



LYRASIS Research

LYRASIS Research Report

The Ohio State University Catalyst Fund Idea Report – Repository Migration Service

PREPARED BY:

Hannah Rosen

Strategist, Research and Scholarly Communication

May 1, 2022



Report Title: The Ohio State University Catalyst Fund Idea Report – Repository Migration Service

Project Abstract:

This report covers the 2021 LYRASIL Catalyst Fund idea proposed by the Ohio State University. LYRASIL staff performed market research, conducting interviews and surveys to determine how current OSS communities are tracking software version usage, as well as to understand the potential for a migration service to alleviate the difficulties in migrating from unsupported OSS versions to current software versions.

© 2022 LYRASIL. IN COPYRIGHT – EDUCATIONAL USE PERMITTED. This Item is protected by copyright and/or related rights. You are free to use this item in any way that is permitted by the copyright and related rights legislation that applies to your use. In addition, no permission is required from the rights-holder(s) for educational uses. For other uses, you need to obtain permission from the rights-holder(s).

This work is licensed under a Creative Commons Attribution 4.0 International License.
<http://creativecommons.org/licenses/by/4.0/>

To request permission to use this content for commercial purposes, please contact: erin.tripp@lyrasis.org

Date of publication [April 2022], LYRASIL Atlanta, Georgia www.lyrasis.org

ORCID iD [Author]: <https://orcid.org/0000-0001-6804-7073> [Hannah Rosen]

Preferred Citation: Rosen, Hannah. “The Ohio State University Catalyst Fund Idea Report – Repository Migration Service.” April 2022, 1-20. DOI: <https://doi.org/10.48609/x3ez-ql04>.

Table of Contents

| | |
|---|----|
| Executive Summary | 4 |
| Phase 1 – OSS Interviews | 5 |
| Technical Barriers | 5 |
| Voluntary Information | 6 |
| Manual Information..... | 8 |
| Preliminary Findings..... | 10 |
| Phase 2 – Fedora Community Survey | 5 |
| Survey Distribution | 5 |
| Demographics | 5 |
| DuraSpace Registry | 6 |
| Current Version Adoption and Migration Services | 8 |
| Conclusions and Recommendations | 17 |
| Acknowledgements..... | 17 |
| Appendix A: Survey Questions | 18 |

Executive Summary

In 2021, The Ohio State University Libraries proposed via the Catalyst Fund that LYRASIS staff perform market research to survey the landscape of open source software (OSS) repositories and determine the total market of users that are running outdated, unsupported repository software of all sorts. The goals of this research project were as follows:

- To better understand the percentage of users who are running outdated OSS repository technology
- To better understand the mechanisms OSS communities use to collect and monitor version usage
- To better understand the barriers users face in updating their repositories to newer versions of OSS software, specifically versions that require significant change and/or data migration
- To provide OSS communities with recommendations for how to better collect data about versions, how to provide more outreach concerning updates that require migrations, and how to ease transitions between versions to keep users from leaving OSS platforms

The first phase of the project began with interviews of various OSS communities. During the second phase, LYRASIS staff utilized insights from those interviews to create and distribute a survey specifically to Fedora community members. While the project started off with the intention of applying more broadly to many OSS communities, for a variety of reasons (which will be explained in more detail later on in the report), the final recommendations were tailored to apply specifically to the Fedora community hosted at LYRASIS.

Some insights from the overall project include:

- Collecting accurate and current usage information from software adopters has not been a priority for the OSS communities that participated in this project. Due to both technical and ideological reasons, current methods of collecting information are frequently decentralized and heavily dependent on self-reporting.
- There are opportunities to more actively track user information amongst OSS users, and specifically for Fedora users through the DuraSpace registry: based on the survey results, users are receptive to more active efforts to gather information from the Fedora community.
- Most survey respondents are not running the most current version of Fedora: they focus on version upgrades when they have resources available to do so, rather than when a new version is released.
- The majority of respondents did not see a migration service as the solution to making version migration easier. Due to the bespoke and/or integrated (as in part of a larger technology stack) nature of Fedora installation, instances are not easily migrated, either internally or by a 3rd party vendor.

It is our hope that the findings and recommendations from this report provide valuable assistance to the Fedora community to make informed strategic planning decisions in the future.

Phase 1

For the first stage of the OSU catalyst Fund Idea project, LYRASIS staff interviewed community managers from four different open source software (OSS) communities – DSpace, Fedora, Islandora, and Samvera – to determine how much data is gathered about version downloads, along with an assessment of the integrity of that data. LYRASIS staff also gathered additional data about how those communities advertise, facilitate and monitor new versions of their software, specifically focusing on major upgrades that require data migration. This information is crucial to understanding how users move between versions, what upgrades have roadblocks to implementation, and which institutions require assistance moving out of unsupported versions of their OSS.

After interviewing community managers from those four OSS communities, LYRASIS staff came to the preliminary conclusion that none of the communities have implemented a strategic plan for tracking usage of their software, let alone what versions are in production around the world – instead, each community relies on a combination of volunteer information and manual research to construct their general understanding of version adoption.

TECHNICAL BARRIERS

None of the communities interviewed currently gather user information (whether contact information, institution name or version implemented) through technical means. This is due in part to certain technological barriers: all four of the OSS communities store their code on GitHub, and therefore allow for direct download of the code, regardless of version, without gathering any information about the user. There is also a limit of usefulness to gathering information from downloads: a person could download the code easily, but it does not necessarily translate to a production system. Trying to gather information through some kind of form during the install has been floated within each of the communities as a way of gathering better information, but that would require more dedicated time from the tech team and is often deemed a lower priority. Tim Donohue from DSpace said, “On the technical side it’s easier to gather the information on the download than the install, but it’s more useful to gather information on the install.” Heather Greer Klein from Samvera believes that adding a tracking component would be a huge undertaking, first to get approved from the community, and then to build in the implementation.

While the technical barriers are not insignificant, the community managers indicated that a bigger barrier to technical tracking is a legacy of philosophical approaches to how open source is defined. Each of the managers indicated that there is hesitancy from parts of the community to requiring information upon implementing their software. There is a desire to remove all barriers to the software and make it open in the purest form, meaning while participation and financial contributions are appreciated, neither is required, and anonymity is part of the overall package. It is worth noting that all four of the community managers indicated that the question of gathering usage data through technical means has been brought up before in community conversations – however, the three older communities (DSpace, Islandora and Fedora), all of whom have gone through or are currently going through major migrations in the past, indicated that leaders within the community have more recently been pressing to re-examine this approach. This indicates to LYRASIS Research Staff that more mature communities are more

willing to change their philosophical approach if it allowed them to better aid users in transitions between versions.

VOLUNTARY INFORMATION

Instead of technical means, all of the communities gather contact and version information through a variety of voluntary and manual methods. The most useful tool the communities use to gather information is some form of a user registry. Both DSpace and Fedora encourage their users to input their version information into the DuraSpace Registry (<https://duraspace.org/registry/>). The registry collects contact information (not publicly listed) and institution, location and software/version information (publicly listed) from voluntary users, and then displays their materials on a map. Islandora has a simpler self-reporting registry (<https://islandora.ca/installations-map>) which also contains a map, identifies institutional uses and links out to publicly available instances. Samvera community managers maintain detailed Implementation pages on their wiki (<https://samvera.atlassian.net/wiki/spaces/samvera/pages/422319621/Samvera+Implementation+In-production>, <https://samvera.atlassian.net/wiki/spaces/samvera/pages/422320492/Samvera+Implementation+In-development>), which detail institutions, user contact information and software versions.

For all of these communities, the primary goal of these registries and implementation pages is not to track version usage for governance and/or migration purposes – instead, each of these user lists is a marketing and outreach tool. All of the communities see the lists as a tool for community building on multiple levels. The lists indicate widespread adoption of the software that could give new potential adopters confidence in joining the community and allow existing community to connect with neighbors. Community members can also use these lists to identify different forms of use cases for the software and/or or provide holistic information for presentations.

These lists are so geared towards outreach, in fact, that only two communities, DSpace and Islandora, indicated that they were used or incorporated into any sort of master list of contacts. DSpace and Islandora have spent more active time encouraging users to register their instances through voluntary forms or outreach initiatives. David Wilcox, former Fedora community manger, indicated that Fedora has not done a lot of active work to encourage entry into the registry, and has therefore seen less participation. Community managers from DSpace, Islandora and Fedora all admitted that the registries are not monitored or regularly updated – users are encouraged to volunteer updates, but none of the managers maintains any illusions that the contacts are up to date. For Samvera, the implementation pages are maintained by staff. Due to the partnership model that Samvera employs, active partners communicate heavily with the community, staff feel more comfortable with the accuracy of the information.

Other lists of users exist within each community, but they are often dispersed amongst their outreach communication channels. Each community appears to have multiple mailing lists, Google groups, and slack domains with multiple channels. These channels contain information, but are often not reliable. For example, many mailing lists contain a mixture of current and potential users. Even for more implementation-focused mailing lists, contact information exists, but is often decontextualized (e.g., personal Gmail addresses or anonymous institutional emails), and therefore not useful for tracking users. In a parallel track, version information is

often offered up in the forms of support questions or user groups in these channels, but none of the community managers indicated that it is purposefully recorded in any form of master list or central database.

MANUAL INFORMATION

To supplement their volunteered information, three of the communities – DSpace, Islandora and Fedora – have undertaken one-time manual investigations to better understand the scope of their user base and the landscape of versions in current adoption. According to Melissa Anez, former Islandora Community Manager, they were able to do a one-time scrape off the internet using a telltale URL (a URL that includes Islandora in the title) that gave them basic institution names and version control, and heavily supplemented the user registry. DSpace did a Google engine search a few years ago and got similar information, which they also used to populate their registry. Fedora sent out a survey during a past IMLS grant that garnered them a one-time snapshot of users. All of these efforts were extremely useful in collecting version information, but were time-intensive and therefore isolated events, not built into any ongoing workflows.

PRELIMINARY CONCLUSIONS

It is clear that collecting accurate and up to date version and contact information from current software users has not been a priority for the OSS communities that participated in the Phase I interviews. While all of them desire better information about adoption of their software, their current methods of collecting information are frequently scattered, manual, and heavily dependent on self-reporting. While this state of affairs is not ideal, the interviews point to several opportunities these communities have to leverage their existing tools to better understand their user base.

Concerted strategies could be developed around combining information from various communication channels, including but not limited to: more routinized monitoring of registries to maintain up to date information, purposefully recording and monitoring information gathered through outreach communication channels, strategically gathering information from the registries and communication channels into centralized databases, and regularly building in manual initiatives into strategic roadmaps. All of these potential initiatives could help OSS communities better understand version adoption across their communities, as well as identify users in danger of falling behind with unsupported version.

Phase 2

For the next phase of the project, surveying users, we decided to restrict our scope to internal LYRASIS programs, to manage costs and time constraints. In addition, the external communities we originally targeted in phase 1 – Islandora and Samvera – did not prove to be ideal candidates for continued surveying, either due to internal staffing issues or fitting within the scope of the survey.

The phase 2 survey was intended to reach both the Fedora and DSpace communities in order to better understand how they would feel about more active updates from the DuraSpace registry, as well as gauging interest for a paid migration service. However, it became clear that this plan conflicted with certain strategic directions the DSpace global community is currently pursuing. Therefore, we decided to only survey the Fedora community – it proved most relevant to the project and removed obstacles trying to coordinate with other communities.

SURVEY DISTRIBUTION

The survey was held between January 19th, 2022, and February 28th, 2022, for a total of six weeks. To avoid GDPR violations, the survey was distributed through the Fedora email list instead of the registry – despite asking questions specifically about the registry, we could not contact non-US Fedora users via the registry since they were never explicitly asked to provide permission to contact them when signing up. That issue, however, does not exist with the Fedora email lists. Announcements were sent out on the following community groups:

- fedora-community@googlegroups.com (745 members)
- scs-fedora-users@googlegroups.com (46 members)
- German-speaking-fedora-users@googlegroups.com (33 members)
- dc-fedora-users@googlegroups.com (84 members)

Weekly reminders were sent out to the groups, along with personalized invitations to invested community members. The survey also offered a gift card prize to two participants. Our goal was to reach 50-100 respondents, or roughly 5-10% of Google Group contacts. Despite our reminders and incentives, only 35 people responded to the survey. The low participation rate should be taken into consideration when analyzing the survey results – the results do represent a very small sample size. This process also indicates an overall difficulty in maintaining engagement with the Fedora community.

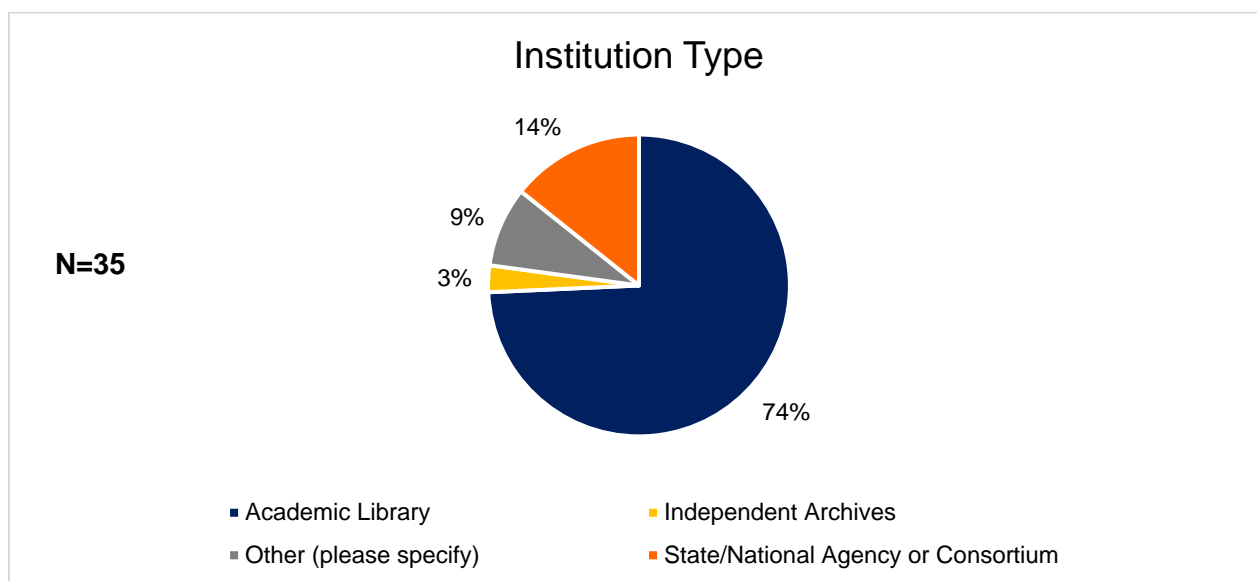
DEMOGRAPHICS

Question 1: Where is your institution located?

| Location | Count |
|---------------|-------|
| Europe | 11 |
| North America | 24 |

The majority of respondents, 69%, were from North America. This should be taken into consideration when looking at privacy components of any policy changes that result from this report – North American institutions are not required to comply with as many privacy regulations as European institutions, and are often more lax in their opinions on information sharing.

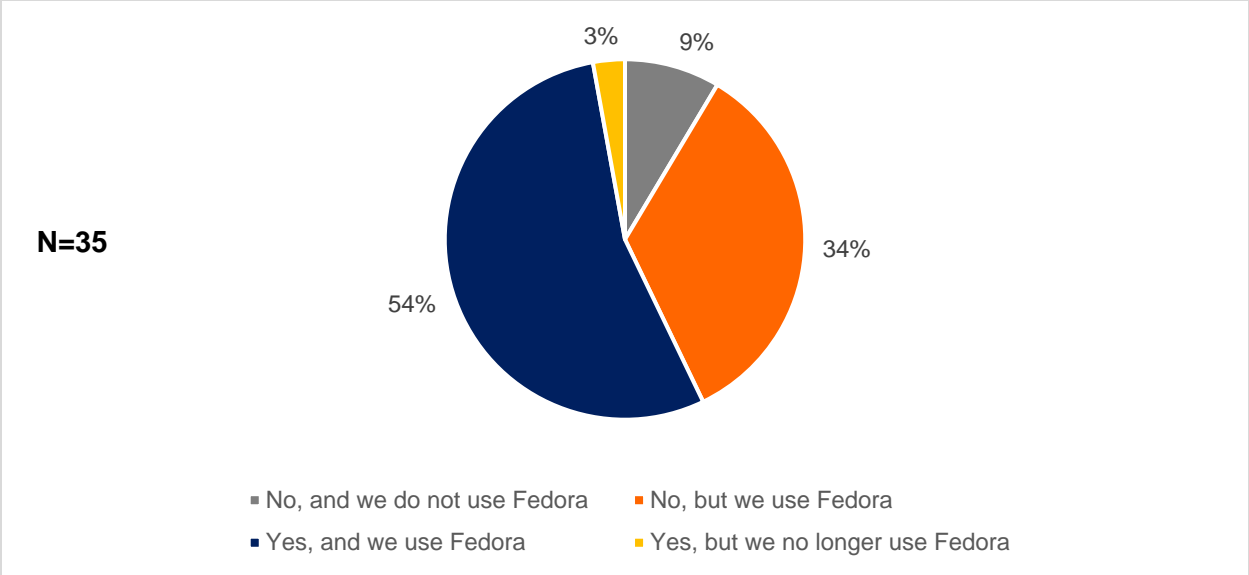
Question 2: What type of institution do you work for?



Almost 75% of respondents represented academic libraries, with the second largest group of constituents representing agencies or consortia.

DURASPACE REGISTRY

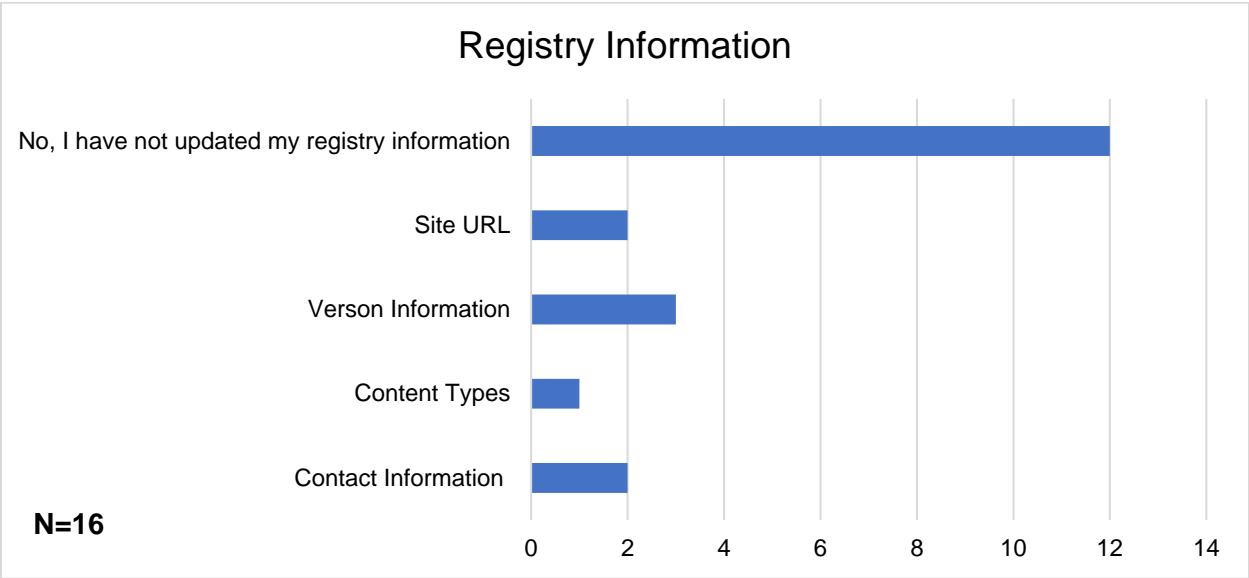
Question 3: Is your institution a member of the DuraSpace Registry? (<https://duraspace.org/registry/>)



This question was firstly intended to understand what percentage of Fedora community members are using the registry – roughly half of the Google Community group members who responded to this survey have registered in the DuraSpace registry.

This question was also intended to funnel participants in the survey to question specifically asking about usage of the registry. The following 5 questions were only directed at users who said yes to question 3.

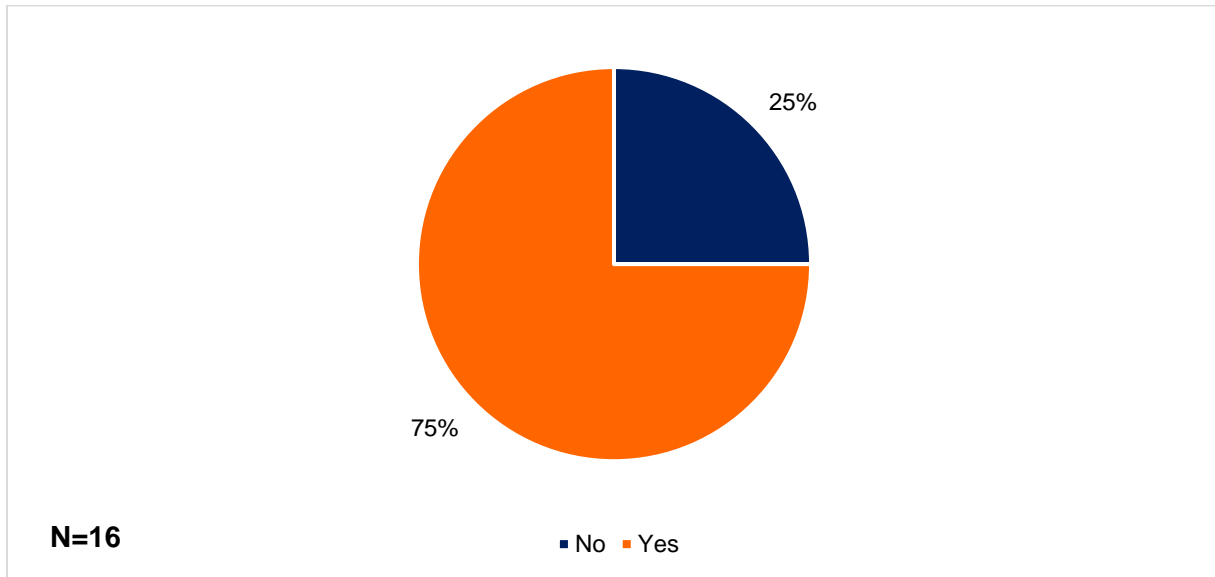
Question 4: Have you updated any of the following pieces of information on the registry since first entering it?



Of the 16 people who responded to this question, 75% said they have not updated their registry information. No more than 3 people had updated their version information or any other individual

registry components. This confirmed our hypothesis that while the registry is an invaluable tool for collecting initial version data, it is currently not useful to gauging version upgrades, since it is rarely updated.

Question 5: Do you see a benefit in being part of the registry?



75% of respondents to this question said they do find the Fedora registry useful. When asked if they could explain the benefits of the registry in more detail, a couple of respondents indicated the importance of demonstrating the strength and breadth of the Fedora community:

- *“I think it is a good idea to create such a registry which helps to build awareness and community around the different software projects.”*
- *“Helping to show the range and extent of Fedora adoption.”*

More people saw the registry as a good community building tool, to see peers and share knowledge about the software:

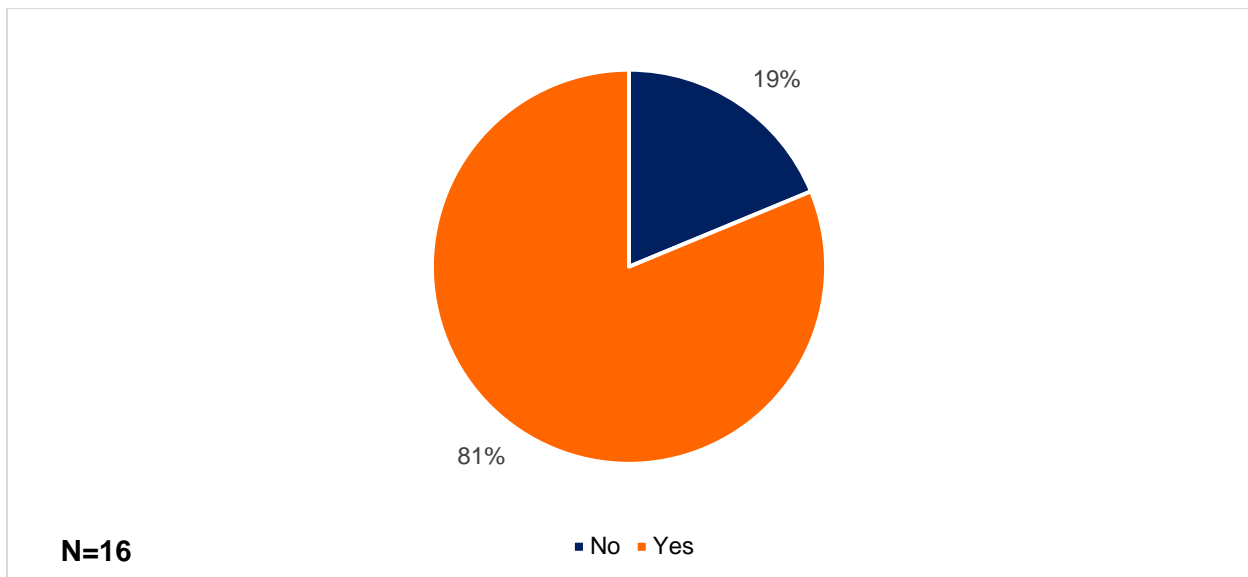
- *“Adding our profile to the list of institutions that use Fedora. I consult the registry regularly, to see how many, what kinds of institutions use Fedora, and where they are located.”*
- *“Provides an index of other peers and aspirants we can contact when need be.”*
- *“Visibility to other institutions interested in sharing knowledge about Fedora.”*
- *“A helpful tool that enables community contacts of other institutions using the Fedora software.”*

These explanations align with the benefits currently touted by Fedora leadership. With this clear demonstration of the benefits of the registry, we wanted to see if the benefits would provide justification for providing more, and more updated information that could help with internal recordkeeping.

Question 6: Would you feel comfortable receiving yearly reminders from the registry to update your information?

All sixteen respondents to this question said they would feel comfortable receiving yearly reminders. This response should encourage the Fedora community to actively ask its users to update their information, which will be crucial in understanding current version adoption.

Question 7: Would you feel comfortable receiving updates about new versions of Fedora through the contact information you provided in the registry?

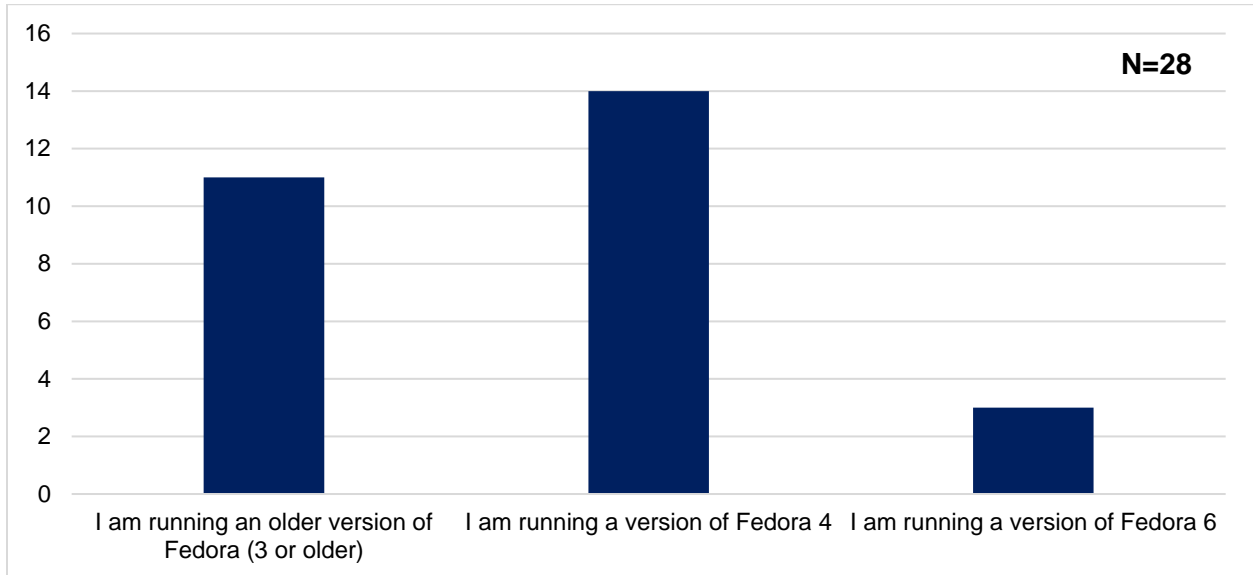


Unlike the previous question about updating institutional information, the responses to receiving version updates through the registry were not universal. They were, however, still overwhelmingly positive, with 81% of respondents saying they would feel comfortable receiving updates about new versions of Fedora through the registry. This response should encourage the Fedora community to actively consider using the registry as a tool for dissemination version information.

CURRENT VERSION ADOPTION AND MIGRATION SERVICES

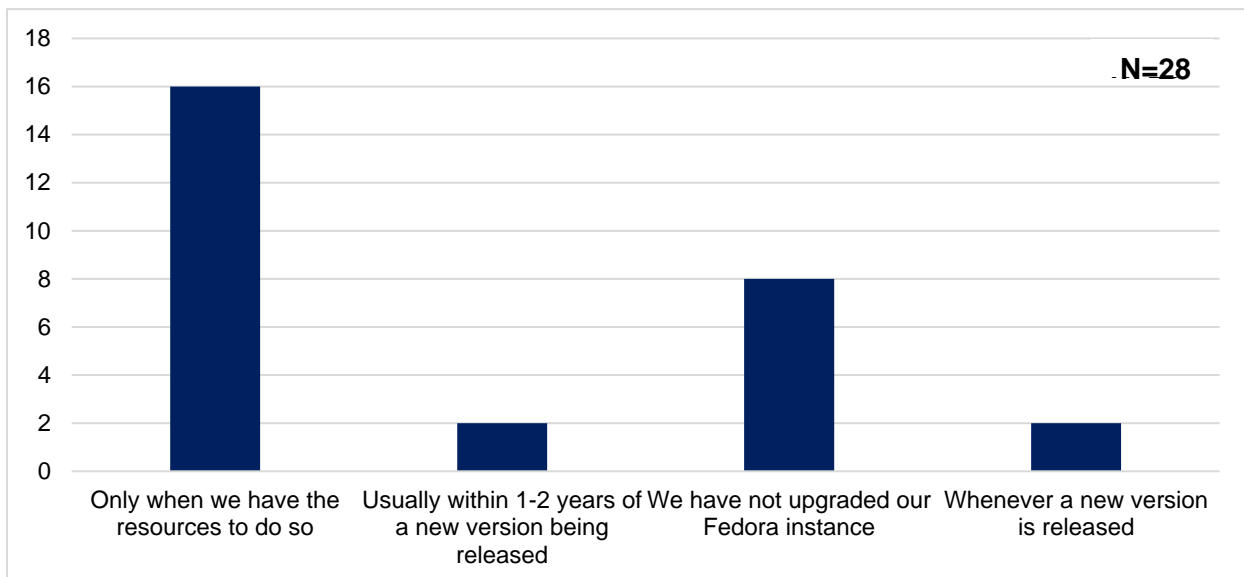
While the registry questions were targeted only to survey respondents who were already on the registry, the second half of the survey was open to all respondents who said they currently use Fedora. These questions were designed to get a rough picture of current version adoption, as well as gauge interest in a migration service that would facilitate version upgrades.

Question 8: What version of Fedora are you currently running?



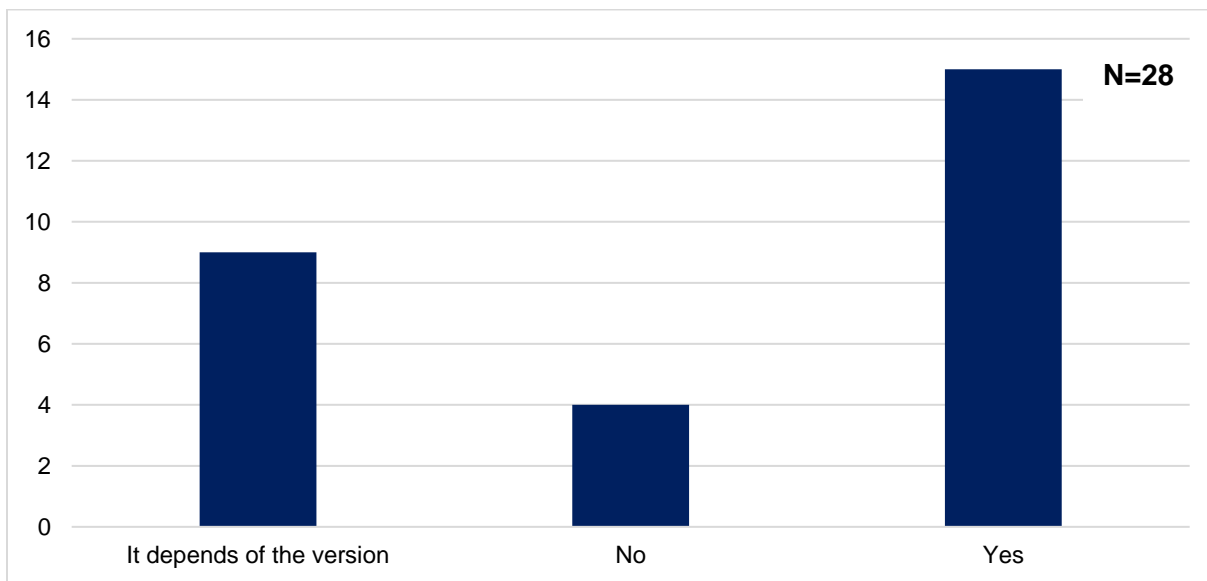
Roughly 40% of respondents are running a version of Fedora 3 or older, while 50% are running Fedora 4. No respondents are running Fedora 5, while the remaining 10% are running Fedora 6. This confirms what Fedora community leaders suspect, that the Fedora 4 migration was a significant enough technical change that it created barriers to adoption. At least according to this survey, it fairly split the community, with few people adopting newer versions.

Question 9: How often do you upgrade your Fedora instance?



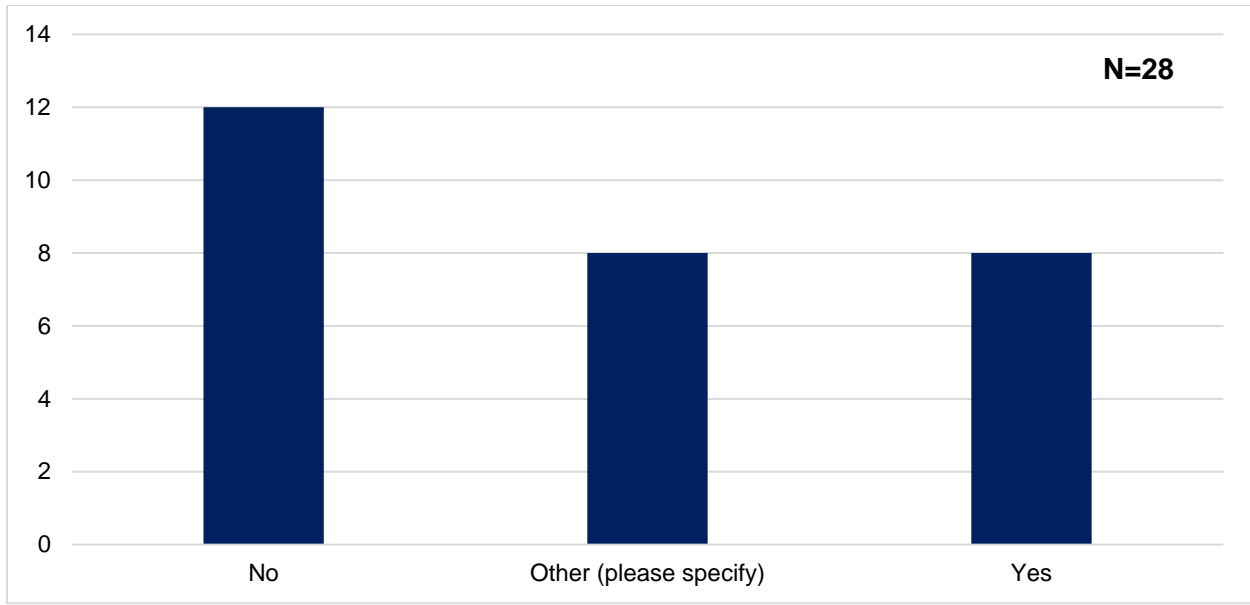
57% of respondents only upgrade their Fedora instance when they have the resources to do so. The second largest group of respondents, 28%, said they have not upgraded their Fedora instance at all. This reflects a disconnect between new version releases and adoption: adoption works on institutional, not community, timelines. The next question delved more deeply into barriers for adoption.

Question 10. Do you feel there is technical expertise available within your institution necessary to upgrade your software when newer versions are released?



53% of respondents said they do feel there is technical expertise available within their institution, while 32% said it depends on the version. This is surprising, based on the previous question's answers – resources are clearly an issue, but technical expertise is not necessarily a barrier in terms of lack of resources. This question could possibly be used to explore ways that Fedora version upgrades could be accompanied by tools and instructions that target other resource issues, such as staff time.

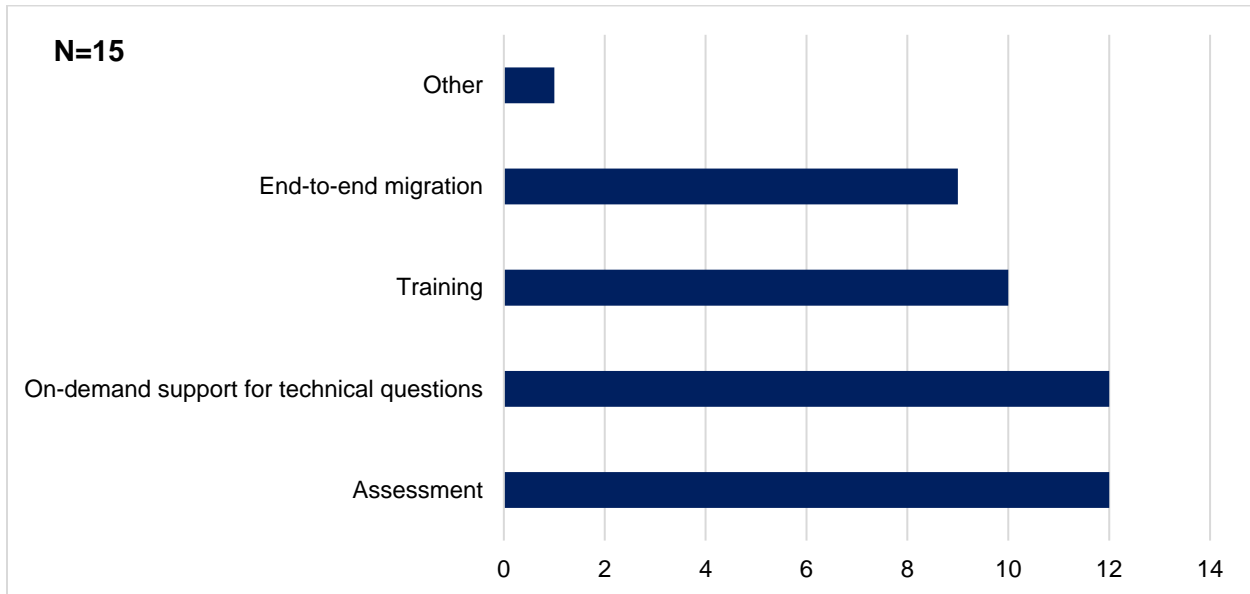
Question 10. Would you be interested in a paid service (administered by Fedora or LYRISIS) to have an outside organization assist you with version upgrades?



This question was one of the main questions of this entire project – could a migration service aid institutions with their Fedora upgrades. The results are decidedly mixed, leaning towards the negative. 42% of respondents said no, 29% said yes, and 28% said Other. A couple of respondents said it would have to be determined by their service provider, while one respondent said they preferred to hire a consultant. More tellingly, several comments under Other indicate that a migration service could not address the complexity of Fedora being implemented within a technology stack:

- *“Possibly, but such an upgrade would likely involve Samvera upgrades as well, so I’m not sure how feasible this would be for us.”*
- *“No, because our Fedora isn’t a stand-alone. It’s part of an ecosystem, and I don’t think we can manage Fedora separately from those components.”*
- *“Perhaps though the problem is not with upgrading Fedora, but rather with maintaining the entire stack. Our current dilemma is trying to get all of our content into F4 so we can move from 4 directly to 6. Getting out of F2 has been a challenge because we needed to build an application on top of F4 to replace the legacy application for F2.”*
- *“Not at this time but I can imagine this being very useful for people running older versions of Fedora or versions not tied to a frontend stack.”*

Question 11: If you said yes, what level of support would you be interested in paying for?



This question was designed to gain more detail about potential services for Fedora and/or LYRASIS to pursue. Unfortunately, this graph is not very illuminating. More respondents answered this question than respondents who indicated interest in a migration service. Also, the results are so close together that they do not demonstrate a clear preference for a single type of service to investigate.

Conclusions and Recommendations

While the sample size for this survey is admittedly small, the results still provide valuable insights for the Fedora team as they undertake long-term strategic planning.

From the beginning of this project, there have been two core challenges: understanding how to better gather information about Fedora usage from the community and understanding how to better assist Fedora users with version upgrades.

For the first challenge, both the interviews and the survey have shown a potential pathway forward in more proactively gathering current information about Fedora users: the DuraSpace registry. The registry has proven to be the most consistent information gathering tool currently in use at LYRISIS OSS communities (and, frankly, with others outside of LYRISIS), serving as a default centralized location for information. It is important to note that the registry has always been a repository of voluntary information: there has historically been concern about creating barriers to adoption and therefore hesitation to use it more actively as a communication tool and/or prompt Fedora users to update their information. However, the survey results demonstrate that there is buy in from respondents – there was overwhelming support for reminders to update entries and to contact users through the registry. It is fair to note that there are privacy concerns that would need to be addressed outside of the U.S., especially in seeking individual permissions to gather information in order to comply with GDPR in Europe, but that should not be a barrier to turning the registry into a more effective tool.

The second challenge, understanding how to better assist Fedora users with version upgrades, is more complicated. A migration service felt like a clear-cut answer to the current problem, but the survey results did not reflect user enthusiasm for this solution. Fedora is often integrated into other software and workflows, and therefore cannot be easily addressed with outside services. The real challenges to users, at least based on the survey results, are limited resources. Unfortunately, the survey did not elucidate what those resources are, although we do know that it does not necessarily mean technical expertise. To aid its community, Fedora governance will need to explore other ways to provide tools to its users that balance out their resource shortfalls.

Acknowledgments

The author of this report would like to acknowledge the contributions of all participants in the survey, who were primarily contacted through the Fedora Community Google Groups. I would also like to acknowledge all the OSS community leaders who participated in the Phase 1 interviews: Melissa Anez, Tim Donohue, Heather Greer Klein and David Wilcox.

Appendix A – Original Survey Questions

Thank you for participating in this Fedora Migration Survey. The goal of this survey is to understand how to improve our communication channels when new versions of Fedora are released, as well as to explore the appeal of various migration services. We ask that only one person per institution respond to the survey. The survey should take approximately 10-15 minutes to complete.

Registry

1. Where is your institution located?
 - a. North America
 - b. South America
 - c. Europe
 - d. Asia
 - e. Africa
 - f. Australia

2. Is your institution a member of the DuraSpace Registry?
(<https://duraspace.org/registry/>)
 - a. Yes, and we use Fedora
 - b. Yes, but we no longer use Fedora [lead to different logic]
 - c. No, but we use Fedora [skip to Version Migration]
 - d. No, and we do not use Fedora [skip to the end of the survey]

[If they said yes]

3. Have you updated any of the following pieces of information on the registry since first entering it? Please select all that apply.
 - a. Contact Information
 - b. Version Information
 - c. Content Types
 - d. Site URL
 - e. No, I have not updated my registry entry

4. Do you see a benefit in being part of the registry?
 - a. Yes
 - b. No

5. If you said yes, what do you see as the main benefit of being part of the registry? [Open-Ended]

6. Would you feel comfortable receiving yearly reminders from the registry to update your entry information?
 - a. Yes
 - b. No

7. Would you feel comfortable receiving updates about new versions of Fedora through the contact information you provided in the registry?
 - a. Yes

- b. No

Version Migration

8. What version of Fedora are you currently running?
 - a. I am running a version of Fedora 6
 - b. I am running a version of Fedora 5
 - c. I am running a version of Fedora 4
 - d. I am running an older version of Fedora (3 or earlier)

9. How often do you upgrade your Fedora instance?
 - a. Whenever a new version is released
 - b. Usually within 1-2 years of a new version being released
 - c. Only when we have the resources to do so
 - d. I don't upgrade my Fedora instance

10. Do you feel you have the technical expertise necessary to upgrade your software when newer versions are released?
 - a. Yes
 - b. No
 - c. It depends on the version

11. What would you say is the biggest internal challenge to upgrading your software when newer versions are released?
 - a. Limited staff time
 - b. Lack of technical expertise
 - c. Lack of institutional support for software hosting
 - d. Other (please specify)

12. Would you be interested in a paid service (administered by Fedora or LYRASIS) to have an outside organization assist you with version upgrades?
 - a. Yes
 - b. No
 - c. Other (please specify)

13. What level of support would you be interested in paying for? Please select all that apply.
 - a. Assessment
 - b. On-demand support for technical questions
 - c. Training
 - d. End-to-end migration
 - e. Other (please specify)

14. [If you said you no longer use Fedora in question 1], could you please tell us what made you decide to stop using the Fedora software? Please select all that apply.
 - a. We found proprietary software of a higher quality
 - b. We could no longer devote enough staff time to maintain the software
 - c. We did not have enough technical expertise to migrate to a newer version
 - d. We found running Fedora too expensive
 - e. We were not interested in the features that would be available in the new version
 - f. Other (please specify)