Invest in Open Infrastructure

Exploratory Research on Governance and Sustainability Practices of Cloud Computing and Open Source Computational Services

The Catalyst Project

Daechan Kim
Rochester Institute of Technology (RIT)
0009-0005-1411-7167

Tania L. Hernández Ortiz
Invest in Open Infrastructure
0000-0002-8046-6024

Shared on July 28, reviewed on August 24, last version August 28, 2023.

Introduction

In May 2023, Invest in Open Infrastructure (IOI) began research activities on governance and sustainability modeling for the Open Cloud Collaborative Project for Latin America and Africa (the "Catalyst Project"). This project is a collaborative effort among six partner organizations[1] to achieve four goals[2]: (1) Deploy and manage open cloud infrastructure for under-resourced communities in Latin America and Africa, (2) Create training and pedagogical content to assist others in using this infrastructure for cloud-based science workflows, (3) Build capacity for technical, pedagogical, and leadership skills within these communities, and (4) Identify a participatory service model to sustain, scale, and generalize impact for global communities.

This exploratory research summarizes our findings on the governance and sustainability practices of 11 cloud service providers, 21 open source computational services, and four fiscal sponsors[3]. We conducted desk research on relevant organizations and services and analyzed them using the Principles of Open Scholarly Infrastructure (POSI) and the Community Cultivation Field Guide. Our prime objective was to spot prevailing practices and instances that could be emulated and augmented by the Catalyst project. Information collected and analyzed is available in the exploratory desk research spreadsheet (for the description of elements observed in the desk research, see Appendix D).

We are sharing this exploratory research with the project manager and project leads in order to check and refine our findings and initial formation activities. Once finalized, this exploratory research and the "Governance and Sustainability Needs Assessment" will serve as the base to guide the next step in our work to facilitate the Catalyst project team and community members as they build a strong governance foundation for this community-driven service. Our efforts in the Catalyst Project align closely with IOI's mission to improve funding and resourcing for open technologies and systems supporting research and scholarship. IOI works to shed light on challenges, conducts research, and works with decision-makers to improve the funding and support of open infrastructure as well as the health of the ecosystem. Working with partner organizations on this project provides us with a valuable opportunity to learn and grow together.

[1] Partner organizations are: 2i2c, Open Life Science (OLS), MetaDocencia, The Carpentries, CSCCE, and IOI.
[2] The project manager is working with the team to draft a mission statement.
[3] See Appendix A. List of organizations and services analyzed

Background

Goals and Scope

We conducted exploratory research on services, organizations, and projects similar to the Catalyst project. We focused on governance and sustainability characteristics to understand the field and identify examples of practices that the Catalyst may replicate or adapt for the planned services.

We followed an empirical-driven approach with a purposeful selection of services and organizations based on IOI's staff's previous experiences and knowledge. This research is intended to present some of the trends and practices we have identified; it is not an exhaustive review of these types of infrastructures and their governance processes. As we conducted this research, we welcomed suggestions from partners in the project. Because of time limitations and scoping, we only incorporated a handful of the many additional services and organizations our partners suggested.

As exploratory research, this report begins to surface and outline the growing complexity of the cloud and open source computational field. Our contribution focuses on exploring existing governance and sustainability practices within the field. We follow the Principles of Open Scholarly Infrastructure (POSI) and Community Cultivation Field Guide as frameworks to help us interpret and organize our observations.

As exploratory research, generalization of findings is limited to the organizations and services studied. Theoretical and conceptual developments on the field's workings are beyond the scope of this work. Below, we open with a brief conceptual framing to help readers understand how we have sought to defragment the complexity of organizations and services.

Key Concepts with a Cooking Analogy

We started this research with a basic categorization of services and organizations. We identified three emerging themes: (i) **cloud computing services**, (ii) **open source computational services**, and (iii) **fiscal sponsors and foundations**. You can find a detailed list of the organizations and services we analyzed in Appendix A.

We understand that these categories only partially capture the complexities of services and organizational arrangements within the field. Therefore, we find it useful to provide a general characterization of cloud computing and open source computational services for

an audience that includes many players from adjacent fields. To make this area as approachable as we can, we follow brief descriptions with a cooking analogy to help explain how we see the distinctions between these services.

Cloud Computing Services

Cloud computing services are organizations or initiatives that provide on-demand, scalable computing resources, including computing power, data storage, and applications over the internet (GoogleCloud, n.d.). Within these providers, they offer three major services:

> **Infrastructure as a Service (IaaS)** is a cloud computing service model that provides users with access to virtual machines, storage, and networking resources. Users can use these resources to deploy their own applications and services. With IaaS, users do not need to purchase or maintain the underlying infrastructure. The cloud provider is responsible for all of the maintenance and updates, so users can focus on using the infrastructure. (revised definition from Google Cloud, n.d.). Examples of IaaS organizations would be Open Access Data Centres and JetStream2.
>
> **Platform as a Service (PaaS)** is a cloud computing service that provides a platform for developers to build, deploy, and manage applications. PaaS providers typically offer tools and services that make it easier for developers to create and deploy applications, including Operating systems, Middleware, Development tools, Deployment tools, and Management tools (adapted definition from Chai et al. 2022). Examples of PaaS organizations are 2i2c, Google Colab, AWS SageMaker, AzureML, and EngageLively - Galyleo.
>
> **Software as a Service (SaaS)** is a cloud computing service model in which software is hosted on remote servers and made available to users over the internet. SaaS applications are typically accessed via a web browser or a mobile app. With SaaS, users do not need to install or maintain the software on their own computers. The software vendor is responsible for all of the maintenance and updates, so users can focus on using the software (revised definition from Chai and Casey, 2022). Some SaaS organizations are JAIRO Cloud, DuraCloud, Read the Docs, Illumidesk, and GitHub Codespaces.

In reality, of course, many organizations provide more than one of these services. For instance, they may offer all three services (IaaS, PaaS, and Saas). Among them are Canada Cloud, European Open Science Cloud / EOSC, HostAfrica, OpenStack, and Saturn

Cloud. Others may offer only two; for instance, OpenAIRE, Coiled.io, Pangeo, and Open OnDemand are hybrid services providing IaaS and PaaS.

Now, the Cooking Analogy

- **IaaS** is similar to ordering ingredients and cooking supplies so you may prepare a dinner for yourself.
- **PaaS** is comparable to using a chef to prepare meals according to your tastes.
- **SaaS** is like eating at a restaurant, where you are presented with a finished dish without having to be involved in the cooking process.

Open Source Computational Services[4]

Packages are tools for analysis (e.g., data manipulation and transposing) that may have a community around them. They are usually code-based tooling for programmers, and you need to know coding to use them. Among these are Matplotlib, Numpy, SciPy, Scikit-learn, and Pandas.

Development and data analysis tooling are resources that provide a graphical user interface tooling for non-programmers. Examples of these tools are Galaxy, Fiji/Image J, and OpenRefine.

Software-oriented communities are spaces to develop, support, and disseminate software, that may review packages mimicking a journal space. Examples of such communities are rOpenSci and Bioconductor.

Infrastructure or orchestration tooling allows programmers to create large-scale, repeatable, and/or high-throughput computational analyses. Within these tools are SLURM and Kubernetes.

Turning back to our Cooking Analogy...

- **Packages** are recipe-based meal kits to prepare food; depending on what you want to cook, you choose a kit and follow the instructions to assemble a meal - but you might also change components to suit your own needs..
- **Development and data analysis tooling** are prepared meals that you just need to heat in the microwave.
- **Software-oriented communities** are your cooking club where you may create new dishes or just share recipes.

[4] Yo Yehudi assisted with these initial definitions and this cooking analogy.

- **Infrastructure or orchestration tooling** are factories, that produce multiple copies of the same meal with relatively little intervention.

Sample Characteristics

To study the governance and sustainability characteristics of the organizations and services in the ecosystem, we studied 11 cloud service providers, 21 open source computational services, and 4 fiscal sponsors and foundations supporting and hosting open source computational services.

As this was an exploratory research initiative, we selected cases based on the IOI staff's previous knowledge of some of the actors in the space. For this reason, the results presented are limited to the services and organizations studied. The study is not attempting to generalize the conditions of the ecosystem as a whole; it focuses on reporting primarily on the evidence collected.

Cloud computing services studied may provide one or three services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). We also included a range of **open source computational services,** including packages, tools, and communities.

Additionally, we studied four **fiscal sponsors and foundations** supporting and hosting open source computational services: NumFOCUS, Python Software Foundation, Linux Foundation, and Digital Research Alliance of Canada. Except for the last one, which is a nonprofit based in Canada, the first three are nonprofit organizations registered in the U.S. NumFOCUS sponsors most of the open source computational services studied in this report; among these are NumPy, Pandas, and Matplotlib.

While we understand there are major differences between cloud computing services and open source computational services in terms of their technical purposes and usage, we also found it helpful to present observations within these two major groups in order to provide an aggregated view of the workings of organizations and services.

Findings

Governance Characteristics

We used the Principles of Open Scholarly Infrastructure (POSI) (Bilder et al. 2020) as the framework for analyzing governance characteristics in both cloud computing and open source computational services. We have selected five out of the seven principles that we believe resonate most with the Catalyst project's expected services.

Those principles include:

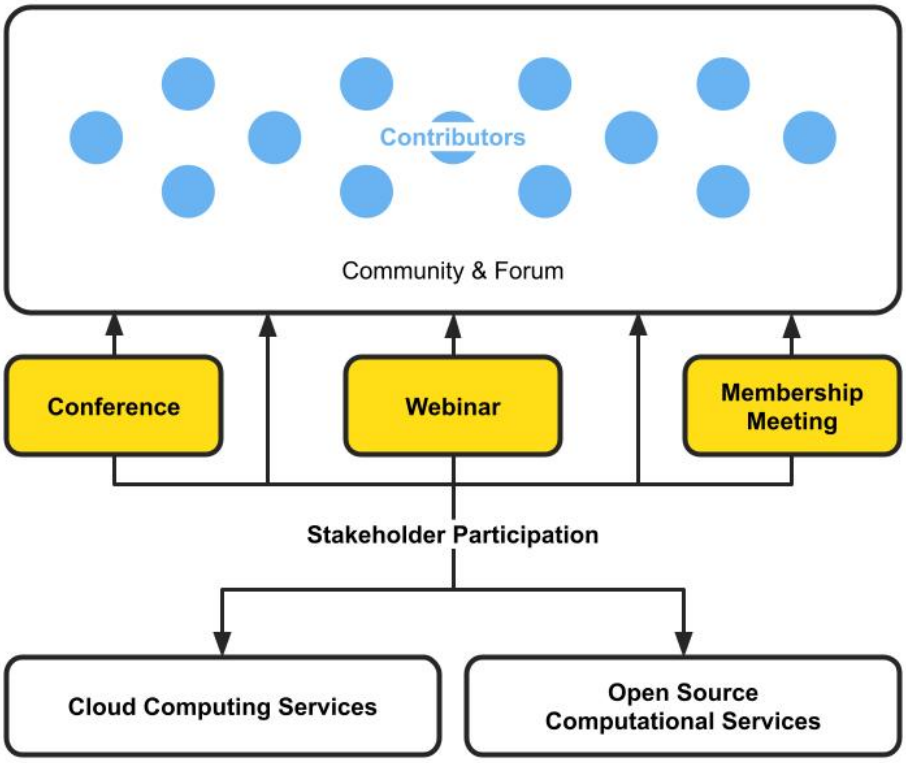
- **Stakeholder governed:** When a board-governed organization includes members from the stakeholder community, it helps increase trust that decisions will reflect the community's consensus and the consideration of diverse interests. This principle is crucial because governance bodies are best situated to serve the communities they represent if community members are involved in the governance creation process.
- **Non-discriminatory membership.** The most suitable approach would be an "opt-in" method that promotes non-discrimination and welcomes any group of stakeholders who express an interest. It is crucial that the governance process is inclusive and reflects the demographics of the membership in day-to-day operations. This principle holds significance as we firmly believe that including diverse voices in organizations is crucial for their success.
- **Transparent operations.** To build trust in the selection of representatives for governance groups, transparent processes and operations should be implemented while keeping in mind the privacy laws. We stand by this principle as we believe that transparency is a stepping stone for accountability practices.
- **Living will.** An effective way to build trust is by providing expectations and plans for winding down an organization, including the conditions under which it would occur, the process to do so, and the preservation of assets and honored principles passed to a future organization. This principle is important to establish and document clear objectives for organizations, the way to achieve them, and the conditions for concluding operations, which includes the achievement of the mission.
- **Formal incentives to fulfill mission and wind down.** Infrastructure serves a specific purpose that can become unnecessary due to technological or social changes. In those cases, the recommendation is for organizations to identify direct incentives to fulfill their mission and close down if feasible. This principle highlights the importance of being alert to changes in the sector, adapting accordingly, and knowing when it is appropriate to discontinue operations.

Stakeholder participation

Within the **cloud computing services**, we identified three profiles of organizations regarding the participation of stakeholders. First, we found organizations that allow stakeholders to participate in their decision-making processes and contribute to the projects by providing online and offline opportunities such as OpenInfra Summit (OpenStack), webinars (OpenAIRE), and regular membership meetings (European Open Science Cloud). Second, we also found cloud providers that invite stakeholders to contribute to their products and services through communities, forums, and projects' GitHub repositories but that present scant evidence of allowing them to participate in their governance. As a third group, we found organizations that provide scant evidence of stakeholder engagement, both in terms of governance and contributions to services. These are often commercial cloud services (e.g.,Open Access Data Centres and HostAfrica) in which we found little evidence of communication channels.

Similar to cloud computing services, **open source computational services** often provide opportunities for stakeholders to contribute to their products and participate in governance. Nonprofit initiatives and organizations like Jupyter Notebook, ROpenSci, and Galaxy have community calls (ROpenSci) and history of meeting minutes (Scikit-learn), and conferences (JupyterCon, Galaxy Community Conference) that can help community members to engage more actively with their projects. Even among for-profit organizations that present scant evidence of stakeholder participation, they still maintain channels for public contributions. For instance, Cocalc and QuantStack stimulate communication between stakeholders in their GitHub repositories (e.g., Cocalc GitHub and QuantStack Github). We also found other for-profit organizations that present scant evidence of stakeholder involvement in decision-making processes and contributions to their products. For instance, Anaconda and QuanSight provide limited evidence of stakeholder engagement.

Image 1. Examples of stakeholder participation



Non-discriminatory participation[5]

Regarding this principle, we focused on finding examples of **cloud computing services** that follow non-discriminatory practices by actively and openly inviting their stakeholders to contribute to the development of the service. We found instances of community interaction that may lead or not to the development of services and stakeholder participation. Webpages like "Become a member" (DuraCloud as part of Lyrasis) and "Individual member" (OpenStack) provide guidance on joining these organizations. Some others provide resources such as the “Developer’s Guide” (OpenStack) and "Developer Document" (Read the Docs) that offer guidance for stakeholders interested in contributing to products as developers. Furthermore, cloud computing organizations host

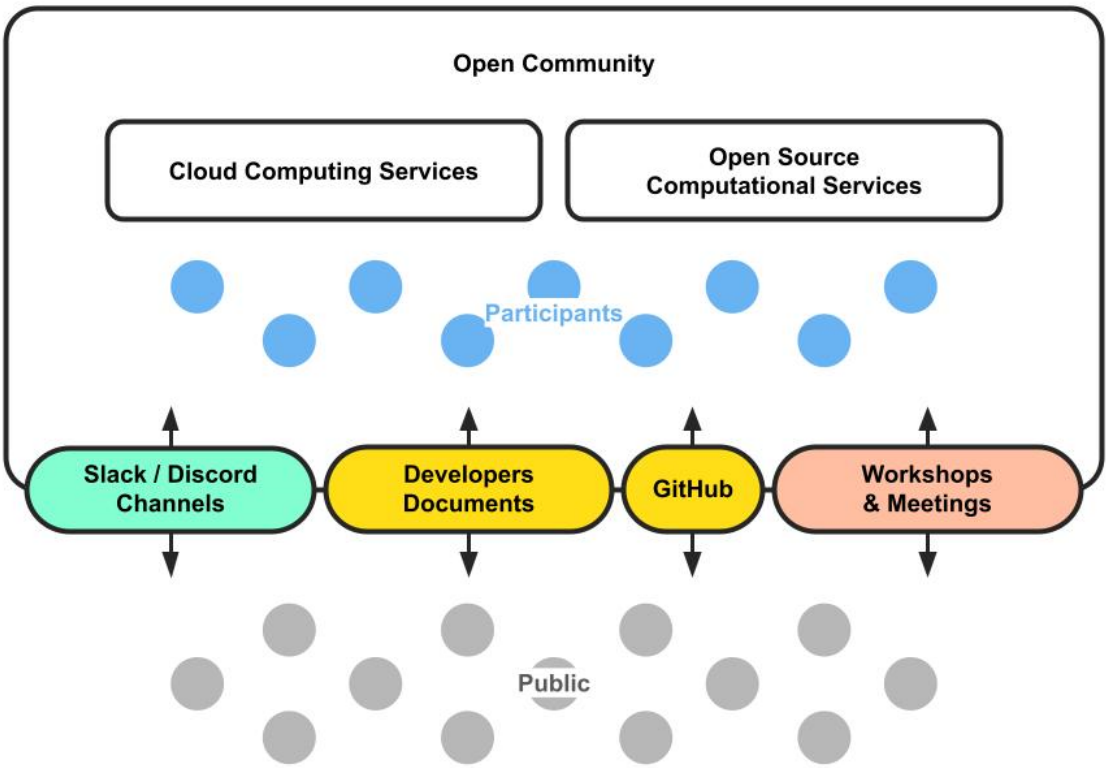
[5] We found it difficult to assess the level of involvement of individuals and communities following POSI, which encompasses non-discriminatory membership. To address this, we analyzed *participation* traces across various channels of involvement rather than focusing solely on *membership*.

events like the "Annual Workshop" (MOC Alliance), webinars (OpenAIRE), and meetup events (OpenStack) to engage with a broader audience of users who have an interest in the project or organization.

We found instances of anti-discriminatory[6] practices in **open source computational services** such as active invitations for stakeholders to join their communities, forums, and GitHub repositories. We found resources like “Contributing” (Jupyter), “Contributing guide” (OpenRefine) or “Contributing guidelines” (Rstudio) pages. Any stakeholders or community members can engage directly with the organizations through Slack channels (SciPy, NumPy), webinars (Galaxy), and mailing lists (Scikit-learn), or find events from public event calendars (ROpenSci). For other services, such as Fiji/Image J, we found limited opportunities for participation; for instance, we only found its publicly available GitHub repository.

Image 2 presents examples of how cloud computing and open source computational services encourage participation.

Image 2. Examples of elements to encourage participation



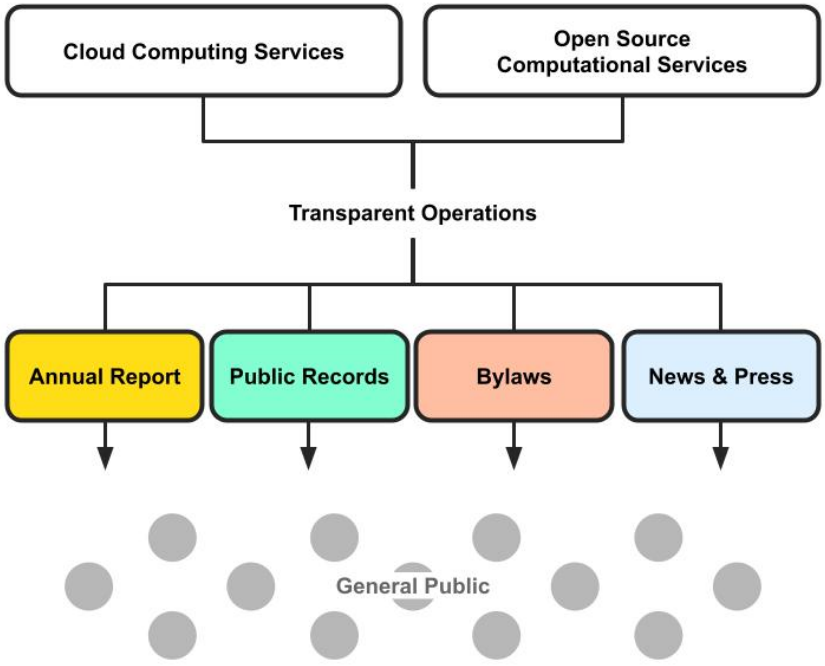
[6] We looked for traces of open and non-discriminatory practices with active strategies for including diverse voices and underserved groups. While many of these services provide a space for participation, we also recognize that they are narrow ecosystems with barriers to access.

Transparent operations

We found differences in the degree **cloud computing services** share details of their operations. Nonprofit and governmental initiatives often provide information about their internal operations. For instance, 2i2c, OpenStack, European Open Science Cloud (EOSC), OpenAIRE, and DuraCloud share documents about the selection of representatives to governance groups (e.g., EOSC Articles of Association and OpenStack bylaws). In contrast, commercial computing services present scant information about mechanisms regarding the selection of governance groups. This is the case of Open Access Data Centre and HostAfrica which share scant information on their internal operations and governance activities.

Among **open source computational services**, we found multiple community-driven projects, including Matplotlib, Scikit-learn, NumPy, and SciPy, that prioritize transparent operations by sharing their governance systems. These projects provide public documents, such as "Bylaws" (for NumPy) or "Governance" (for Scikit-learn, SciPy, and Matplotlib), which outline clear regulations regarding the selection of representatives for their governance groups. Commercial services, such as Anaconda and QuanSight, often provide scant information regarding the selection process for their governance groups and operations in general. Image 3 presents some examples used by organizations and services to provide information on their operations.

Image 3. Examples of transparency in operations in selected cloud computing services and open source computational services



Living will

When analyzing cloud providers, we focused on finding plans and insights addressing the conditions under which services and organizations would be wound down or “sunset” and what processes they would follow when those conditions arise. Overall, we found scant evidence of such provisions across **cloud computing services**. We did find that they recognize the importance of providing details on their goals and fostering trust among stakeholders by publicly sharing documents that outline their organizations' visions, future plans, and resources for other organizations that share common goals. These commitments can be found on their websites, exemplified by resources like "Mission and Strategic Plan" (DuraCloud as part of Lyrasis), "Vision, Service Provision, and Role in the EOSC Governance" (European Open Science Cloud), or the "Developer's Guide" (OpenStack). Additionally, GitHub repositories often serve as valuable archives of the organization's assets. However, commercial services, such as Open Access Data Centre and Host Africa, generally do not provide specific plans encompassing the organizations' past, current, and future endeavors.

Unlike the cloud computing services, **open source computational services** offer shared assets that can result in the development of sunsetting plans or living wills. They share their assets in various channels, including source codes in GitHub repositories, community platforms, and blogs to archive their progress and provide visibility into their ongoing developments. Additionally, they share resources such as event calendar (Jupyter), contributing guide (StackSpin, Matplotlib), or developers’ resources (Bioconductor) to facilitate communication with external stakeholders. Commercial services, like Anaconda and QuanSight, present limited open information that restricts the sharing of assets with outside contributors.

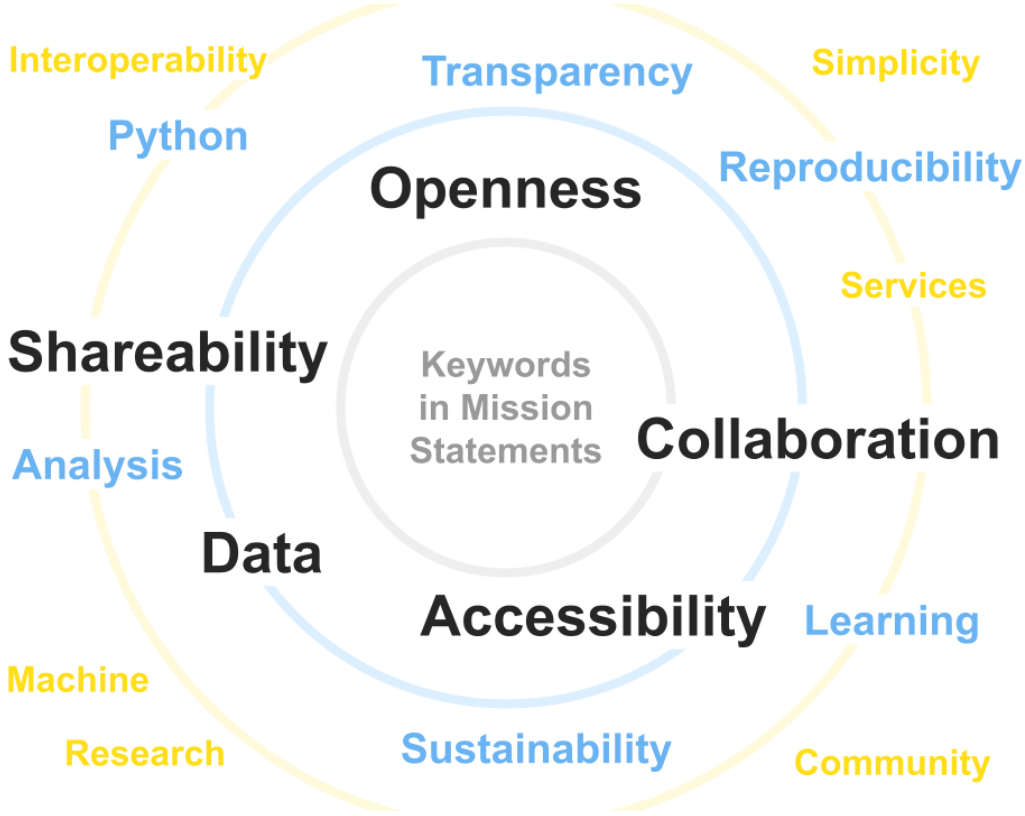
Formal incentives to fulfill mission & wind-down

We explored the aims and incentives of organizations and service providers as well as their expectations around winding down. Given the design of our desk research, we found it challenging to observe incentives and plans to wind down by only looking at websites. An in-depth analysis of this principle would require primary data collection, such as interviews and surveys with representatives and participants of the studied services. Further qualitative research is out of the scope of this report at this time.

By looking at mission statements on websites, we found that most organizations and services report descriptions of their purpose and overall objectives. **Cloud computing services** emphasize elements such as openness, accessibility, simplicity, interoperability, and research-focused services (e.g., OpenAIRE, OpenStack, and MOC Alliance). **Open source computational services** focus on elements such as openness, collaboration, accessibility, shareability, reproducibility, transparency, and sustainability (e.g., CoCalc

and Galaxy). In Image 4, we present some keywords organizations and services use in their mission statements.

Image 4. Keywords in mission statements of selected cloud computing and open source computational services



Sustainability Characteristics

We used the Community Cultivation Field Guide (Skinner, 2018) as the framework for analyzing sustainability characteristics in both cloud computing and open source computational services. We focus our attention on those elements and conditions that most resonate with the Catalyst project's ecosystem and aspirations.

These include:

- **Vision:** Communities should share information about their identity, purpose, and methods to achieve their goals. This is important to establish clarity and direction for communities' activities.
- **Infrastructure.** Communities should provide details on the methods of communication they employ, the tools they use, and how they monitor their progress. Developing procedures is important for communities to formalize and enhance their organizational structure.
- **Finances.** Communities should create proper structures for financial management and fiscal planning. Proper financial planning is important for communities to ensure their financial sustainability.
- **Engagement.** Communities should identify the best way to recruit members, create committees, and maintain consistent community involvement. Maintaining the engagement of stakeholders is important for communities to establish legitimacy in their activities.
- **Governance.** Communities should develop governance structures by creating documents such as bylaws and plans to train leaders. The governance structure will allow communities to have direction, be stable, and enable growth.

Vision

We studied how organizations and services articulated their core problems and mechanisms to assess progress, such as strategic planning. We concretely studied vision statements and the availability of documentation regarding strategic planning. While mission statements are common for organizations and services, that was not the case for vision statements. Among the handful of **cloud computing services** that present vision statements, they have a range of aspirations going from providing the structure and resources that will enable collaboration (e.g., MOC alliance) to establishing and operating world-class, client-centric data center facilities (e.g., Open Access Data Centers). Regarding strategic planning, we found cloud computing services outlining multi-annual roadmaps and implementation plans (e.g., European Open Science Cloud: implementation and multi-annual roadmap).

For **open source computational services**, we also found a handful of services providing vision statements that range from advancing science through collaboration (e.g., CoCalc) to general values in terms of data analytics and manipulation of software such as accessibility, flexibility and easiness for usage (e.g., Pandas). Regarding strategic planning, services provide scant information on long-term goals; we only found roadmaps (e.g., Pandas and Galaxy).

The two images below aim to present some examples of the aspirations and plans of studied organizations and services. Image 5 presents common words used in the handful of mission statements that we analyzed. Image 6 outlines examples of ways in which organizations and services plan and keep track of activities.

Image 5. Keywords in vision statements of selected cloud services and open source computational services

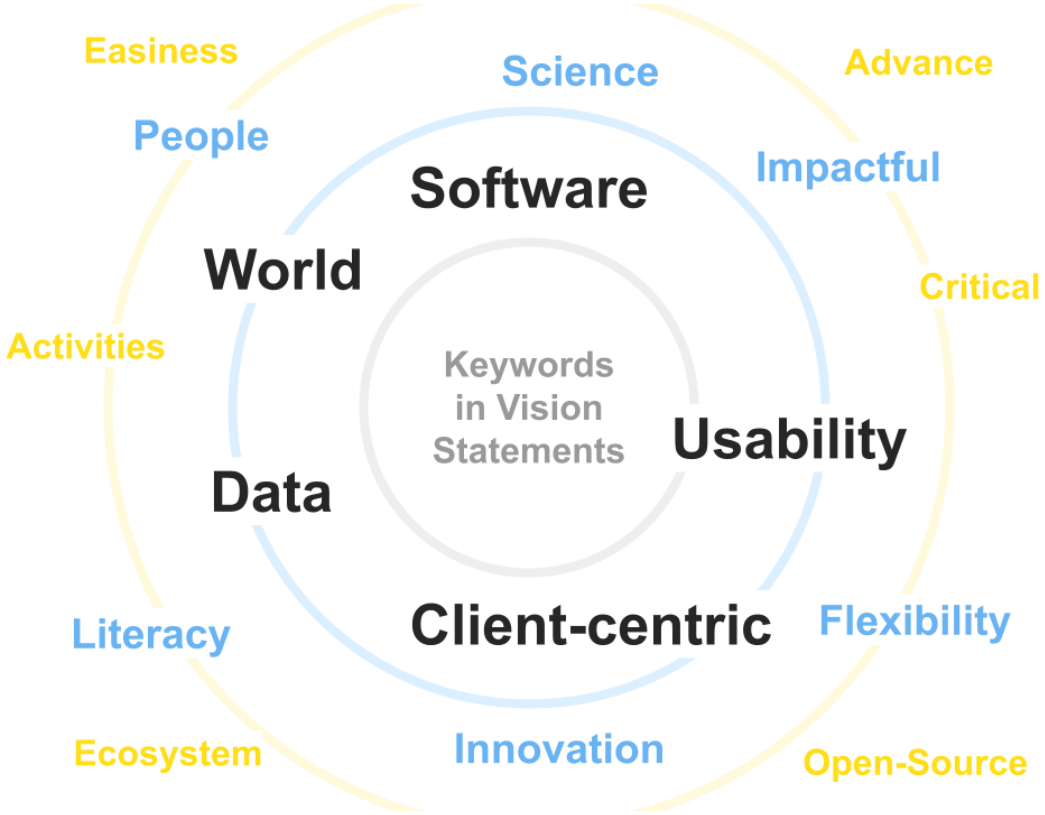
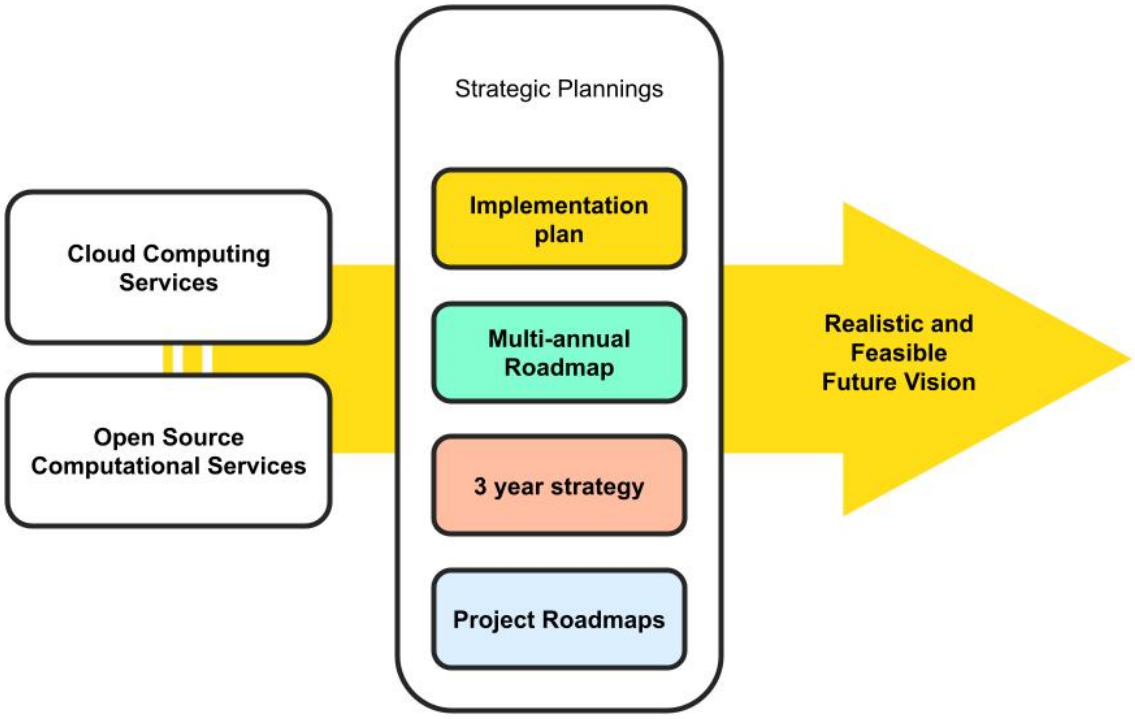


Image 6. Examples of strategic planning in selected organizations and services



Infrastructure

For infrastructure, we studied established administrative and communication structures made public by the organizations and services analyzed. To understand their administrative structures, we reviewed organizational charts and documents that provide insight into their internal decision-making processes. For communication structures, we focused on their regular communication channels. Other than bylaws and articles of incorporation (see Transparent operations subsection), overall, **cloud computing services** provide scant information on their organizational charts. We found a handful of services reporting details on their organizational structure (e.g., organizational chart). For instance, OpenAIRE provides information on its governance structure and organizational chart. Other services only list names and roles of key personnel with limited information on decision-making processes and structures.

Among the **open source computational services**, we found better examples of public documentation outlining organizational structures. For instance, Jupyter, Galaxy, Scikit-learn, Matplotlib, and SciPy, provide public information regarding roles and responsibilities (e.g., Galaxy governance and planning structure and Scikit-learn governance and decision-making).

Finances

To study this aspect, we look for public financial information such as annual reports, financial statements, and grant information. We found disparities in financial information due to the diverse maturity of organizations and services studied. For instance, some **cloud computing services,** such as DuraCloud (Lyrasis), make their financial information available through public annual reports and financial statements. Others, such as 2i2c, share their financial statements on their Google Drive platform, although access is limited to authorized individuals. Other providers present information on grants received or the funding they provide. This is the case of the European Open Science Cloud, which provides public information about its calls for proposals (e.g., EOSC funding).

Open source computational services often provide scant public information about their finances in annual reports. Exceptions to this are QuanSight which provides some details on financials in its annual report (2022). It seems that visibility on grants received is gaining popularity among open source computational services. For instance, Jupyter and OpenRefine publicly share information about the grants they have received through blog posts (e.g., grants of Jupyter and grants of Open Refine).

Engagement

When analyzing organizations and services, we look for ways of engagement, such as community participation, open meetings and events, and ways to solicit and provide feedback. We found that many **cloud computing services** actively promote community participation by sharing their event schedules on public event calendars. This is the case with services such as OpenStack (with OpenInfra Summit), EOSC (Upcoming events), OpenAIRE (Event Calendar), MOC Alliance (Workshop schedule), and Open Access Data Centres (Events). Some providers take additional steps by organizing events for their users and communities like the Infra Summit (of OpenStack) and Contribution Calls (DuraCloud Meetings).

For **open source computational services**, hosted conferences play a vital role in fostering communication with users and communities. Jupyter, Anaconda, Posit, Galaxy, NumPy, SciPy, Bioconductor, and Kubernetes either hold their own annual conferences or actively participate in public conferences (e.g., JupyterCon and Posit Conference). Some organizations that do not engage in conferences frequently, including ROpenSci, Open

Refine, and Scikit-learn, share their event schedules publicly to encourage contributions and participation from the community (e.g., ROpenSci Events).

Governance

In terms of governance elements for the sustainability of services, we look for both administrative structures, and regulations and policies that rule organizations and services.

Among **cloud computing services**, we found little evidence of usage of bylaws or similar documentation with provisions on elections of representatives, terms, and turnover. Examples of organizations providing details on their governance structure are 2i2c (e.g., Steering Council Description), OpenStack (e.g., Bylaws of OpenStack Foundation), and the European Open Science Cloud (e.g., Articles of Association). Scant evidence of governance documentation (such as bylaws) may be indicative of weak administrative structures and a major limitation for organizations studied. Another layer to such complexity is the diversity of organizational arrangements. For instance, services or sponsored projects may rely on the governance structures of larger organizations or sponsors.

In terms of policies, given the sensitive nature of handling client data, most cloud computing services publicly state privacy policies on their websites (e.g., Privacy Policy of Open Access Data Centers). Privacy policies often include provisions around data collection, use of information, information not collected, and sharing information (e.g., Privacy Policy of Read the Docs).

We also found scant evidence of bylaws or governance documentation among **open source computational services.** Moreover, organizational arrangements add to this complexity. For instance, sponsorship of projects may require them to have a governance structure and assorted documentation. For example, most NumFOCUS sponsored projects have governance guidelines (e.g., Matplotlib's Main governance document and Pandas' Project Governance).

We also looked for regulations and policies used in open source computational services. We found that services usually adapt or refer to the policies of their fiscal sponsors (e.g., services part of NumFOCUS). Privacy policies are also common practice among open source computational services. Such policies include provisions around the use of information, data retention, and third-party sharing (e.g., Privacy Policy of matplotlib). Some specialized software services for interactive computing, such as Apptainer and

SLURM, also provide security policies with provisions around problem identification and triage, embargo period, and advisory public release (e.g., Security Policy of SLURM).

Formation Activities for the Catalyst Project

We have used POSI, Community Cultivation Field Guide, and desk research on governance and sustainability models and practices to suggest formation activities for the Catalyst project. If you want more information about initial governance and sustainability guidelines, please check out the Next Steps section in the "Governance and Sustainability Needs Assessment" report. There, you will find detailed recommendations and a plan for addressing these areas through a series of working sessions to create Governance and Sustainability Structure and Norms.

Table 1. Governance Formation Activities

Areas	Governance Formation Activities
Stakeholder participation	• Inform and let onboarded communities be part of the decision-making processes in order for the project to serve community needs. As soon as Latin American and African communities are onboarded, we propose to engage them with project activities and host regular meetings for them to provide input in the design and implementation of the services intended in this project.
Non-discriminatory membership	• To establish clear mechanisms for stakeholders to express their interest in participating in the Catalyst project and let them take part in all facets of its work and governance. • To ensure representation of participating communities in all decision-making processes.
Transparent operations	• Establish transparent processes and operations in general. • Enhance external accountability to incentivize partners to deliver sound outputs and services.
Living will	• Define a plan addressing the conditions under which the project would be wound down and defining processes for doing so. For the Catalyst, this would mean a plan for local tech infrastructure, social infrastructure, and funding sustainability to ensure the smooth continuation of the work.
Formal incentives to fulfill mission & wind-down	• Once the mission of the project is agreed upon, the Catalyst project should work to align partner incentives to fulfill the goals of the project.

Table 2. Sustainability Formation Activities

Areas	Sustainability Formation Activities
Vision	• Identify and document the core problem the Catalyst project is hoping to address. • Establish how to assess progress regularly and how to recalibrate as necessary. • Establish and document project charters for all affiliated work.
Infrastructure	• Document current dependencies (e.g., activities around the provision of services). • Document risk mitigation and exit strategies for those dependencies.
Finances	• Establish and clearly document administrative costs (overhead for running the services). • Establish initial pilot pricing for services/products in order to assess the minimal financial resources needed for the continuity of services.
Engagement	• Foster relationships within Latin American, African and other global minority communities that are starting the services. • Establish and facilitate subgroups and regular meeting schedules.
Governance	• Document governance procedures in order to understand the decision-making process, blockers, and areas to recalibrate. • Establish and grow community leadership by encouraging stakeholder groups to be part of decision-making processes. • Establish policies (e.g., privacy policy and conflict of interest policy).

Acknowledgments and Contributors

We thank the Chan Zuckerberg Initiative (CZI) for funding the “A Collaborative Interactive Computing Service Model for Global Communities” project, now renamed by project members as the Catalyst project.

IOI’s team members who participated in reviewing the report include Dr. Katherine Skinner (Research Lead) and Kaitlin Thaney (Executive Director).

The report was edited by Yo Yehudi (Executive Director at OLS) and James Munroe (Product and Community Lead at 2i2c).

Daechan Kim, a student at the Rochester Institute of Technology (RIT), conducted desk research, authored initial drafts of findings sections, and created images for this report. Aboli Deepak Shete (also a student at RIT) gathered definitions and provided descriptions of cloud computing services, drafted Appendix D, and conducted desk research on additional organizations and services suggested by partner organizations. We would like to express our gratitude to Michael Nolan, Assistant Director of Open@RIT, for his assistance in overseeing the students’ work and providing valuable input during our research.

We welcome corrections, revisions, and suggestions regarding any and all aspects of this report.

Appendices

Appendix A. List of organizations and services studied

Cloud computing services

- 2i2c
- Canada Cloud
- DuraCloud
- European Open Science Cloud / EOSC
- HostAfrica
- JAIRO Cloud
- MOC alliance
- Open Access Data Centres
- OpenAIRE
- OpenStack
- Read the Docs

Fiscal sponsors and foundations

- Digital Research Alliance of Canada
- Linux Foundation
- NumFOCUS
- Python Software Foundation

Open source computational services

- Anaconda
- Apptainer
- Bioconductor
- Cloud Native Computing Foundation
- CoCalc
- Fiji[7]/Image J[8]
- Galaxy
- Jupyter project
- Kubernetes
- Matplotlib
- Numpy
- OpenRefine
- Pandas
- QuanSight
- QuantStack
- rOpenSci
- RStudio/Posit
- Scikit-learn
- SciPy
- SLURM
- StackSpin

[7] Fiji is an image processing package distribution of ImageJ.
[8] Currently in its second version, Image J2 is a software for processing and analyzing scientific images.

Appendix B. Analytical Approach

We first define key concepts for understanding cloud computing and open source computational services. We tried to look for widely used definitions in the field that help to guide common understanding in the project.

- **Cloud infrastructure**: a system that includes both hardware and software components to enable cloud computing (wmware, n.d.).
- **Cloud service provider** (CSP): an organization or initiative that provides on-demand, scalable computing resources, including computing power, data storage, and applications over the internet (GoogleCloud, n.d.).
- **Open cloud service**: an online service delivered on-demand over the internet without the user needing to use their own hardware (OneCommons, 2022).
- **Open source computational service**: a program or service in which source code is made available for use or modification for users and developers (Terrell, n.d.)
- **Interactive computing**: software that accepts input from users as it runs (Wikipedia, 2023).

Once the list of key concepts was identified, we conducted desk research on organizations and services providing or supporting cloud computing and open source computational services. Organizations were identified by IOI's team based on previous knowledge of the ecosystem (See the complete list in Appendix A). Websites were our main source of information for this research. We started with the *about us* information and then looked for specific information on governance (such as decision-making processes and practices) and sustainability (such as vision and service infrastructure).

We follow an interactive research process with various iterations of refining key concepts, reviewing the list of organizations and services under study, and analyzing collected information. We also plan to refine our observations based on the feedback provided by project leads and the program manager.

The main limitation of this summary is that we found scant information on services developed on, targeted, or tailored for Latin American and African communities. We hope to include services in such contexts based on the recommendations from partner organizations and future working sessions.

Appendix C. Bibliography

Bilder G., Lin J., & Neylon C. (2020), The Principles of Open Scholarly Infrastructure, retrieved [08/22/2023], https://doi.org/10.24343/C34W2H

Chai, W. & Casey, K. (2022). Software as a Service (SaaS). TechTarget.

Chai, W., Brush, K. & Bigelow, S. (2022). What is PaaS? Platform as a service definition and guide. TechTarget.

Eghbal, N. (2020). Working in public: the making and maintenance of open source software. Stripe Press.

GoogleCloud (n.d.). What is a cloud service provider?

Google Cloud (n.d.). What is IaaS?

OneCommons (2022). Open Cloud Services.

Raymond, E. (1999). The cathedral and the bazaar. Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media

Skinner, K. (2018). Community Cultivation, A Field Guide. Educopia Institute.

Terrell Hanna (n.d.). Definition. Open source. TechTarget.

Wikipedia (2023). Interactive computing.

WMware (n.d.). What is Cloud Infrastructure?

Appendix D. Description of elements observed in the desk research

The table below provides a description of the elements observed.

Heading	Number	Column name	Description
1. General characteristics	1.1	Type of actor	Describes the type of the organization or services: - e.g., cloud providers/service, fiscal sponsors/ foundations, open source computational services.
	1.2	Organizational structure	Describes the type of arrangement - e.g., formal nonprofit organization, a collective, or a social enterprise.
	1.3	Open source organization	Checks if the organization is open source, meaning it typically provides resources, infrastructure, and community support to facilitate collaboration and the sharing of software source code.
	1.4	Community driven	Checks if the organization is Community Driven, meaning it is established and operated by members of a particular community or group of individuals.
	1.5	Product type	Describes the product type of the organization in terms of whether it is a Platform, Project, Commercial Company, or other.
	1.6	Main services/products	Describes the main services or products designed and provided by the organization.
	1.7	Managed by / legal and fiscal organization	Provides the name of the organization that manages the initiative or that functions as a fiscal sponsor.
	1.8	Geographical area/continent of clients/users	Provides the geographical region or continent where clients or users are located.
	1.9	Disciplinary focus	Describes whether the services target specific disciplines (e.g., computer science, social sciences.)
	1.10	Main funders	Provides the names of foundations or governmental agencies supporting open infrastructures.
	1.11	Who are the users?	Provides the profiles of users of these services.

Heading	Number	Column name	Description
	1.12	Open cloud? Yes/No	Checks if it is an open cloud service.
2. Stakeholder participation	2.1	Official 1	Provides the Name and affiliation of the president or chair of the board.
	2.2	Official 2	Provides the Name and affiliation of the second name listed in the governance body.
	2.3	Official 3	Provides the Name and affiliation of the third name listed in the governance body.
	2.4	Availability of information about the election of officials: yes/no	Checks if the information about the election of officials is available or not.
	2.5	Election of officials	Provides a brief description of how they elect officials and has links to website documentation.
	2.6	Characteristics of board members	Provides a brief description of regularities on affiliations. For instance, if they mostly come from academia or other nonprofit organizations or even companies.
	2.7	Academic focus board: yes/no	Checks if there is an academic focus board. Yes, if ~50% or more of board members are affiliated with universities.
	2.8	Regional diversity of the board	Provides a brief description of diversity based on affiliations.
	2.9	Participation of stakeholders	Provides a description of events, community calls, and conferences to include and get to know stakeholders and users of services.
	2.10	Stakeholder governed: yes/no	Checks if an organization is stakeholder governed based on POSI: Stakeholder Governed – a board-governed organization drawn from the stakeholder community builds more confidence that the organization will take decisions driven by community consensus and consideration of different interests.
	2.11	Stakeholder governed justification	Provides a brief description justifying the yes/no for Stakeholder governed.

Heading	Number	Column name	Description
3. Non-discriminatory membership	3.1	Statements/ descriptions of how people can participate and contribute to the organization.	Describes or outlines ways individuals can participate and contribute to the organization.
	3.2	Non-discriminatory membership: yes/no	Checks if an organization has non-discriminatory membership based on POSI: Non-discriminatory membership – we see the best option as an “opt-in” approach with a principle of non-discrimination where any stakeholder group may express an interest and should be welcome. The process of representation in day-to-day governance must also be inclusive of governance that reflects the demographics of the membership.
	3.3	Non-discriminatory membership justification	Provides a brief description justifying the yes/no for Non-discriminatory membership.
4. Transparent operations	4.1	Availability of documentation on processes or operations: Yes/No	Checks if there is a presence of documentation regarding processes or operations.
	4.2	Documentation’s description	Provides a brief description of the documentation found.
	4.3	Transparent operations: yes/no	Checks if an organization has Transparent Operations based on POSI: Transparent operations – achieving trust in the selection of representatives to governance groups will be best achieved through transparent processes and operations in general (within the constraints of privacy laws).
	4.4	Transparent operations justification	Provides a brief description justifying the yes/no for Transparent Operations.
5. Living will	5.1	Living will: yes/no	Checks if an organization has Living Will based on POSI: Living will – a powerful way to create trust is to publicly describe a plan addressing the condition under which an organization would be wound down, how this would happen, and how any ongoing assets could be archived and preserved when passed to a successor organization. Any such organization would need to honor this same set of principles.

Heading	Number	Column name	Description
	5.2	Living will justification	Provides a brief description justifying the yes/no for living will.
6. Formal incentives to fulfill mission & wind-down	6.1	Mission statement	Provides the mission statement of the organization.
7. Vision	7.1	Vision Statement	Provides the vision statement of the organization.
	7.2	Do they have documentation associated with strategic planning? Yes/no	Checks if there is a presence of documentation regarding strategic planning.
	7.3	Justify strategic planning answer	Provides a justification for strategic planning documentation.
8. Infrastructure	8.1	Do they have documentation on administrative structures? Yes/no	Checks if there is a presence of documentation regarding administrative structures.
	8.2	Justify administrative structure documentation	Provides a justification for administrative structure documentation.
	8.3	Describe the cloud infrastructure they use	Provides an overview of the cloud infrastructure utilized by the organization.
9. Finances	9.1	Do they have any public information about their finances? Yes/no	Checks if there is any publicly available information regarding the organization's financials.
	9.2	Justify financial information	Provides a justification for the financial information.
	9.3	Do they have any public information about grants received? Yes/no	Checks if there is any publicly accessible information regarding the grants the organization has received.
	9.4	Justify grant information	Provides a justification for the grant information.

Heading	Number	Column name	Description
10. Engagement	10.1	Do they host meetings /events/conferences to engage with their users/communities? Yes/No	Provides a list of community engagement events arranged by the organizations for their users and community.
	10.2	Justify the description of community engagement	Provides a justification for the community engagement information.
11. Governance	11.1	Do they have descriptions of roles and responsibilities within their organizations? Yes/No	Checks if there are any available descriptions outlining the roles and responsibilities within the organization.
	11.2	Justify the description of roles and responsibilities	Provides a justification for the description of roles and responsibilities information.
	11.3	Do they have policies? Yes/No	Checks if there are any existing policies in place.
	11.4	Justify the description of policies	Provides a justification for the policies information.