have access to mobile internet, according to a report by GSMA.[77].

The access to mobile internet and smartphones is crucial not only for developers' ability to take advantage of OpenEPI but also for ensuring that the solutions developed on the platform can find a market and be effectively utilized within local communities in sub-Saharan Africa.

To address challenges related to low internet penetration and the particularly vulnerable position of women in accessing the internet, a notable development project was launched aimed at empowering smallholder farmers in Kenya. This initiative, supported by Mozilla and run in collaboration with the University of Westminster in the USA, Moi University, and the Technical University of Kenya, focuses on innovative solutions to bridge the digital divide.[78]

One of the key components of this project is the development of an audio chatbot that does not rely on internet connection, making it a viable solution in areas with limited internet access. This chatbot, developed with the input from rural smallholder women farmers in Kenya, serves as an alternative source of agricultural information. By leveraging Mozilla's Swahili datasets, the chatbot is designed to be accessible on both basic feature phones, commonly referred to as "kabambes," and smartphones. This approach not only circumvents the issue of internet access but also ensures that women, who often face greater barriers to internet use, can benefit from critical agricultural information and support.

This project exemplifies how collaborative efforts between different stakeholders, such as technology companies, development agencies, and local educational institutions can produce tailored solutions that address specific challenges faced by marginalized communities. By

[77] GSMA Intelligence, *The Mobile Gender GAp Report 2023,* 2023, https://www.gsma.com/r/wp-content/uploads/2023/07/The-Mobile-Gender-Gap-Report-2023.pdf
[78] University of Westminister, *University of Westminster contributes to East African voice technology project funded by Mozilla,* 23.09.2022, https://www.westminster.ac.uk/news/university-of-westminster-contributes-to-east-african-voice-technology-project-funded-by-mozilla

focusing on the needs of smallholder women farmers in Kenya and utilizing technology in an innovative way, it is possible to work around the challenges caused by the digital divide.

Lack of local competence and skilled professionals
A further risk that may impair widespread uptake and use of OpenEPI, is the scarcity of technically skilled individuals and professionals capable of recognizing and acting on the potential of open data (readiness). For OpenEPI to reach its full potential, users must possess a certain level of proficiency in programming languages, data analysis and web development skills. The lack of such competence in the developing world, and in sub-Saharan Africa especially, poses a multifaceted risk to the platform's adoption and efficacy.

An additional challenge related to app development, is the need for developers to stay abreast with the continuous advancement of operating systems (OS) like Android and iOS. This constant evolution requires developers to constantly update and expand their technical skill set. A scarcity of local technical expertise forces reliance on external consultants or partnerships, which can be costly and unsustainable in the long term. Moreover, this dependency may lead to solutions that are not fully aligned with local contexts or needs.

The Norwegian Meteorological Institute is cooperating with the Department of Climate Change and Meteorological Services (DCCMS) in Malawi to develop a weather app. A major challenge for the project has been to get hold of local professionals with updated knowledge about API's and mobile-based app development. Addressing the competence gap will not only be crucial for the successful adoption of OpenEPI but more generally for ensuring that the developing world can fully participate in and benefit from the global data revolution.

Competition from existing data platforms
The presence of competition from established open data platforms in the climate and nature domain represents an external risk to the widespread adoption of OpenEPI. Already existing platforms provide extensive climate data, which could potentially saturate the market and challenge the entry and uptake of a new platform.[79] Success of OpenEPI hence depends on its ability to offer something unique or complementary to what is currently available. By identifying and leveraging its unique strengths and opportunities for collaboration, OpenEPI can establish itself as a valuable resource in the open nature data landscape.

7.4.3 Mitigation strategies

We see several measures and strategies as relevant to avoid or reduce the effect from the above mentioned risks:

- **Engage with a wide variety of stakeholders to reduce the risk of data gaps:** To minimize the risk of delivering data that may not meet the needs or relevance criteria of users seeking to develop innovative solutions for local climate challenges, it's imperative for OpenEPI to establish partnerships across a broad spectrum of stakeholders. This includes engaging with organizations and data providers at the international, regional, national, and local levels, as well as directly collaborating with the developers who will utilize the service. Such a comprehensive approach ensures

[79] World Resource Institute (WRI), *Overview of 100 climate data platforms*, 2023, https://www.wri.org/data/overview-100-climate-data-platforms

a diverse input in curating data that is both pertinent and valuable, enhancing the platform's utility in facilitating the creation of targeted solutions to address climate-related issues effectively.

- **Stakeholder engagement:** Establishing partnerships with key stakeholders, including local governments, NGOs, universities and data providers, is vital for advocating for and promoting the platform. Utilizing stakeholder networks to gather insights and feedback helps tailor the platform to meet user expectations and needs more effectively. Actively involving local communities, experts, and organizations in data collection and validation processes enhances the platform's credibility and relevance by ensuring the data reflects local realities and needs.
- **Defining performance metrics:** Open data projects are better positioned for success when they develop and monitor metrics of impact to inform management and iteration. These metrics inform management and iteration, fostering a user feedback loop that is essential for continuous improvement.
- **Providing adequate training and support:** Investing in user outreach, training, and support initiatives to encourage user adoption. Offer a range of support resources, such as tutorials, guides, and FAQs, tailored to different user skill levels. Organize workshops and training sessions to build capacity among users, particularly focusing on developing technical skills.
- **Improving usability and technical functions:** Improving usability and technical functions is fundamental to maximizing user uptake and impact. OpenEPI must deliver a seamless, intuitive user experience, supported by high-quality APIs and, eventually, innovative AI-driven functionalities. Staying updated with the latest technologies and continuously refining the platform’s features based on user feedback ensures it effectively addresses user needs.
- **Ensuring reliable access to data:** Ensuring reliable access to data is crucial. OpenEPI should consider establishing SLAs with their most prominent data providers. Establishing SLAs not only enhances the platform's reliability but also its appeal to a broader user base, including startups and other commercial entities. SLAs can serve as a foundation for building a more robust ecosystem around open data, characterized by higher levels of trust, innovation, and collaboration. Furthermore, incorporating performance metrics and regular review processes within SLAs can ensure continuous improvement in data quality and service delivery.
- **Addressing technical infrastructure barriers:** To address challenges faced by end-user application developers, OpenEPI must focus on adaptations for low-bandwidth environments, including implementing location-specific access to APIs to ensure efficient and minimal data transfer. This concern for optimizing mobile usability is primarily a task for the application developers, as OpenEPI itself serves as a data platform aimed at these developers, rather than the end users. Additionally, OpenEPI should offer targeted training for developers on crafting effective solutions for low-bandwidth conditions. Strengthening partnerships with universities, technology companies, and organizations like the Gates Foundation and Mozilla Foundation will also facilitate the integration and market adoption of these solutions within local communities.
- **Bridging the digital divide and gender gap:** To mitigate the lack of competence and to bridge the digital divide, a comprehensive strategy involving multiple stakeholders is essential. This strategy could include investing in local education and training programs tailored to developing data science and analytics skills, fostering

partnerships between academia, industry, and government to create an innovation ecosystem. (Partner with educational institutions and tech communities to offer training programs and resources focused on developing local talent and technical skills. Create mentorship and internship programs to provide hands-on experience and skill development in data utilization and application development.) Additionally, initiatives aimed at demystifying data science and making it more accessible to non-experts could help in broadening the base of users capable of engaging with OpenEPI effectively. Further, collaborate with local initiatives and community groups capable of reaching out to local users. Have a focus on reaching underrepresented groups, including women.

- **Differentiating from competitors:** Identify unique value propositions that set OpenEPI apart from existing platforms, such as local innovation focus, specialized data sets, unique analytical tools and features. Focus on cross-cutting themes and integrations that address user needs.

7.5 Privacy risks

Privacy concerns probably rank among the most cited worries over opening up data. Potential privacy harms can arise even from the release of ostensibly anonymized personal data. A lack of sophistication in anonymization or aggregation techniques can result in the inadvertent release of personal data. Additionally, in some instances information that itself poses no privacy concerns can be combined with other openly available datasets and lead to unexpected disclosure or inference of personal data.

7.5.1 Unintended processing of personal data

OpenEPI's primary focus is on publishing data that is devoid of personal data, such as geospatial, weather, and soil data. However, it is important to note that OpenEPI may also distribute datasets containing personal data, both intentionally and unintentionally. For instance, agricultural information, encompassing crop details, field data, and related content, may potentially include personal data such as farmers' names, addresses, and locations. Furthermore, farm data may involve indirectly identifiable information, such as a combination of farm activities and details about crops and livestock. Farmers often have limited control over their farm data, and this creates concerns related to safeguarding the farmers' privacy rights.[80] The same challenges related to privacy and data protection may arise when processing forests and deforestation data as these data may be linked to land ownership.

7.5.2 Local data may contain personal data

Further, OpenEPI is specifically aimed at local developers and start-ups. To maximize the utility of the platform for local developers, integration of data like weather and satellite information with local datasets can prove essential. Local data sets may often contain personal data. If the data sets are anonymized, there is a risk that anonymized data may be re-identified when combined with other data sets, such as location data.

[80] Kaur, Jasmin and Seyed Mehdi Hazrati Fard, Mohammad Amiri-Zarandi, Rozita Dara, *Protecting farmers' data privacy and confidentiality: Recommendations and considerations*, Front. Sustain. Food Syst., 19 October 2022, Sec. Social Movements, Institutions and Governance, Volume 6 - 2022, https://doi.org/10.3389/fsufs.2022.903230

7.5.3 Sharing personal data with third parties

OpenEPI will most likely also process personal data about its users as part of managing the platform, for instance through an implemented identity management system and a user feedback solution. The platform will also process personal data about its users if it uses cookies and tracking technologies for data analytics and user experience optimization purposes. There is a risk that personal data may be shared with third parties without the consent and knowledge of the users.

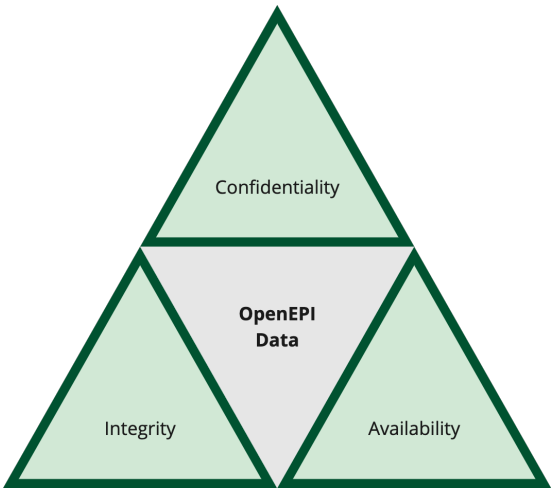
7.5.4 Mitigating strategies

We see the following strategies for mitigation:

- **Ensuring GDPR compliance:** OpenEPI should prioritize adherence to the General Data Protection Regulation (GDPR), regardless of its geographic or organizational setup. This entails establishing comprehensive privacy policies and data governance frameworks to safeguard personal information.
- **Privacy policy implementation:** OpenEPI's privacy policy should transparently detail the collection, use, and management of personal data, including the types of data collected (e.g., personal details, usage statistics) and their usage. It should address the use of cookies and tracking technologies, third-party data sharing conditions, data anonymization practices, storage durations, security measures, and data disposal procedures. The policy must clearly define users' rights to modify or delete their data and provide channels for inquiries and rights execution. Regular reviews and updates of the privacy policy are essential to maintain its relevance and compliance.
- **Robust data governance and anonymization:** A stringent data governance system is vital for minimizing privacy risks. OpenEPI should rigorously assess datasets for personal information and apply advanced anonymization techniques before publication. In cases where data cannot be anonymized, alternative compliance measures must be explored. Whenever possible, data about individuals should only be included in open datasets if explicit consent has been given. Conducting Privacy Impact Assessments (PIAs) helps identify and mitigate re-identification risks and potential privacy infringements when correlating datasets. Continuous auditing of data usage and a solid breach response strategy are crucial. Appointing a dedicated data steward ensures ongoing attention to data quality and privacy concerns.

7.6 Security and safety risks

A platform that employs open source software and technology to distribute open data, is subject to various technical security and safety risks. While these risks are common across many software types, they present unique and complex challenges within the open source ecosystem. Predominantly, these challenges revolve around the critical principles of the CIA triad - confidentiality, integrity, and availability, which are essential for maintaining the security and functionality of any information system.



Confidentiality: Ensures that information is accessible only to those authorized to view it, protecting it from unauthorized disclosure.

Integrity: Integrity safeguards information from unauthorized alterations, ensuring that data remains accurate and complete throughout its lifecycle.

Availability: Availability guarantees that data and systems are accessible to authorized users when needed, crucial for maintaining continuous operation.

7.6.1 Confidentiality risks

In the OpenEPI platform, which primarily distributes openly licensed data, the significant confidentiality risks center around protecting user information. This includes protecting details about user activity, such as which datasets users access and how often, as well as preventing the identification of individual users. Additionally, confidentiality extends to any feedback mechanisms implemented on the platform, ensuring that user inputs or suggestions do not inadvertently reveal personal identifiers or usage patterns. If such information is not properly secured, it could be exploited for targeted advertising, phishing attacks, or other malicious purposes, leading to serious privacy breaches.

7.6.2 Integrity risks

Integrity risks within the OpenEPI platform are extensive, impacting both data managed internally, and data obtained from external sources. As OpenEPI collects and distributes climate and nature data, the potential for internal corruption or manipulation threatens the platform's reliability and utility. This, in turn, affects end users who rely on the accuracy of this data.

Issues internal to OpenEPI such as misconfigurations or bugs in the software that processes incoming data can lead to errors, resulting in inaccurate data outputs. These issues may include incorrect data mapping, transformation errors, or complications during data aggregation. A key aspect of internal processing in OpenEPI involves simplifying source data to enhance usability for developers. However, this process carries the risk of oversimplifying the data and losing critical data properties.

Additionally, if OpenEPI's systems are compromised by malware or hacking, the integrity of the data could be intentionally sabotaged, including unauthorized changes by attackers aimed at manipulating data for fraudulent purposes.

The platform's heavy reliance on third-party data providers also introduces inherent risks concerning the accuracy and completeness of the information received. Data might be inaccurately recorded or processed due to human error or technical faults within the third-party systems. Furthermore, this data could be intentionally altered for various reasons, such as political bias, financial gain, or sabotage. Moreover, errors during data transmission from third parties to OpenEPI could potentially lead to data corruption or loss.

7.6.3 Availability risks

The operation of OpenEPI is heavily dependent on third-party data providers. Any interruptions in their services, whether due to technical failures or logistical issues, can obstruct the steady flow of data into OpenEPI, impacting the platform’s functionality.

A challenge to the availability of OpenEPI is the risk of Denial of Service (DoS) attacks, which can overwhelm the platform with excessive traffic, disrupting normal operations and access. Software glitches and issues within the platform's infrastructure can also lead to unexpected downtimes, directly impacting user access.

Additionally, as OpenEPI relies entirely on cloud vendors for its hosting and operational infrastructure, it faces risks associated with these dependencies. Cloud service outages, policy changes by vendors, or the potential discontinuation of services can significantly affect the stability and accessibility of the platform.

7.6.4 Mitigation strategies

To effectively mitigate risks related to confidentiality, integrity, and availability, OpenEPI should employ a variety of strategies tailored to address specific vulnerabilities:

- **SSL to prevent interception of data**: Secure Socket Layer (SSL) encryption is crucial for protecting data in transit between OpenEPI and its users. By encrypting the data exchanged over the network, SSL prevents potential eavesdroppers from intercepting and deciphering sensitive information.
- **Encrypt data at rest**: OpenEPI should ensure that all stored data is encrypted using strong encryption protocols. This practice protects data from unauthorized access while it is stored on disk, safeguarding information against breaches and exposure.
- **Anonymous usage statistics**: OpenEPI should collect usage statistics in an anonymized format. This approach allows the platform to gather necessary operational data without compromising the privacy of individual users.
- **Checksums for datasets**: OpenEPI should utilize checksums to verify the integrity of datasets. By calculating and comparing these checksums, the platform can detect any unauthorized changes to the data since it was last verified, ensuring the data’s accuracy and trustworthiness.
- **Collaboration with domain experts at data providers**: To minimize the risk of oversimplifying data, OpenEPI should collaborate closely with domain experts at third-party data providers. This ensures that the complexities and nuances of the data are preserved and accurately represented.
- **DDoS prevention at edge**: To combat the threat of Distributed Denial of Service (DDoS) attacks, OpenEPI needs to implement DDoS mitigation techniques at the

network edge. These measures are designed to filter out malicious traffic and prevent it from reaching the platform's infrastructure, thus maintaining service availability.

- **Use multiple cloud vendors**: It is advisable for OpenEPI to enhance its resilience and ensure continuous service by distributing its operational load across multiple cloud vendors and availability zones. This approach significantly mitigates the risk of single points of failure and facilitates seamless failover, maintaining service availability even if one part of the infrastructure becomes compromised or unavailable.

7.7 Data ethics and environmental risks

Data ethics risks refer to potential ethical concerns and issues related to the collection, processing, storage, and use of data. OpenEPI will offer many different datasets, covering a range of topics and issues. Could datasets be combined and used in ways that can cause harm or exacerbate existing power asymmetries? Could it be used to target, profile or prejudice vulnerable populations? Could the platform's computation activities have adverse impacts on the environment? Addressing data ethics risks requires a comprehensive approach, including legal frameworks and organizational policies that prioritize ethical considerations in data handling and decision-making processes.

7.7.1 Unintended use of data

When publishing open data, there is a risk that the data may be used for purposes that was not intended.[81] The term "dual use" is used to describe how a particular technology or knowledge has the potential to be used for both beneficial and harmful purposes. It implies that a technology or innovation has applications that can serve both constructive and potentially destructive ends. When combining different datasets, the risks of harmful dual use may increase. Datasets that are initially harmless can be combined in ways that allow them to be used for malicious purposes. During our interviews with various open global data platforms, one platform identified the most significant risk as the potential misuse of their data for purposes that were never intended. For instance, biodiversity data containing information on endangered species can be combined with satellite data to develop applications and measures aimed at protecting vulnerable flora and fauna. However, this same knowledge can be leveraged to pinpoint the locations of valuable animal and plant species, enabling targeted exploitation. Similarly, data pertaining to river courses and soil conditions can offer insights into suitable areas for new agricultural development. However, this knowledge also carries the risk of being utilized by hostile neighboring states with intentions to undermine or sabotage regional business development.

These examples demonstrate that with open data, although the original intention may be to deliver a greater public good, the unintended consequences may have harmful effects on specific environmental conditions or single out vulnerable communities or individuals for exploitations. However, to deny unintended consequences is to deny access to innovation. The solution is therefore to perform due diligence to understand the data landscape well enough to identify potential hazards. This can be done within the platform by bringing

[81] Deloitte, *Open data driving growth, ingenuity and innovation*, 2012, https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

together a diverse group, including protagonists as well as critics, to perform “red teaming” - to examine potential data use cases from various perspectives, thereby identifying and addressing risks before they materialize.[82]

7.7.2 Risk to vulnerable populations

Climate change disproportionately impacts vulnerable communities, highlighting the urgent need for comprehensive data collection to shape policy decisions and develop targeted mitigation strategies. These communities, as identified by the UN Agenda 2030, include the economically disadvantaged, women and girls, children and youth, the elderly, people with disabilities, migrants, refugees, rural communities, indigenous peoples, individuals living with HIV, and other groups considered vulnerable within specific national contexts.[83]

The management of data related to these populations is inherently complex, requiring robust safeguards to ensure its responsible utilization. Although data can empower these groups, it also holds the potential to perpetuate societal inequities and reinforce established power hierarchies. For indigenous communities worldwide, the notion of open data is particularly contentious. The prevailing principles in the open data movement focus primarily on enhancing data accessibility and sharing, often neglecting the importance of addressing power imbalances and historical injustices. This singular focus on data dissemination creates challenges for indigenous peoples who advocate for greater sovereignty over the use and application of their data for communal benefit. This includes the assertion of rights to utilize indigenous data in ways that resonate with their cultural perspectives and to capitalize on opportunities within the knowledge economy.[84]

Fact box

The CARE principles

- **Collective benefit:** Data ecosystems shall be designed and function in ways that enable indigenous peoples to derive benefit from the data.
- **Authority to control:** Indigenous peoples’ rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables indigenous peoples and governing bodies to determine how indigenous peoples, as well as indigenous lands, territories, resources, knowledge and geographical indicators, are represented and identified within data.
- **Responsibility:** Those working with Indigenous data have a responsibility to share how those data are used to support indigenous peoples’ self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to indigenous peoples.
- **Ethics:** Indigenous peoples’ rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

[82] Deloitte, *Open data driving growth, ingenuity and innovation*, 2012, https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.p

[83] United Nations, *Transforming our world: the 2030 Agenda for Sustainable Development,* https://sdgs.un.org/2030agenda

[84] Hudson M, Carroll SR, Anderson J, Blackwater D, Cordova-Marks FM, Cummins J, David-Chavez D, Fernandez A, Garba I, Hiraldo D, Jäger MB, Jennings LL, Martinez A, Sterling R, Walker JD, Rowe RK, *Indigenous Peoples' Rights in Data: a contribution toward Indigenous Research Sovereignty,* Front Res Metr Anal, 2023, May 4;8:1173805. doi: 10.3389/frma.2023.1173805.

In 2019, The Global Indigenous Data Alliance (GIDA) developed the CARE Principles for Indigenous Data, to complement the FAIR principles. The CARE Principles for Indigenous Data Governance are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination.[85]

7.7.3 Environmental risks

Though digital technologies are critical drivers to combat climate change and advance global sustainability, a growing body of research underscores the significant carbon footprint and environmental impact of digital activities. The operation of data platforms like OpenEPI involves the processing and storage of vast quantities of data, potentially also coupled with the development and implementation of AI algorithms. These computational activities consume energy and require hardware, contributing to embodied emissions that can have adverse effects on the climate. Recognizing and addressing these environmental implications is essential in ensuring the responsible use of digital technologies for sustainable and climate-conscious practices.[86]

The architectural design of an IT platform and the selection of cloud services significantly influence the environmental impact of the project. Strategic architectural choices, including the adoption of efficient frameworks and technologies, can greatly reduce the energy consumption of servers and data centers. Additionally, opting for cloud services that emphasize sustainability—such as those powered by renewable energy sources or equipped with advanced energy-saving technologies—can further decrease the environmental footprint.

All major cloud vendors provide tools for calculating the carbon footprint of their solutions. These tools offer valuable insights into the environmental performance of the platform, enabling more informed decisions to enhance sustainability.

7.7.4 Mitigating strategies

Addressing data ethics risks requires a comprehensive approach, including legal frameworks and organizational policies that prioritize ethical considerations in data handling and decision-making processes:

- **Implement the CAIR-principles:** The CARE Principles and FAIR Principles should be engaged simultaneously. The CARE Principles approach is not only relevant for the processing of data about indigenous people, but of vulnerable people in general. This is important to ensure that data is used without ignoring the contexts in which data exist, and to create opportunities for self-determination and self-governance for the said groups.
- **Conduct risk assessments:** Perform risk assessment to reduce the risks associated with the dual use of nature and climate data. This process involves identifying potential vulnerabilities, assessing the likelihood and impact of different types of

[85] Global Indigenous Data Alliance (Gida), *CARE Principles for Indigenous Data Governance*, https://www.gida-global.org/care
[86] GPAI, Climate Change & AI: Recommendations for Government Action, 2021, https://www.gpai.ai/projects/climate-change-and-ai.pdf

misuse, and developing strategies to minimize risks. For example, a risk assessment might identify that detailed climate data could be used to exploit natural resources in vulnerable ecosystems. In response, data anonymization techniques and ethical guidelines can be established to ensure that data sharing promotes sustainability and conservation efforts, rather than contributing to environmental degradation.

- **Conduct ethics audits:** Engage ethics expertise to assess whether the release of certain types of data is ethical. Regularly conduct audits to ensure that data release and usage are not disproportionately affecting marginalized communities.
- **Be transparent:** Be open (in annual reports or other documents) on details on who is using the data, for what purpose, and what steps are being taken to ensure compliance with the CAIR-principles.
- **Educate users:** Educate the users of OpenEPI about responsible data use and the associated risks. Increase data ethics literacy so that the users can better understand the limitations and risks associated with using the data.
- **Measures to minimize carbon footprint:** Actively monitor and manage daily carbon emissions using cloud vendors' tools to track and calculate the carbon footprint. Make this information public. This transparency enables informed decision-making and operational adjustments. Minimize data duplication and prioritize efficiency in solution design to reduce environmental impact effectively.

7.8 Technological infrastructure risks

The technological infrastructure of OpenEPI plays a crucial role in its functionality and service delivery, encapsulating various elements from multi-cloud deployment to complex system orchestration. This chapter highlights some of the risks associated with maintaining and scaling such an infrastructure, critical for understanding how these factors could impact the platform's effectiveness and reliability.

7.8.1 Complexity of infrastructure

OpenEPI's technological infrastructure is defined by its use of multiple cloud vendors and the routing between them, adding layers of complexity that pose several operational risks. The platform employs Kubernetes for container orchestration across cloud environments. While Kubernetes offers a standardized orchestration solution, minor differences in implementation among vendors can complicate deployment, management, and service interoperability.

Additionally, OpenEPI integrates advanced data orchestration with Dagster and leverages parallelism with Dask to enhance processing capabilities, which currently are operational on a single cloud vendor. However, a full realization of OpenEPI might necessitate running these services across multiple cloud vendors to meet scalability and resilience requirements.

Including these advanced tools introduces complexity into the system, especially considering the potential need to expand these functions across multiple cloud environments in the future. Managing this complexity is crucial to minimize the risk of system failures, downtime, or technical glitches that could significantly impact the platform's users.

7.8.2 Vendor risks

OpenEPI's reliance on external cloud vendors and other third-party service providers introduces risks that could impact the platform's operational reliability and availability.

In an effort to reduce dependency on any single provider, OpenEPI has diversified its technology suppliers. Although this strategy decreases reliance on individual vendors, it introduces complexity in managing multiple vendors and integrating diverse technological solutions. This complexity requires diligent oversight to ensure that OpenEPI remains adaptable and not overly constrained by any specific vendor’s technological limitations.

Variability in the performance and quality of services among different vendors can lead to a non-uniform user experience and potentially impact the overall effectiveness of the OpenEPI platform.

7.8.3 Open Source

OpenEPI’s commitment to utilizing and releasing open source software demonstrates its dedication to fostering innovation, collaboration, and transparency. This approach aligns with modern software development practices but introduces several risks that must be managed to safeguard the platform.

OpenEPI’s use of open source software and external libraries brings inherent risks, including security vulnerabilities and the potential for rapid technological obsolescence. The maintenance levels for these open source projects can vary significantly, with some not promptly addressing security flaws, potentially leaving OpenEPI and its users exposed. Furthermore, as technological advancements continue at a fast pace, OpenEPI could face challenges if it relies on technologies that become outdated or unsupported, necessitating timely updates or the search for new solutions that could disrupt the platform’s operations. Unlike proprietary systems with dedicated support, OpenEPI depends on community-driven support or its internal capabilities, requiring proactive management and substantial resources to ensure continued effectiveness and resilience against disruptions.

While releasing its source code promotes an open environment, it also introduces risks. Exposing the code publicly can lead to potential exploitation of any undiscovered vulnerabilities by malicious actors. Furthermore, this transparency allows competitors to view and possibly replicate proprietary approaches or techniques used within OpenEPI, potentially eroding its competitive edge.

7.8.4 Technical Debt

Technical debt is an inevitable aspect of software development, often compared to taking out a loan. Just like financial debt, technical debt can serve a useful purpose, allowing development teams to push forward quickly and adapt rapidly to changing needs or market demands. However, also like a loan, technical debt comes with "interest" — the additional costs and complexities that accrue over time if the debt is not managed properly.

In the fast-paced world of technology, taking on some level of technical debt is often unavoidable. This strategic move can enable a platform like OpenEPI to launch new features

more quickly or meet critical deadlines. In these scenarios, technical debt is consciously incurred with a clear understanding of why it’s being accepted and what the repayment plan will be. It acts as a calculated risk, underlining the necessity of balancing speed in innovation with long-term code maintainability and system stability.

Just as with financial debt, technical debt requires a plan for repayment. For OpenEPI, this means embedding practices into the development lifecycle that prioritize the regular refactoring of code and the updating of systems. Part of this proactive approach includes staying current with the latest versions of libraries and technologies used within the platform. Keeping software up to date is crucial to avoiding security vulnerabilities and maintaining compatibility with new software features and improvements.

While technical debt can be beneficial under certain circumstances, it must be managed meticulously to prevent it from spiraling out of control. Unaddressed technical debt can accumulate "interest", making future changes more difficult and time-consuming, and potentially leading to increased bugs, system downtime, and a decrease in code quality. As more resources are diverted to manage the accruing interest — in the form of patches, workarounds, and increasingly complex maintenance efforts — less time is available for new developments and essential innovations.

7.8.5 Availability of competence

The sustained success and evolution of the OpenEPI platform depends on the ability to secure and retain the necessary technical expertise. As the platform grows and technologies advance, finding, acquiring, and maintaining the right talent becomes increasingly challenging, posing significant risks to the platform's capacity to innovate and stay competitive.

One of the risks facing OpenEPI is the scarcity of skilled professionals with expertise in the technologies that the platform employs. As new technologies emerge and existing ones evolve, the demand for skilled professionals often outstrips supply. This talent gap can lead to difficulties in recruiting individuals who have the necessary skills to develop, maintain, and enhance the platform effectively.

The availability of competent technical staff directly impacts OpenEPI’s ability to maintain, troubleshoot, and upgrade its systems efficiently. Inadequate expertise can lead to delays in addressing technical issues, slower implementation of new features, or even the inability to fully embrace new technological opportunities. This can hinder the platform's ability to adapt to changing technological landscapes or user needs, potentially affecting user satisfaction and the platform’s overall relevance.

7.8.6 Robustness and scalability

The infrastructure underpinning OpenEPI must not only support operational demands required by the pre-project but also scale effectively to accommodate future growth and technological advancements. This is particularly challenging given the extensive data volume and processing power required for handling nature and climate data on a global scale.

Several key risks associated with the robustness and scalability of OpenEPI's infrastructure need to be considered to ensure the platform's long-term viability and efficiency.

Handling massive data volumes

One of the principal challenges for OpenEPI is the management of vast amounts of environmental data. As the platform aims to provide comprehensive insights into climate and nature-related phenomena, the data volume it needs to process, and to store, is enormous. This represents significant risks in terms of database management, data retrieval speeds, and overall system responsiveness. Ensuring the infrastructure can handle large datasets without performance degradation is critical.

Ensuring technical availability

Uptime is important for OpenEPI, particularly if the platform is to be relied upon by developers and researchers for making reliable applications with nature and climate data. Downtime can have significant repercussions, not only disrupting user operations but also diminishing trust in the platform's reliability. Ensuring consistent uptime requires a resilient infrastructure that can handle failures and automatically recover without affecting data availability or system performance.

Cost of infrastructure

Scaling the infrastructure to meet growing data and processing demands invariably leads to increased overall costs. These include expenses related to data storage, computational resources, and energy consumption, among others. Balancing these costs while maintaining high service levels poses a financial risk, especially as OpenEPI expands its capabilities and user base.

7.8.7 Mitigation strategies

Strategies for meeting the risks related to the technological infrastructure are:

- Implementing **centralized management tools** that can handle deployments, monitoring, and operations across multiple clouds can reduce complexity. In the pre-project *Pulumi* has been used for configuration management, and *ArgoCD* has been used as a continuous deployment tool.
- Where possible, **standardize the environments** across different vendors to minimize the variations in Kubernetes implementations. This could involve using the same configurations, and third-party tools across environments.
- Design the architecture to **simplify and decouple components** as much as possible. Using microservices architecture can help isolate functions and reduce dependencies between different parts of the system.
- Strategically **distributing services** based on themes across different cloud vendors can optimize resource utilization and enhance service performance.
- Conduct regular **security audits** and adopt proactive security practices to identify and fix vulnerabilities.
- Implement **automated scanning tools** for dependencies and vulnerabilities to ensure continuous monitoring and immediate identification of potential security issues.

- Foster a **culture of continuous learning** and adaptation within the organization to maintain technical agility and resilience.
- Construct the platform using **industry-leading and widely recognized technologies**.
- Invest in **continuous professional development** to keep the team's skills relevant, adapting to new technologies and methodologies as they arise.
- Integrate **refactoring into daily tasks**, whether maintaining existing features or developing new ones. Educate the technical team to view this as a standard practice, helping to prevent the onset of the broken window syndrome by maintaining consistently high standards in code quality.
- Utilize cloud services that provide **scalability and flexibility**, enabling the infrastructure to adapt seamlessly to fluctuating demands without substantial initial investment. This approach allows for both expansion during peak usage and cost-effective scaling down as demand wanes.
- Regularly review and **optimize infrastructure costs**, employing strategies such as reserved instances for predictable workloads, scaling down during off-peak times, and using cost-effective data storage options, like hot and cold storage.
- Design the infrastructure with **built-in redundancy and failover** solutions to minimize downtime and ensure data is consistently available.

7.9 Lack of transparency

Transparency is important to build trust and allows users to understand the limitations and potential biases of the data. Lack of transparency in open data platforms can lead to several significant risks, undermining OpenEPI's credibility, effectiveness, and user trust.

Developers and other users are less likely to trust and rely on a platform that does not provide clear information about data sources, update processes, data provenance and accuracy. This lack of trust can hinder user adoption and engagement. It is also important to be transparent about all important operations of the platform, including the work on data quality, what specific competencies OpenEPI houses, the governance structures, and the decision-making processes. Especially, many will be much interested in understanding the criteria for prioritizing from the backlog of datasets, suggested by developers. Openly addressing challenges and how they are managed can also build trust. Generally, there is also a need for transparency regarding funding and financing of OpenEPI.

To mitigate the lack of transparency risks, open data platforms can adopt several strategies, including:

- Comprehensive **documentation** and **metadata**: Providing detailed documentation and metadata about the data, including its source, collection methodology, update frequency, and any limitations or biases, can help users better understand and interpret the data accurately. To provide metadata above the strictly necessary could be time-consuming, but perhaps rewarding in the long run.
- Open **communication channels**: Establishing open channels for communication and feedback allows users to ask questions, report errors, and suggest improvements, fostering a sense of community and collaboration. In OpenEPI, this will be particularly important, as long as the developers' feedback is an important part of the service improvement mechanism.

7.10 Gender equity risks

OpenEPI is confronted with risks related to gender equity. These risks are rooted in existing gender disparities, biases in how data is collected and interpreted, and the varying levels of accessibility to data and to necessary technological resources. Achieving gender balance within OpenEPI is identified as a strategic priority, and addressing gender equity risks is considered crucial for ensuring that the platform fulfills its strategic objectives.

7.10.1 Representation bias

OpenEPI is dependent upon data collected by a wide group of different organizations (data providers). Data collected and made available by these data providers may not adequately represent women, especially if the data collection processes overlook or underrepresents women. If OpenEPI processes this data further, without having a conscious reflection upon the challenges related to bias, misrepresentation or data gaps in the data they process and aggregate further, this can lead to skewed insights and decisions that do not reflect the needs or realities of women.

7.10.2 Access inequality

In numerous developing economies, particularly in Africa, significant gender disparities exist in terms of access to technology and the internet. As previously mentioned, in sub–Saharan Africa women are 36 percent less likely to have access to mobile internet than men.[87] This disparity could hinder women from taking advantage of the products and services developed based on data from OpenEPI. Further, given the fact that there is a predominance of male developers across both developed and developing nations, OpenEPI faces the challenge of attracting a substantial user base of female developers with the requisite skills and competencies in the field.

7.10.3 Mitigation Strategies

There are some important mitigation measures that could be taken, for instance:

- **A gender balanced organization:** Gender balance in the OpenEPI organization (including management, strategic boards and other decision structures) is crucial for several reasons, particularly when considering the response to user needs, data selection and data aggregation processes. Ensuring female representation in the core business processes of the OpenEPI entity is essential to achieve data diversity and address the planetary crisis effectively. Women may have unique insights and priorities related to climate change, nature, and measures for climate change adaptation that can enrich the data creation process and the design of the services from the platform. Their involvement can help identify data gaps and capture nuances that might be missed in a less diverse staff. A gender balanced organization will help to ensure that OpenEPI's data is comprehensive, unbiased, and relevant to all.
- **Promote access and participation through partnerships:** Implement initiatives aimed at increasing data science/data analytics and software development skills among female students and in local, female led start-ups. This could include training

[87] GSMA Intelligence, *The Mobile Gender GAp Report 2023,* 2023, https://www.gsma.com/r/wp-content/uploads/2023/07/The-Mobile-Gender-Gap-Report-2023.pdf

programs, scholarships, hackathons and community-based projects that encourage women to take advantage of OpenEPI to develop new digital services and solutions. OpenEPI should also collaborate with local organizations and women's groups to gain insights and support for selecting data that promote and secure gender equality and female perspectives on climate related challenges.

- **Policy and guidelines:** Be aware of the need for OpenEPI policies and guidelines to support gender equality. This includes guidelines for equitable data selection, aggregation and distribution, as well as mechanisms to address potential biases in data and algorithms.
- **Monitoring and evaluation:** Develop metrics and regularly monitor and evaluate OpenEPI's impact on gender equality. Use these insights to adjust strategies and practices to better serve the needs of all genders.

7.11 Legal and liability risks

The legal jurisdiction under which OpenEPI will operate remains undetermined. Nonetheless, irrespective of the jurisdiction it eventually falls under, the platform must proactively address and mitigate several legal and regulatory risks.

7.11.1 Liability risks

For open data platforms, the risks related to liability are particularly pertinent due to the nature of open data sharing and the diverse uses of this data. Given the potential impact of the use of the data on economic, social, and environmental outcomes, the implications of inaccuracies, misuse, or misinterpretation can be significant. Establishing clear liability and accountability for the accuracy, completeness, and usability of data shared on the platform is hence crucial. Clarifying responsibilities for any downstream risks, including misuse, manipulation, or other harms arising from the use of data from the platform, is essential.

OpenEPI must develop comprehensive terms of use and disclaimers that clearly outline the responsibilities of the platform and its users. This should include statements that users bear the risk of using the data and that the platform does not guarantee data accuracy or completeness. OpenEPI should further regularly consult with legal counsel to stay informed about changes in data protection laws and other relevant regulations that affect the platform. This will help the platform adapt its policies and processes in compliance with legal requirements.

7.11.2 Data ownership risks

The legal owner of the data is not always the legitimate owner of the data. A typical example of this dilemma is seen in disputes over land ownership data, where government entities may assert ownership over the data, while indigenous populations argue that the data rightfully belongs to them due to their historical and cultural ties to the land. Such conflicts highlight the disparity between legal and moral claims to data ownership, particularly in scenarios where state interests clash with those of individual groups or communities.

In developing countries, the absence of comprehensive legal frameworks for data sharing exacerbates these challenges, leading to uncertainties and potential legal disputes. This situation underscores the necessity for platforms like OpenEPI to implement robust

mechanisms for verifying the ownership of data they collect and utilize from other data providers and platforms. Establishing clear and transparent terms of use, licenses, and permissions for data sharing is crucial in navigating these legal complexities and in mitigating conflicts related to data ownership and access. Further, maintaining open channels of communication with users to promptly address concerns and disputes can help resolve issues before they escalate into legal challenges.

7.11.3 Intellectual Property Rights (IPR) risks

Navigating the complexities of intellectual property rights (IPR) in the context of sharing open data is crucial for OpenEPI. IPR issues encompass a wide range of legal considerations, including licensing, copyright, and ownership, which are critical to respect the rights of data creators and avoid potential legal disputes.

Firstly, it is essential to implement a robust licensing framework that clearly defines the terms under which data can be used, shared, and modified. Open data licenses, such as those provided by Creative Commons, offer varying levels of openness and restrictions, allowing data providers to specify how their data may be used by others. These licenses can facilitate the legal sharing of data while protecting the rights of the original creators, ensuring that users are aware of and comply with the terms of use.

Secondly, understanding and respecting copyright laws is crucial in the dissemination of open data. Copyright laws vary significantly across jurisdictions and can affect everything from raw data to derived products and visualizations. OpenEPI should ensure that data shared on the platform does not violate copyright laws, which may require obtaining permissions or licenses from copyright holders or relying on data that is in the public domain.

Furthermore, establishing clear ownership agreements is vital, especially when dealing with data that may have multiple stakeholders or sources. This is particularly relevant for nature and climate data, where the provenance of data can be complex, involving contributions from governments, research institutions, indigenous communities, and crowd sourced data. Clear agreements help delineate the rights and responsibilities of all parties involved, minimizing the risk of conflicts and ensuring the ethical use of data.

In addition to these measures, OpenEPI should also engage in proactive communication and education efforts to inform data providers and users about IPR considerations. This could include providing resources and guidelines on selecting appropriate licenses, understanding copyright implications, and negotiating ownership agreements.

Moreover, as the legal landscape around open data and IPR continues to evolve, staying informed about legal developments and adapting policies and practices accordingly will be crucial for OpenEPI. Engaging legal experts specializing in IPR and open data can provide valuable insights and guidance, ensuring that the platform remains compliant with existing laws and best practices in the field of open data sharing.

7.11.4 Mitigating strategies

The key mitigating strategies for handling liability, data ownership, and intellectual property rights (IPR) risks for OpenEPI can be summarized as follows:

- **Clear liability and accountability:** Establish clear terms of use and disclaimers that outline the responsibilities of both the platform and its users. This includes statements that users bear the risk of using the data and that the platform does not guarantee data accuracy or completeness.
- **Legal compliance and consultation:** Regularly consult with legal counsel to stay abreast and informed about changes in data protection laws, copyrights laws rights and other relevant regulations. Adapt policies and processes in compliance with these legal requirements to mitigate liability risks.
- **Data ownership clarification:** Implement robust mechanisms for verifying the ownership of data collected and utilized from other providers. Establish transparent terms of use, licenses, and permissions for data sharing to navigate legal complexities and mitigate conflicts related to data ownership and access.
- **Intellectual property rights (IPR):** Develop a robust licensing framework that clearly defines the terms under which data can be used, shared, and modified. Utilize open data licenses, like those offered by Creative Commons, to facilitate legal sharing and protect the rights of original creators.
- **Copyright Compliance:** Ensure that data shared on the platform does not violate copyright laws, which may involve obtaining permissions or licenses from copyright holders or relying on public domain data.
- **Proactive communication and education:** Engage in communication and educational efforts to inform data providers and users about IPR considerations. Provide resources and guidelines on appropriate licenses, copyright implications, and negotiating ownership agreements.

7.12 Reputational risks

Maintaining public trust is crucial for the reputation and success of OpenEPI. If users lose confidence in the reliability or management of the data, they may be hesitant to use or contribute to the initiative. Reputational risks associated with open data platforms stem from various factors, including data accuracy, privacy concerns, and the ethical use of data. These risks can significantly affect the trust and credibility of the platform among users, stakeholders, and donors. Managing these risks effectively is crucial for the sustainability and impact of open data initiatives.

Open climate data initiatives, primarily targeting developers in developing countries, face unique reputational risks. These risks are not only linked to the inherent challenges of managing open data platforms but to the potential sensitive nature of some environmental data (discussed above), and the specific needs and vulnerabilities of developing countries. Addressing these concerns effectively is crucial for maintaining the credibility and impact of OpenEPI, in addition to the risks mentioned earlier in this chapter.

7.12.1 Eurocentrism

Eurocentrism presents a significant reputational risk for global open data platforms, especially those aimed at users in developing countries. This risk arises when platforms, either through their data selection, governance structures, or user engagement strategies, reflect a predominantly European perspective that may not align with the realities, needs, or priorities of developing countries. The risk of eurocentrism can also follow from building data ethics principles and guidelines on European regulation and perspectives and not including perspectives from developing countries. Such an orientation can marginalize non-European culture, local knowledge systems and data priorities, making the platform seem irrelevant or insensitive to local contexts. The perception of Eurocentrism can erode trust among target users and stakeholders in developing countries, who may view the platform as catering more to European interests than to fostering global inclusivity and diversity.[88]

Mitigating this risk requires a conscious effort to incorporate a wide range of perspectives in data collection, platform governance, and community engagement, ensuring that the platform truly serves the global community it aims to support. Prioritizing inclusivity and diversity in this manner not only enhances the platform's relevance and utility for developers in developing countries but also strengthens its reputation as a genuinely global open data initiative.

7.12.2 Only selected groups reap benefits from the platform

If OpenEPI becomes a platform whose benefits are predominantly accessible to a privileged few, it risks sustaining reputational harm and perpetuating inequalities. This concern is intricately linked to the broader issue of the digital divide, particularly pronounced in sub-Saharan Africa where only a small fraction of the population has access to smartphones, and an even smaller subset to mobile internet connections. Recent trends indicate that despite technological advancements, a significant digital and economic gap remains, hindering equitable access to climate mitigation solutions. For instance, innovative technologies such as smart agriculture apps have the potential to transform lives, yet their reach is often limited to those with the necessary digital tools and literacy.[89]

To mitigate these risks and ensure the sustainability (included long term funding) of the platform, OpenEPI must proactively ensure that the products and services developed through its data are equitably distributed across all societal segments. This involves designing inclusive strategies that address the barriers to access and use of climate mitigation solutions, particularly in underserved communities. Initiatives could include partnerships with local organizations to deliver technology training and targeted efforts to bridge the gender gap in technology use.

These initiatives can be effectively advanced through strategic collaborations with organizations like Norad and other development aid agencies, which have an established presence in the targeted regions, along with valuable contextual insights and a robust local network. Leveraging their on-the-ground experience and expertise can significantly enhance

[88] Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre

[89] Verhulst, Stefaan G. & Andrew Young, *Open Data for Developing Economies*, 2017

the reach and impact of these initiatives, ensuring they are tailored to meet the specific needs and circumstances of the communities served. Such partnerships can also facilitate access to additional resources, amplify advocacy efforts, and foster a more coordinated approach to bridging the digital divide and promoting equitable access to climate mitigation solutions.

7.12.3 Global power dynamics and dependencies on Big Tech Platform-Vendors

Relying on big tech platform-vendors for infrastructure, software, or services introduces another layer of reputational risk, particularly in the context of global power dynamics. This dependency can make the open data platform subject to the whims of these vendors, including changes in terms, pricing, or the discontinuation of services. Moreover, it can raise concerns about data sovereignty, as data hosted on platforms controlled by entities in powerful countries might be subject to foreign law, surveillance or control.

This situation can exacerbate feelings of distrust and vulnerability among the users of open data platforms in developing countries, who may perceive that the data provided, and the digital infrastructure are under the control of foreign corporations.[90]

Such dependencies also highlight concerns about the digital divide and the reinforcement of existing inequalities in the global technology landscape. If local developers and secondary users of OpenEPI are dependent on foreign tech giants to take advantage of the platform and services developed based on the platform, it can limit their autonomy and hinder the development of local digital ecosystems.

More on this risk also in the section above on technological risks, and in previous chapters.

7.12.4 Mitigating strategies

Addressing reputational risks requires a nuanced understanding of global power dynamics and a commitment to fostering an open data ecosystem that is equitable, resilient, and respectful of the diverse needs and contexts of its users. By implementing the following management practices, in addition to the measures already mentioned in this chapter, OpenEPI can secure and maintain trust among its users, stakeholders, and donors, ensuring its long-term success and impact:

- **Monitor and evaluate impact:** Continuously monitor the platform's impact and share these evaluations with the public. Demonstrating positive outcomes and learning from feedback can enhance the platform's reputation and will help in building a sustainable platform.
- **Develop a crisis response strategy:** Be prepared to respond swiftly and effectively to any issues that could harm the platform's reputation, such as data breaches or controversies over data use. Transparent and accountable handling of such crises is crucial for maintaining trust.

[90] Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre

- **Equity and inclusion focus:** Proactively ensure equitable distribution of OpenEPI's data-driven products and services by designing inclusive strategies to overcome access barriers in underserved communities, including partnerships for technology training and efforts to close the gender technology gap.
- **Strategic collaborations:** Enhance initiative impact through strategic collaborations with development aid agencies like Norad, leveraging their regional presence, contextual knowledge, and networks to tailor solutions to local needs.
- **Diversification of services and vendors:** Avoiding reliance on a single vendor and instead using a mix of services from different providers, including local or regional providers that may offer more relevant and tailored services.
- **Data sovereignty measures:** Implementing measures to protect data sovereignty, such as hosting data within the country or region it pertains to and ensuring legal protections for data stored on foreign servers.
- **Capacity building:** Investing in local capacity building to develop indigenous technologies and platforms can reduce dependency on big tech vendors and foster a more equitable global digital ecosystem.

8. Conclusions and recommendations

8.1 Summing up - what have we done?

Over some months, we have explored the feasibility of an open data platform aiming at innovation of climate change mitigation and adaptation related solutions. We have explored data sources, and we have explored what such a platform could technically look like and have described a blueprint for this - with an emphasis on describing a vendor-independent technology stack. In addition to strictly following the specifications for open source and open data, the key motivation has been to make the technical arrangements of an open platform independent of a particular technology supplier. Further, we have assessed how to organize and implement the necessary organizational structure for operating and govern the platform, data and related services to developers. As part of this, we have suggested an open policy on data and technology, supporting the suggested arrangement.

As part of our study, we have assessed the somewhat immature “market” for open data, the landscape of existing initiatives and stakeholders, and discussed how to reach our goals - in the light of a ToC (theory of change) stating the intended impact on climate change resilience and a boost of local innovation in many countries and regions.

As part of the project, we have also established data sets and prototyped solutions where open data are exploited in building simple end user applications related to climate change. Thereby and by interviews and document studies we have gathered insight in user needs and the mechanisms of local innovation in sub-Saharan countries.

8.2 Our recommendations to Norad

OpenEPI is a comprehensive and ambitious concept in many ways, with great opportunities, but also with risks. As a general conclusion, we think the world needs OpenEPI. We therefore recommend that Norad implements OpenEPI as a concept, by targeted and tailored, subsequent and long-term grants. This recommendation comes with a number of prerequisites:

- The platform should be established with the most important datasets first, and with a defined core of services that there is good reason to believe will be demanded by innovators in countries struggling with climate change adaptation and mitigation.
- We have explored and demonstrated a vendor independent technology stack supporting open access to relevant data for climate change mitigation. Our study provides a technical blueprint for a full scale OpenEPI and should follow as specifications and strong recommendations for the potential grant recipients.
- We recognize long term data storage of crucial data as an imperative part of a future, global open data infrastructure, but we have not assessed this specifically in this study. This must be done in a dedicated, follow-up assessment. The building of a global, digital infrastructure for long term data storage could be addressed in a specific, targeted and multi-annual grant. Other possible arrangements could also be feasible, including the building of a permanent agency or national data storage body under international supervision.

- More differentiation on which country to address is needed. Our findings imply there are varying needs and indeed varying potential for user uptake among developer communities (as well as end users) in different regions and political contexts.
- The platform must be governed under a regime that includes strict policies on data quality, metadata standards, standardized API specifications, documentation of relevance, service specifications, and recommended use areas for data and data products provided by the platform.
- As we argue for a distributed OpenEPI organization, there is a need for a core organization governing the digital infrastructure, policies and legal and administrative arrangements. This could be established under Norwegian government and legislation, but it can as well be established elsewhere, under a multi-annual grant from Norad. Furthermore, domain-specific initiatives that follow the OpenEPI concept may well be assigned to organizations other than the core unit and be funded by more thematically limited calls. Independent of who is receiving grants to cooperate with OpenEPI in such a distributed layout, the core unit should seek partners internationally to expand its capabilities and reach.
- At start-up, the overall OpenEPI organization will be at a minimum 22 full-time equivalents, with technical, data analysis, domain specific and communication expertise. The number of FTEs at the full-scale implementation is estimated to be about 60.
- Based on our explorations of available open data and user needs, and on Norad's priorities of domains, the data sets in the initial setup of the platform should be selected from data sources related to weather, flood, soil health, vegetation coverage and biodiversity.
- An international advisory board should be established to provide the business with ongoing support for assessments of ethical aspects of the activities, prioritization of new datasets and data products, and for strategic advice on the platform's role in different ecosystems. The board should have participants from academia, from the user community (developers and innovators in LMIC countries) and from the open movement.
- The business will gradually be expanded, as new data and topics are included, new aggregated data products are incorporated, more services are added, and the user volume grows. The services provided by the organization should be focused on the mentioned data at first but expand and scale to new data domains and advisory services on components, open source code, architecture, and related topics.
- A partner programme should be established to ensure academic support in the research community in the areas of environment, agtech and climate adaptation, but also to strengthen the user acceptance, reach and financial support.
- Most important, a parallel programme for systematic demand stimulation should be initiated by Norad, through marketing activities, training materials and other pedagogical instruments, support services and agreements with academia and incubator environments in relevant geographical areas.
- The necessary funding of approximately NOK 35-45 million per year in 2025, increasing to approximately NOK 90-100 million in a few years, must be secured by designing the necessary grant programme in Norad. Funding and financing mechanisms must be further assessed, when the first call for proposals is closed and Norad knows what kind of organization will oversee the initial building of the core unit.

- There must be a proper recognition and handling of the above identified risk factors, i.e. both technical (open source vulnerability, complexity, supplier-agnosticity, scalability), organizational (responsibilities, robustness, relational, financial, etc.), ethical and legal nature.

By our recommendations, we are aiming at a scalable, open and functional technical platform, providing crucial data and technology support for new services and digital solutions for climate mitigation and climate change adaptation. We think the time frame for the gradual implementation of OpenEPI will be from 2025 to 2028. The focus in the start will be on open data of relevance for agriculture, fisheries and food production in sub-Saharan countries, but eventually the geographical focus will be extended to the global arena. The platform will gradually also provide more services, for instance advisory services on licensing and policy issues, data standards, or how to build flexible solutions based on open source technology.

OpenEPI will become not only a technical platform for data access, but also a platform-of-platforms, and a widely recognized organizational entity with their support and advisory services. As such, OpenEPI will also serve as an open policy “caretaker” and set the agenda for openness, transparency and inclusion. It is important that this role is exercised in close consultation with and fully complementary to the efforts of Digital Public Goods Alliance, avoiding duplication of roles and resources.

The core entity that runs OpenEPI must from the very beginning put in place a functional unit that constantly follows up the risk factors. The unit must carry out thorough assessments both during the establishment of the platform and related services, and during OpenEPI’s running operations. This will be crucial in itself - to manage known risks, but also for maintaining the credibility and trust on which this type of service will be completely dependent. In the chapter on risk, such a function is described in more detail.

Crucial for a successful development of OpenEPI as a digital public good, as an organization and as services for innovation purposes, will be the emphasizing of stakeholder’s engagement and involvement. A central part of this will be extensive communication efforts, both targeting the most relevant data providers, funding sources and professional bodies of climate and environmental expertise, but also more generally - towards the open movement, media, NGOs and national authorities and actors across the globe. We therefore suggest an extensive marketing and communication strategy as part of the implementation roadmap.

The greatest risk of failing with OpenEPI as a concept is the uncertainty of how the user uptake will develop in the developer communities of those countries and regions we intend to reach. Referring to our initial theory of change, and the long term success criteria for OpenEPI that we have described, we can’t be sure OpenEPI meets the needs of those countries which need the most support for meeting climate change. OpenEPI will not be able to secure the intended impact of strengthened resilience, democratic access, improved innovation capacity and economic prosperity. OpenEPI must therefore be combined with other measures and programmes suited for strengthening the demand side and the capabilities of the varying local innovation systems in different countries and regions.

9. Appendices and annexes

9.1 Stakeholders covered by or interacted with in this study

Stakeholder	Web	Domain	Description
GBIF	https://www.gbif.org/	Biodiversity	An international network and data infrastructure funded by the world's governments and aimed at providing open access to data about all types of life on Earth.
Digifarm	https://digifarm.io/	Earth Observation, Soil	Norwegian ag-tech start-up using AI to provide solutions for accurately detecting field boundaries and seeded acres in precision farming.
GEO (Group on Earth Observations)	https://earthobservations.org/	Earth observation	Is creating a Global Earth Observation System of Systems (GEOSS) to better integrate observing systems and share data by connecting existing infrastructures using common standards. There are more than 400 million open data resources in GEOSS from more than 150 national and regional providers
World Resource Institute (WRI)	https://www.wri.org/	Forest, Ocean	Produce data sets, data products and data-based tools, freely available through their open data commitment: https://data-api.globalforestwatch.org/
CGIAR	https://www.cgiar.org/food-security-impact/one-cgiar/	Soil	A leader in agricultural science and innovation for development
ISRIC Data Hub	https://data.isric.org/geonetwork/srv/eng/catalog.search#/home	Soil	World Data Centre for Soils (WDC-Soils) provides a focal point for soil-related collections and information services, and to ensure their long-term preservation and archiving. Our holdings include soil specimens, country documentation, and geo-referenced databases.
Varda SoilHive	https://www.soilhive.ag/	Soil	Varda is an ag-tech start-up enabling farm & field data sharing. The platform is a collaborative platform where you can find relevant soil data for your projects, create your own queries, generate tables and maps, and easily save and share them.
ISDA	https://www.isda-africa.com/isdasoil/developer/	Soil (Africa)	Data API: https://www.isda-africa.com/isdasoil/ iSDA's vision for spatial agronomy combines the use of cutting edge remote sensing data, combined with AI and decision science.

OpenET	https://openetdata.org/	Soil, Earth Observation	OpenET uses best available science to provide easily accessible satellite-based estimates of evapotranspiration (ET) for improved water management across the western United States.
Meteorologisk institutt	https://api.met.no/	Weather	Yr.no, interviewed in the context of experiences from field work on weather forecasts and user cases in African countries.
Arkivverket	https://www.arkivverket.no	Documents	The National Archives of Norway, interviewed in the context of long term data storage
BarensWatch	https://www.barentswatch.no/	Environment	An open information system with services for end users, presented in the portal www.barentswatch.no. In addition, a shielded monitoring system contributes to the efficiency of operational efforts.
Digitaliserings-direktoratet	https://www.digdir.no	Digitalization	The Norwegian Agency on data and digitalization policy, interviewed in the context of open data policy measures.

9.2 Interview guide - data portals and platform owners

Name of organization:

Date of interview:

Name and title of respondent:

Role and responsibilities in the organization:

Theme	Answer
Background Can you provide a brief overview of your open data platform (and services) and its primary objectives? What motivated the establishment of the platform?	
Open Data Principles Implementation How have you implemented the open data principles? Do you adhere to a specific open standard for representing data (eg. CC, Open API)? Do your data providers need to adhere to some defined principles/criteria, for instance domain specific metadata schemes? How do you follow up/verify that data providers adhere to your policy/criteria? Is all software being part of the platform open source? Is the data open to everyone (free access or payment options? log in solutions?) How does your platform ensure transparency in data sourcing and processing?	
Data governance and management How is data governed and managed on your platform? (what is most important to have in place, what is most challenging)? Data Collection and Sources - How do you collect data for your platform? - Are there specific sources or partnerships that contribute to the data on your platform? - How do you recognize information gaps/need of new data sources and how do you prioritize between user needs for data?	

Data quality - What measures are in place to ensure the quality and accuracy of the data? - How often is data updated? - Are there community feedback mechanisms in place to improve data quality? Data use: - How do you monitor data usage on the platform? - What metrics do you use? (KPI for data use?) Long-term data storage - For how long is data stored? - How do you manage long-term data storage?	
Governance issues How is the platform managed, governed, and organized? (What is the legal status? Does it fall under a specific jurisdiction?) In your opinion, what are important prerequisites to secure sustainable, long-term governance and management of an open global data platform? What kind of support, and from whom, is important? Major source of funding? Robustness of funding - long-term fixed/secure funding? What type of key competence is needed to run the platform? (domain expertise, IT/data expertise, marketing, community engagement) Scalability: What are your experiences on dimensioning the competences and services necessary for platform operations? Liability: how is responsibility manganged between you and the data providers (eg in terms of liability for data quality). Your responsibility vs data holders? How is the day-to-day management of the platform organized, eg. how many employees are needed to run the platform? Any metrics/KPI's measuring the impact of the platform? Challenges: - What aspects are most challenging when it comes to managing a successful open data platform? - What challenges have you encountered? - What common pitfalls must be avoided?	

Stakeholder engagement Can you describe the ecosystem around your data platform (eg. major stakeholders and entities influencing you and your initiative)? Who are the primary stakeholders having benefits from your open data platform? How does your platform (and your organization) encourage collaboration with other organizations or platforms in the same domain? How do you engage with users - and potential users - of your data platform? How do you identify their needs and possible information gaps? What measures have you put in place to stimulate use (eg. online learning resources, easy-to-use guidelines)? Incentives for use (eg. payment for contributions or feedback)	
Platform architecture Describe the architecture and functionality/user interface API's etc. Use of AI Storage technology	
Legal and regulatory issues How do you handle data validation and verification processes? Eg to clarify/validate the legal ownership of the data? How do you address data privacy and security concerns? What jurisdiction does the platform fall under, do you need to follow GDPR for instance?	
Risks What do you perceive as the greatest risks? (eg. reputational risk, cyber threats, misuse or manipulation of data, data bias, user adoption) Are there some specific risks connected to being an open source platform? How have the risks been mitigated? Liability: the greater the responsibility the greater the risk	

Future Developments Are there any upcoming features or improvements planned for your open data platform? (data, ai, functionality, areas) How do you envision the evolution of your platform in the context of Digital Public Goods/Open data/FAIR principles? How do you perceive the future for open global data platforms? Will AI and new ways of sharing data, make them obsolete?	
Final Thoughts Is there anything else you would like to share about your platform and how you manage it?	

9.3 OpenEPI Open policy

OpenEPI Open policy - Beta version - February 2024

In general, it is often the case that copyrightable works produced for general and non-profit purposes, such as those for climate change adaptation and mitigation, are not openly licensed and disseminated widely to the public. We have found that many stakeholders and other members of the public are generally not aware of the resources created as a result of aid programmes, research projects and other international initiatives. Intentionally or unintentionally, the resources are often created and disseminated locally or disseminated to limited audiences. Even when the resources are known to exist, stakeholders and the public are not sure how to access them, what usage rights or permissions are necessary to use them, or how to obtain those rights or permissions.

The Open Policy we are adopting here addresses these key problems. It explicitly gives permission to the public to access, reproduce, publicly perform, publicly display, and distribute the copyrightable work; prepare derivative works, and reproduce, publicly perform, publicly display and distribute those derivative works; and otherwise use the copyrightable work, provided that in all such instances attribution is given to the copyright holder.

We believe that the implementation of this policy will result in significantly enhanced dissemination and use of deliverables created and provide stakeholders and members of the public with a simpler and more transparent framework to access, use, and modify these deliverables for the benefit of their communities.

Open Policy

This open policy and its requirements apply to all data created or funded by the Open Earth Policy Initiative (OpenEPI) and all data and other works distributed or made available from other sources, on the OpenEPI data portal, or in any way referred to or recommended by OpenEPI.

As a digital innovation platform and digital public good, OpenEPI has adopted an open policy that:

- Enables the unrestricted access (except for an attribution requirement) and reuse of all collected or produced digital data created or funded by OpenEPI, and all data linked or referred to, or disclosed, reused, recombined, recommended or shared by OpenEPI, including any underlying data sets in projects that collect or produce data, or are in any way made available by OpenEPI.
- Enables the unrestricted access and reuse of software, algorithms and models created or funded by OpenEPI, and all software, algorithms and models linked or referred to, or disclosed, reused, recombined, recommended or shared by OpenEPI.
- Enables the unrestricted access and reuse of all peer-reviewed published research created or funded by OpenEPI, including any underlying data sets.
- Enables the unrestricted access and reuse of educational resources created or funded by OpenEPI.

The policy will apply both to the deliverables themselves and any support materials necessary to the use of the deliverables.

OpenEPI applies the necessary reviewing and endorsement mechanisms for the data, software, algorithms, models and research publications linked or referred to, or disclosed, reused, recombined, recommended or shared by the platform.

This policy is effective from March 2024.

As new practices emerge, this policy and guidelines will be reviewed and updated as needed. Any questions or suggestions should be sent to contact@openepi.io

More details and guidelines on Open Policy:
Open Policy for Data and Underlying Data
Open Policy for Software, Algorithms and Models
Open Policy for Research
Open Policy for Educational Resources
Availability statements for data, software, algorithms and models, research, and educational resources
FAIR Principles and Open Data
Why is access to underlying data important?
Exclusions to the Open Policy

Open Policy for Data and Underlying Data

Data encompasses all primary data, associated metadata, all relevant aggregated data, and any additional relevant data necessary to understand and assess the data in question. For research this must include any data needed to replicate and reproduce the reported study findings in their totality.

Data can be compiled into any open data file type, including any necessary access instructions, code, or supporting information files, to ensure the file(s) can be accessed and used by others.

The data must be available for reuse under the following conditions:

- Data must be dedicated to the public domain using the Creative Commons Zero (CC0) Public Domain Dedication or licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license.
- Data must be available in a publicly documented machine-readable format and through application programming interfaces (API) and bulk download.
- Metadata and data must be available in an open, vendor-independent format.
- In the case of research articles, there must be immediate open access to the underlying research data upon publication of the article(s) or launch of the project.
- The dataset must have a persistent and unique identifier, such as a DOI (digital object identifier) to facilitate linking and citation.
- Provide long-term storage and preservation for the data.
- All documentation, learning resources and any other content related to the data must be made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license.

Compliance is a requirement of referring to, covering by or referencing by OpenEPI. This open policy applies to all data and underlying data. Compliance will be continuously reviewed and authors, data producers and data curators will be contacted by the OpenEPI team when data used or referred by OpenEPI are found non-compliant.

More specific preconditions or activities around pre-publication data planning (in the case of research projects), data collection, analysis, storage, sovereignty, informed consent, interoperability, and the use of domain specific standards (for instance for specific data types and metadata schemes) are reviewed and evaluated individually by the OpenEPI team.

Open Policy for Software, Algorithms and Models

The Open Policy for OpenEPI requires that relevant entities aiming at endorsement by OpenEPI, openly license and make publicly available all underlying software source code, algorithms, models, components and any training data sets to replicate production ready versions of the technologies developed.

The technology must be available for reuse under the following terms:

- Software must be published under an open source software license approved by the Open Source Initiative (OSI).
- All source code, algorithms and models must be openly available for reuse from a publicly available repository like Github or Gitlab.
- The documentation included in the publicly available repository must give all the necessary information needed for re-implementation of the source code.
- Producers of software, algorithms and models shall retain sufficient copyright in their work to ensure they have the necessary rights to openly license and share these works under the open policy.
- All training data (for machine learning) must be dedicated to the public domain using the Creative Commons Zero (CC0) Public Domain Dedication or licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license.
- Algorithms and models must be made available under standardized technical formats.
- All documentation, learning resources and any other content related to the technology must be made available under Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license.

This policy applies to all relevant Software, Algorithms and Models created or funded by OpenEPI. Compliance with this Open Policy is a requirement for receiving OpenEPI support or being linked or referred to, or disclosed, reused, recombined, recommended or shared by OpenEPI. Compliance will be continuously reviewed, and responsible providers or authors will be contacted when they are non-compliant.

Open Policy for Research

Research subject to the Open Policy should contain the following elements:

- All OpenEPI relevant research, including articles accepted for publication, will be published under “Open Access” terms. All research created or funded by the OpenEPI, including articles accepted for publication, shall be published under the CC BY 4.0 license, or an equivalent license. This will permit all users to copy, redistribute, transform, and build on the material in any medium or format for any purpose (including commercially) without further permission or fees being required.
- Authors of OpenEPI research shall retain sufficient copyright in their research to ensure that articles accepted for publication are deposited into an open-access repository and published under the CC BY 4.0 or an equivalent license.
- Publications and underlying data will be accessible and open immediately. All OpenEPI relevant research, including articles accepted for publication, shall be available immediately at publication, without any embargo period.
- The research data must be dedicated to the public domain using the Creative Commons Zero (CC0) Public Domain Dedication or licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license.
- Each accepted article must be accompanied by a Data Availability Statement that describes where any primary data, associated metadata, original software, and any additional relevant materials necessary to understand, assess, and replicate the reported study findings in totality can be found.
- OpenEPI can require that underlying data supporting the accepted article shall be immediately accessible and open upon article publication. Entities applying for OpenEPI compliance are encouraged to adhere to the FAIR principles to improve the findability, accessibility, interoperability, and reuse of digital assets.

- Publications will be available in a public open access repository. Accepted articles shall be deposited immediately upon publication in PubMed Central (PMC), or in another open access repository, with proper metadata. In addition to depositing the article in an open access repository, entities applying for OpenEPI compliance are encouraged to deposit their accepted article in a subject specific or institutional repository of their choice.
- Preprints of submitted manuscripts are encouraged. While not needed to fulfill the Open Policy requirements, entities applying for OpenEPI compliance are encouraged to deposit research consisting of their submitted manuscript, and its subsequent versions, on a preprint server under the CC BY 4.0 license.
- Compliance is a requirement for all research funded by OpenEPI. Compliance will be continuously reviewed, and authors will be contacted when they are non-compliant. Non-compliant research will not be encompassed by the OpenEPI data portal or by other OpenEPI channels.

Open Policy for Educational Resources

Educational resources subject to the Open Policy should contain the following elements:

- Educational resources will be published and shared under open terms. All educational resources applying to OpenEPI compliance shall be published under the CC BY 4.0 license, or an equivalent license. This will permit all users to copy, redistribute, transform, and build on the material in any medium or format for any purpose (including commercially) without further permission or fees being required.
- Producers of educational material shall retain sufficient copyright in their work to ensure they have the necessary rights to share these works under the CC BY 4.0 license, or equivalent license.
- Educational resources will be accessible and open immediately. All educational aspiring to be OpenEPI relevant shall be available immediately at publication, without any embargo period.
- Educational resources will be made available in an open repository. Educational resources relevant for OpenEPI shall be deposited in an open educational resource’s repository, with proper metadata tagging.
- Compliance is a requirement for all educational resources created or funded by OpenEPI. Compliance will be continuously reviewed, and authors will be contacted when they are non-compliant. Non-compliant educational materials will not be encompassed by the OpenEPI data portal or by other OpenEPI channels.

Availability statements for data, software, algorithms and models, research, and educational resources

Data sources providing materials covered by this Open Policy must include an Open Availability statement. This pertains to all projects and initiatives that have open research, data, software, algorithms and models, and educational resources associated with the project.

This statement must be added to the title page of the research manuscript prior to submission for all types of research, and to the title page or landing page of the educational resource(s). For general data and software, the availability statement must be published on the organization's website and on any public repository used as host.

The availability statement must include:

- A description of the open data and their metadata, or the software, algorithms and models developed, underlying data to research articles, and/or educational resources.

- A direct reference to where the data, source code, algorithms or models, research data or educational resources can be found for re-use.
- A direct reference to any documentation of the above mentioned resources, including a list of data and content assets and the open license (or public domain dedication) on each (linked to the relevant open license).
- If relevant, a direct link to a reference implementation.

The availability statement should not refer readers or reviewers to a data responsible contact or an author to obtain the data or resources but should instead include the applicable details where the data, software, algorithms, models and/or other content can be found.

FAIR Principles and Open Data

OpenEPI endorses the FAIR Data Principles as a framework to promote the broadest possible reuse of collected data.

In addition to the direct requirements described in this document, all entities aspiring to be OpenEPI compliant are encouraged to follow the FAIR data principles. They are guidelines to improve the Findability, Accessibility, Interoperability and Reuse of digital assets.

The FAIR data principles emphasize machine-actionability as we increasingly rely on computational support to distribute, handle and manage data, due to rapid increases in generated volumes and complexity.

The FAIR data principles:

- Support knowledge discovery and innovation
- Support data and knowledge integration
- Promote sharing and reuse of data
- Are discipline-independent and allow for differences in discipline
- Help data and metadata to be 'machine readable', supporting new discoveries through the harvest and analysis of multiple datasets.

The principles stress that data must be retrievable without specialized or proprietary tools or communication methods, and that data should be released with a clear and accessible usage license. Individuals and organizations that put FAIR data principles into practice may do so under a variety of data usage licenses. In other words, FAIR does not necessarily imply Open; data can be FAIR and shared under restrictions. OpenEPI requires data to be dedicated to the public domain using the Creative Commons Zero (CC0) Public Domain Dedication or licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0), or an equivalent license, and for all partners to follow the FAIR principles.

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

Using such a repository and identifier ensures that your dataset continues to be available to both humans and machines in a usable form in the future. To aid discoverability, data should also be followed by metadata appropriate for both finding (for instance by search engines) and in more detail describing the data.

The content and format of metadata is often guided by the specific discipline and/or repository, through the use of a standardized metadata scheme.

When depositing data in a repository, it is important that you fill in as many fields as possible as this information usually contributes to the metadata record(s).

In some cases, especially when using a discipline-or domain-specific repository, the submission of specific metadata files alongside the data may be required.

Actions:

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation. Data supporting partner research aspiring to be OpenEPI compliant should be openly published under the CC0 public domain dedication or the CC BY 4.0 license - both of which facilitate legal data reuse.

In these cases, OpenEPI has policies in place to allow the publication of articles associated with such data, while still maintaining the appropriate level of security. For guidance, please see the Availability statements section above.

Actions:

- A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
    - A1.1 The protocol is open, free, and universally implementable
    - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually needs to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. Interoperable data can be compared and combined with data from different sources by both humans and machines – promoting integrative analyses.

To bolster interoperability, data should be stored in a non-proprietary open file format and described using a standard vocabulary (where available). In some cases, the preferred file formats and vocabularies will be dictated by the repository you choose to host your data.

Actions:

- I1. (Meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow the FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable

The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. Data that is

findable, accessible, and interoperable is generally fit for reuse. On occasion, the inclusion of additional documentation alongside the data may be required to ensure that the data are understandable and thus reusable.

As a rule, someone who is not familiar with the data should be able to understand what it is about, using only the metadata and documentation provided. By extension, the same practices that enable data reuse also support reproducibility.

Actions:

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
    - R1.1. (Meta)data are released with a clear and accessible data usage license
    - R1.2. (Meta)data are associated with detailed provenance
    - R1.3. (Meta)data meet domain-relevant community standards

What resources are available to help make data FAIR?

This list of resources can provide best practices and guidance to support providers aiming to make data FAIR:

- GO FAIR: FAIR Principles
- F1000 Getting Started Guide – Simple steps and best practices to follow to make data FAIR and Open when publishing a research article.
- How to Make Your Research Data FAIR – Explanation of FAIR principles and translated into practical information for researchers.
- Output Management Plan Template – Guidelines on FAIR Data Management and OMP template example
- Metadata Standards Directory – Online catalog that can be searched for discipline-specific standards and associated tools.
- FAIRSharing.org – A curated and searchable portal of data standards, databases, and policies across many scientific disciplines.
- Fairsfair.eu - European Open Science Cloud support on FAIR.

Why is access to underlying data important?

Providing access to all data, including underlying data is key in fulfilling OpenEPI's goal of rapid and free exchange of knowledge, research and ideas.

When we remove barriers to data sharing and reuse, innovators, researchers and decision makers can freely benefit from open data in their own data and knowledge driven processes, and the scientific community can freely build upon each other's work.

Access to all data allows for:

- Barrier-free and timely access to data
- Reassessment of current data interpretations and analysis
- Ability to verify, reproduce, and reuse data in new ways
- Data provenance and preservation
- Innovation of new data products and services, for the good of humanity and the SGD efforts.

Exclusions to the Open Policy

Work that must be openly licensed under the CC BY 4.0 license includes both new content created and existing, pre-OpenEPI content, modified to be OpenEPI compliant. Only work that is developed, in full or in part, by the entity aspiring to be OpenEPI compliant, is required to be licensed under the CC

BY 4.0 license. Pre-existing copyrighted materials licensed to, or purchased by the entity from third parties, including modifications of such materials, remain subject to the intellectual property rights the organizational entity receives under the terms of the particular license or purchase.

Generally, this policy implementation change will not apply to datasets that contain personally identifiable information (PII), datasets that contain information of relevance for national security interests, or review articles and other works of synthesis or opinion/analysis where the entity is invited to contribute on a specific topic. There are many valid reasons to restrict data access. In addition to the above mentioned reasons, access may be restricted in cases where consent has not been given for release, when the data can be defined as confidential commercial information, or in situations where there are sound public reasons (e.g. protection of endangered species, archaeological sites). The use of anonymization techniques, data sharing agreements, and safe havens where data can be accessed in controlled and secure circumstances (e.g., data trusts) are key in such cases.

That being said, the greatest benefits come when data is open and complies with both the FAIR and CARE principles (for further reading on CARE: https://www.gida-global.org/care), supporting the widest possible reuse, and reuse at scale.

When specific legal or ethical restrictions prohibit public sharing of a data set, the entity aspiring for OpenEPI compliance must indicate how others may obtain access to the data.

9.4 Cross-cutting issues and perspectives

Gender balance

Gender balance in OpenEPI is not just a matter of representation; it is a strategic imperative. It will be an important cross cutting topic in the feasibility study and the pre-project will suggest specific programs and activities focused on gender balance and empowerment of women as part of the OpenEPI ecosystem.

This focus ensures that the platform's data is comprehensive, unbiased, and relevant to all, and it supports the initiative's vision of fostering global and local innovation to address the planetary crisis of climate change, nature loss, and pollution. By embracing gender diversity in data creation and aggregation, OpenEPI can effectively fulfill its mission to provide high-quality, openly accessible data that benefits a wide range of stakeholders, from governments and investors to local communities and individuals.

Gender balance in the Open Earth Platform Initiative (OpenEPI) is crucial for several reasons, particularly when considering the data creation and aggregation processes. Ensuring female representation in these processes is essential to achieve data diversity and address the triple planetary crisis effectively.

Diverse Perspectives

Gender balance promotes the inclusion of diverse perspectives and experiences. Women may have unique insights and priorities related to climate change, nature, and pollution that can enrich the data creation process. Their involvement can help identify data gaps and capture nuances that might be missed in a less diverse team.

Data Relevance

Gender-balanced teams are more likely to consider the needs and concerns of both men and women when designing data collection strategies and analysis. This ensures that the data gathered is relevant to a wider range of stakeholders, including different genders and communities.

Mitigating Bias

Gender bias can inadvertently creep into data collection and analysis. By having a gender-balanced team, there is a higher likelihood of recognizing and addressing potential biases, resulting in more accurate and unbiased data.

Gender-Specific Impacts

Climate change and environmental issues often affect genders differently. For example, women may be disproportionately affected by climate-related disasters or have specific roles in natural resource management. A gender-balanced approach makes it more likely that these gender-specific impacts are adequately captured and addressed in the data.

Enhanced Credibility

A gender-balanced team enhances the credibility of the data and the platform itself. It signals a commitment to inclusivity, diversity, and equity, which can foster trust among data users, stakeholders, and investors.

Broader Reach

Women often play essential roles in local communities, particularly in developing regions where OpenEPI is focusing its efforts. Their inclusion in data creation and aggregation processes can help reach and engage with underserved populations more effectively.

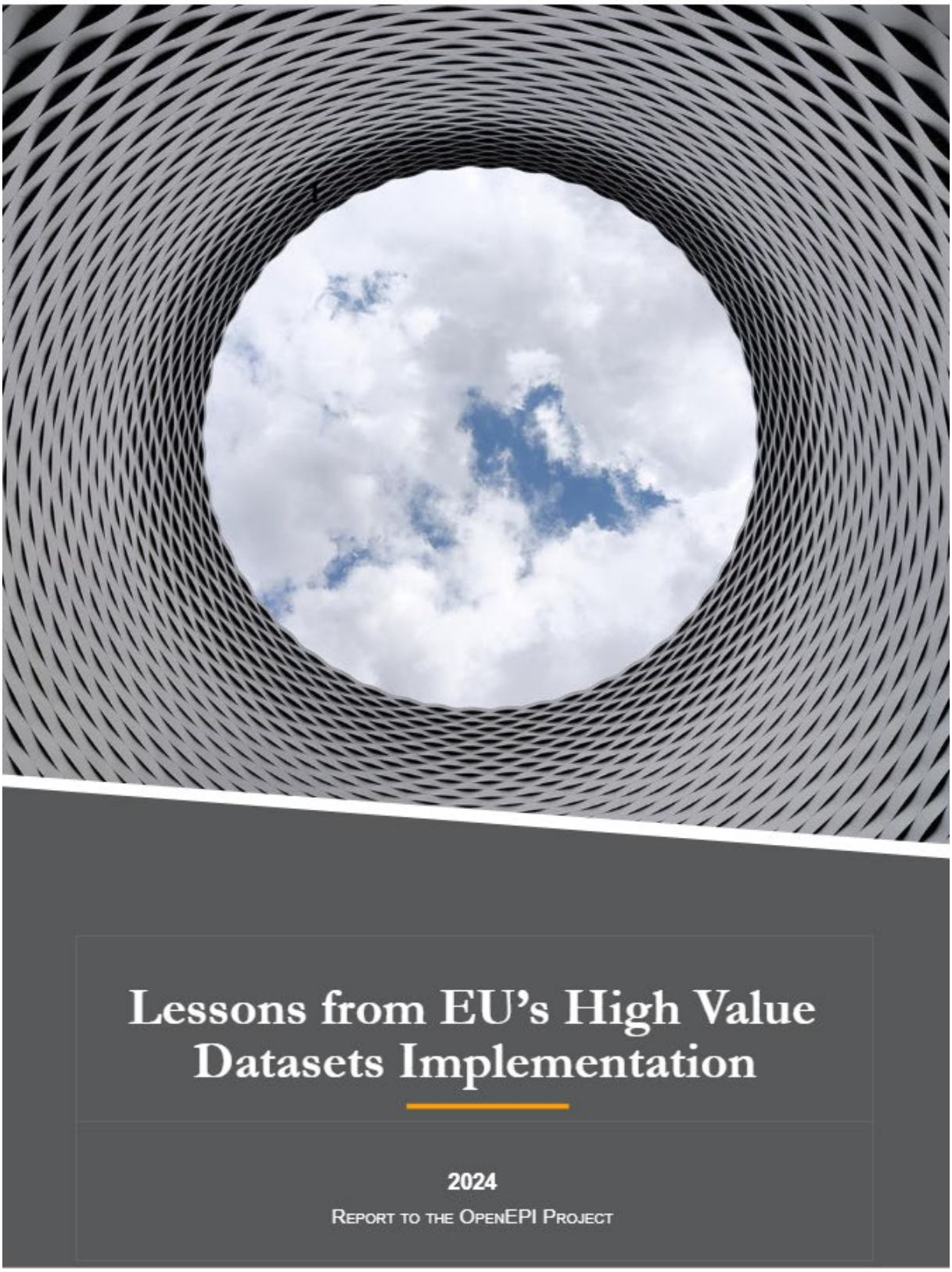
Innovation

Diverse teams are known to foster innovation and creative problem-solving. By including women in the process, OpenEPI can tap into a broader talent pool and potentially discover new approaches to data collection, storage, and accessibility.

Long-Term Sustainability

Achieving long-term sustainability for the OpenEPI platform is one of the key goals. Gender balance promotes a balanced and inclusive approach, which can lead to more sustainable solutions and long-term support from diverse stakeholders.

9.5 Report from Agenda Kaupang on EU's HVD



OPPDRAGSGIVER:	**The OpenEPI project**
RAPPORTNUMMER:	**R1022543**
RAPPORTENS TITTEL:	**Lessons from EU's High Value Datasets Implementation**
ANSVARLIG KONSULENT:	**Kjersti Nordskog, Agenda Kaupang**
FOTOGRAFI I RAPPORT:	**Pexels.com**
DATE:	**19.2.2024**

Preface

During the last decades, the European Union (EU) has developed a rights-based policy thinking on the data governance field that many consider to be the most mature and developed to date. These data management policies are created within a specific geographical, political, and cultural area, building upon several decades of previous cooperation, funding, regulations and implemented acts. The potential future OpenEPI-project will operate in a very different context, but important lessons can still be learned from the implementation of EUs data policy work.

This report investigates the EUs High Value Data (HVD) policy to identify its possibilities and limitations in a European context, and what parts of the policy could be relevant and applicable on a global scale: What elements of the HVD policy are relevant to adopt into a global scale and framework for an open data infrastructure?

This report is written by Kjersti Nordskog, Agenda Kaupang AS. Thanks to Gjermund Lanestedt at Capto AS for input and comments along the way. Also, thanks to the informants from five countries and six public bodies, for sharing their experiences with implementing HVD and giving their advice to the OpenEPI-project.

Thanks to the OpenEPI-project for this interesting task, and the opportunity to dive into the HVD-policy.

Oslo, Norway, February 2024.

Contents

1 Background

A set of data governance principles is needed for the open data, and the open infrastructure for climate mitigation data, that Norad is looking to establish with the OpenEPI project.

During the last decades, the European Union (EU) has developed a rights-based policy thinking on the data governance field that many consider to be the most mature and developed to date. However, the data management policies of the EU are created within a very geographically, politically, and culturally specific area, building upon several decades of previous cooperation, funding, regulations and implemented acts. A blueprint copy of the EU data policy to the global scale is not an option. Instead, lessons must be learned from the steps taken by the EU Commission and its member states (MS) and by the policies implemented in the different member states.

This report will investigate the EUs High Value Data (HVD) policy to identify its possibilities and limitations in a European context, and what parts of the policy could be relevant and applicable on a global scale: What elements of the HVD policy is relevant to adopt into a global scale and framework for an open data infrastructure?

1.1 A reader's guide

The second chapter describes the methods used for creating this report. Chapter number three is dedicated to HVD, and explains what it is, how the European Commission describes it through the implementing act, how it should be understood, and what data is included in the act. It further addresses the status of implementation and discusses what the legislative documents say about long-term storing of the high value datasets. Chapter four, "Lessons from the field", describes the findings from the interviews with representatives from countries implementing HVD. The chapter describes challenges and experiences with implementing HVD in the different countries so far, and the informants advise for the OpenEPI-project. In chapter five, Agenda Kaupang's recommendations to the OpenEPI-project are listed. The report finishes off with a list of documents, articles and websites that might be of relevance to the reader who wants to know more about the topics discussed in this report.

2 Methods

This report is made up of document studies and interviews.

2.1 Document studies

To describe the HVD-policy, this report relies heavily on documents published by the EU commission. These documents are all freely available online and have been identified and accessed through google searches. All documents are listed in the reference chapter at the end of the report. Additional documents, websites and online resources that are thought to be relevant are listed in the last chapter, “Additional resources identified”.

2.2 Interviews

Five interviews have been conducted. They all took place as video conference calls during October and November 2023. During some of the interviews more than one person participated, so in total 8 individuals were interviewed.

- Two Scandinavian countries, in total three institutions and six informants
- A third Scandinavian country gave a brief statement on the topic via email
- One Eastern European country, one institution, one informant
- One Southern European country, one institution, one informant

All the interviewees represented countries thought to be quite mature and relatively advanced in their data management procedures, and therefore also in their implementation of HVD.[91] This is important to keep in mind while reading the rest of this report.

The findings from the interviews are presented in the chapter 4 “Lessons from the field”. The very small number of interviewees must be kept in mind when reading and interpreting the results, as they are by no means representative for the experience held by all the member states and others, implementing HVD.

[91] This assumption is made based on the countries’ scores in the European Maturity Report survey.

3 What is HVD?

The list of High-Value Datasets is set up under the Open Data Directive (ODD) (read more about ODD in the box to the left) and consists of a list of datasets considered to have important benefits for society. The list was defined by the European Commission, and published in January 2023, after having been under development for several years. It is specifically the expected benefit for the environment and the economy that lies behind the selection of datasets, according to the European Commission.

HVD is divided into of six categories:

- Geospatial
- Earth observation and environment
- Meteorological
- Statistics
- Companies and company ownership
- Mobility

HVD includes only public sector data, data which is of particular interest for creators of value-added services and applications. The data must be made available for reuse, free of charge for the end user.

In the January 2023 press release, the Commission points out that the thematic range of HVD may be extended later, for example to reflect developments in technology and market.

The increased availability of public data that comes with HVD is thought to increase the possibilities for entrepreneurship and result in creation of new companies. The datasets are relevant for SMEs and others who want to develop new products and innovative solutions. The commission especially highlights the mobility datasets, geolocation of buildings, meteorological observation data, radar data, air quality data and

Open Data Directive

The Open Data Directive (ODD) is an amendment to the PSI-directive of 2003. The aim of ODD is to fully exploit the potential of public sector information.

"*The public sector in Member States collects, produces, reproduces, and disseminates a wide range of information in many areas of activity, such as social, political, economic, legal, geographical, environmental, meteorological, seismic, touristic, business, patent-related and educational areas. Documents produced by public sector bodies of the executive, legislature or judiciary constitute a vast, diverse and valuable pool of resources that can benefit society. Providing that information, which includes dynamic data, in a commonly used electronic format allows citizens and legal entities to find new ways to use them and create new, innovative products and services. Member States and public sector bodies may be able to benefit from and receive adequate financial support from relevant Union funds and programmes, ensuring a wide use of digital technologies or the digital transformation of public administrations and public services, in their efforts to make data easily available for re-use*»

ODD states that public data should be available free of charge. Some public bodies, and their data, are exempt from this requirement, but this does not apply to any of the datasets defined as High Value Datasets.

The directive entered into force on 20 June 2019.

soil contamination data as having a potential for innovation, research, better-informed policy making and fighting climate change.[92]

EUs creation of Common European Data spaces is expected to increase the demand after, and use of, the datasets included in HVD.[93]

3.1 The purpose of the HVD

The Implementation Regulation (EU) 2023/138 states that: «*The main objective of establishing the list of high-value datasets is to ensure that public data of highest socioeconomic potential are made available for reuse with minimal legal and technical restriction and free of charge.*»

This should be achieved by harmonising the implementation of re-use conditions in a technical specification that states that the data should be available in a machine-readable format and via application programming interfaces (APIs). These requirements are chosen because they support the FAIR-principles of findability, accessibility, interoperability and reusability, and are thought to strengthen the open data policies in the member states.

The EU commission encourages its member states to go even further than the regulation requires them to in all thematic categories, «…*especially the category 'companies and company ownership', Member States are encouraged to go beyond the minimum requirements with respect to the scope of datasets and arrangements for re-use*».

3.2 Exemptions from the HVD-principles

There are some exemptions from the requirements set in HVD. These are identical with the exemptions from the ODD. The requirement to make HVD available free of charge does not apply to:

- Libraries, including university libraries, museums, and archives.
- Member states can exempt individual public sector bodies, if they have requested so and if it is "*in line with the criteria set out in the Directive from the requirement to make high-value datasets available free of charge for a period not exceeding two years from the date of the entry into force of this Implementing Regulation*».

3.3 Data that contains personal data

Some of the datasets that are included in HVD can contain personal data. Implementing Regulation documents states clearly that when making datasets available for reuse also involves processing of personal data, this should be done in accordance with Union law.

Member states can complement data in HVD with other public sector information they might have, when it's thematically related and considered high value based on the criteria under Article 14(2) of ODD (Directive (EU) 2019/1024). When this includes information that constitutes personal data, this information must be assured to be necessary, proportionate, and must genuinely meet objectives of general interest.

3.4 Creative Commons licences are recommended

The intention of ODD is to promote the use of standard public licences that are available online. Creative Commons (CC) licences are identified as an example of such recommended licences. CC licences are developed by a non-profit organisation and have become a leading licence for public

[92] European Commission Press Release 20 January 2023: Commission defines high-value datasets to be made available for re-use

[93] European Commission: Common European Data spaces consulted February 19th, 2024

sector information, research results, educational resources, and the cultural domain across the world. For this reason, the Implementing Regulation refers to the most recent version of the CC licence suite, CC 4.0. It is possible to use a licence equivalent to CC, but it must not restrict the possibilities for re-using the data (12).

3.5 On geospatial data specifically

In Article 2(3) the Implementing Regulation states that for datasets in the geospatial, earth observation and meteorological categories Directive 2007/2/EC – The INSPIRE-directive, shall apply.

3.6 Other requirements

A few other requirements from the Implementing Regulation that might be of relevance are:

- **API, bulk download and metadata:** HVD should be made available in machine-readable formats via APIs corresponding to reasonable needs of the re-users. Some datasets (indicated) should also be made available as bulk downloads. Terms of use, quality of service criteria on performance, capacity and availability of the API should be set out and published. The terms should be available in recognised, open, human-readable and machine-readable format. A point of contact should be assigned for questions and issues related to the API. The datasets should be denoted as high-value datasets in their metadata description. (Article 3)
- **Arrangements for reuse**: exemptions to HVD requirements granted by Member States shall be published online. The obligations imposed on HVD also apply to existing machine-readable HVD created before the date of application of the HVD regulation. (Article 4)
- **Reporting**: Reporting of measures will be taken out after 2 years of implementation (Article 5).

3.7 What are the datasets in scope for HVD? (Annex 1)

In this chapter the datasets in scope, as listed in Annex 1 of the Implementing Requirement document are listed:

3.7.1 Geospatial

The datasets included in INSPIRE data themes Administrative Units, Geographical Names, Address, Buildings, Cadastral Parcels as described in Annex I, II and III of Directive 2007/2/EC (INSPIRE-directive) plus Reference parcels and Agricultural Parcels as defined in Regulation 1306/2013 and 1307/2013 are also included in HVD.

The datasets granularity, geographical coverage and the key attributes are listed in the table below. If datasets are not available at the scale indicated in the table below but are available at higher spatial resolution(s), they shall be provided at the available spatial resolution.

Table 1 Granularity and key attributes for the data included from INSPIRE

Datasets	Administrative units	Geographical names	Addresses	Buildings	Cadastral parcels	Reference parcels	Agricultural parcels
Granularity	All levels of generalisation available with a granularity up to the scale of 1:5 000. From municipalities to countries; maritime units	N/A	N/A	All levels of generalisation available with a granularity up to the scale of 1:5 000.	All levels of generalisation available with a granularity up to the scale of 1:5 000.	A level of accuracy that is at least equivalent to that of cartography at a scale of 1:10 000 and, as from 2016, at a scale of 1:5 000, as referred to in Article 70(1) of Regulation (EU) 1306/2013.	A level of accuracy that is at least equivalent to that of cartography at a scale of 1:10 000 and, as from 2016, at a scale of 1:5 000, as referred to in Article 70(1) of Regulation (EU) 1306/2013.
Geographical coverage	Single or multiple datasets that shall cover the entire Member State when combined.						
Key attributes	Unique identifier; Unit type (administrative or maritime unit); Geometry (6); Boundary status; National identification code; Identification code of the upper administrative level; Official name; Country code; Name in multiple languages (only for countries with more than one official language) including a language with Latin characters, when feasible.	Unique identifier; Geometry; Name in multiple languages (only for countries with more than one official language) including a language with Latin characters, when feasible; Type.	Unique identifier; Geometry; Address locator (e.g. house number); Thoroughfare (street); name; Administrative units (e.g. municipality, province, country; Postal descriptor (e.g. post code); Date of last update.	Unique identifier; Geometry (footprint of the building); Number of floors; Type of use.	Unique identifier; Geometry (boundary of cadastral parcels or basic property units (7)); Parcel or basic property unit code; A reference to the administrative unit of lowest administrative level to which this parcel or basic property unit belongs.	Unique identifier; Geometry (boundary and area); Land cover (8); organic (9); Stable landscape elements (10) ("EFAlayer"); areas with natural/specific constraints.	Unique identifier; Geometry (boundary and area of each agricultural parcel); Land uses (crops or crop groups); Organic; Individual landscape element; Permanent grassland.

- **Arrangements** for publication and re-use

- CC by 4.0 licence, equivalent or less restrictive open licence
- In a publicly documented, union or internationally recognized open, machine-readable format
- Available through API and bulk download
- Available in its most updated version

- **Metadata** within scope of INSPIRE data themes at a minimum set out in (EC) No 120/2008[94].
- For Reference parcels + agricultural parcels Member States shall take into consideration the ongoing implementation of Directive 2007/2/EC as well as the obligation foreseen by Article 67(3) and 67(5) of Regulation (EU) 2021/2116.

3.7.2 Earth observation and environment

The thematic category “Earth observation and environment” includes earth observation, space based or remotely sensed data, as well as ground-based or in-situ data, environmental and climate data within the scope of INSPIRE in the listing below, and defined in Annex I-III, or data produced or generated in context of the legal acts listed in Table 2 below.

The most up-to-date datasets as well as historical versions of datasets available in machine-readable format at all levels of generalisation available up to the scale of 1:5 000 covering the entire Member State when combined are included. If datasets are not available at this scale but are available at higher spatial resolution(s), they shall be provided at the available spatial resolution.

Furthermore, consistent with and without affecting the relevant access regimes as defined in Directive 2003/4/EC, the earth observation and environmental thematic category includes all “Environmental information”, as defined in Article 2 of Directive 2003/4/EC.[95]

[94] Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

[95] Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC

INSPIRE DATA THEMES (as defined in Annexes to Directive 2007/2/EC)

Hydrography (I)

Protected sites (I)

Elevation (II)

Geology (II)

Land cover (II)

Orthoimagery (II)

Area management / restriction / regulation zones & reporting units (III)

Bio-geographical regions (III)

Energy Resources (III)

Environmental monitoring Facilities (III)

Habitats and biotopes (III)

Land Use (III)

Mineral Resources (III)

Natural risk zones (III)

Table 2 Environmental data to be included in the HVD-list

ENVIRON-MENTAL DOMAIN	Legal acts laying down the key variables
Air	Articles 6-14 of Directive 2008/50/EC of the European Parliament and of the Council (15), Articles 7 of Directive 2004/107/EC of the European Parliament and of the Council (16)
Climate	Articles 18(1), 19, 26(2), 39(3) of Regulation (EU) 2018/1999 of the European Parliament and of the Council (17), Article 26 of Regulation (EC) 1005/2009 of the European Parliament and of the Council (18)
Emissions	Article 24, 32, 55, 72 of Directive 2010/75/EU, Article 21 of Directive 2012/18/EU of the European Parliament and of the Council (19), Article 10 of Council Directive 91/676/EEC (20), Article 7 of Regulation (EC) No 166/2006 of the European Parliament and Council (21), Article 18 of Regulation (EU) 2017/852 of the European Parliament and of the Council (22), Article 10 of Directive (EU) 2016/2284 of the European Parliament and of the Council (23
Nature preservation and biodiversity	Articles 4, 9, 12 of Directive 2009/147/EC of the European Parliament and of the Council (24), Articles 4, 6, 16, 17 of Council Directive 92/43/EEC (25), Article 24 of Regulation (EU) 1143/2014 of the European Parliament and of the Council (26), Data for the nationally designated protected areas inventory – (CDDA), National biogeographical regions
Noise	Articles 4, 5, 7, 10 of Directive 2002/49/EC of the European Parliament and of the Council (27)
Waste	Article 15 of Council Directive 1999/31/EC (28), Article 18 of Directive 2006/21/EC of the European Parliament and of the Council (29), Article 10 of Council Directive 86/278/EEC (30), Articles 15-17 of Council Directive 91/271/EEC (31), Article 13 of Regulation (EU) 2019/1021/EU of the European Parliament and of the Council (32), Article 15 of Commission Recommendation 2014/70/EU (33)
Water	Articles 15-17 of Directive 91/271/EEC, Article 13 of Directive 2006/7/EC of the European Parliament and of the Council (34), Articles 5, 8, 11, 13, 15 of Directive 2000/60/EC, Articles 3-6 of Directive 2006/118/EC of the European Parliament and of the Council (35), Article 5 of Directive 2008/105/EC of the European Parliament and of the Council (36), Articles 17, 18 of Directive 2020/2184/EU of the European Parliament and of the Council (37), Articles 3-8, 10 of Directive 2007/60/EC of the European Parliament and of the Council (38), Articles 6-11, 13, 14, 17-19, 26, 27 of Directive 2008/56/EC of the European Parliament and of the Council (39)
Horizontal legislation	Articles 15, 18 of Directive 2004/35/EC of the European Parliament and of the Council (40), Article 8 of Regulation (EU) 2020/852 of the European Parliament and of the Council (41)

- **Arrangements** for publication and reuse:
  - The licence should be CC by 4.0 or equivalent
  - Data shall be available in a recognized open, machine-readable format
  - Data shall be available in APIs and bulk download
  - Datasets also within the scope of INSPIRE directive shall contain minimum the metadata elements required by this directive
  - The datasets shall be described in a complete and publicly available online documentation describing at least the data structure and semantics

- The datasets shall use Union or internationally recognised and publicly documented controlled vocabularies and taxonomies, where available

3.7.3 Meteorological

The meteorological thematic category contains datasets on observations data measured by weather stations, validated observations (climate data) weather alerts, radar data and numerical weather prediction (NWP) model data with the granularity and key attributes listen in the table below:

Tabell 3 Datasets in scope for the topic meteorology

Datasets	Observations data measured by weather stations	Climate data: validated observations	Weather alerts	Radar data	NWP model data
Granu-larity	Per weather station, full temporal resolution	Per weather station, full temporal resolution	Alerts, 48 hrs or more ahead	Per radar station in the MS and national composite	Minimum 48 hrs ahead in 1hr steps, national, at 2.5km/best available grid
Key attributes	All observation variables measured	All validated measured observation variables; daily average per variable		Reflectivity, Backscatter, polarization Precipitation, wind, and echotops	Deterministic and/or ensembles if available, for meteorologically relevant parameters and levels
Format	BUFR, NetCDF, ASCII, CSV, JSON	NetCDF, JSON, CSV	XML (Cap or RSS / Atom), JSON	HDF5, BUFR	GRIB (or NetCDF)
Update frequency and timeliness	Every 5–10 minutes in real time for automated stations, hourly unvalidated for all stations, for the last 24 hrs	Daily validated hourly (and better temporal resolution) and daily average observations data; all digitised historical data	As issued or hourly	Near real time in 5 minutes intervals (or available shortest interval)	Every 6 hrs, or better temporal resolution, from the last 24 hrs.

- **Arrangements** for publication and reuse
  - The licence should be CC by 4.0 or equivalent.
  - Data shall be available in a recognized open, machine-readable format.
  - Data shall be available in APIs and bulk download.
  - Datasets also within the scope of INSPIRE directive shall contain minimum the metadata elements required by this directive.
  - Update frequency as in the table above
  - The datasets shall be described in a complete and publicly available online documentation describing at least the data structure and semantics.
  - The datasets shall use Union or internationally recognised and publicly documented controlled vocabularies and taxonomies, where available

3.7.4 Statistics

This thematic category includes statistical datasets (except microdata) related to various reporting obligations on the topics listed in the box below:

Datasets regulated by specific legal acts:

Industrial production

Industrial producer price index breakdowns by activity

Volume of sales by activity

EU International trade in goods statistics – exports and imports breakdowns simultaneously by partner, product and flow¨

Tourism flows in Europe

Harmonised Indices of consumer prices

National accounts – GDP main aggregates (see Tables 6-7 below for variables in scope)

National accounts – key indicators on corporations (see Table 8 below for variables in scope)

National accounts – key indicators on households (see Table 9 below for variables in scope

Government expenditure and revenue

Consolidated government gross debt

Environmental accounts and statistics

Population, Fertility, Mortality

Details on these topics, what datasets they specifically refer to, with the associated legal acts and detailed specifications are listed in Directive (EU) 2019/1024[96], but not included in this report, due to their level of detail and extensiveness.

- **Arrangements** for publication and reuse
  - The datasets shall be made available as frequently as the corresponding legislation requires.
  - The licence should be CC by 4.0 or equivalent.
  - The data shall be made available in CSV, XML (SDMX), JSON or in another recognized open, machine-readable format.
  - The data should be available through APIs and bulk download
- Metadata describing the data should be in a "well-developed structured file containing at least a description of the statistical data, the statistical concepts, methodologies and information on data quality"
- "The datasets shall be described in a complete and publicly available online documentation describing at least the data structure and semantics".

[96] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)

- Where available; the datasets should use “Union, or internationally recognized and documented controlled vocabularies and taxonomies”.

3.7.5 Companies and company ownership

Basic company information: key attributes

Name of the company (full version; alternative names when applicable);

Company status (such as when it is closed, struck off the register, wound up, dissolved (as well as the date of these events), economically active or inactive as defined in national law);

Registration date;

Registered office address;

Legal form;

Registration number;

Member State where the company is registered;

Activity/activities that are the object of the company, such as the NACE code.

Company documents and accounts, accounting documents, which include:

Financial statements (incl. the list of participating interests, subsidiary undertakings and associated undertakings, their registered office address and proportion of capital held), audit reports.

The datasets in scope are datasets containing basic company information and company documents and accounts at individual company level and with the key attributes:

- **Arrangements** for publication and reuse
  - The datasets should be made available for re-use:
    - Without undue delay, after latest update
    - With the licence CC BY 4.0 or equivalent or less restrictive open licence, and when relevant with the additional conditions for the re-use of personal data
    - In a Union or internationally recognised open, machine-readable format, with complete metadata,
    - Through APIs and bulk download
    - At individual company level
  - Datasets should be described in a complete and publicly available online documentation, describing at least the data structures and semantics.
  - Where available; the datasets should use “Union, or internationally recognized and publicly documented controlled vocabularies and taxonomies”.

3.7.6 Mobility

The mobility thematic category includes datasets within the scope of the INSPIRE data theme “Transport networks” as set out in Annex I to Directive 2007/2/EC, at all levels of generalisation available up to the scale of 1:5 000 covering the entire Member State when combined. If datasets are not available at the scale of 1:5 000 but are available at higher spatial resolution(s) (91), they shall be

provided at the available spatial resolution. The datasets include as key attributes national identification code, geographical position as well as links with cross-border networks, where available: INSPIRE DATA THEME – Transport Networks

For member states, to whom Directive 2005/44/EC on harmonised river information services on inland waterways applies, data on inland waterways, are also included. The details on the data concerning Directive 2005/44/EC are not included in this report.

- Arrangements for the publication and re-use of transport network-datasets
  - The datasets shall be made available as immediately after the latest update
  - The licence should be CC by 4.0 or equivalent or less restrictive open licence
  - The data shall be made available in a Union or internationally recognized open, machine-readable format.
  - The data should be available through APIs and bulk download
  - The data should be in their most up to date version
- Metadata describing the data should include minimum the metadata elements defined in Regulation (EC) No 1205/2008
- "The datasets shall be described in a complete and publicly available online documentation describing at least the data structure and semantics".
- Where available; the datasets should use "Union, or internationally recognized and documented controlled vocabularies and taxonomies".[97]

3.8 Status on implementation of HVD

There has been no reporting on the HVD implementation yet, so we do not know what progress the different countries are making in their implementation process. There are several reasons why the implementation can progress differently, and at different paces, in different member states, for example technical debt, open data and digitalization maturity, access to resources, open data mindset or complexity in the government structures.

Despite the lack of knowledge on implementation progress now, the EU publishes an annual report on open data maturity: This report tells us, through self-reporting from the European countries, how they are performing on four different indicators:

- Policy – what are the countries' open data policies and strategies.
- Impact – open data reuse and the impact that these data make.
- Portal – features with the national portal that enhance accessibility and support interaction with the open data community.
- Quality – the mechanisms that ensure the quality of the data/metadata.

The latest report, from 2023, was based on a self-assessment survey completed by 27 EU Member states, 3 EFTA countries (Norway, Switzerland, Iceland) and 5 candidate countries (Albania, Bosnia and Herzegovina, Montenegro, Serbia, Ukraine). Based on scores for each country, four groups are created, and countries are ranked as either trend-setters, fast-trackers, followers or beginners.[98]

The general result from the 2023 Report is shown in this Country Maturity Map:

[97] EU 2023/138 High value data directive legal document

[98] Data.europa.eu (2024) The 2023 Open Data Maturity report has been released Consulted February 16th2024

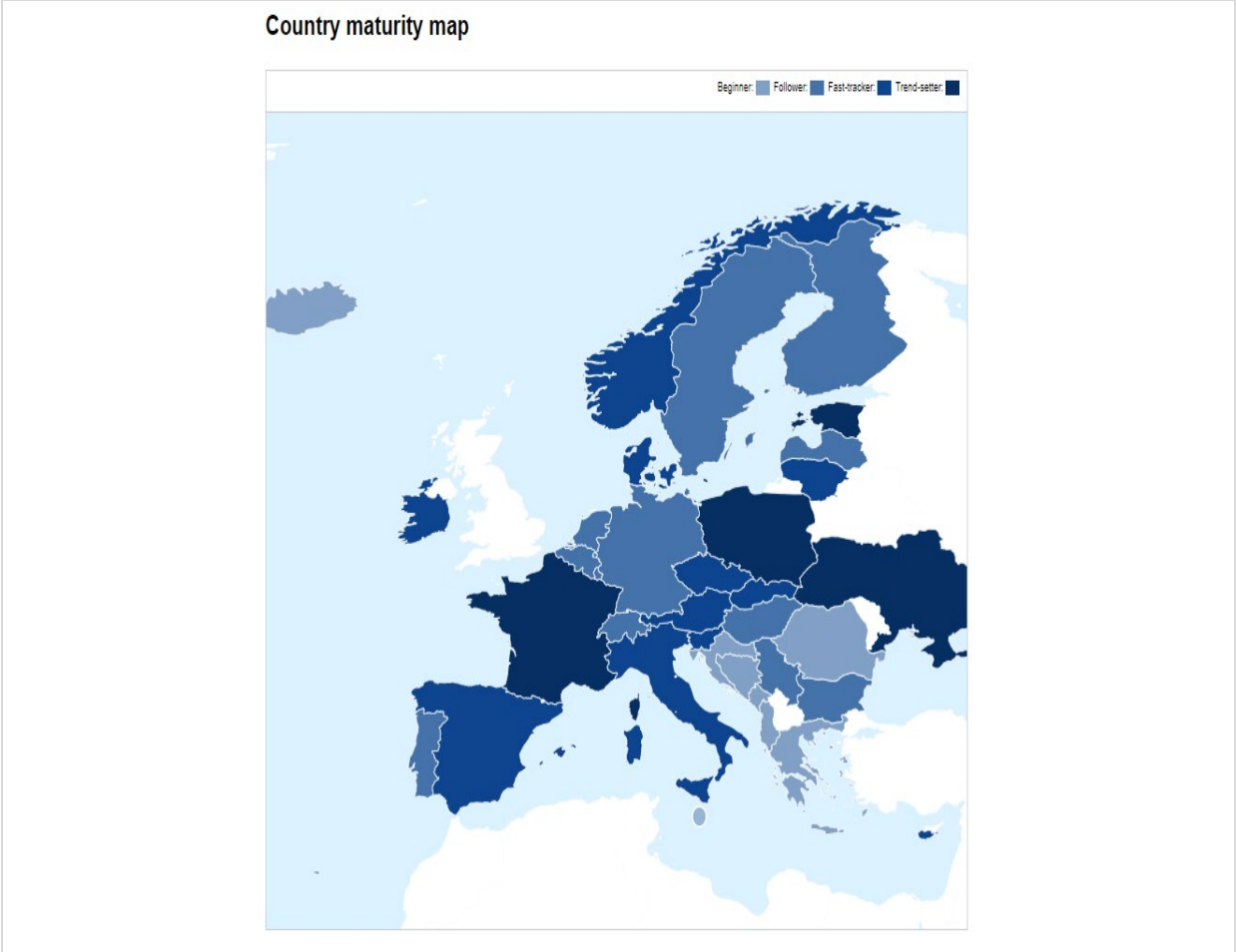


Figure 1 Country maturity map from the Open Data Maturity Report 2023

The EU's Data.europa.eu team has carried out the maturity report since 2015, though the methods used for measurements have been developed since then. In 2022, a focus was introduced to measure the countries' level of preparedness for the future implementing regulation on high-value datasets. This means that some of the questions in the self-assessment report are directly linked to the implementation of HVD.[99] According to the 2023 report, "*In 2023, 22 Member States (81percent) reported that they were making progress on ensuring the interoperability of datasets, particularly high-value ones, in their country. This reflects the preparatory work done by 17 countries (63 percent) in 2022*." The report goes on to comment on some of the measures taken in the individual countries.[100]

3.9 What does the directive say about long-term storage?

According to article 4.2 of the HVD legislative document, historical versions of the datasets should be available, as long as they are machine-readable: "*To facilitate the availability of datasets for re-use covering longer periods of time, the obligations imposed under this Regulation shall also apply to existing machine-readable high-value datasets created before the date of application of this Regulation».* From this we also understand that datasets that predate the directive, should be made available according to HVD legislation, as long as they are machine readable.

The HVD legislative document further states that historical versions available in a machine-readable format are required for the datasets belonging to the "Earth observation and environment" category (in chapter 2.1 Datasets in scope, under chapter 2 Earth Observation and Environment). This is not

[99] Data.europa.eu Measuring Open Data Maturity

[100] Data.europa.eu 2023 Open Data Maturity Report

explicitly said for the data belonging to the other categories. Also, there is no further instructions on how often historical versions should be stored, or how long historical versions should be kept.

The ODD legislative document does not place any requirements on the storage of data, neither on where the data should be stored, nor for how long non-current versions of the datasets should be stored for future reference. However, ODD states that «*Public sector bodies shall not be required to continue the production and storage of a certain type of document with a view to the re-use of such documents by a private or public sector organisation*» (Article 5 Available formats: 4). We understand this to mean that organisations who no longer need a certain type of data, are not obligated to continue to produce them – or store them – for the purpose of reuse by others.

4 Lessons from the field

This part of the report is based on findings from the interviews conducted in the autumn of 2023, as described previously.

4.1 Communication surrounding the implementation

We assume that there is an advantage for member states that have the HVD-requirements aligned with already existing national policies.

All the countries already have had several years and put significant effort into implementing the INSPIRE-directive, in addition to adding the directive to their national legislation. The countries in this brief survey confirm that having implemented the INSPIRE-directive already, had helped to implement HVD. This must be seen as a significant advantage for the implementing countries, as some of the thinking, the mindset and organisation associated with the data management in accordance with European standards.

4.2 Implementing process

The countries were asked what the most challenging part of implementing HVD was. One country highlighted meteorological data, because of the frequency by which the data were produced, and due to the specialist analysis, which was done on them prior to publishing.

The business registry also caused challenges for the eastern European country, as the sale of these data were previously the basis for financing and maintaining the dataset and covering the agency's cost.

One of the countries is not an EU-member, and thus has not started to implement the directive yet, as the ODD has not been implemented in the country's legislation. However, the country stays vigilant and observes what challenges and solutions the other member countries are experiencing, hoping to learn from them and their implementation practices. The same was also the case for the INSPIRE-implementation, where non-EU-members in practice got an extended deadline for implementing the directive, and therefore could learn from the process in the other countries first.

One country states that the biggest challenge concerning the implementation of HVD, is the fact that the responsibility for the data that HVD covers, is very distributed in the country. The responsibility lies in a very high number of municipalities, provinces and regions, and a variety of central administrations with specific responsibilities.

Representatives from another country stated that the progress was slow, mainly because they were waiting for the implementing act, and because they had posed technical questions to the commission that had not been answered. Also, lack of resources to do the actual implementation, slowed down the process for them.

There are also examples of EU policy being in direct conflict with the national policy:

One member country has not formally started the implementation at all, since their national government has not given the "go ahead" yet, and by that also not allocated any resources for the purpose of implementing. The staff at the agency of digitalization says that they do pay attention to what is being said and done about the HVD implementation process in the other member states, but that they have yet not started working on the matter themselves.

One country also explains how the implementing requirements are in direct conflict with national policies, which leads to doubts about what data to implement. Specifically, this is with regards to data about poverty, where the national parliament has defined a different metrics to measure poverty, than the Eurostat. This puts the implementing agency in a difficult situation, as they cannot go against the

decision of their own parliament and use a definition that the same politicians have explicitly rejected. Not implementing EUs definition will be a failure to implement. At the same time, having two definitions coexisting is also not an option.

Open data mindsets

There are varying degrees of open data mindset reported in the countries interviewed. Lack of open data mindset is especially an obstacle to the implementation of the legislation when the responsibility for the implementation is distributed to many different government bodies. Having a data financing model that is based on the sales of data, also is an explicit inhibitor of open data mindsets, one of the informants points out.

Support in the implementing process

On the topic of receiving support in the implementing process, especially from the commission itself, we find that the different countries have diverse impressions and experiences. The support either comes from the commission and the EU themselves, or from other member countries through more or less formal meeting points and networks of cooperation:

One informant expresses discontent with the implementing support provided by the commission, and claims the commission leaves it up to the countries themselves to do all the “heavy lifting” and does not follow up on their own end.

Another informant participated in working groups on the topic, and say that these are helpful, but that it is mostly the input from the other member states, rather than the commission itself, that proves to be helpful for them.

One country has identified several obstacles in the implementation (technical aspects, coordination, publishing methodology etc) but had, at the point of interview, not yet presented these questions to the commission, and could therefore not say how the support and follow up from the commission would be. However, help from the commission will be of importance for overcoming these obstacles.

“*Our impression is that the commission wants to help*”, says the informants from one country. However, they find the answers they get too diplomatic and careful, and the answers do not really serve any purpose for the member states that posed the question in the first place. The people that are in place to answer the questions do not have the required technical understanding or skills, and the informants’ suspects that they are afraid to say something that is incorrect. The informant’s reflection on the reason for this, is that it is the natural result of the size of the political body that the EU has become. The different member states have different levels of ambition, wishes and requests in the implementing processes, depending on how mature they are within the field. According to the informants, some states want more standardisation, and some want less. Some countries do not want to make changes that are purely for formal purposes but do not serve a pragmatic purpose. Yet others just want the efforts to implement HVD to be as small and demanding as little resources as possible. Giving clear answers and guiding under such circumstances are hard, states the informants.

Many of the countries that came late to the EU and started implementing the INSPIRE-directive with a delay, compared to the other member countries, find the process of implementing the INSPIRE-directive very hard, and this will also transfer to the HVD.

One informant states that in practice, implementing the INSPIRE-directive is impossible for these member states. THE HVD has less strict and heavy requirements than INSPIRE, but the recommendation is to not make the policy so strict and/or detailed that implementing it cannot and will not be done, in practice.

We find that there are reasons to believe that the questions related to support in implementation could have been answered differently, had the process of implementation been more advanced at the time the questions were asked.

4.3 DCAT and technology choice

The DCAT and Geo-DCAT APIs are considered a useful and successful feature for data sharing in the EU, and several of the informants' advice to use it to exchange data between data catalogues. One indeed says that OpenEPI should make use of the DCAT-APIs mandatory, along with core vocabularies (but maintain users and user needs in the process of creating them!)

Another informant points out how new technology brings on new modes of sharing data and emphasizes that the policy should not be too specific regarding what technology should be used. With the INSPIRE-directive, the legal documents were very explicit on what technology was acceptable, and that is now proving to be a problem for them, according to the informant. As the technology has developed, and more modern and attractive technologies are available, changing the legal documents allowing these technologies to be used, have proved challenging. The informant thinks HVD has learned from the INSPIRE-experience, as the wording in HVD is less specific: "*Unlike INSPIRE, the HDV Act does not mention specific technologies. I think that is wise.*"

4.4 Demand-side for data

Several of the countries that are presented in this survey express that there is a "problem" with the demand side for the data that are presented to the users. One country recommends prioritising the data that has actually been requested from the users, and that are thought to add value to the economy and society.

In one country, which is quite mature and advanced in their data sharing practices, there has been a noticeable increase in data usage since HVD made the data more available. Especially the business registry, which previously was not available free of charge, has been put to good use by the private sector.

4.5 Licences

Several of the countries previously had national licences on their data but has now changed to the CC BY 4.0 licence that is requested by HVD. All the informants agree that making an internationally known standard default is beneficial for data sharing internationally. The HVD states that CC BY 4.0 or similar, or less restrictive, can be used.

4.6 Storage of data

The countries interviewed for this report state that the HVD data is stored on premise or other places. At least one of the countries has a government cloud that can be used.

4.7 Making available historical versions of the datasets

For a variety of uses, having access to historical versions of the different datasets will offer great value. One informant says that the different national agencies have different points of view on storing old data. Some datasets are overwritten every year, while others may be available in archive. For very large datasets, like spatial data, keeping historical versions would make the amount of data too large to practically manage and store. However, one country has a strategy to store semi-old data, and an archive for some of the older data. The official recommendation is to store the data as long as practically possible.

A second country states that all data are not equal, and some of the data makes more sense to keep than others, like the population register. Further, they state that every five years, data must be handed over to the national archive. This is a requirement by law. However, the national archive has a lot of data too, so they also delete data they do not consider to be worth saving.

A third country has a requirement that certain data, especially statistical data, meteorological data, and some environmental data like air quality data should be stored, but there is no requirement for them to be published – only a recommendation.

4.8 Language

The availability of information, instructions, licences and similar documents in the language actually used in the country where the data policy is implemented is highlighted by informants from two different countries.

4.9 Data buyouts

Two countries mentioned that they had to do data buyouts because of ODD and HVD. In one country the buyout was of the business register, and the buyout is completed. In the other country, the process of buyout is underway, as the public bodies in question are looking to find new models for financing the data they produce. A third country had a large data buyout approximately 10 years ago, after open and free data became a political priority. No further buyouts were therefore necessary for them regarding the implementation of ODD and HVD.

In the case of Norway, an implementation of ODD and HVD would require a buyout of large amounts of basic geodata, and no alternative financing of these data is yet established.

4.10 Data security

The recent shift in the geopolitical climate over the last few years has made some countries, and some groups within countries, less hesitant to share data openly, says one of the informants. ODD and HVD do not accept that certain data is protected by authentication mechanisms, while previous data sharing directives, like the INSPIRE-directive, did permit this. According to one informant representing one country, a growing concern with openly publishing waterways-data has emerged. The informant refers to the destruction of gas pipelines in the Baltic Sea in September 2023, as a response to what might happen when infrastructure data becomes too openly available.

4.11 Take-aways from the EU to the global scale

The informants' advice to the OpenEPI project, when it comes to open data sharing efforts, are the following:

- Start with the most important elements, datasets or objects. Use these to prove that data sharing is useful and well worth the time and resources placed into it. Show how users can benefit and make value from the data. Once this is in place, the project should generalise and scale up.
- Define use cases. One informant states that the biggest problem with INSPIRE and HVD is the lack of use cases: Be specific on who the users are, and what the data are supposed to be used for.
- Place as few requirements as possible on the subjects sharing their data. The more difficult, the less successful the policy will be.
- Obligatory sharing of the data a project produces (according to certain standards) could be a requirement to receive funding for the project.
- The ideology in the EU is likely different from the global level. Be aware of this.
- The rationale surrounding digital public goods (DPG) is not taken for granted everywhere and by everybody and may be opposed.
- The EUs focus on HVD is the growth of the member states' economy and the businesses across the Union, plus transparency in the economy of the member states. This focus differs from supporting local innovations based on open data.

- Data security has surged during the last few years, and control with who uses data, and for what purposes, is now more on people's and state's agendas. This could slow down the eagerness to disseminate data openly.
- Be aware that the demand for open data may not be there just because the data has been made open. Parts of being successful in an endeavour like OpenEPI also requires stimulus to the demand-side.
- For users of small languages, access to implementing acts, guidance documents, specifications etc in their own language is an important element for implementing success.
- Open data mindsets might not be fully developed. For example, one informant pointed out that some organisations thought sharing data meant setting up a viewer to let people visually browse their data (for example in a web browser). It needs to be made clear what sharing data entails.
- Lack of societal trust makes sharing data difficult in many countries.
- Make the OpenEPI policy concrete: mandate the use of metadata standard DCAT-API and core vocabularies, and make sure users and user needs are involved.
- Look to the research and science domain and how they share their data. There are perhaps lessons to be learned, for example regarding storage of data resulting from finished projects or in the use of metadata standards.

5 Recommendations

5.1 The realism in EUs HVD implementations plans

It appears to be a certain degree of complexity involved in the implementation of the HVD in the European countries, even though the countries all should have a fairly good basis for starting the job, based on the previous work done by the EU on the field, and the data sharing-maturing process we must assume has taken place during the last few decades.

One of the informants pointed out that there appears to be less complexity in HVD than the previous INSPIRE-directive, and that this probably is a lesson learned from the implementation process of the INSPIRE-directive. The focus for HVD appears to be more on the data sharing and open licences, and less on technical data models. The implementing countries have just started the process, and we still do not know how successful the implementing will be, or if the goals stated by the commission are indeed reached. It is also not clear to us how the commission intends to support the countries in the implementing process.

5.2 The implementing processes

As we have seen from the interviews that form the basis for this report, implementing data policies that are supposed to fit several countries can be challenging. All the countries have different cultural, political, economic, and social backgrounds, which causes them to have a variety of challenges meeting the requirements of the European commission.

At the point of these interviews, the process of implementation had just started in the European countries, and none of the countries had finished the process. The different countries had all identified some challenges, but the help they report having received from the EU commission is little at best. This can, of course, change as the implementing process continues the years to come. We do not yet know if the HVD policy will work according to the intended goals.

At the same time, it is worth adding that some of the countries interviewed here are some of the, supposedly, most mature European countries, when it comes to data sharing. Many of them have already fully implemented the INSPIRE-directive, and therefore have an advantage when it comes to implementing data sharing policies. Also, by being EU-members and implementing earlier directive, the open data mindset and understanding of why implementing common policy is useful and necessary, is, assumably, also more advanced than in other regions of the world.

This report shows that the EU-member states struggle to implement the data policies decrees approved by Brussel. Despite, assumingly, having the skills, resources and support to do so, it does take time and consume a lot of resources within each individual country. Being a member state by choice, and over a long period of time, probably makes policies easier to implement in each country, especially since the participation in the union also represents several benefits for the member states.

5.3 Open data mindset

Having success with data sharing requires that there are sufficient levels and a mature open data mindset in the organisation that is to share its data. We know that even in countries that are relatively mature when it comes to sharing data, from the European informants that the open data mindset is varied, which inhibits the data sharing success.

Several elements can inhibit open data mindsets, for example:

- Lack of resources to prepare data for sharing.
- Lack of metadata following the data
- Not familiar with the technology for sharing data.
- Afraid that the data will be misunderstood/interpreted wrong, if released to “the crowd”, which will reflect badly on the organisation of origin.
- Data contains personal information that cannot be published.
- National law is unclear as to which data can be published and not. Unclear how data could be anonymized before publishing.
- Data sharing is not a prioritised task for the organisation's management, or the government body regulating the organisation.
- Data has flaws or is incomplete, and the organisation is hesitant to share.
- The data is already being shared, but not free of change. Changing this would require a new financial model for the organisation/data.
- The data could be of strategic importance for the organisation.
- The management does not see the value and strategic importance of the data they possess.
- Unsure of what data is of relevance and use for others.
- Exist within a culture and/or society with low trust and/or where data and information sharing is not the norm.

This list is not exhaustive but illustrates that the reasons behind a lacking open data mindset can be diverse and caused by several factors. At the same time, how to best address it to build an open data mindset could also vary, depending on what is causing it. Knowing exactly what the problem is, can also sometimes be a problem.

There might have been a shift in points of view on open data during the last years, due to a shift in geopolitical climate. The recent changes to international politics might work to reverse some of the open data thinking and -work that has been done in Europe the last decade.

5.4 Technology choice

The technology for dissemination and use of data is rapidly changing and is constantly being improved. Being very specific as to what technology should be used might be tempting, to ensure good interoperability. Web-based technology for data sharing is changing rapidly and could soon be considered obsolete or less than optimal. Still, several of the informants for this report recommend requiring the use of the Data catalogue Vocabulary (DCAT)-API, as part of the future OpenEPI data policy.

The DCAT API allows for interoperability between data catalogues published on the internet, and *“…enables a publisher to describe datasets and data services in a catalogue using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues. This can increase the discoverability of datasets and data services. It also makes it possible to have a decentralised approach to publishing data catalogues and makes federated search for datasets across multiple sites possible using the same query mechanism and structure*”.[101]

[101] W3.org (2024) Data Catalogue Vocabulary (DCAT) Version 3 consulted February 16th, 2024

5.5 Stimulation of the demand-side of data

Experience from Norway[102], along with the European countries interviewed in for this report, shows that the demand-side of public data might not be impatiently waiting to access this data, to use it in creative and innovative new solutions. In fact, the demand-side for the data that OpenEPI looks to disseminate, might also need some stimulation, for the project to reach its goal. Finding out what data is likely to provide the most use and effect for the demand-side, and prioritising making this data available first, is one way to address this problem. Other measures, such as “advertising” for the data, giving courses in how to access the data and how to do analysis based on them or could be other options. Further, giving out awards for creative use or grants to people with good ideas for the use of the data could be done, but the real effects of these suggestions are unknown.

5.6 Licences

National licences, despite having an open disposition, can be perceived as non-open because they are unknown to the user. Also, having to read through legal documents in a foreign language, is in effect an inhibitor of the FAIR-principles[103]. OpenEPI should therefore aim at requiring internationally well-known data licences, for example the Creative Commons licence.

5.7 Data storage

There are several options for the actual storing of data. Some organisations still store their data physically, on premise, in servers that they own and are responsible for operating. This method is considered somewhat old fashioned, as different types of clouds have taken over for the servers.

There are some obvious risks associated with having all your data stored on servers on premise, such as risk for virus or malware, which could compromise the data, and the servers themselves are at risk of being damaged, for example in case of fire.

Storing data in the cloud means that the actual storing facility is located somewhere else, and access to your own data is through user interfaces on the internet. This could be less costly than running one’s own servers, and the actual data security mechanisms are in the hands of the cloud provider. Upscaling and/or downscaling the amount of storage needed is also relatively easier, and less costly, than acquiring and installing new servers. The largest cloud storage and cloud computing providers today are typically US based companies such as Amazon or Microsoft, who operate under US laws, for example when it comes to personal data. Some countries are developing their own national clouds, to provide a national option, in addition to the US clouds.

Some data are considered too important, or too sensitive, to be stored in cloud storage solutions overseas. Data redundancy, having data stored in more than one place, is considered favourable, to assure that access to the data is always available.

5.8 Making available historical versions of the datasets

Making historical versions of the datasets available to users could be of critical importance for many users and represents possibilities for analysis and data usage that are requested by many different types of users. However, even in European countries, access to historical data is varied. In some cases, the data no longer exist, they have been deleted as a new version of the dataset is produced. These data must be considered lost forever. In other cases, historical versions of datasets are stored somewhere, but they are not (easily) available for users outside of the owning organisation. Making these data available for external users could in some cases be costly, as the data could be stored in

---

[102] Regjeringen.no (2022) Utredning av dataøkonomien i offentlig sektor (consulted February 16th 2024)

[103] FAIR is an acronym for Findability, Accessability, Interoperability, Reuse. Read more about the FAIR-principles at go-fair.org

formats that are no longer used, and/or where technology to read the data is no longer readily available. Also, the methods and standards used to collect the data could have changed over the years, making digitalization, publication, and comparison of the old and new versions of the same data challenging.

Today, the nature of the data seems to be of relevance, when deciding if they are published. Statistical datasets are typically produced annually, and the results from previous years are of importance to track development over time. The historical versions of these data are se datasets are usually relatively likely to be available.

Other datasets change weekly, daily, or more frequently, up to thousands of times per day. Typical examples here are certain types of geospatial data and weather data. Keeping historical versions of all changes to these datasets can certainly be a challenge and be very costly. One possible solution to this problem could be to store the dataset as it is, on a predefined time and date during the year, and let this dataset represent the full year.

Technologically, there are several options available for publishing historical versions of the same datasets and making them available for download or through web services.

The size of the dataset, the nature of the data and the available resources should be considered when deciding on the best way to store and distribute historical versions of datasets.

5.9 Language

Some of the informants in this investigation point out the importance of having critical data sharing information available to stakeholders in their native and/or preferred daily language. It is worth noticing that this is highlighted in countries that already have a relatively high level of education among its inhabitants. When it comes to countries that have many minority languages, and lower levels of education among its inhabitants, there is reason to believe that the language barriers might be even higher. In many counties and societies, we also know that illiteracy is a problem, and having the necessary information only available in written form, or not in the preferred language, could represent a showstopper for implementing the policy. For the illiterate, or barely literate, information could be made available in the form of cartoons, videos or as audio files. This requires the project to have access to translators and editors to make the information available. This represents a significant extra expense for a project like OpenEPI.

In the future, there might be reason to think that artificial intelligence could help with the translation and facilitation of information, but this is still not the case. Also, the smaller languages in the world are currently not available on the internet to the degree that is necessary for reliable language models to be developed.[104]

5.10 A shift in the international point of view on open data sharing

During the recent years, a growing awareness of the importance of open and FAIR-data has emerged in the European region, and countries such as Denmark made the political decision of offering all public data free of charge[105]. However, events taking place in 2022 and the following years appear to have shifted slightly the point of view surrounding what data should be shared and what should not. More data might be considered sensitive, and thus should be protected in one way or the other, than was the case just a few years ago. The findings presented in this report shows that this already is the case in countries implementing the HVD, and as the process of implementing carries on, this is a topic that might surge. We do not yet know how much of an issue this will be for the HVD implementation, and how the EU commission might choose to address the issue.

---

[104] Codastory.com (2023 When AI doesn't speak your language) Consulted February 19th, 2024

[105] Digitaliseringsstyrelsen.dk (2024) Betaling for data Consulted February 19th, 2024

Europe has up until recently been spared from serious armed conflicts within its geographical borders, and during the previous decades of peace, a mindset of sharing and cooperations has been developed. This might not be the case in all parts of the world, and the preexisting mindset of sharing, openness and will to cooperate across country borders might not be equally present in other geographical regions. This could present a challenge for the OpenEPI-project, aiming to implement its policy globally.

5.11 Summing up – recommendations to the OpenEPI-project

Only a handful of interviews with countries implementing HVD is the basis of this report. This will of course affect the validity of the findings reported. However, all the countries included are presumably relatively mature when it comes to data sharing and open data, which can indicate other countries have other, and bigger, challenges than described in this report.

The advice given by the informants in chapter 4.11 could be well worth considering, while going further with the OpenEPI-project:

- Start small, establish good examples and relations. Starting small, and then scaling up as the project progresses and the maturity increases among involved partners and stakeholders could be beneficial. At the same time, this allows for the project to also establish a good foundation and develop use cases and successful examples that can be important for the project going further. This will also allow for the project to establish "ambassadors" who can help with communication and promotion of the project.
- Don't make fulfilling the requirements too hard. Putting very severe requirements upon the data that is to be shared, and the people who are to do the actual sharing, can inhibit the progress and success of the project. Building open data, and open data mindsets has to be done gradually, and more profound requirements should rather be introduced later on. The project should also have capacity and ability to follow up and support stakeholders implementing the requirements.
- The open data mindsets might not be equally developed everywhere. Europe is a relatively affluent region where cooperation within the European Union has decades of history, and where there is a relatively high level of societal trust. This probably makes implementing data sharing policies easier here, than in other countries and regions of the world. The project could meet unwillingness to share data (openly).
- Communicate clearly. What is the purpose of the project, and why is sharing data important? The project should also have a good and clear response to the question "*what's in it for me*?"
- Open dissemination of data could be considered a national threat. The geopolitical situation in the world has changed significantly during the last few years. In Europe critical voices question if all types of data should be shared openly. Thus, the political situation may affect the OpenEPI-project.
- The demand-side probably needs stimulation. Experience from Europe tells us that the demand-side for openly shared data might not be there automatically. The demand-side will probably need stimulation, also in the case of the OpenEPI-project. The importance and scope of this task should not be underestimated. At the same time, it is essential for the project to be a success.
- Information should be in languages people use. All the relevant information from the project, being instructions on how to share data, or how to use data, should be available in the language that the users *actually* know and use in their daily life.
- Implement the DCAT-API. This will allow the data to flow from the OpenEPI-portal to another portal, and vice versa, allowing for the project's datasets to have a much larger reach!
- Take inspiration and learn from others. In addition to the data policies mentioned in this report, from the European Union, a lot of thinking around data management and data policies are also done in the academic research and science domain, and important input could also be had from there.

6 Additional resources identified

6.1 On HVD and its implementation in Europe

- Data.europa.eu is the official portal for European data. Here you can find European datasets and data catalogues from 36 countries, along with news and updates on the development of open data in the European union.
- Data.europa.eu also publishes the annual Open Data Maturity Report, which tracks the development achieved in the field of open data in Europe on an annual basis.
- Data.europe.eu has developed an introductory course on open data, describing what it is, and what the benefits of open data are: Introducing open data (5 lessons).
- Data.europe.EU's visual overview of High-Value datasets explaining the characteristics of the HVD.
- In this 2023 Article, Eline Lincklaen Arriëns from the consultant house Capgemini describes what high value datasets are, and gives examples of what they can be used for: The EU rules for high-value datasets have changed – how are European countries keeping up?
- Research article by Anastasija Nikiforova "Towards enrichment of the open government data: a stakeholder-centered determination of High-Value Data sets for Latvia"

6.2 On the INSPIRE-directive

- THE INSPIRE geoportal gives access to all the data provided under the INSPIRE directive. The portal allows access to data and metadata and monitors the availability of the datasets in scope.
- The INSPIRE Knowledge Base provides information, policy context, legislative information and news on the INSPIRE directive. This is also a site to access the different tools developed for the implementation and utilization.

6.3 Relevant legal documents

- The HVD implementing regulation, available in several languages: Commission implementing regulation (EU) 2023/138 of 21 December 2022 laying down a lift of specific high-value datasets and arrangement for their publications and re-use
- The INSPIRE-directive, available in several languages: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
- The Open Data-Directive, available in several languages: Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)

6.4 Technology

- More information on the DCAT-API, including vocabulary overview and specification at w3.org

6.5 From the Norwegian Context

- Consultation on the Open Data Directive, June 2022 with consultation input from several public sector bodies and other stakeholders (only available in Norwegian)

6.6 Other relevant aid projects

- DHIS2 is a global public good transforming health information management around the world. More than 80 countries worldwide use DHIS2 for collecting and analysing health data. The project is coordinated by the University of Oslo, Norway.
- DHIS2-based management information systems are also used for climate, demographic and agricultural data to improve food security in Malawi
- Malawi Digital Plant Health Service (MaDiPHS) works to establish a digital agricultural plant health service, based on internationally developed digital systems.

Open Earth Platform Initiative - Feasibility study and pre-project
September 2024