The Global Dataverse Community Consortium

Supporting Dataverse Repositories Around the World

Finding a Path to Sustainability for the Dataverse Community

A 2024 report by the Dataverse sustainability working group.

October 16th, 2024

Authors: Dieuwertje Bloemen (KU Leuven), Dimitri Szabo (INRAE), Rene Belso (DeiC), Philipp Conzett (UiT The Arctic University of Norway), Stefano Iacus (Harvard University), Julian Gautier (Harvard University), Vaidas Morkevicius (Kaunas University of Technology/LiDA), James Myers (GDCC/QDR), Pedro Principe (UMinho)

Contributors: Chris Baars (DANS), Oliver Bertuch (Forschungszentrum Jülich GmbH), Jamie Jamison (UCLA Data Science Center), Steve McEachern (Australian Data Archive), Asbjørn Skødt (University of Copenhagen)

The Dataverse® Project

I. Table of Contents

II. Executive summary

The Dataverse Project is an open-source software project that has cultivated a growing and lively community of collaborators, contributors and implementers. In 2023, the Dataverse community formally recognized the need to address the challenges the Dataverse community has been facing due to its continuous popularity and growth. From this, the Dataverse Sustainability Working Group was established and decided an in-depth analysis of the current state of affairs was necessary to identify and find a path to sustainability for the Dataverse community. The mission of the working group, as endorsed by the Global Dataverse Community Consortium (GDCC), was to provide evidence-based, community-reviewed, and consensus-based recommendations on how to sustain and grow the Dataverse community. The group started its work in August 2023 and ran until October 2024, ending with the publication of this report.

The working group divided its work across two work packages. One to analyze the current state of affairs of Dataverse's sustainability and the second to gauge the possibilities to generate monetary support for Dataverse sustainability by providing more added values in exchange for e.g. membership fees.

Work package one was completed by performing a phase analysis and organizing several workshops as described in the It Takes a Village (ITAV) framework. During the phase analysis, the categories of governance, technology, resources and engagement were examined, resulting in maturity levels of 1, 2.5, 1, 2 out of three, respectively. What consistently stood out during these analysis sections, was the current invaluable investment by the Dataverse Project Team at IQSS at Harvard University in Dataverse, which resulted in the working points being mainly in the area of formalizing and documentation of processes and policies.
After this analysis, workshops were organized to identify critical catastrophes and to afterwards explore the necessary steps to prevent these from happening. The phase analysis in combination with the workshops results in the definition of the following 10 recommendations for the Dataverse community to improve its sustainability:

1) Establish a new legal organization for GDCC
2) Establish a formalized governance of GDCC
3) Clarify/demonstrate the value/benefits of the GDCC membership and strengthen the value exchange
4) Formalize the financial model and fee contributions/membership model
5) Diversify income streams for GDCC
6) Establish a functional strategy for the Dataverse software
7) Formalize and set up a decision-making process
8) Make it easier to deploy and upgrade the application and integrations
9) Formalize and document the communication strategy for high-priority items
10) Help grow, manage and sustain a community-supported ecosystem of Dataverse-integrated tools, services, solutions, and add-ons

For work package two (added values), a survey was carried out early 2024, where the things found to be most valuable by the respondents were: the support for software installation/migration and upgrading, the integration of long-term preservation in Dataverse, members' ability to prioritize getting a bug fix or feature request into a release, and

sustaining the Dataverse software and associated services and tools. The latter shows the overall attribution of value from the community members to sustainability for Dataverse. These results can help provide options to establish recurrent income for the Dataverse community, though further feasibility studies are necessary per item before implementation or offering of these options can be considered.

In conclusion, there is a need for formalization that is currently lacking, for which this working group has proposed 10 recommendations. In order to move forward with these recommendations, dedicated human resources that can focus on drafting an action plan will be necessary. The working group urges individual Dataverse installations to explore how they can invest in the provision of these dedicated human resources and in the sustainability work that is necessary. Especially since the continued sustaining of the Dataverse software and associated services and tools was voted as being one of the most valuable value exchanges available for the respondents of the survey performed in work package two. And as the community grows, formalization of the mechanisms and incentives to contribute to sustainability will be critical to address the above-mentioned 10 recommendations.

1. Background

"The Dataverse Project is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data [and other research resources] available to others, and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility."[1]
Since its conception in 2006 at Harvard University's IQSS, the open-source software has cultivated a growing and lively community of collaborators, contributors and implementers. Being a community-driven open-source software (OSS), the Dataverse Project faces many of the opportunities and challenges that are common in the OSS world. "The Dataverse community and its diversity is a critical factor in the long-term sustainability of OSS, ensuring the software's ability to upgrade, adapt and grow to meet new needs and evolve with advances in technology (Arp & Forbes, 2022, p. 6)"[2]. In recent years, this community has grown considerably. To help support the community, the Global Dataverse Community Consortium (GDCC) was established in 2018 by UNC, Harvard and DANS to provide international organization to existing community efforts and to provide a collaborative venue for institutions to leverage economies of scale in support of Dataverse repositories around the world. The GDCC has since grown to include organizations such as the Australian Data Archive, INRAE, UiT The Arctic University of Norway, and many others as its members.[3]

In 2023, the GDCC steering committee and Dataverse community recognized a need to establish a working group to address challenges the Dataverse community has been facing due to the continuous growth of the Dataverse software deployments across countries, domains, and types of organizations.
To meet this need, the Dataverse sustainability working group (WG) was established in September 2023 to map out the needs and challenges of the community.

The mission of the Sustainability WG was to provide evidence-based, community-reviewed, and consensus-based recommendations on how to sustain and grow the Dataverse community. These guidelines are intended to help the community establish new and/or adapt existing strategies, organizational structures (e.g., through an organizational model), guidelines, tools, routines and other resources that are necessary to empower the global Dataverse community to ensure that the Dataverse software be maintained and developed in a way that benefits the user community as a whole in a long-term perspective.

The working group's aim, as defined in its charter[4], is to deliver recommendations in the form a report(s), addressing the following issues:

- **Mission and vision:** What purpose should the community have? What should the community envision to achieve in the next 5-10 years?
- **Strategic objectives:** What objectives should the community have to carry out its mission and to achieve its vision?

---

[1] *Dataverse Project - About page.* (n.d.). Retrieved July 10, 2024, from https://dataverse.org/about
[2] Conzett, P. (2022). Dataverse Community Survey 2022 – Report. *Septentrio Reports, 1.* https://doi.org/10.7557/7.6872
[3] *GDCC - Members page.* (n.d.). Retrieved July 10, 2024, from https://www.gdcc.io/members.html
[4] Dataverse sustainability working group charter: https://docs.google.com/document/d/17zp7hBy4OeprpZ4cL2YwuhpRL9li-7j_OCjYE0MYC1k/

- **Means and resources:** What means and resources should the community use to achieve its strategic objectives?
- **Implementation:** What plan and roadmap should the community have to achieve its objectives?

The working group finished its work in September 2024 with the finalization of this report, ensuring all the community could join meetings when relevant and were kept up-to-date on the work in progress during the run of the working group.

2. Working group work packages

At the start of the working group, two main work packages were decided on;

WP1: Analyze the current state of affairs of Dataverse sustainability
WP2: Gauge the possibilities to generate monetary support for Dataverse sustainability by providing more added values in exchange for e.g. membership fees

With the first work package being identified as having the highest priority. The second work package can be seen as part of the first. Though, due to the immediate recognition of the need for this exploration of added values to offer by the working group, it was decided to pick this up separately and early on.

For work package 1, the analysis of Dataverse's current sustainability, the first step was to decide what framework or previous work done by other communities to follow. For this, the working group evaluated frameworks and examples such as Community Canvas[5], Open Source Guides[6], Lyrasis Community Programs (e.g. DSpace)[7], It Takes A Village[8], EC Guidelines for creating sustainable open source communities[9], Community Cultivation Resource Library[10], Apache Software Foundation[11], and the Eclipse Foundation[12].

After having a preliminary look at these frameworks and examples, it was decided that the It Takes a Village Framework by Lyrasis was the best fit for the objective of this working group and would cover the relevant areas of interest in-depth while also providing clear guides and workshop guidelines.

Next to this initial analysis of possible frameworks to evaluate the overall sustainability, work package 2 also made headway by deciding to organize a community consultation to get

---

[5] *Community Canvas*. (n.d.). Retrieved July 12, 2024, from https://community-canvas.org/
[6] *Open Source Guide*. (n.d.). Retrieved July 12, 2024, from https://opensource.guide/
[7] *Lyrasis - Community Programs.* (n.d.). Retrieved July 11, 2024, from https://lyrasis.org/community-programs/
[8] *It Takes A Village - ITAV: Open Source Software Sustainability.* (n.d.). Retrieved July 14, 2024, from https://itav.lyrasis.org/
[9] *European Commission - Open Source Observatory (OSOR): Guidelines for creating sustainable open source communities.* (n.d.). Retrieved July 11, 2024, from https://joinup.ec.europa.eu/collection/open-source-observatory-osor/guidelines-creating-sustainable-open-source-communities
[10] *Educopia - Community Cultivation Resource Library.* (n.d.). Retrieved July 11, 2024, from https://educopia.org/cultivation/
[11] *Apache Software Foundation - How The ASF Works.* (n.d.). Retrieved July 11, 2024, from https://apache.org/foundation/how-it-works/
[12] *Eclipse Foundation - About page.* (n.d.). Retrieved July 11, 2024, from https://www.eclipse.org/org/foundation/

input from the current Dataverse installations on what they would find valuable if it were offered by a body such as the GDCC.

3. WP1: Sustainability analysis with It Takes a Village

The It Takes a Village (ITAV) framework points as its first step to do a first phase analysis to identify the current state of affairs of the sustainability of an open-source project and community. After this phase analysis, gaps were identified and workshops were organized to lay the foundational work for recommendations on how to bridge the identified gaps and to improve the overall sustainability of the Dataverse Project.

3.1. Phase analysis

ITAV's phase analysis groups a series of elements and characteristics that should be present in a sustainable open source community. The elements are grouped across four categories: governance, technology, resources and community engagement and allow for a step by step evaluation of the current state of affairs of the overall sustainability. To determine in which phase of sustainability a community is for each group, an evaluation needs to be made on whether the element is currently present or not, or if it's in progress of being implemented or developed. After this one-by-one evaluation of each element, the overall phase of a category of elements is calculated and can be adjusted if the collaborators feel the phase estimate is not indicative of the current state of affairs.

The resulting phase determination can range from one to three, with one indicating a less mature set-up, and with three indicating the most mature set-up.

3.1.1. Governance

The governance category of the phase analysis contains elements such as having a well-communicated governance model, providing documentation around existing governance policies, community engagement, and other elements having to do with the community structure and organizational structures.
Overall the analysis led to a calculated phase determination of one, being less mature, with the value being upheld after a brief discussion. An important note on the governance phase determination is that it's mainly the policies and formalization that is missing, not so much the overall efforts of different parties of the community and the community as a whole.

3.1.2. Technology

Instead of focussing on governance models and structural work to do with the community set-up, the technology category focused on elements such as how community input and feedback is collected and processed, how stakeholder and community engagement is set up and roadmapping, and decision-making to do with the software itself, such as new features and functionality coming in or methods of evaluating existing features and functionality.
Compared to the governance, this phase determination scored higher with a value of two, though after discussing the elements in depth with the working group, it was decided, that it's likely more somewhere between two or three as the difference between two and three for this category is mainly a lack of formalization of most elements described and evaluated. With there being a lot of maturity overall in the Dataverse Project in this category, just the

documentation of some things is missing and could therefore not be scored higher per element.

3.1.3. Resources

Resources is a category that focuses a lot on the financial sustainability of the community and the project, by including elements such as necessary human resources, financial planning, working on partnership and elements to do with the business model and income streams.
Overall this category's determination scored a one, and after a brief discussion, this determination was upheld by the working group as being correct.

3.1.4. Community Engagement

The last category, community engagement, contains elements such as stakeholder groups, outreach, communication and engagement plan(s), community activities, and dedicated staff for engagement.
In this category, the calculated phase determination was one, though the working group decided it was skewing too low, and therefore the final determination was set to two, with a similar note as for the technology category. Namely, that it's again about the ambiguity of formalization versus informal realization that doesn't allow for the working group to say that certain elements are present in this analysis. There was even a brief discussion if this category should perhaps even be scored a three due to the extensive informal work already being done in regards to community engagement, by, among others, the Dataverse Project team at IQSS at Harvard University.

The figure below summarizes the maturity level of Dataverse according to the ITAV analysis carried out by the Sustainability WG:

	Maturity Level		
Category	**1**	**2**	**3**
Governance			
Technology			
Resources			
Engagement			

3.2. Workshops

Once the phase analysis exercise was completed, the next step was to organize a catastrophizing exercise as described in the It Takes a Village framework. The main category chosen for the first workshops was governance, as it seemed from the elements across the different categories that the need for formalization and policies could also improve the determination of the other categories. Though during the catastrophizing exercise, it was

decided not to focus too narrowly on just governance and draw the scope wider where relevant.

The first workshop was about listing possible worst case scenarios that could happen to the Dataverse project and community, such as losing funding, divergent factions within the community or there being a sudden high-security issue for all installations that needs to be addressed. The brainstorm of these worst-case scenarios was captured and these were then grouped together to find common threads or common causes to some of these issues. Grouping them resulted in four different so-called *critical catastrophes*:

1. Loss of GDCC financial income streams and loss of key human resources currently at work for the GDCC (e.g. loss of support from its members, or loss of key technical personnel).
2. Loss of key administrative support. Without a separate legal entity for GDCC, the ability to manage memberships, contracts, and grants depends on volunteer efforts and is subject to the policies and priorities of the organizations willing to manage GDCC finances.
3. Differing view on the way forward for Dataverse (e.g. creation of PRs for features that aren't in line with what the Dataverse project software is intended for, or the community disagrees about future development of the software, leading to multiple spin-off forks).
4. A sudden need of technical capacity for all Dataverse installations (e.g. implementation of a new authentication set-up requires technical skills not all installations might have available, or there being a critical security issue that necessitates all installations to upgrade to the latest version as soon as possible).

Once these four critical catastrophes were defined, for each of them, the current possible response was noted down and with that gaps in and issues with the current approach were identified.

From these critical catastrophes, current approaches and their gaps, a next step in the workshop was to identify actions that could be taken to avoid or mitigate these catastrophic situations. With this exercise, it's understood that full coverage of all gaps and issues is near impossible, so a focus was to define actions that were realistic and could be strived towards. Though, notably no ideas or possible initiatives were excluded if there was uncertainty on its feasibility to ensure opportunities for improvements weren't overlooked. The corresponding actions were broad in scope and ranged from very specific actions to more strategic goals. For each above mentioned critical catastrophe, corresponding broad descriptions of the different actions are the following:

1. Establishment of a different type of legal organization for the GDCC, and establishment of a formalized governance of the GDCC.
2. Strengthen and clarify the added value of a GDCC membership, and formalization of the financial model of the GDCC and its membership model.
3. Formalization of a functional strategy for the Dataverse project, and outlining a democratic and transparent decision-making process.
4. Making Dataverse easier to deploy and upgrade, and formalize and document the communication for high-priority items.

During following workshops these actions were further refined and translated into 10 recommendations for the Dataverse community that can be found in the similarly named section further down in this report.

4. WP2: Value exchange analysis

As previously mentioned, work package two was quite a bit smaller in scope, but as can be seen from action point 2 from work package 1, as predicted there is a need to identify what added value GDCC membership could provide to partners, as this would be a welcomed source for more income for GDCC and therefore could make the community and software more financially sustainable.
Against that background, it was decided to run a community consultation to determine what kind of wants and needs current installations may have would be worth the investment of a GDCC membership to them, considering that currently a relatively small portion of Dataverse installations are members of the GDCC.
The survey was a first step to find where the current focus of needs and wants of the community is and what they would deem valuable additions or exchanges if provided by the Dataverse community, possibly in return for a membership fee. The included value propositions were not a priori tested for feasibility, but rather collected as ideas and options to gauge the overall options that could be of interest to Dataverse installations. It was decided to postpone the feasibility analysis and think through possible practical implementations of the proposed value exchanges at a later date to ensure that no time would be spent on propositions that might not even be deemed as valuable by the participants and to keep the door open for value propositions that might be high in cost, but could be signaled as of interest to all, possibly making the higher investment worth it. By keeping the door open to all ideas, this also ensured responding installations would supply their own ideas without necessarily feeling the need to consider feasibility before submission of ideas.
This set-up of the included value propositions not being set-in-stone options but purely being a brainstorm exercise was clearly communicated to the participants.

In this context, an initial brainstorm with the working group was organized where ideas for added values were listed and organized across a couple of categories. The proposed ideas were then cleaned up to avoid duplication and grouped in the following final categories:

1. Technical & Development Support (6 propositions)
2. Community Governance (11 propositions)
3. Sustaining Software (2 propositions)
4. Training, Communication & Outreach (9 propositions)
5. Member’s Cost Reduction (3 propositions)
6. Organizational & Certification Support (6 propositions)

These were then translated into a survey, which left room for more input and ideas by community members and the respondents were asked to rate each possible added value across each category. The proposed added value ideas were grouped in a series of categories. For each proposed added value, the respondents had to indicate it being: not valuable to us, slightly valuable to us, valuable to us, fairly valuable to us, or very valuable to us.

With each category being fronted by the same phrase: “Rate these proposed value exchange ideas. They are suggestions of added value to a community membership. What would make the community membership valuable to your installation, if they were offered?” With this set-up, we could ensure comparison across the categories. Ranking the value propositions would have made this much more difficult.

The survey was sent out early January 2024 to the Dataverse Project community across the usual channels, which includes email, Google Groups, Zulip and Slack. The survey closed after a one-month run and in total 31 installations participated with;

- 15 European Dataverse installations
- 8 South American Dataverse installations
- 6 North American Dataverse installations
- 2 Asian Dataverse installations

4.1. Data Collected

During the survey conducting period a total of 33 responses were collected using the Microsoft forms tool. 2 of these responses were by installations that had also sent in another response to the survey, these answers were deduplicated by taking the midpoint answer if the answers differed between rated questions. So, a total of 31 unique installations filled in the survey. All respondents were from existing Dataverse installations.

The 31 installations represent 27% of the known 115 Dataverse installations as registered in the overview provided on the Dataverse Project website (https://dataverse.org/). Of these 31 installation respondents, 45% indicated being current members of the GDCC.

4.2. Results

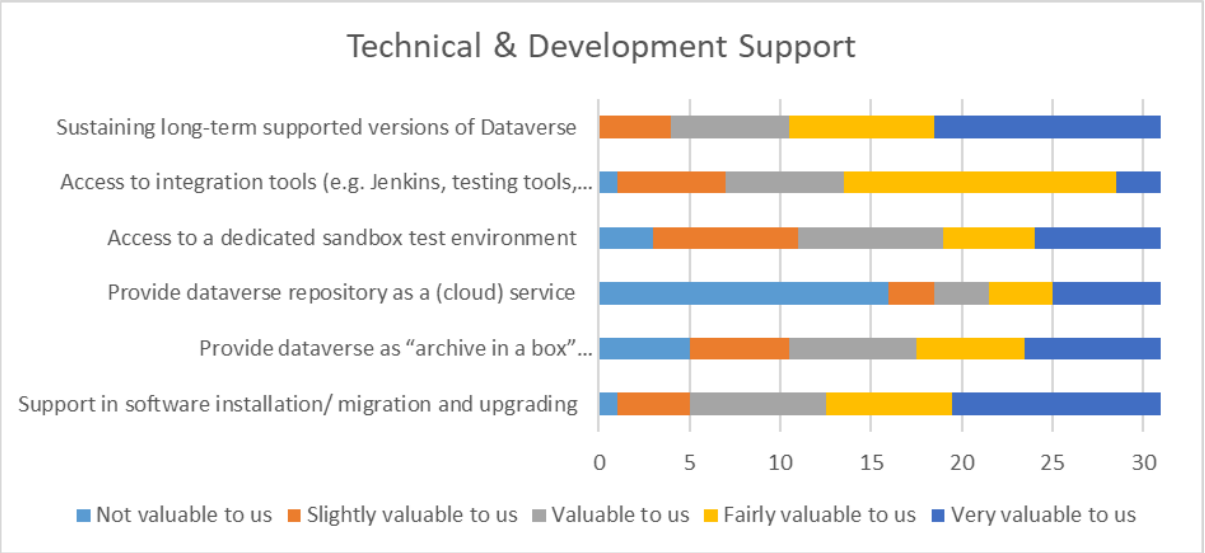
4.2.1. Technical & development support category

Under the technical & development support category, the following proposed value exchange ideas were included:

1) Sustaining long-term supported versions of Dataverse
2) Access to integration tools (e.g. Jenkins, testing tools, tests)
3) Access to a dedicated sandbox test environment
4) Provide Dataverse repository as a (cloud) service
5) Providing Dataverse as “archive in a box”
6) Support in software installation/migration and upgrading

With the results being as shown in figure 1. Which point to the first and last idea of this category receiving the strongest value evaluation, while providing Dataverse repository as a (cloud) service was voted as not being valuable by more than 50% of participants. Though, note that the respondents are all current community members with active Dataverse installations, thus pointing to already having a running set-up and therefore not needing this service.

Figure 1: Responses on the added value ideas in the technical & development support category



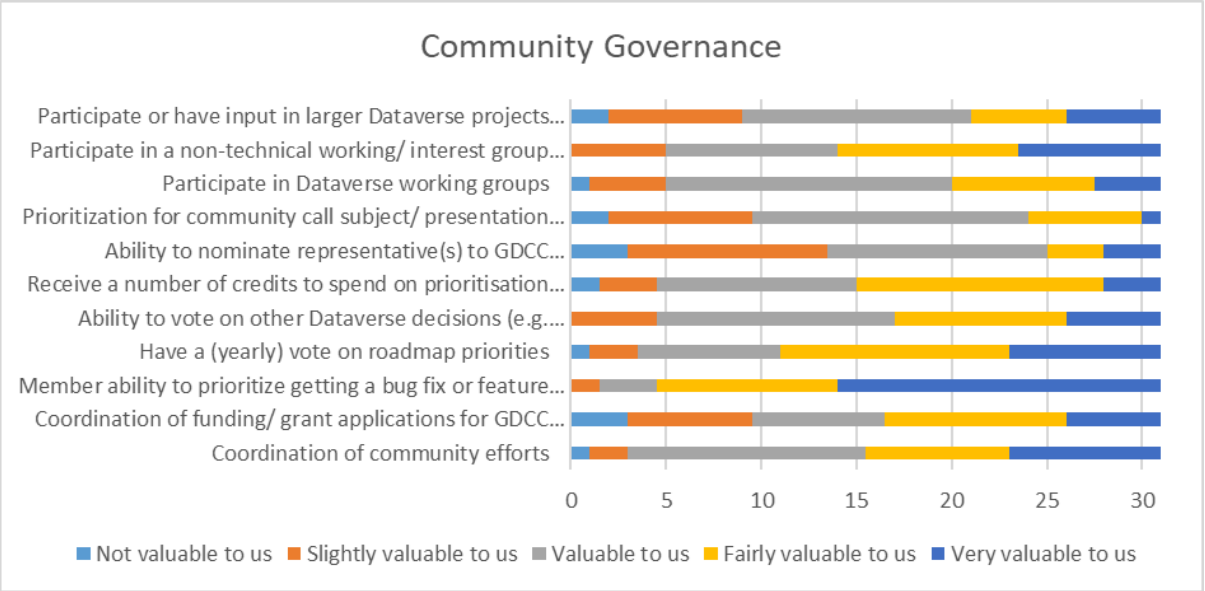
4.2.2. Community governance category

Within the community governance category, the following ideas were included:

1) Participate or have input in larger Dataverse projects (e.g. SPA planning)
2) Participate in a non-technical working/ interest group that provides input on matters such as UX/ Metadata etc. from the user's or admin's perspective.
3) Participate in Dataverse working groups
4) Prioritization for community call subject/ presentation proposals
5) Ability to nominate representative(s) to GDCC governance groups
6) Receive a number of credits to spend on prioritization of a specific PR for the GDCC developers
7) Ability to vote on other Dataverse decisions (e.g. location of the Dataverse Community Meeting)
8) Have a (yearly) vote on roadmap priorities
9) Member ability to prioritize getting a bug fix or feature request into a release
10) Coordination of funding/grant applications for GDCC members
11) Coordination of community efforts

The results of this category are shown in figure 2. As the figure shows, the idea of being able to prioritize getting a bug fix or feature request into a release is voted as being very valuable or fairly valuable by almost all installations. Voted least valuable were option 1, 4, 5 and 10. Community coordination, voting on roadmap priorities and participation in a non-technical working/interest group were also deemed valuable to most participants.

Figure 2: Responses on the added value ideas in the community governance category



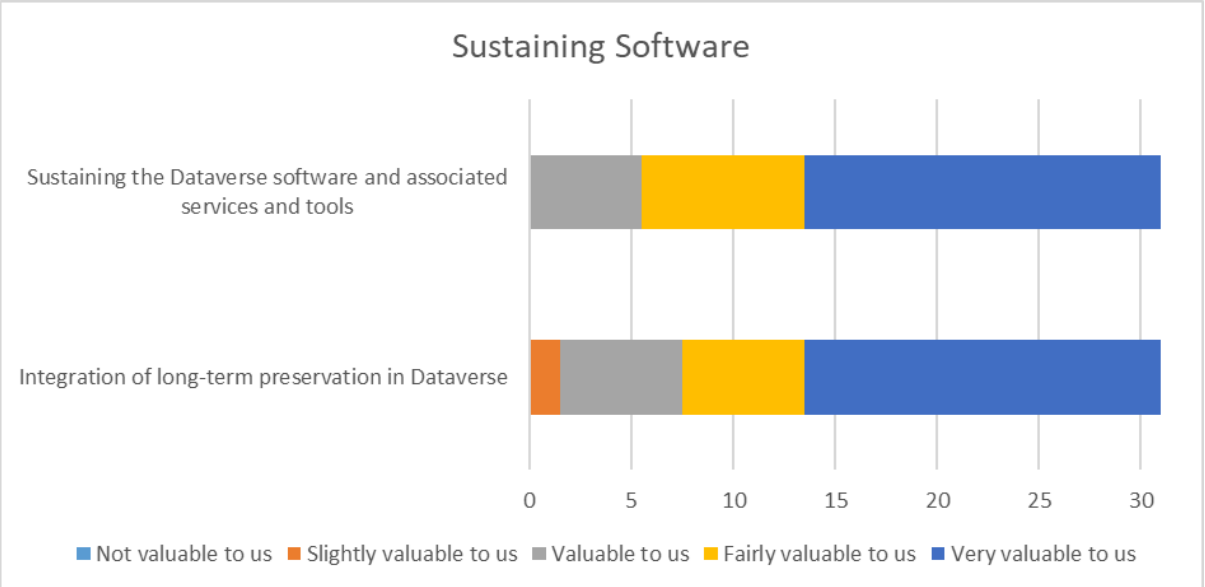
4.2.3. Sustaining software category

The section on sustaining software included the following two propositions, possibly pointing to a need for more detailed options for this section, as both added values are relatively abstract and general.

1) Sustaining the Dataverse software and associated services and tools
2) Integration of long-term preservation in Dataverse

The results (shown in figure 3) point to a general sense of importance of these two points. Neither of these options were voted as being not valuable by anyone and the majority for both suggestions voted them as being very valuable.

Figure 3: Responses on the added value ideas in the sustaining software category



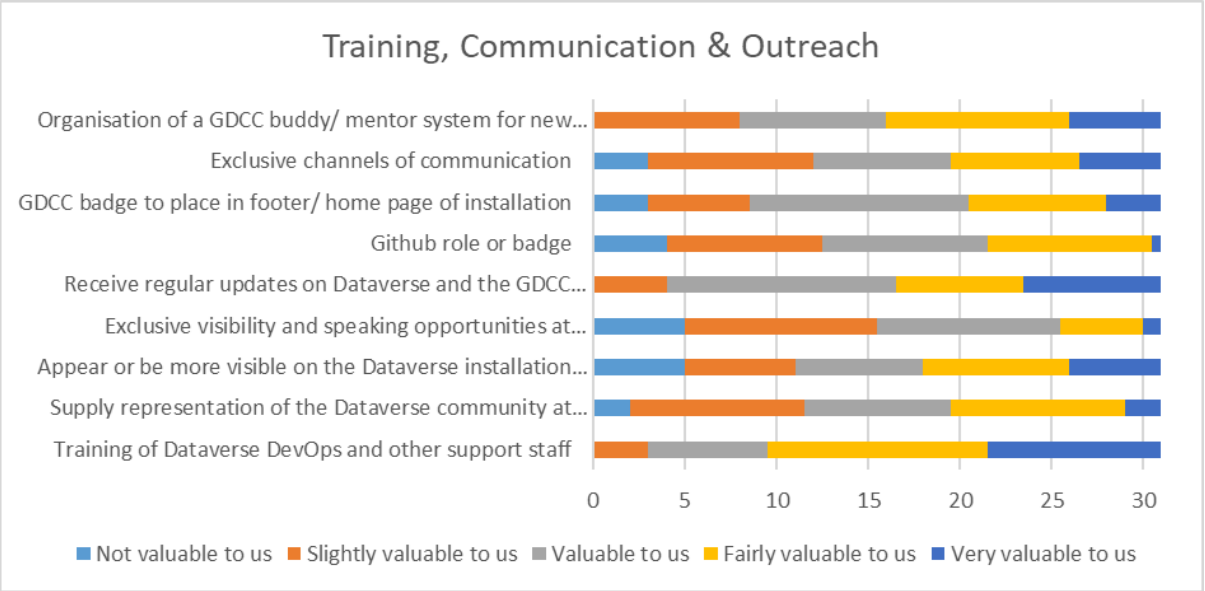
4.2.4. Training, communication & outreach category

The category on training, communication and outreach was quite varied and included the following varying options;

1) Organization of a GDCC buddy/ mentor system for new members/ installations
2) Exclusive channels of communication
3) GDCC badge to place in footer/ home page of installation
4) Github role or badge
5) Receive regular updates on Dataverse and the GDCC between events
6) Exclusive visibility and speaking opportunities at Dataverse Community Meeting or other events
7) Appear or be more visible on the Dataverse installation map
8) Supply representation of the Dataverse community at conferences
9) Training of Dataverse DevOps and other support staff

As seen in figure 4, most added value ideas were voted across both sides of the spectrum. The three slight outliers were idea 1, idea 5 and idea 9, with all three generally being received as valuable to most, with no one indicating those as being not valuable at all. Idea 4, 6 and 7 scored as less valuable in comparison to the other options for this section.

Figure 4: Responses on the added value ideas in the training, communication & outreach category



4.2.5. Member's cost reduction category

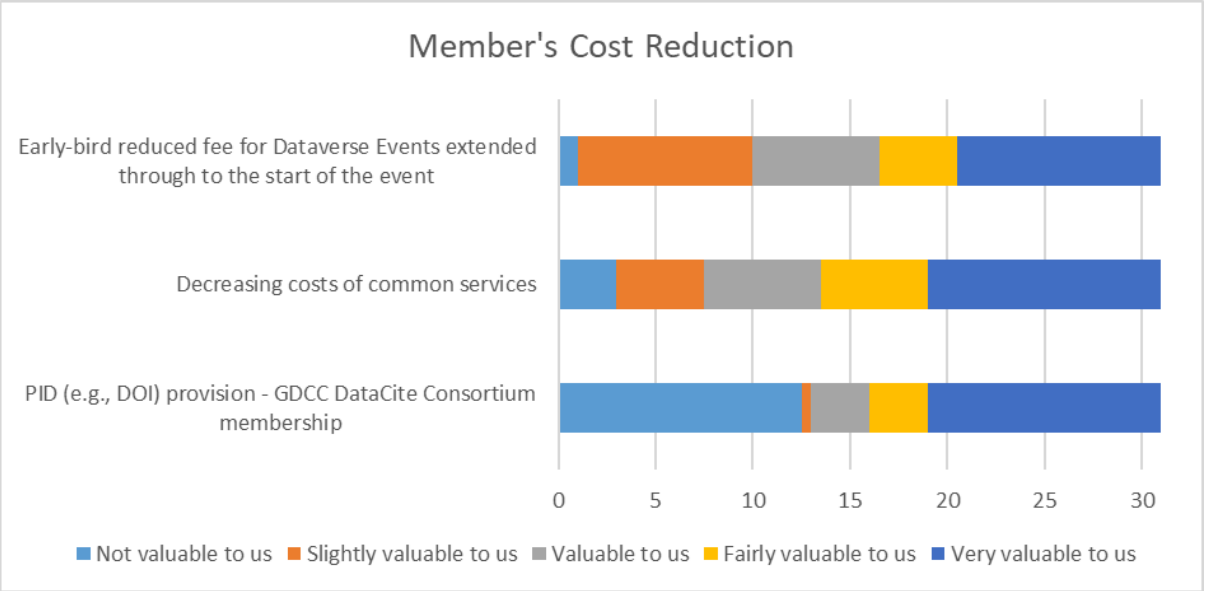
A section named "Member's cost reduction" focused on ideas/services that could be provided by the GDCC to lower the overall cost of running a Dataverse installation. With the included ideas being:

1) Early-bird reduced fee for Dataverse Events extended through to the start of the event
2) Decreasing costs of common services
3) PID (e.g., DOI) provision - GDCC DataCite Consortium membership

As shown in figure 5, the answers for this section were quite varied. Though again, it should be noted that all respondents already have a Dataverse installation, and therefore perhaps already have a DOI membership individually. In addition to this value evaluation exercise, an open question was also asked under this section asking for other common services the respondent would be interested in for the Dataverse community to offer to reduce costs of members. Here, 8 participants provided answers, with the answers including:

- First level support.
- Technical support to fix bugs and run updates. Integration of additional metadata standards.
- Cloud hosting with SaaS model.
- Support from the network itself.
- A more uniform approach to the GDCC discount provided for community meetings.
- Establishment and management of a common permanent fund to run the annual meeting at reasonable costs for the whole community.
- Fee reduction when we present work at the event and in training.

Figure 5: Responses on the added value ideas in the member's cost reduction category



4.2.6. Organizational & certification support category

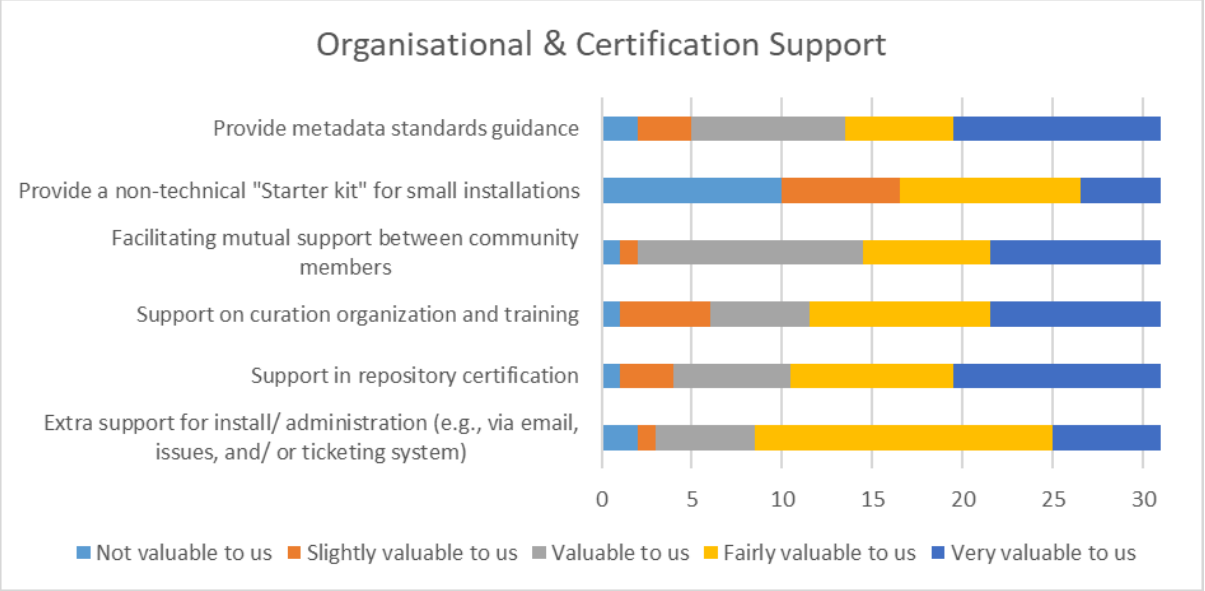
Under the section called “organizational & certification support”, the following added value ideas were included:

1) Provide metadata standards guidance
2) Provide a non-technical "Starter kit" for small installations
3) Facilitating mutual support between community members
4) Support on curation organization and training
5) Support in repository certification
6) Extra support for install/ administration (e.g., via email, issues, and/ or ticketing system)

Figure 6 shows the distribution of the value evaluation of the ideas, with the non-technical “starter kit” being voted least valuable. Again, this should be interpreted in the context of the

respondents representing already running Dataverse installations, for whom the need for/value of a starter kit is obviously less.

Figure 6: Responses on the added value ideas in the organizational & certification support category



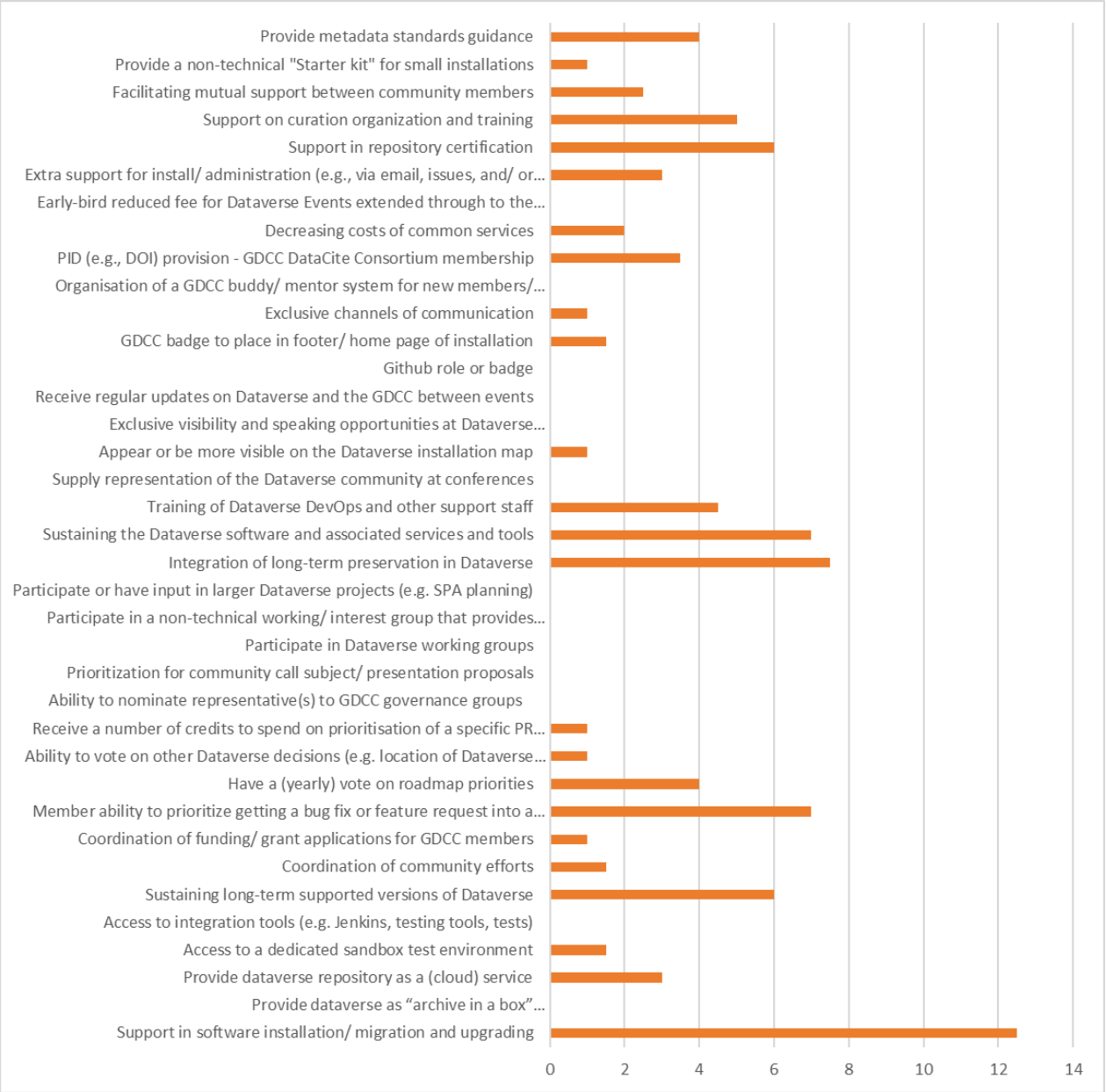
Noteworthy is that most other ideas all score more valuable on average than the previous categories, with providing metadata standard guidance and support in repository certification scoring especially high.

4.2.7. Priorities & other remarks

In order to get an overview of those value exchange ideas that are valued most by the respondents across all categories, an extra section was provided at the end that asked respondents to select their top 3 added values in no particular order. On top of that, an open-ended question was also provided to inquire after other ideas that a community membership could or should offer.

The result of all respondents providing their top 3 added value ideas is shown in figure 7, with support in software installation/migration and upgrading with 12.5 votes, Integration of long-term preservation in Dataverse with 7.5 votes being the top 2, while member ability to prioritize getting a bug fix of feature request into a release, sustaining the Dataverse software and associated services and tools sharing third place by both getting 7 votes.

Figure 7: Count of added value ideas mentioned in the top three of respondents.



On the optional question to supply other ideas, there were four responses provided. With the input mentioning a paid service to outsource DevOps work, a dockerized quick-start version of Dataverse, democratized access to Dataverse for institutions with limited resources, and access to paid services that could be provided by other members of the community.

4.3. Conclusions

The survey showed that overall the things found to be most valuable by respondents were the support in software installation/migration and upgrading, the integration of long-term preservation in Dataverse, members’ ability to prioritize getting a bug fix or feature request into a release, and sustaining the Dataverse software and associated services and tools.

The latter shows the overall attribution of value from the community members to sustainability for Dataverse.

However, it should be noted once more that the responses collected are only from already existing Dataverse installations. There were no respondents from non-Dataverse-using institutions. Therefore, the survey results don't necessarily reflect their needs or perceived added values. A good example of this is the suggestion of *providing Dataverse as a service* being scored relatively low, which makes sense in the context of all respondents already having an up and running Dataverse instance, which they likely set up themselves. It would perhaps be interesting to also gather input from other types of stakeholders, though this was outside of the scope of this exercise.

With the outcome of this survey, one of the action points in the It Takes a Village Analysis can partially be addressed, though more work is necessary. As previously mentioned, none of the proposed added values were investigated for their feasibility, so therefore a feasibility analysis will be necessary before any of these could potentially be offered to generate more income for the GDCC. Some of the top-voted options also require considerable organizational input, development work and governance, which can be taken on-board with the other recommendations resulting from the overall phase analysis previously discussed.

5. Dataverse sustainability recommendations & action points

In order to condense the evaluation down to actionable tasks, the following ten recommendations were established based on the previously mentioned work. It's a consolidation of the different steps, such as the phase analysis and the catastrophizing exercise. For each recommendation, the general ITAV category and the prioritization by the working group is indicated. The prioritization is either top or high priority, due to the previous pruning and narrowing of the scope down to these 10 key recommendations
Note, these recommendations are based on the It Takes a Village work done by the working group, where the main conclusions were that there is a strong informal community, with the invaluable efforts of the Dataverse Project team at IQSS at Harvard University playing a key role in both the software and community's day-to-day operation having lead to the success of Dataverse as one of the leading data repository softwares. These recommendations mainly focus on the need for formalization that is currently lacking, and shouldn't be read without first acknowledging the current investments and work by teams, such as the previously-mentioned Dataverse Project team at IQSS, in the community and software's functioning.

5.1. 10 key recommendations for Dataverse sustainability

Cf. notes final WG meeting
Prioritization of recommendations (phase analysis & catastrophizing)

1. *Establish a new legal organization for GDCC*

Category: Governance
Priority: top priority

The establishment of a new legal organization for GDCC is a recommendation stemming from the first and second critical catastrophes summarized as the loss of key human resources currently at work at GDCC (e.g. the entire steering committee steps down, or loss of key technical personnel), and the loss of GDCC financial income streams (e.g. loss of all paying members, or loss of support from its members).
The current approach to tackle these catastrophes were deemed insufficient as they would be based on voluntary efforts of people currently part of the community and a reliability on one key figure in the current GDCC financial and structural set-up. As such, the need was defined for a new legal organization to house or host the GDCC. As part of this recommendation, it is advised by the working group, based on the It Takes a Village analysis to hire the required dedicated personnel that can focus on doing the necessary research into different organizational set-ups and do the preparatory work necessary to initiate this new legal set-up for the GDCC. It will be necessary to analyze different organization types and structures, which can result in the GDCC being a stand-alone legal entity or becoming part of another existing organization. These organization types and structures will have to be explored as well as options to engage a fiscal sponsor or nonprofit to possibly serve as a home or sponsoring organization as is suggested in the ITAV framework to improve overall sustainability.
It is suggested by the working group, that once the research and exploratory phase is done, the community is consulted on what option is preferred by the majority of the community. This can be in the form of a vote or other type of stakeholder consultation to ensure every community member has had the opportunity to provide input on this matter.

2. *Establish a formalized governance of GDCC*

Category: Governance
Priority: high priority

Formalizing the governance of the GDCC is another recommendation stemming from the two identified catastrophes mentioned in recommendation 1. This recommendation particularly stems from a need to stabilize and strengthen the financial funds on which the GDCC is working. This to ensure that any catastrophes resulting in sudden loss of funding can be mitigated to a certain point and has a clear set of roles responsible in GDCC for these matters to address as needed.
The working group's advice to achieve this recommendation is to continue the It Takes a Village work where it is advised to organize a committee charter activity to formalize charters and governance of the organization. On top of this, the documentation of the governance and accountability will prove to be key to formalize the inner working of the GDCC and ensure transparency to the entirety of the community. Resulting from the phase analysis, further exploration of governance models is recommended by the working group, after which the chosen governance model(s) should be shared with the different stakeholders and checked for legibility and transparency before implementation. As part of the necessary formalized governance documentation, the establishment of a succession plan is also strongly suggested to further ensure the sustainability of the community.
And to ensure proper future functioning of the GDCC going forward, the election procedure of the GDCC steering committee members and the technical advisory board members (and possibly for other types of bodies) should be decided on and formalized.

To implement this recommendation, the working group recognizes that this requires considerable human resources input, as the majority of the work necessary is documentational and organizational in nature paired with extensive stakeholder consultation.

3. *Clarify/demonstrate the value/benefits of the GDCC membership and strengthen the value exchange*

Category: Governance
Priority: top priority

The need to strengthen the income potential of GDCC is based on the ITAV critical catastrophe where the GDCC would lose financial income streams (e.g. loss of all paying members, or loss of support from its members). As discussed in work package 2, the majority of Dataverse installations are currently non-paying members of the community. The survey pointed to the majority of non-member either not knowing of the GDCC or not seeing the added value of becoming a paying member. Though, overall, the respondents point to sustaining the Dataverse software and associated services and tools being one of the most important added values the GDCC offers or can offer.
As part of this recommendation, the working group advises the establishment of transparent structure and clear expectations around contributions to all the GDCC and Dataverse stakeholders and to provide a cohesive overall message as to the purpose and goal of the GDCC and provide policies that guide and foster engagement, such as contributor guidelines, codes of conduct, onboarding policies, etc. These are partially in place, though not all are findable to all community members. These recommendations focus on making the purpose of the GDCC more clear, though the matter of resources cannot be seen separate from this, as one strongly depends on the presence of the other.
The working group also sees the need for an exploration of mechanisms and processes that could enable community members to help each other, as well as blogs, easy to update documentation, forums and arenas for the active community that are all easily findable. Though, it is noted that there are already mechanisms in place, with the Dataverse Project team at IQSS at Harvard University being the main provider of support. This can be seen as an opportunity to build on the knowledge and experience of the Dataverse Project team at IQSS to broaden the teams and people available to provide support for Dataverse installations in an interchanging set-up.
All of these parts can help clarify the value of a formalized open-source community, as the added value of the GDCC currently doesn't seem to be clear to all members of the community.

4. *Formalize the financial model and fee contributions/membership model*

Category: Governance
Priority: high priority

This recommendation ties in strongly with the previous and the following recommendation. There is a need to review the financial model and fee model of the GDCC to explore more sustainable options. As part of this, the working group identified the need to establish a model to cover the GDCC's operational costs, which are needed to manage and run the GDCC on a day-to-day basis. Two types of funds defined as crucial to be established for this are contingency funds and operational funds.

On top of the organizational side, the GDCC should also explore other types of members' contributions, such as automatic renewal of pluriannual payments, tiered memberships, doing a call for funds to the community to ensure the sustainability, or organizing a call for donations of sorts. A further exploration of the added values proposed in the outcome of the survey from work package 2 should also happen to explore the feasibility of each suggestion to see if the GDCC can provide more added value and therefore attract more paying members and possibly increase its membership fees. All this in an effort to increase the funds necessary to stabilize and strengthen the funds of the GDCC. In parallel to this, the working group advises the GDCC to determine the core set of people and skills required to support the program priorities. This is necessary to develop a plan to gather the funding to provide for these people and skills. Once this is defined, it can be used to communicate and adjust the financial expectations of members to improve the overall sustainability.
A necessary note to add to this recommendation is the need for this financial model to allow for improved participation from users in developing economies to ensure equitability. As though the main focus should be to increase and stabilize the funding of the GDCC, this should not come at the cost of less resource-heavy institutions or installations to ensure it remains a true global community, as its name reflects.

5. *Diversify income streams for GDCC*

Category: Governance & Resources
Priority: high priority

Another recommendation attached to the critical catastrophes related to sudden loss of income, is this need for diversification in income streams. Currently the main income of the GDCC consists of its membership fees and individual ad-hoc consultancy income. As might be clear from the previously mentioned recommendations, there is also a significant need for human resources to make the GDCC more sustainable, and as part of this the working group strongly recommends the GDCC to explore funding opportunities to get temporary or long-term funding to provide in these resources, such as the NSF POSE program. On top of this exploration of funding opportunities via funder institutions, the exploration of attracting long-term dedicated resources is also necessary (though again, there are human resources necessary to execute this exploration). As part of the phase analysis, a suggestion is to explore the creation of a network of registered service providers that contribute to the monetary sustainability of the GDCC. This can in part be done by exploring partnerships with existing leading organizations in the Dataverse community or other open-source communities that might be able to provide steady sources of income and/or in-kind contributions.

6. *Establish a functional strategy for the Dataverse software*

Category: Technology
Priority: high priority

The sixth and seventh recommendations are the outcome of the critical catastrophe summarized as there being differing view on the way forward for Dataverse (e.g. creation of PRs for features that aren’t in line with what the Dataverse project software is intended for, or the community disagrees about future development of the software, leading to multiple spin-off forks).

A first step to mitigate this risk is to document a definition of the functional perimeter of the Dataverse software, as in; what should and shouldn't be or come into the Dataverse software. Alongside this, there is a need for a plan to evolve the technology in such a way that is stable and sustainable. To do so, the working group advises the community to carry out a core community needs assessment on a regular basis to understand how core community members' needs have evolved since the software project began and discuss any necessary adaptations stemming from this with the different stakeholders. In addition, a broader community consultation, including non-members and potential members, could be insightful.
As part of this functional perimeter, the working group also sees the need to establish and/or improve cohesive overall messaging, and policies that guide and foster engagement such as contributor guidelines, codes of conduct, onboarding policies, etc. As previously mentioned, these are currently available, though could be strengthened and made more transparent.
Once the perimeter and functional strategy has been defined and documented, it will need to be shared with the community to ensure everyone is aware of this and can be pointed to if there are certain features deemed inappropriate or not in-scope of becoming part of the Dataverse code.

7. Formalize and set up a decision-making process

Category: Technology
Priority: top priority

As mentioned in recommendation 6, this recommendation is meant to mitigate the risk of divergent Dataverse subgroups. In order to ensure the community stays on the same track as much as possible, it will be necessary to make the decision-making process more formalized. The working group also sees the need for the decision-making process to consciously pivot more towards a functionality focus on the end-users to avoid losing sight of the end-users due to the most prominent and active community members being developers and possibly losing this point of view at times.
The Dataverse is currently moving to a more modular set-up. With this more modular set-up, the working group advises the community to set-up mechanisms of consensus across the community for what the available "recommended" or "core" modules are. Perhaps in parallel with this, there is also a need for coordination of modules while in development, though the organic nature of certain innovative work shouldn't be nipped in the bud by this coordination.
A working point stemming from both the phase analysis and the catastrophizing exercise, is to make the decision-making more democratized and transparent, by establishing a formal and documented strategy for engaging the community in decision-making. The need for establishing and publishing policies for code contribution, technical road mapping, strategic planning, policy decision-making etc. is identified.

8. Make it easier to deploy and upgrade the application and integrations

Category: Technology
Priority: high priority

This eighth recommendation is based primarily on the critical catastrophes having to do with the loss of human resources at GDCC or IQSS and there being a sudden need of technical capacity for all Dataverse installations. A first way to decrease the risk of this critical

catastrophe is to provide Dataverse as a service, as this may diversify the income for GDCC, leading to more stable resources. On top of the added stability for the GDCC, the working group also suggests to make Dataverse or Dataverse as a service more commercially appealing by exploring the improvement of the software's/service's marketing to better compete with other data repository softwares currently available, e.g. in settings where organizations acquire repository services through competitive tenders/biddings. With this, there could be work done to expand the community involved in the program and to involve new users in program growth and development.
The working group also suggests that the community provide technical (e.g. CI/CD, containerized Dataverse) training and certification. Though, it was noted in the group that this kind of training could also be of interest to non-technical profiles such as admins or curators of a Dataverse instance. However, the focus in this recommendation is on lowering the threshold for new(er) installations.
While there already is a Dataverse containerization working group, a need to prioritize and possibly fund a production-ready Dataverse containerized set-up was identified as helping to implement this recommendation. The exploration of supplying long term support (LTS) versions could also offer more sustainability of Dataverse installations with a lack of technical staff or who work with minimal staff.
As mentioned in the Dataverse as a service paragraph, the Dataverse software could benefit from looking at its competitors. One of the items to perhaps improve is the release notes to ensure non-technical staff can also understand what's new in a release. This can be alongside the technical release notes as currently already being supplied in a well-documented and legible manner. With these more non-technical release notes, could also come a focus on how to configure certain features, media and use cases to show certain set-ups or features.
In general, the working group believes the Dataverse community could benefit from more engagement with participants of all backgrounds and skills (e.g. librarians and non-developers) to ensure the code continues to grow in the direction that the end-users would like. This could also lead to an increased engagement of community members early with testing and documentation, though it should be noted that the recently launched UX working group seems to be making head-way on this already.

9. Formalize and document the communication strategy for high-priority items

Category: Resources
Priority: high priority

This recommendation also stems from the catastrophizing exercise and points primarily to the need for the formalization of the communication strategy. There is currently already a set-up in place, but it's not very well-documented and uses the same list of email addresses as for example community surveys. This is possibly not an issue, though it has been noted that some of these email addresses are no longer functional, pointing to a possible risk of not being able to reach all registered Dataverse installations if a high security risk is signaled. On top of this, due to the open-source nature of the software, there are installations out there that aren't known to the community, so they would also potentially miss any crucial information. Two working points are to better manage the email subscriptions for high-priority communication and to make the security contact for Dataverse more visible in both GitHub and the documentation.

10. Help grow, manage and sustain a community-supported ecosystem of Dataverse-integrated tools, services, solutions, and add-ons

Category: Resources
Priority: high priority

This recommendation is related to the more modular future of Dataverse as well as the overall open-source character of the software. To ensure certain quality control and ensure sustainability of certain tools or features, the working group advises the community to establish workflows and methods to help support Dataverse-integrated tools, services, solutions and add-ons. This is not understood as having to support all parts of the ecosystem that are out there, but create and manage community-supported and community-driven implementation, such as for example a standardized French version that is available to be used by others and endorsed as the one supported by the community.

The following table provides an overview of the 10 key recommendations:

#	Recommendation	Category	Priority
1	Establish a new legal organization for GDCC	Governance	Top
2	Establish a formalized governance of GDCC	Governance	High
3	Clarify/demonstrate the value/benefits of the GDCC membership and strengthen the value exchange	Governance	Top
4	Formalize the financial model and fee contributions/membership model	Governance	High
5	Diversify income streams for GDCC	Governance, Resources	High
6	Establish a functional strategy for the Dataverse software	Technology	High
7	Formalize and set up a decision-making process	Technology	Top
8	Make it easier to deploy and upgrade the application and integrations	Technology	High
9	Formalize and document the communication strategy for high-priority items	Resources	High
10	Help grow, manage and sustain a community-supported ecosystem of Dataverse-integrated tools, services, solutions, and add-ons	Resources	High

6. Conclusion

The GDCC identified the need to address challenges the Dataverse community has been facing due to the continuous popularity and growth of the Dataverse software deployments across countries, domains, and types of organizations in 2023. This challenge had to be addressed to be able to continue to grow without introducing risks to the overall sustainability

of the Dataverse project. From this need, the Dataverse sustainability working group was created and decided an in-depth analysis of the current state of affairs was necessary to identify and find a path to sustainability for the Dataverse community. In this context, on the one hand, an It Takes a Village Analysis was performed, while on the other hand, options to generate additional monetary support for Dataverse sustainability by providing more added values in exchange for e.g. membership fees were explored. This exploration was done in the form of brainstorming sessions and a survey, which resulted in a series of added values being identified that could potentially strengthen the monetary income necessary for sustaining Dataverse. A next necessary step before implementation will be to analyze these added value propositions for their feasibility.

As found during the It Takes a Village Analysis of the Dataverse project performed by the working group, the Dataverse community currently has many strengths stemming from its informal nature. There are many building blocks already in place to build a strong foundation for sustainability, with the invaluable commitment to support the software and community by the Dataverse Project Team at IQSS at Harvard University playing a key role. The analysis indicates that the project features a solid level of technological maturity and characteristics of a well-established open source community, particularly in terms of community engagement. However, it also highlighted areas for improvement, mainly in formalizing and documenting processes and policies, from which the group proposed and prioritized ten recommendations.

In order to further work on the sustainability project, it is clear that more investment by human resources will be necessary. Some of this work cannot be done by a group of volunteering members of the community, as was done for the above-mentioned work due to the practical nature of some of the recommendations and the time investment required of said volunteers, though such a group can continue to support the work done where necessary. This working group highlights the need for dedicated human resources that can focus on drafting an action plan for each of the 10 recommendations. As was indicated in the value exchange survey, one of the most valued items on the list was to sustain the Dataverse software and associated services and tools. As such, this working group urges current installations to explore options to invest in the provision of such dedicated human resources and sustainability work necessary for the community and the software, as they are part of the community that would benefit from continued support for the years to come. It should be noted that the Dataverse community has benefited from significant support from many institutions. As the community grows, formalization of the mechanisms and incentives to contribute to sustainability will be critical to address these previously-mentioned recommendations.

7. References

Apache Software Foundation - How The ASF Works. (n.d.). Retrieved July 11, 2024, from https://apache.org/foundation/how-it-works/

Community Canvas. (n.d.). Retrieved July 12, 2024, from https://community-canvas.org/

Conzett, P. (2022). Dataverse Community Survey 2022 – Report. *Septentrio Reports, 1.* https://doi.org/10.7557/7.6872

Dataverse Project - About page. (n.d.). Retrieved July 10, 2024, from https://dataverse.org/about

Eclipse Foundation - About page. (n.d.). Retrieved July 11, 2024, from https://www.eclipse.org/org/foundation/

Educopia - Community Cultivation Resource Library. (n.d.). Retrieved July 11, 2024, from https://educopia.org/cultivation/

European Commission - Open Source Observatory (OSOR): Guidelines for creating sustainable open source communities. (n.d.). Retrieved July 11, 2024, from https://joinup.ec.europa.eu/collection/open-source-observatory-osor/guidelines-creating-sustainable-open-source-communities

GDCC - Members page. (n.d.). Retrieved July 10, 2024, from https://www.gdcc.io/members.html

It Takes A Village - ITAV: Open Source Software Sustainability. (n.d.). Retrieved July 14, 2024, from https://itav.lyrasis.org/

Lyrasis - Community Programs. (n.d.). Retrieved July 11, 2024, from https://lyrasis.org/community-programs/

Open Source Guide. (n.d.). Retrieved July 12, 2024, from https://opensource.guide/