



# Mapping the Scholarly Communication Landscape 2019 Census

Katherine Skinner

EDUCOPIA  
INSTITUTE  
*Community Cultivators*

## Publication Notes

Title: Mapping the Scholarly Communication Landscape 2019 Census

Author: Katherine Skinner

Editors: Nathan Brown, David Lewis, Brianna Morrow, Mike Roy, Melanie Schlosser

Publisher: Educopia Institute, 235 Peachtree Street NE, Suite 400, Atlanta, GA 30303

Cover Image Credits: bharath g s, Unsplash.

Copyright: 2019

This publication is covered by the following Creative Commons License:

### **Attribution-NonCommercial 4.0 International**

You are free to copy, distribute, and display this work under the following conditions:



Attribution – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner. Specifically, you must state that the work was originally published as *Mapping the Scholarly Communication Landscape 2019 Census* and you must attribute the copyright holder as Educopia Institute.



Noncommercial – You may not use this work for commercial purposes.

Any of these conditions can be waived if you get permission from the copyright holder. Your fair use and other rights are in no way affected by the above.

The above is a human-readable summary of the full license, which is available at the following URL:

<https://creativecommons.org/licenses/by-nc/4.0/>

Suggested Citation: Skinner, Katherine. *Mapping the Scholarly Communication Landscape 2019 Census*. (Atlanta, Georgia: Educopia Institute, 2019).

## Table of Contents

Executive Summary	3
Introduction	6
Section 1: Methodology	10
Data Sources	10
Data Analysis	14
Section 2: Findings	14
A. General Overview	14
B. Participation Trends	15
C. Dataset Observations	16
D. Demographic Trends	17
E. Vision and Scoping	19
Mission, Vision, Values	20
Strategic Plans	21
Market Analysis	22
F. Technical Development and Design	23
Codebase Management and Contributions	23
Hosting and SCRs	24
SCR Updates, Licenses, Standards, and Interoperability	25
Development Process	26
Migration and Preservation	27
G. Administrative and Financial	28
Staffing	28
Financial and HR Documentation	29
Revenues and Expenses	31
H. Community Engagement	33
I. Governance	34
Section 3: Recommendations	35
Appendices	40

## Executive Summary

This report documents the design, methods, results, and recommendations of the 2019 Census of Scholarly Communication Infrastructure Providers (SCIP), a Census produced by the “Mapping the Scholarly Communication Infrastructure” project team (Andrew W. Mellon Foundation; Middlebury College, 2018-19). The SCIP Census was created to document key components comprising the organizational, business, and technical apparatuses of a broad range of Scholarly Communication Resources (SCRs) – the tools, services, and systems that are instrumental to the publishing and distribution of the scholarly record.

Using [Community Cultivation – A Field Guide](#) (Educopia, 2018) as a framework, we designed a Conceptual Model detailing the impact and outcomes the SCIP Census would address. We then produced and tested a survey instrument with 123 questions that delves into an SCR’s mission, vision, and scoping; technical development and design; administrative and financial scaffolding; community engagement activities; and governance model. The instrument took between 1-3.5 hours for each SCR respondent to complete; variability in time was largely based on the structure, complexity, and availability of an SCR’s organizational, fiscal, and technical information.

We conducted the Census through direct invitations, contacting just over 200 identified scholarly communication resource providers by email to participate. The Census remained open for a condensed, month-long collection period (February 18-March 22, 2019). More than 60 SCRs responded to us during this period, and more than 40 tools, services, and platforms ultimately participated in the Census.

Our team also researched basic information about 96 additional SCRs, creating a Composite dataset that combined this researched data with a few fields of the respondents’ anonymized data from the Census. This Composite dataset provides a system-level view of the broad range of SCR tools, services, and platforms in use today, including their purposes, founding dates, locations, and other basic information that could be quickly compiled by our team. It complements the deeper information about the technical, fiscal, and organizational mechanisms of SCRs today that the Census dataset provides.

The Census and Composite datasets provide a crucial lens through which we can now begin to do three things: 1) increase understanding of the range of forms, functions, structures, and models represented by SCRs across our system today; 2) formally assess some of the factors that influence the sustainability and “fit-for-purpose” of SCRs, and 3) identify concrete tasks and activities that specific SCRs might engage in to improve their stability over time.

Our findings include the following, each of which is elaborated upon in the report:

- We need a standardized taxonomy for the various functions performed by SCRs. It is currently difficult to differentiate between the broad range of functions offered by SCRs. It is also challenging to understand which steps are common in scholarly communications and publishing workflows, and what SCR choices might work for each of these steps.

- SCRs operating within nonprofit and hosted environments report ongoing challenges in raising and sustaining appropriate levels of funding to enable them to build and maintain services over time. These SCRs need additional support if they are to be viable options for institutional use.
- Connected to the above, sunsetting in our scholarly communication technical environment is often considered a sign of failure. Instead, we need to welcome it as a sign of a healthy overall environment. We also need to further explore the value of mergers, migrations, and other mechanisms that may provide the necessary administrative, fiscal, and social infrastructure to help support the technical development and maintenance SCRs require. Scaled, leveraged efficiencies (e.g., multiple programs hosted by a single entity with shared leadership and staffing) may help to bring needed expertise while also maintaining a lower overhead.
- SCRs need guidance, mentorship, training, and opportunities to refine their visions, technical platforms and design, financial and HR models, community engagement and outreach practices, and governance frameworks, as well as the decision-making processes that undergird each of these elements. This need applies particularly to several key areas of development:
  - Vision and Strategy. The Census evidenced that many SCRs lack clarity in their expressions of their purposes and goals. This is quickly mendable through specific, targeted investments in business practices that are well understood and documented across a wide variety of fields.
  - Technical Development and Design. Findings that stood out included the high variability in the number and type of software developers that currently participate in SCRs and the challenges to code contribution that exist in some environments, including Open Source Software projects and programs.
  - Financial and Staffing. Of all of the areas of concern that have been highlighted in this report, none is more compelling than the financial self-descriptions provided by respondents. Many SCRs report that they have low-to-no financial reserves. Most also do not reconcile their books on a regular schedule, and most lack the basic checks and balances that keep businesses safe from both accidental and purposeful financial misreporting.
  - Community Engagement and Governance. Deeper evaluation into current community engagement and governance strategies is needed at an individual SCR-level, but the collated and aggregated results from the Census show that most SCRs are engaging in a range of community-building activities and all responding SCRs prioritize in-person events as one part of their approach. We must work harder to ensure that governance bodies regularly evaluate the financial health of the organizations they are empowered to serve, and that external structures help to train both these Boards and staff members to do functions (e.g., accounting for revenues, not just expenditures) that simply are not business-as-usual within most academic environments.

This report begins with an introduction describing the motivation and rationale behind this research. It defines what we mean by “scholarly communication infrastructure” and “Scholarly Communication

Resource” and describes the overall goals, not just of this initial project effort, but of the broader trajectory that we are undertaking in the “Mapping the Scholarly Communication Infrastructure” project.

Our methodology is then described in detail, including our data sources and data framework. It provides an analysis of the data gathered to date and points to a series of data visualizations produced by Data Researchers Nathan Brown and Brianna Morrow (TrueBearing Consulting) that can be adjusted and controlled by users to see different views of the anonymized data and to answer different questions using the data.

The remainder of the report documents our findings to date and our recommendations for a larger and ongoing effort to assess the stability of scholarly communication infrastructure components, including recommendations for concrete actions to strengthen and ultimately enhance the sustainability of the infrastructure upon which we increasingly depend. The report closes with suggestions about next steps that a range of prospective partners and affiliates might undertake together in the future.

Educopia Institute and TrueBearing Consulting greatly appreciate the opportunity to conduct this research on behalf of Middlebury College and the “Mapping the Scholarly Communication Infrastructure” team, and we look forward to our future involvement in the next phases of work.

## Introduction

Since the advent of the World Wide Web in the 1990s, the production and dissemination of scholarship has transformed from almost entirely print-based to predominantly digital-based environments and workflows. Facilitating this rapid, three-decade shift in production, a wide range of publishers, distributors, libraries, research centers, and IT divisions have spawned an even wider range of tools, services, and platforms to support digital scholarship in its myriad forms. These technical components, now ubiquitous in scholarly publishing, have developed and evolved in an era of high opportunity, low coordination, and minimal standardization.

The technologies upon which these tools, services, and platforms rely (including software codebases, operating systems, hardware, and other components) become outmoded rapidly, due in large part to the unrelenting pace of industry-based development. To simply remain viable year-by-year, most digital components require significant investments in maintenance, updating, and integration. Many scholars, librarians, and publishers have built tools during landscape shifts that render those tools obsolete before coding is complete. Others have produced tools, built user communities, and celebrated success...only to find that *maintaining* the tools and user communities is more challenging and expensive than creating them.

Project directors often garner ample support while creating new tools (in terms of funding sources, community energy, publicity, and institutional backing). This support significantly wanes for most when they seek to sustain those same tools. A well-documented “Valley of Death” stretches between soft-funded projects and sustainable programs. Without deep knowledge of how to build a support community, and how to manage such non-technical elements as finances, communications, engagement, and governance, most project directors in our field find the bridge between grant funding and ongoing operational funding very difficult to cross.

The net result is that many scholarly communication tools and services wither, not due to shortfalls in demand or shortcomings in the products, but rather due to a lack of attention to and “know-how” in organization, community, and market building.

## Survival Skills

Even in our relatively unpredictable early-digital arena, a small number of tools, services, and platforms have achieved ongoing operational stability against challenging odds. A few of these have now carried forward for more than two decades. What can we learn from these longer-lived tools, services, and platforms? What models have they used for technical development, governance, fundraising, and community engagement? Are there factors and characteristics in these that reliably predict the successes or failures of our scholarly communication innovations? If we knew more about the business and technical models used by today’s scholarly communication tools, services, and platforms, would it be possible to integrate these into a stronger, more stable infrastructure?

These questions are not new; indeed, they have spawned many research projects and programs over the years, including the *Sustainability Implementation Toolkit* (Nancy Maron, 2014), *It Takes a Village* (Gemmill-Arp et al, 2018), *Community Cultivation – A Field Guide* (Educopia, 2018), The Socio-Technical

Sustainability Roadmap (Visual Media Workshop at the University of Pittsburgh, 2018), and Building Financial Resilience in the Digital Infrastructure for Research and Teaching in the Humanities (Nonprofit Finance Fund, 2019).

Similar questions have also led to a plethora of “environmental maps” and “landscape scans” and “workflow diagrams” that have tried to document different portions of the scholarly communication environment, including Bianca Kramer and Joroen Bosman’s 2018 “Mapping the open science ecosystem: Looking at tools and platforms,” Brian Lavoie *et al*’s 2014 “The Evolving Scholarly Record,” John Maxwell’s 2019 “OSS Landscape Scan,” Alejandro Posada and George Chen’s 2017 “Activities across the research workflow,” and Herbert Van de Sompel’s 2004 “Rethinking Scholarly Communication.”

As evidenced by these and other mapping exercises undertaken over the past few years, scholarly communication infrastructure (tools, services, platforms) are incredibly hard to track, identify, compare, or understand *en masse*. Many stakeholders strive to build coherence and fuse these often isolated elements into a system, beginning with improving documentation about what components exist, what they do, and how (and whether) they interoperate.<sup>1</sup>

Our project team evaluated a broad range of these sources in September 2018, in order to better understand existing research and sources, and also to help frame our own questions about today’s scholarly communication infrastructure providers—and about what we considered “in scope” for the Census.<sup>2</sup>

### Defining Scholarly Communication Resources

Determining what elements were in and out of scope for the Census was a process that included weeks of debate and refinement between the project team and Advisory Board. We acknowledge that the result is still imperfect, and itself is indicative of the need for a stronger, shared, and eventually standardized taxonomy to categorize scholarly communication tools and platforms.

Much of our conversation hinged on how best to differentiate between existing, “blurry” terms used in the scholarly communication realm and establish clear scoping boundaries for “Scholarly Communication Resources” (SCRs). The project team determined that for the purposes of this study, SCRs would be defined as organized groups building, offering, or using tools, platforms, and services in ways that enable active engagement and participation in the scholarly communication and publishing process. We intentionally included SCRs with both open and proprietary software code bases, for profit and nonprofit orientation, and a range of implementation models. Given the library-based project team’s desire to better understand the landscape of tools, platforms, and services with which they interact, we further defined SCRs as tools, platforms, and services that enable the library to actively engage, not as a user or conduit to users, but also as a contributor or conduit to contributions.

---

<sup>1</sup> See e.g., the Invest in Open Infrastructure (IOI) initiative, the Joint Roadmap for Open Science Tools (JROST), and the Open Platform Initiative as just a few of the more recent examples.

<sup>2</sup> For more information about this process, please see David Lewis, “Scholarly Communication Resources: A Literature Review” (forthcoming, 2019).



Again, we recognize the limitations in this definition. There is not yet a satisfactory taxonomy for the tools, platforms, and services that comprise our growing scholarly communication “infrastructure,” and our project team was not charged with solving that larger issue; we do mark it as a challenge that needs attention, and we also mark that this unsolved challenge made our own work difficult.

For an example, take our team’s discussion of four well known tools, platforms, and services: Public Knowledge Project’s Open Journal Software (OJS), HathiTrust, Directory of Open Access Journals (DOAJ), and JSTOR.

- OJS is a platform used to produce and publish content;
- HathiTrust is a hybrid that includes both content (objects) and a platform through which content can be stored, accessed, and preserved;
- Directory of Open Access Journals (DOAJ) is a service that indexes Open Access Journals that meet specific criteria (infrastructure);
- JSTOR is a platform that provides content that already exists and is used to conduct research (content).

In the context of this study, the first three are in scope; the fourth is not. The distinction here rests in part on what the library’s role is in each. For OJS and HathiTrust, a library can have an active role (e.g., contributing content to HathiTrust or using OJS to publish a journal). The library can use the tool, platform, or service to produce or include an output. Similarly, in DOAJ, libraries play an active role in contributing submissions (and some also serve as reviewers in this environment) for journals that meet the criteria. Libraries do not have that “contributor” capacity when they subscribe to JSTOR; instead, a JSTOR subscription is a licensing arrangement whereby the library and its users pay to receive content.

In other words, when our project team invokes the term “tools, services, and platforms” above, we are intentionally describing a subset of the whole, with which libraries can actively engage. In our work, we have tried to differentiate that infrastructure from the *content* that we know tools, platforms, and services help to produce, disseminate, and preserve. Of course, most SCRs participate in a range of functions, as evidenced by the examples above. If an SCR engaged in functions that were in scope for the Census, we did not eliminate them for also participating in out-of-scope functions (e.g., content creation).

As our team built the Census, we used the following documentation to guide and scope the inclusion/exclusion of tools, services, and platforms:

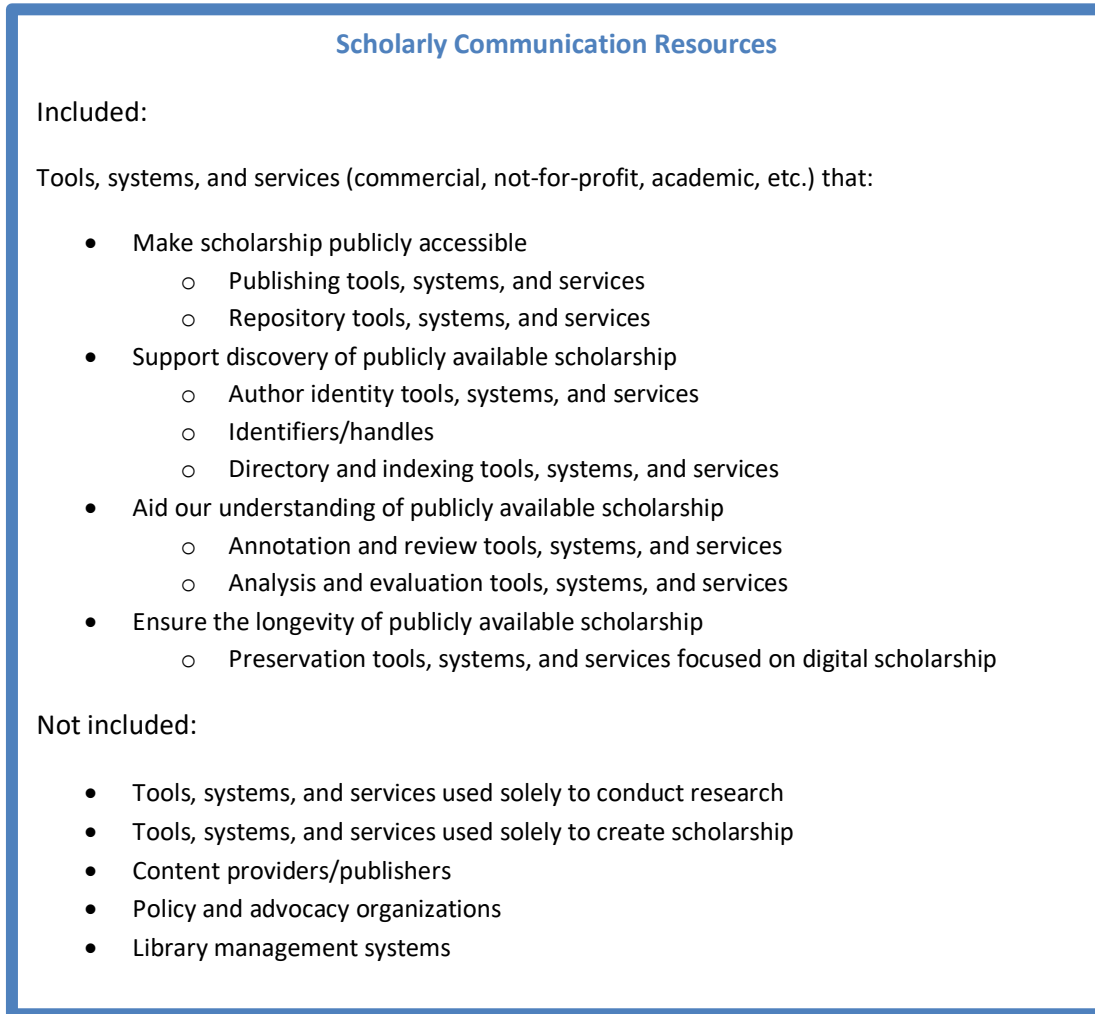


Figure 1: Scholarly Communication Resources Scoping

### Surfacing Hidden Models

With the SCIP Census, we have launched a field-wide experiment in how to surface elements of the very real, but often hidden models of operation that undergird today's scholarly communications infrastructure, element by element. In this experiment, we are relying on voluntary self-reporting by directors of SCRs about sensitive details concerning their business models and their financial histories. They have been incentivized to share this information by several key conditions:

1. Their data will be included in a systemwide view that will inform the field about the average, mean, and range of models and practices underway.
2. They will be able to compare their own data to that average, mean, and range of models and identify potential improvement areas based on this information.
3. Their data's anonymity will be protected as only the *aggregated* data will be analyzed, reported on, and offered as an open, anonymized dataset.

## SCIP Census in Context

The SCIP Census intends to provide a more transparent view into the technical, organizational, and financial capacity of scholarly communication tools, platforms, and services, both as individual (anonymized) elements and as a system. This information will help us identify ways to improve the resilience of individual tools, services, and platforms, and the system as a whole. It will also inform and inspire collaborations that will work towards these ends. We hope that this SCR landscape work could support the development of social norms and standards among various constituencies that may help to protect the interests of universities, scholars, and the many research communities and publics that they serve.

This Census is the first deliverable in the broader “[Mapping the Scholarly Communication Infrastructure](#)” project (Andrew W. Mellon Foundation, 2018-20) that seeks to understand the current level of investment in the infrastructure necessary for digital scholarship, to document stakeholder attitudes about these sorts of investments, and to identify promising strategies for encouraging greater investment by colleges and universities.

To promote and inform future work, this synthesis of the project’s findings and recommendations includes 1) details about the research methodology and instruments; 2) observations based on the methods we have deployed, 3) analysis of the data we have gathered to date, and 4) priorities for further data collection and for ongoing use of the data.

We greatly appreciate the Andrew W. Mellon Foundation’s support of this project.

## Section 1: Methodology

### DATA SOURCES

The SCIP Census has employed multiple data collection methods, including a survey and web-based research to surface existing documentation. This section provides a brief overview of the scope, focus, and general information gathered by each. For the full SCIP Census survey instrument please see Appendix A; for the abridged instrument used for web-based research and the Composite dataset, please see Appendix B.

### Documentation Review

The project team evaluated a range of relevant documentation, including bibliographies, curated lists, landscape/environmental scans, scholarly publications, and the Educopia Institute’s [Community Cultivation – A Field Guide](#). We used these sources to inform the construction of the project’s data model. This research also helped us to better understand what a range of important stakeholders, including librarians, archivists, publishers, editors, and scholars have documented regarding the scholarly communication ecosystem.

## Data Model

We invested the first few months of project work in documenting and refining our data model – an essential foundation for the success, not just for this project, but also to future endeavors that seek to document and assess factors influencing the success of a “Scholarly Communication Resource” or an infrastructure element in the scholarly communication ecosystem.

Our data model (see Appendix C) uses the framework provided by [Community Cultivation - A Field Guide](#) (Educupia Institute, 2018) as its foundation. This publication synthesizes more than a dozen years of Educupia’s experience with and study of community formation and evolution, drawing upon a range of relevant disciplinary approaches (e.g., social movement theory, sociology, organizational psychology, business) and theorist perspectives (e.g., Paul DiMaggio, Pierre Bourdieu, Elinor Ostrom, Yochai Benkler, Karl E. Weick, Glen Carroll, Michele Lamont, Mancur Olson, and Keith Provan).

At the core of *Community Cultivation – A Field Guide* is a model and framework designed for use in understanding, assessing, and guiding community development and maturation. The model identifies five growth areas (vision, infrastructure, finances and HR,

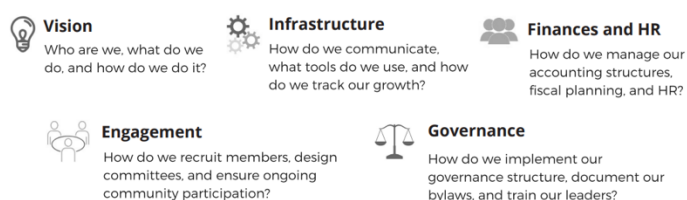


Figure 2: Community Cultivation Growth Areas (Educupia 2018)

engagement, and governance—see Figure 2) to which communities need to attend as they form and evolve. It then tracks growth markers for each of these five growth areas across four lifecycle stages of community evolution (Formation, Validation, Acceleration, Transition – see Figure 3). For each growth area and each lifecycle stage, there are specific activities and tools that communities can engage in or deploy in order to foster their own growth, stability, and resilience over time.

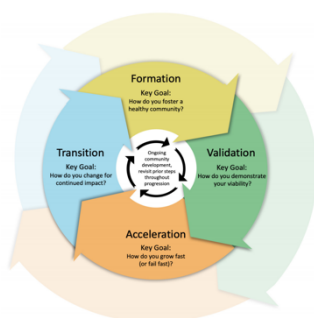


Figure 3: Community Cultivation Lifecycle Stages (Educupia 2018)

Our project team used this model extensively, culling important fields and establishing/defining a common set of inputs and outcomes based on its framework. We found that our work intersected with and complimented the work of the [Joint Roadmap for Open Science Tools](#) (JROST), and during the development process, we combined our efforts to build a single Census instrument that has served both this project and the emerging [Invest in Open Infrastructure](#) (IOI) network.

The resulting data model was then vetted by two “Mapping the Scholarly Communication Infrastructure” project advisory committees, and also by the IOI and JROST teams, and edited according to their guidance.

We sought to strike a balance between gathering enough data to understand and analyze the landscape, but not so much that it would deter participation. The data model we finalized in January 2019 contains 283 fields that focus primarily on documenting information about each responding Scholarly Communication Resource’s vision; technical infrastructure and design; administrative, finance, and

human resources structures; community engagement; and governance. The data model is available as Appendix C of this report.

## Data Policy

As we finalized the data model, we also formalized and documented our data policies for the instrument, as shared below in Figure 4.

**Data Privacy**

Individual and aggregate data and documents shared by respondents will only be used for the following purposes and under the following circumstances.

**All data:**

- Analysis by the “Mapping the Scholarly Infrastructure” (<https://scholarlycommons.net/map-plan/>) project team and the Joint Roadmap for Open Science Tools (JROST) (<https://jrost.org/>) project team.
- Delivery to the respondent of their own SCR’s data output (including a dashboard view of that data)
- Aggregated data will be analyzed, reported on, and offered as an open, anonymized dataset for reuse (no individual SCR’s data will be identifiable)
- SCR names of all responding SCRs will be collated and included with the dataset; individual respondent names will not be shared

**Data for which additional permissions are granted:**

Respondents can explicitly grant permissions to the research team to do the following:

- Share their individual SCR’s response publicly
- Share their individual SCR’s response to specific other groups conducting related surveys: Global Sustainability Coalition for Open Science Services (SCOSS), OSS Landscape Scan effort (MIT/SFU)

As per our data policies, we will not share your responses with any other group without your explicit, written permission

Figure 4: SCIP Census Data Privacy Statement

## Survey

The SCIP Census instrument was developed as an online survey (launched on the SurveyMonkey platform). This census instrument was developed by Katherine Skinner, Melanie Schlosser, Nathan Brown, Mike Roy, David Lewis, Dan Whaley, and Brianna Morrow. We are grateful to all those who contributed to and helped to improve this survey, including: Raym Crow, Christina Drummond, Heather Joseph,

Vanessa Proudman, Jessica Meyerson, Pierre Mounier, Kristen Ratan, and Danielle Robinson. Please see Appendix A: SCIP Census Instrument.

The survey included binary, open text, multiple choice, and Likert questions; it also provided prompts to upload files or share web addresses for relevant documentation. Based on the survey logic, respondents were prompted with up to 123 questions. Instructions to respondents specified that they could stop and resume the survey at will, and we provided respondents with a link to a PDF of the full survey to help them gather information. Once a survey was completed, survey responses could not be revised.

Questions within this survey were designed with the expectation that the respondent would be a SCR manager, either in part or wholly responsible for directing and managing the SCR. Each survey response corresponded to one SCR, and institutions that host multiple SCRs were expected to complete the survey multiple times, once for each SCR they had (e.g., DuraSpace Foundation would be asked to contribute separate entries for DSpace, DuraCloud, Fedora, and Samvera). Respondents were asked for a range of demographic, business, technical, financial, and governance details, including information about the SCR's legal structure and tax status, what parts of the research lifecycle the SCR serves, what documentation and processes the SCR has created and maintained, what stakeholders work with the SCR, what its technology environment looks like, how code is created and contributed, and how the SCR is funded. Respondents were also asked for their names, roles, and email addresses where they could be reached.

The survey was launched in SurveyMonkey and an invitation to participate was sent to more than 150 individuals identified as the director/manager of an SCR. The invitation specified that the survey was intended as a collection mechanism for a project seeking to assess factors influencing the sustainability and “fit-for-purpose” of Scholarly Communication Resources (SCRs)—tools, services, and systems in order to guide development of and investments in scholarly communication infrastructures. The invitation also specified that the survey would take between 1.5-3 hours to complete, including time spent researching and/or asking questions of other members of the SCR team. The survey instrument was officially open by invitation only from February 18-March 22, 2019.

A total of 39 individuals completed 43 surveys. Another twelve surveys were started by seven additional respondents, but not completed, and these were eliminated from the dataset accordingly.

## WEB-BASED RESEARCH

Our project team supplemented these full survey responses through gathering data via web-based research to fill in an abridged version of the SCIP Census instrument containing 11 questions and 48 fields (see Appendix B: Abridged instrument). This data collection effort was limited to 96 non-responding SCRs (to whom invitations were sent, but from whom we received no response in the four-week response period). This information was collected by PI Mike Roy (Middlebury) in March-May 2019. This data is maintained separately from the SCIP Census Respondent dataset.

## DATA ANALYSIS

The research team has employed qualitative data and quantitative, descriptive statistics as its primary analytical tools. This approach has provided us with a rich, comprehensive view of the data gathered in the project; it has also enabled us to begin to raise a range of questions.

Data analysis of the datasets was conducted by Educopia Institute and TrueBearing Consulting in close partnership with the Principal Investigators, Mike Roy and David Lewis.

Our research yielded two core datasets: 1) the SCIP Census Dataset, based on the answers provided directly by SCIP Census respondents, and 2) the Composite Dataset, based on both the respondent answers and data gathered by the project team about additional, non-responding SCRs from web-based research. All data collected by our team was normalized by Katherine Skinner and Mike Roy to address inconsistencies and data entry errors.

Data visualizations have been used to highlight important findings for the two major datasets built in this project. A set of Tableau dashboards created by TrueBearing Consulting is referenced throughout our findings, and is available here:

- [SCIP Census Dataset](#)
- [Composite Dataset](#)

These dashboards can be queried directly by the viewer to explore questions at will. They are designed such that a user can generate mappings and charts about different parts of the datasets.

## Section 2: Findings

### A. GENERAL OVERVIEW

One of the most important findings of this project is that many SCRs have now demonstrated both the capacity and desire to contribute information to a field-wide study of SCRs. A wide range of SCRs, including ones from a variety of tax and legal statuses, willingly contributed information about their business operations, revenue sources and levels, technical roadmaps and releases, and strategic plans with our project team. The transparency demonstrated by those that completed the full instrument in a limited, one-month collection period, was very high, and perhaps unprecedented in our field to date. Many additional SCRs indicated that although they could not

...a remarkably wide range of SCRs, including ones from a variety of tax and legal statuses, willingly contributed information about their business operations, revenue sources and levels, technical roadmaps and releases, and strategic plans.

participate during the one-month collection period, they would appreciate having an opportunity to participate at a later time.

Respondents represented a variety of SCRs, both in terms of form (from informal and young projects to 20+ year incorporated or hosted entities) and function (publishing, hosting and access, discovery, evaluation and commenting, and archiving and preservation). They also represented a wide range of geographical bases, including the US, Canada, UK, Denmark, France, Italy, and Austria. The annual 2018 budgets for the 15 SCRs that fully reported on financials ranged from \$100,000 to \$5,864,000 in revenues and from \$91,645 to \$5,764,000 in expenditures; those that did not share direct financial data regularly cited that their budgets are primarily or entirely dependent upon grant funding or that they are for-profit entities that do not share this information with external parties. The number of FTE reported as currently supporting the SCRs varied from 0 to 1,200.

Respondents identified time, distributed responsibilities, and documentation locations as the greatest challenges to their participation. As one respondent shared, “It is difficult for just one person to have all the information.” Of our respondents, 20 reported the Census took two hours or less to fill out, and 11 reported that it took more than three hours.

Our project team will be further evaluating participation opportunities in the near future, including through co-hosting an open collection instrument on a rolling basis with the Invest in Open Initiative (IOI). We hope that it may be possible for a group to coordinate the collection of this data regularly (every other or every third year), enabling comparisons back to a baseline and measurement of change over time.

## B. PARTICIPATION TRENDS

The institutions that participated in the SCIP Census represent a broad range of institution types and sectors, and (unsurprisingly) a strong English-language bias. Responding SCRs were largely located either on the West Coast or East Coast of the US/Canada (the California/Pacific Northwest Coast or the East Coast) or in Western Europe. This is consistent with the invitations that were extended to approximately 200 SCRs primarily in these same geographical and national contexts.

Many of those who participated in the SCIP Census filled out all of the fields relevant to their SCR. Nonprofit players (35 total) shared everything or provided specific reasons for not sharing. For-profit and B Corp SCRs (8 total) tended not to share market-sensitive information. For-profit and B Corp players either did not share their fiscal information, or only provided it in general and non-numeric terms (e.g., three for-profit SCRs cited “it generates a surplus” and one added “in the millions” but none shared specific figures). Similarly, most for-profit and B Corp SCRs did not share copies of documentation (e.g., strategic plans, codes of conduct, product roadmaps), though one did volunteer to share these via phone “if necessary,” signaling a willingness to participate more fully under controlled circumstances.

A large number of SCRs also contacted us to express their interest in the project and their regrets that they could not participate in February/March 2019. Those that contacted us in this way typically pointed



to one of two factors: 1) the need for a longer timeframe for responding, ideally several months given the complexity of the information we sought, or 2) the need for an easier mechanism for submitting information. We plan to include many of these SCRs in an open iteration of data collection that is currently underway.

## C. DATASET OBSERVATIONS

Here, we report separately on the two datasets created within this Census project.

### SCIP Census Dataset

The project team anonymized all data for external reporting, including the Census Report and Data Visualizations. Anonymization was achieved through excluding SCR names in the Census Report and through only making publicly available the aggregated data for visualizations. All participants in the Census are receiving a copy of their own data and a static visualization showing their data against the aggregate, using the average, mean, median, and/or range of responses (depending on the question's structure).

Respondent data required some normalization; this was largely completed by Educopia and TrueBearing. Partial answers were removed, and open-answer questions were normalized where possible (e.g., to standard representation of country name; to a set of semi-standard answers such as annually, quarterly, monthly, bimonthly; or to move web addresses and URLs into a separate field). Some questions yielded a large range of bulky descriptions without enough cohesion to warrant coding. For example, answers to an intentionally open question, "What is the SCR's primary output(s)?," revealed the diversity of SCR understandings of what an "output" is. This data was not normalized for quantification, but rather was analyzed through a qualitative lens.

With 283 total fields of data, this rich dataset provides a broad set of analysis points, including many correlations that deserve further study as the number of respondents grows. We have laid the foundation for a much broader understanding of the field's current practices and the ways these differ across specific categories of SCR.

Currently, with 43 respondents, we are able to use the dataset and visualizations to improve our understanding of 1) the forms, functions, structures, and models SCRs use today; 2) some of the factors influencing sustainability and resilience of SCRs; and 3) what tasks and activities specific SCRs might engage in next to improve their stability. We hope that as the respondent pool increases, additional projects will be able to analyze the data to identify additional trends, challenges, and opportunities both at the individual SCR level and at the field level. The "Mapping the Scholarly Communication Infrastructure" project team will seek an answer to whether and how to continue this effort into the future.

### Composite Dataset

As described above, our team collected abridged information about 96 additional SCRs. These SCRs were selected out of the initial SCRs invited to take part in the Census; prioritization was based on the definition

of SCRs (see above), the availability of web-based information to our project team, and by consensus of the project team.

The 48 fields of data collected in this effort corresponded directly with 48 of the SCIP Census's 283 fields. In May 2019, Educopia Institute and TrueBearing Consulting combined the web-based collection data with the relevant 48 fields of the SCIP Census, yielding a total of 139 abridged entries.

The Composite Dataset primarily gives an overview of the following information for each SCR entry: 1) geographic location; 2) founding year; and 3) What part(s) of the research lifecycle it serves.

## D. DEMOGRAPHIC TRENDS

### SCIP Census Respondents

The survey gathered a range of demographic information from respondents. The SCRs that participated in the SCIP Census represent a broad range of institution types, sectors, and functions. Of the 43 respondents, 27 reported they are hosted by another entity (13 by academic institutions and three by foundations), 15 reported they are a standalone, incorporated entity, and one reported that it is not an entity. The majority of respondents reported a non-profit orientation (36), with seven taxable for-profit organizations responding.



Figure 5: SCIP Census, SCR locations

As depicted in the map-based overview (see Figure 5), responding SCRs were largely located either on the West Coast or East Coast of the US/Canada (the California/Pacific Northwest Coast or the East Coast) or in Western Europe. More specifically, about half of the respondents were from the US (21 respondents), and the other half were from Canada (2 respondents) or Europe (UK, 5 respondents; Denmark, 2 respondents; and one respondent each from Greece, Switzerland, Italy, France, and Austria). This corresponds to the invitation list, which included mostly US, Canadian, Australian, and European entities.

The respondents represent a wide swath of scholarly communication activities, and these centered primarily on “Discovery” functions, closely followed by “Archiving/Preservation,” “Publishing,” “Hosting and Access,” and “Evaluation and comment” functions as seen in Figure 6. Far fewer were

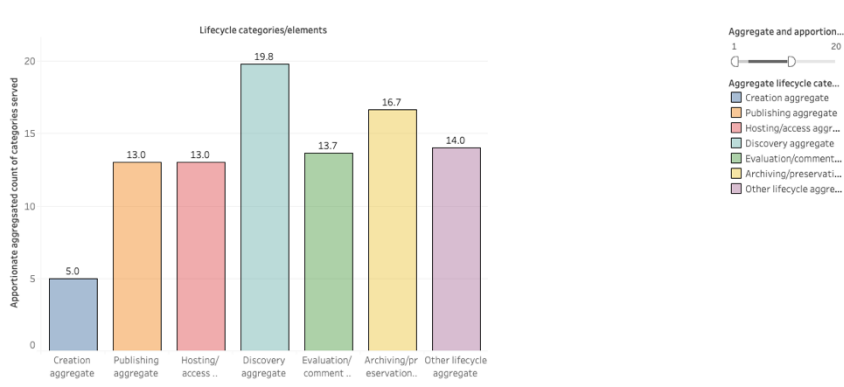


Figure 6: SCIP Census, SCR Publishing Functions

involved in “Creation” tasks, and almost all of these were non-profits, with only one for-profit focusing on computation and machine learning. For-profit actors cited heaviest involvement in “Publishing” and “Hosting and Access” functions, especially review, submission, and pre-print services. Note that many SCRs cited multiple functions; all SCRs citing “Creation” also cited one of the other SCR publishing functions within the project’s taxonomy.

Of the 43 respondents, all but four cited their roles as a Director, Officer, Founder, or a Manager. Three of these exceptions listed academic library affiliations and job titles. Respondents can be assumed to be familiar with, and often the primary person responsible for, the operations of the SCR.

Census respondents were asked to state their SCR’s primary output(s); the resulting open-text answers were so divergent that normalization proved impossible. Most SCRs reported several outputs, including software code, published content, and a platform or environment; some also cited a collection of content or a service they provide to others.

**...18 respondents reported founding dates between 11 and 27 years ago, or between 1991 and 2009, demonstrating their relative longevity in this swiftly changing industry.**

A total of 38 SCRs provided a founding date. Of these, more than a quarter reported that they were less than five years old (10), and more than half reported they were less than 10 years old (20 respondents). Fully 18 respondents reported founding dates between 11 and 27 years ago, or between 1991 and 2009, demonstrating relative longevity in this swiftly changing industry.

Looking more closely at the 18 SCRs that have sustained their operations for more than 10 years, we see a range of characteristics and practices underway, not one consistent equation for sustainability. Only two of these longer-lived entities had a for-profit orientation; the rest reported non-profit orientations. Seven of the entities are formally incorporated (one as a for-profit corporation, one as an LLC (or equivalent), one as a CIC, and four as nonprofits), while the rest are hosted by an academic institution (6), a non-profit organization (3), a foundation (1) or claimed no legal structure (1). They represent a broad range of geographical locations (e.g., Switzerland, France, Germany, Canada, Greece,

UK, and US), and a broad range of functions, including publishing preprints, journal articles, books, blogs, and ETDs; indexing content; preserving content; and creating DOIs.

The differences between each of the 18 SCRs that have operated for a decade or more are even more dramatic when we look more closely at the business documentation, technical procedures, and other characteristics. While 15 of these have a documented vision, mission, and values statement, only 11 of these have (or have had in the past) a strategic plan, and only eight have a review process for that strategic plan. Only nine report having conducted a market analysis. Just over half (10 respondents) have in-house staff managing their SCR's code (5 said no and 3 declined to answer) and only 10 have a product development roadmap (6 do not and 2 declined to answer). Five of these SCRs do not start the year with an approved budget (11 do and 2 declined to answer). They are funded through a wide variety of revenue streams, including member fees (10 respondents), service fees (7 respondents), private foundation grants (5 respondents), private donations and government grants (6 respondents), in-kind contributions (4 respondents), host subsidies (3 respondents), donations from corporations or high net worth individuals/family foundations (3 respondents), contracts/consulting (3 respondents), and subscription fees (1 respondents) subsidies.

This high level of diversity in organizational, technical, and fiscal characteristics between even the long-lived SCRs marks the lack of conventions and proven pathways to success within the SCR arena. More fully documenting some of the many and variable examples of how to successfully build and sustain an SCR over multiple decades (e.g., building on the "It Takes a Village" project's case studies) may help to illuminate what combinations of variables and approaches have worked well and provide a set of prospective models that new SCRs may mine for ideas without trying to adopt any one model as though it is a roadmap to success.

In the next five sections, we dive more deeply into the data reported by respondents in the Census. These sections are organized according to our data model, which identifies five core components in which an SCR can demonstrate competency and resilience: Vision and Scoping, Technical Infrastructure and Design, Administrative and Financial, Community Engagement, and Governance. For more details about our data model, please see Appendix C.

## E. VISION AND SCOPING

Among the key components assessed by the Census is the strength of the SCR's "vision," or the maturity of the SCR's definition of what the SCR is, what it does, and how it sets and accomplishes goals. Correlations between mission/vision strength and organizational health have been shown from the 1980s forward; a mission's alignment with its stakeholders, its

SCRs vary in their knowledge of basic community and organizational documentation forms...which shows a need for outreach and education

cost/revenue model, and with the needs of the field of practice also correlate strongly with organizational longevity.<sup>3</sup>

The Census seeks to assess an SCR's work in vision and scoping through raising questions about core pieces of organizational/community documentation: an SCR's mission, vision, and/or values; its strategic plan; its market analysis, and its Code of Conduct (or other written community standards). It also seeks information about the way the SCR engages with its stakeholders (leadership, membership/clients, etc.) in producing this shared vision. Our analysis of the responses is largely qualitative, and it relies heavily on the documents that SCR respondents shared with our team.<sup>4</sup>

The Census clearly demonstrates that SCRs vary in their knowledge of basic community and organizational documentation forms. In particular, many of the examples submitted by SCRs had only vague connections to the questions asked, which shows a need for outreach and education in our sector about these documents.

Just having these documents does not in any way guarantee sustainability for an organization or program; it does indicate that the organization or program has made an investment in planning and studying its vision and scoping and how those relate to the surrounding environment or market.

### Mission, Vision, Values

Most respondents (36) said that their SCR has a documented mission, vision, and/or values statement(s). Four respondents said they do not have these, and three respondents chose not to answer this question. Only around half of respondents (20) cited having a review process for their mission, vision, and values documentation; 14 had no such process, and nine chose not to answer. Most of the SCRs that cited a review process also stated that their documentation was last updated within the last five years (with 24 overall citing that their documentation was last updated in 2017 or later. All of the for-profit entities reported having a documented mission, vision, and/or values statement.

Fully 21 respondents shared their current documentation with our team either through a link or an upload. Of the 21 documentation examples shared with the team, many were vague and philosophical. Only seven of these examples provided some level of concrete and comprehensible statement about the SCR's vision (what the world looks like if it accomplishes its goals) and of those, only four tied their mission (the actions they take in order to accomplish their vision) back to their vision. Some differences were notable between younger SCRs (ones that have only operated for five years or less) and older SCRs (those with 10 or more

---

<sup>3</sup> See e.g., Amran, A., Lee, S. P., & Devi, S. S. (2014). The influence of governance structure and strategic corporate social responsibility toward sustainability reporting quality. *Business Strategy and the Environment*, 23(4), 217–235 and Macedo, I. M., Pinho, J. C., & Silva, A. M. (2016). Revisiting the link between mission statements and organizational performance in the non-profit sector: The mediating effect of organizational commitment. *European Management Journal*, 34(1), 36–46; note also that there are calls for additional research into these relationships, e.g. O'Shannassy, T. F. (2017). Associate editor reflections on the progress in and future of strategic management research in *Journal of Management & Organization*. *Journal of Management and Organization*, 23(4), 473–482.

<sup>4</sup> We note that the simple presence of these documents is not as powerful as evidence that such statements are known by employees and community members and guides their actions. Future assessments might include additional engagement with an SCR, including a brief interview or focus group, that could include deeper questions.

years of operations). In particular, the older SCR's had more elaborate and complex documentation that often was harder to parse, and the younger SCR's either lacked this documentation altogether or had more concise statements.

The processes respondents reported using to create their SCR's mission, vision, and/or values statements were as diverse and uneven as the documentation itself. Five SCR's named their founders as the sole authors; three others cited leadership/management as the sole author. Many described involving employees and team members (nine respondents) and another nine respondents cited involvement of their members or communities in either the drafting or the review process. The most common pathway for producing these statements was an SCR's Board or Steering Committee or other governance group undertaking this work on behalf of the SCR (15 respondents).

Having a consultant involved seems to have impacted the SCR's documentation; these outputs were sharper, more memorable and direct, and clearly connected the mission to the vision.

Looking closely at the ways the documentation was produced in combination with the documentation itself, several things correlate strongly with documentation clarity (though without statistical relevance due to the small respondent size). Having a consultant involved in the process (three respondents) seems to have impacted the SCR's documentation; these outputs were sharper, more memorable and direct, and clearly connected the mission to the vision. These statements also had stronger assessment potential, with clearer ways of measuring and demonstrating the SCR's progress towards its mission. Likewise, the involvement of employees and community members seems

to correlate to stronger documentation and statements that can be affirmed and assessed. Statements authored by founders or leadership often hinged on philosophical aims and gave fewer measurable ways of establishing when or whether progress is being made.

In public policy, sociology, business, and other disciplines, a lack of clarity and shared understanding of an institution's mission and vision has been shown to be a vulnerability for a wide range of organizational types.<sup>5</sup> The lack of consistency and strength in the mission/vision/values statements of many reporting SCR's indicates an area where investments in education, training, and the production of these elements may help to strengthen the longevity of organizations and programs.

### Strategic Plans

Compared to the mission/vision/values documentation, fewer respondents (28) cited having a documented strategic plan, and more respondents (11) cited that they did not have a strategic plan. Four respondents declined to answer this question. Of the 28 respondents with a strategic plan, only 21 reported having evaluation criteria, and only 22 cited having a review process. Only one respondent cited having a strategic plan, but no mission, vision, and/or values.

---

<sup>5</sup> Ibid.

Eleven respondents included either a link to their strategic plan or an uploaded document. The forms and content over the eleven examples varied greatly, from general directions documents to reiterations of the mission and vision; only six of these fit the established standard for a strategic plan (with clear strategies, goals, and outcomes). These six were well-structured, clear plans that map the strategic goals directly to the mission of the organization and that provide clear evaluation criteria for goals/objectives that could be measured and assessed. Almost half instead linked to project documentation, presentations, or a project management software instance (e.g., Trello, showing all of the development work underway, but without any strategic component visible). While these show forethought and planning, they had no visible established strategic approach, and all but one lacked concrete goals and objectives.

Quick-change environments (which SCRs often are) may neither need nor want bulky, multi-year strategic plans. In more nimble/agile environments, these are often intentionally eschewed in favor of more streamlined strategic directions documentation. This may contribute somewhat to the lower numbers of respondents reporting strategic plans and to the diversity of examples shared through the Census. However, we did not see evidence that the organization/program leaders responding to the Census were intentionally producing lean, flexible strategic directions documentation—indeed the “strategy” aspect was absent from most of the documents submitted or cited by respondents.

The Census responses suggest that many SCR leaders have not prioritized strategic goal documentation. This points to another potential liability for sustainability and longevity of the organization and/or program, as studies consistently have shown that planning and performance are linked, and that planning that is more, rather than less, strategic, leads to stronger outcomes and performance.<sup>6</sup>

### Market Analysis

While 25 SCRs responded that they have conducted a Market Analysis, descriptions of these analyses varied wildly. Similar to our findings with strategic plans, the respondents that contributed or described a market analysis ranged from those citing full market analysis processes to those who describe only how they have surveyed their own client bases or held a community discussion at a forum. For example, one respondent shared about its process: “It included an environmental scan, a rigorous review of both competitors and potential strategic affiliates, and it was conducted by a consultant and a subcommittee that involved both Governance voices (Steering Committee) and others within the community. We surveyed the community as part of this process as well, identifying other groups/tools they turn to.” In striking contrast, another volunteered “We run regular regional workshops and provide multiple opportunities and channels for our community to provide input and feedback.” Only three of the 20 respondents that used an open text box to describe the form and focus of their market analysis mention any market exploration activities such as researching potential new or expanded offerings or researching competitors or alternative providers. The five responding for-profit SCRs all cited conducting a market analysis in the last two years and descriptions from these respondents provided evidence of a

---

<sup>6</sup> See e.g., Andrews, R., G. A. Boyne, J. Law, and R. M. Walker. 2012. *Strategic Management and Public Service Performance*. New York: Palgrave Macmillan; and Elbanna, S., R. Andrews, and R. Pollanen. 2016. “Strategic Planning and Implementation Success in Public Service Organizations.” *Public Management Review* 18 (7): 1017–1042. doi:10.1080/14719037.2015.1051576.

multifaceted approach involving explorations of the overall environment, competitors, prospective clients, and potential service expansions.

The unevenness in responses across each of these documentation bodies indicates there is also unevenness in understandings of these core organizational and business document forms, perhaps particularly (though not solely) among the SCRs with a non-profit orientation. As with the other vision and scoping elements described above, this points to an opportunity for strengthening the scholarly communications tools, platforms, and services arena through promoting education and training around the value of undertaking right-sized, right-timed planning. In particular, SCR leadership needs to understand and plan for such things as: market development (including actually building a market for an emergent tool, service, or platform), market adoption rates (e.g., anticipating differences between early adopters and second and third-phase adopters), and investment strategies (e.g., what/when to invest in building the tool vs. building the market).

## F. TECHNICAL DEVELOPMENT AND DESIGN

The Census assessed the infrastructure and design of SCRs in order to determine the health and stability of the technical environments they provide. This included seeking to establish if the technology is current and appropriate, if technical debt is low, and if obsolescence factors are established and purposefully avoided. We also asked questions designed to probe the alignment between the technology infrastructure and other elements in the ecosystem.

We assessed the health and resilience of SCR’s technical design through a range of questions about the SCRs’ technical planning and implementation work. We asked SCRs about the status of the software they use and maintain, including who manages and integrates the codebase, how contributions are made, whether they maintain a roadmap, what the update cycle looks like, what standards and licenses they use, their interoperability with other systems, the form(s) of their outputs, and whether they are pursuing preservation for the code and/or outputs of the SCR. Ultimately, the Census tries to assess whether the technology is current and appropriate, whether the SCR is planning its technical development in proactive ways, and how interoperable or self-sufficient a system is.

### Codebase Management and Contributions

Most respondents (29) indicated that in-house staff manage their codebases, and most also indicate that the code they develop relies on a small number of staff. On average, seven paid staff members contribute to the codebases of 36 projects; of these, 20 reported having five or fewer code-oriented staff members, 12 reported having between 6-15 staff contributors, and four reported they have more than 15 staff members working on the code (with 25 being the highest number reported). Notably, three of

Approximately how many coders currently contribute to the code base?

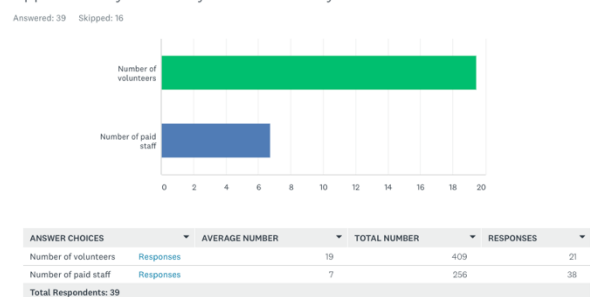


Figure 7: SCIP Census: Number of Code Contributors



these four respondents reporting more than 15 software developers on staff represented for profit organizations.

Volunteer numbers fluctuated much more wildly for the SCRs, with most SCRs reporting seven or fewer volunteers (13 respondents), and five respondents reporting that between 30 and 106 volunteers contribute to their SCR code bases. Five SCRs reported having two or fewer coders contributing to the code base, likely indicating these SCRS have a stronger dependency on individual coders.

Available pathways for non-staff members to contribute code to be considered for inclusion in the codebase included GitHub (15 respondents), direct communication with the team (four respondents), or through being part of the community (one respondent). Ten respondents reported that their SCR does not allow any external contributions to the codebase. Particularly given the heavy leaning towards “open source” code, the low number of SCRs with clear contribution opportunities or with any volunteer code community may be a point for concern. Increasing participation in the code creation process may be a worthwhile aim at the field level to ensure the SCR landscape is robust and adaptable, and could help to mitigate the dangers of code knowledge being in very few hands and minds.

Most respondents cited that their code is currently best characterized as “Live” (33 respondents), with one respondent in “Alpha,” two in “Beta,” and one “No longer supported.” Most were last formally evaluated within the last year (28 respondents) or in the last two to five years (seven respondents); only two respondents cited that they had never formally evaluated the code.

The low number of code-oriented FTEs in many of the reporting SCRs (especially those with a non-profit orientation) may be cause for concern, which becomes especially apparent when this answer is cross-tabulated with questions about funding channels. SCRs reporting the lowest number of in-house FTEs working on the code base also tended to report fewer revenue streams and high dependencies on grant funding from federal and/or foundation sources. Grant funding tends to be tied to “innovation” oriented work, or work that privileges new development trajectories and activities rather than code maintenance and responsiveness to user requests. Sustaining code-based tools, platforms, and services requires some level of coding energy to be dedicated to maintaining and updating the code, and programs that do not have established “services” revenues may find operational funds challenging to raise and sustain.

Based on the responses, SCRs may need additional research and/or training and guidance in specific areas that “tech incubators” and “accelerators” often provide in more commercially centered development, including 1) understanding when and how to use (or even to allow) volunteers vs. hiring contractors vs. hiring FTEs, and 2) establishing and recalibrating the balance between product development and product maintenance.

### Hosting and SCRs

Most SCRs reported that they are “flying blind” in terms of how many users they have engaging with their tool, platform, or service. Of the responding SCRs, 22 respondents host an instance of their SCR as a platform or service that others can subscribe to or use. In response to a question about how many users they have, SCR reports varied from “14 organizations” to “1 million api calls/day,” with most of the respondents (six out of eleven) simply citing “unknown” or stating “other organizations” without

quantifying. Some SCRs (16 respondents) also reported that they provide their code for other organizations to host on their own servers. Nearly half of these respondents (seven) do not know how many organizations currently host the code; those that do cite between “2” to “300+” organizations.

The high variability in these answers may result in part from the many different output types that SCRs report that they provide. A question regarding what the SCR’s primary output(s) are received a tremendous range of responses, from code to discovery platforms, from journals to datasets, and from training programs to standards (see “Demographic Information” above).

### SCR Updates, Licenses, Standards, and Interoperability

Responding SCRs reported maintaining regular update cycles for their code. Most respondents (12 respondents) cited an iterative process of updating-as-necessary. Others reported daily (3 respondents), 1-2 week sprints (5 respondents), 1-3 months (5 respondents), or 2-3 times annually (4 respondents). Most respondents (29 respondents) provided the dates of the last two formal software releases, almost all of which were dated in 2017-2019.

SCRs are using nearly 10 different software licenses for their code, including Apache 2.0 (9 respondents), MIT (9 respondents), GPL (6 respondents), BSD 2-clause (1 respondent), BSD 3-Clause (3 respondents), , OSS (3 respondents), and CC-By (1 respondent). Three respondents cited their software as “proprietary.”

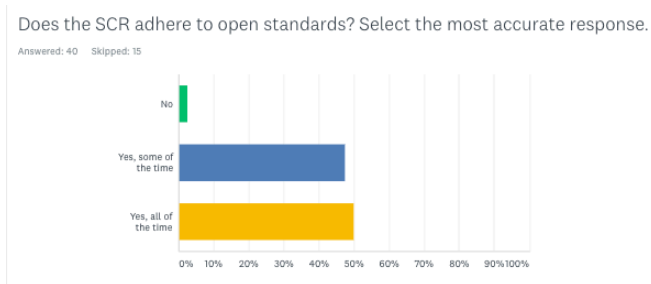


Figure 8: SCIP Census: Open Standards Adherence

As depicted in Figure 8, most SCRs also report that they adhere to open standards either “All of the time” (19 respondents) or “some of the time” (16 respondents); only one respondent reported not using any open standards. The 47 standards contributed in response to an open text question about across software, hardware, data, infrastructure, service, and protocols. We know the standards provided are just a sample of the ones currently used by respondents; we include them here as a partial list in Figure 9.

ATOM	css	IIIF	Marc/MarcXML	openUrl	TRAC
bijson	csv	ISO16363	MEDRA	PCDM	W3C
Bootstrap 4	DOI	JATS	MODS	Rails 5.2	WARC
CASRAI	Dublin Core	JSON API	OAI-PMH	RESTful API spec	WCAG 2 AA
CHORUS	EAD	KBART	OAIS	RFC1807	Web Annotation
CMIS	HTML5	Keepers	OAuth	Shibboleth	Webpacker
COUNTER	http	LOC headings	Open API spec	Solr 7.2	XML-TEI
CC0	https	LOCKSS	OpenAIRE	SWORD	

Figure 9: Open Standards Cited

Only ten respondents reported that their SCR does not interoperate with any other SCR systems. As with the question about open standards, an open-text answer about interoperability yielded a huge list and several “as many as we can” answers. Among the more than 70 systems cited by respondents were (spelling/representation here as provided by respondents):

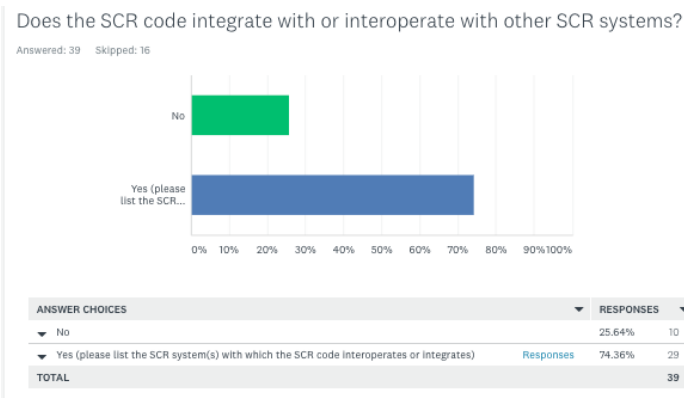


Figure 10: Interoperability

AGU ESSOAr, Altmetric, Archivematica, ArchivesSpace, Aries Editorial Manager, ARPHA, arXiv, Austrian Academy of Sciences LinkedCat Repository, BASE, BenchPress, bioRxiv, Blacklight, CALM, Chronos, CLOCKSS, ContentDM, CORE, CrossRef, DOAJ, DPLA, Dspace, DataCite, DuraCloud, Editoria, eJournal Press, eLife Libero, Erudit, Europe PMC, EZID, Fedora, Fulcrum, Google Scholar, Hyku, Hypothes.is, Isidore, Janeway, Keepers Registry, LOCKSS, Lucene/SOLR, MBOX, Merritt Preservation Repository, Mirador, NERD, netX, OJS/PKP Harvester,

Open Science Framework (COS), OpenAIRE, OpenDOAR, OpenWayback, ORCID, Outlook, PLOS, PubMed, PyWb, Rebus reader, RePeC, Ringgold, ROpenSci, Rosetta, ScholarOne, Scopus, SHARENotify, SharePoint, Sheridan Press, Sherpa/RoMEO, SimplyE, Spotlight, SWORD, SXF link resolvers, Symplectic Elements, TMS, Universal Viewer, VOSViewer, WikiData, and WorldCat.

Perhaps the most interesting thing about this interoperability self-reporting by SCRs was that fully 63 of these 75 cited tools appeared on the list only one time. Of the 12 that did appear more than once, five had only two citations each (BASE, DSpace, Fulcrum, Janeway, SharePoint), two had three citations each (DOAJ, Europe PMC), two had five citations each (Hypothes.is, OJS), one had six citations (ORCID), and one had eight citations (CrossRef). Based on these responses, there appear to be few common integration priorities shared by current scholarly communication tools, platforms, and services today.

### Development Process

Product development roadmaps are used by most reporting SCRs (29 respondents), and only nine respondents explicitly reported that they do not use a product roadmap. As with other forms of documentation, the product roadmap documentation submitted by respondents varied significantly across SCRs, though in this case, there were two main forms: 1) Trello boards or JIRA-like environments with small “tickets” or short descriptions of small development tasks, sometimes grouped under broader categories, and 2) descriptive paragraphs, often focused primarily on the broader aims rather than the specific development steps required by these aims. About half of the respondents report that they maintain open, publicly available roadmaps (16 respondents); just under half do not host these openly (14 respondents).

... there appear to be few common integration priorities shared by current scholarly communication tools, platforms, and services today.

The development and maintenance of these roadmaps involves a mix of their: staff/team (23 respondents), community (8 respondents), leadership (6 respondents), and users (3 respondents), often with a combination of several of these categories. Respondents report that prioritization and selection of features and repairs comes in multiple forms as well, depending in part on the number of coders and requesters engaged with the SCR. Some report that they use consensus to make decisions (particularly for SCRs with fewer coders); others have editors or product managers make prioritization decisions (especially for SCRs with many staff or volunteer coders). Unsurprisingly, respondents also reported that some of their prioritization decisions are based on funding, particularly in grant-funded projects.

### Migration and Preservation

Most SCR's (29 respondents) report that they have some type of mechanism (e.g., an API) to help users migrate off of their infrastructure; seven do not have such a mechanism. Most (29 respondents) also report that they are pursuing "preservation," though they do not all share the same definition of preservation, judging by their open-text descriptions of this work.

Some respondents report they are currently thinking about preservation or assessing their options. Some are currently putting aside a preservation fund to ensure resilience and longevity; some cited the use of backups or GitHub or Zenodo as their preservation vehicle. Some of the respondents are SCRs that perform preservation functions; these not only described the preservation they do for others, but also described the efforts they make to preserve their code, documentation, and other materials. Seven respondents mention subscribing to LOCKSS, CLOCKSS, Portico, APTrust, and Merritt as one part of their preservation approach. For those that are not currently preserving their SCR's code or data or assessing their preservation options (10 respondents), answers as to why they are not ranged from "making our code publicly available is good enough," or that they have "sufficient backups" to responses that they are "out of funding" or "too busy writing code."

Have you taken measures to ensure the long-term preservation of the data and/or software associated with the SCR?

Answered: 40 Skipped: 15

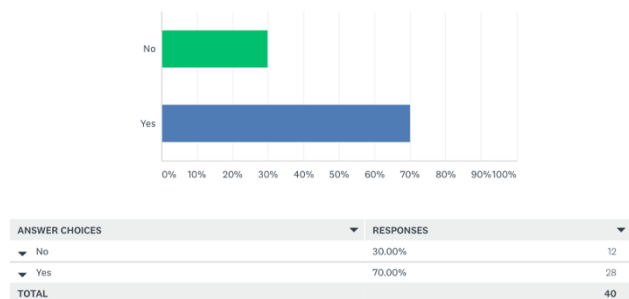


Figure 11: Preservation

As with the "Vision and Scoping" measures, the data gathered by the Census about the Technical Infrastructure and Design of SCRs today paints a picture of relatively divergent practices and expectations. From the low numbers of coders and technical development staff and volunteers to the lack of consistency in roadmap development practices, and from the lack of alignment seen in the "interoperability" answers to the diversity of licensing flavors cited, the current tenor of development in SCRs seems to be relatively insecure and *ad hoc*.

## G. ADMINISTRATIVE AND FINANCIAL

The Census sought to establish a solid baseline understanding of the approaches to administrative and financial management of SCRs. Of particular concern was the level of knowledge and planning demonstrated by SCRs in these areas. For example, do SCRs know their cost of operations? Do they engage in planning and budgeting activities? Do they support staff, and if so, how stable and mature is that support? The Census queries SCR's about their practices and documentation, including through several multi-year questions, in order to begin to evaluate and assess their administrative and financial practices.

### Staffing

Most SCRs reported some level of paid staff support, funded through a mixture of SCR earnings (25 respondents, with 6 reporting zero), grant funding (23 respondents, with 9 reporting zero), or in-kind contributions from partner institutions (16 respondents, with 9 reporting zero). The total number of staff paid through SCR earnings ranged from 0-1,200, with most citing between .5 and 52 employees (see Figure 12).

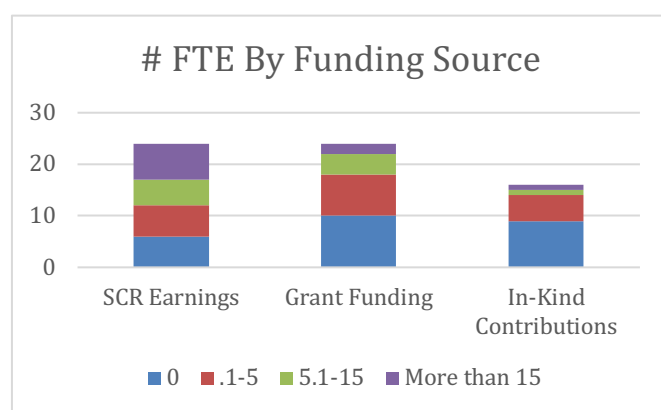


Figure 12: Number of Staff Paid by Funding Source

Removing one outlier (with 1,200 FTE), 18 SCRs report funding a total of 297.25 FTE through SCR earnings. Nearly a third of those (99 FTE) are employed by for-profit entities, and approximately two-thirds (198.25 FTE) are employed by entities with a non-profit orientation. Most of the SCRs that employ more

than 15 employees pay for these employees through SCR earnings rather than through grant funding or in-kind contributions, as depicted by Figure 12.

A total of 14 SCRs reported funding 102.4 FTE through grant funded sources. Ten SCRs reported having “zero” FTE paid through grant-funded sources, which is a higher number than those reporting “zero” FTE paid through SCR earnings or in-kind contributions. There were two SCRS that reported paying more than 15 employees from grant funds, with one reporting 19.5 FTE, and another reporting 34 FTE currently covered by grant funding.

A total of 16 respondents cited that they fund 209.45 FTE through in-kind support, but one outlier (the same one reporting 1,200 FTE supported by SCR earnings) accounts for 192 of these in-kind FTE.

Most SCRs that report hosting staff members host at least some of these remotely (28 SCRs out of 35 respondents to this question). Nine SCRs report that their staff are 100% remote, and another four report that 50% or more of their staff members work from remote locations.

## Financial and HR Documentation

Documentation practices around financial information and staffing/HR information vary among the responding SCRs. As shown in Figure 13, out of 32 respondents, most reported having some standard forms of staffing documentation and support, including job descriptions for all employees (27 respondents), an organizational chart (25 respondents), and regular staff evaluations (23 respondents). Just over half of those who responded also reported having an employee handbook (21 respondents) and a professional development budget (17 respondents). Roughly a third of respondents that reported they

Which of the following documents are in place for this SCR? Select all that apply.

Answered: 32 Skipped: 23

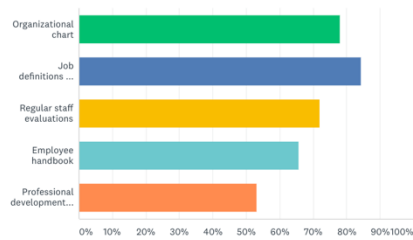


Figure 13: Staffing Documentation Maintained for the SCR

do not have these basic documentation components in place. Notably fewer Census-takers responded to a question regarding some of the standard forms of financial documentation a program or organization would be expected to maintain (and in most cases would be required to maintain if supported by any federal funds). As depicted in Figure 14, of the 26 respondents, a fair number reported having a bank account (20 respondents), earned revenue of some form (20 respondents), financial reserves (18 respondents), and some type of accounting software (18 respondents). More than half of the Census respondents reported that they do not maintain an Accounting Manual, adhere to Generally Accepted Accounting Principles (GAAP), maintain accrual-basis accounting, or engage in audits (all of which are required by federal agencies for all recipients of grant-or contract-based federal funding).

As with other measures, we do not assume a perfect correlation over time between any one of these elements and long-term business success. However, we do anticipate that fiscal maturity and responsibility are among the most critical markers of sustainability. As such, the numbers in Figure 14 below raise significant concerns. The most common point of stress and failure in nonprofit businesses is fiscal insecurity and mismanagement, and symptoms of this often include mistaking bookkeeping for financial planning and not having appropriate checks and balances that ensure multiple people “see” the fiscal picture and ensure its health and accuracy.

Which of the following does the SCR maintain? Select all that apply.

Answered: 26 Skipped: 29

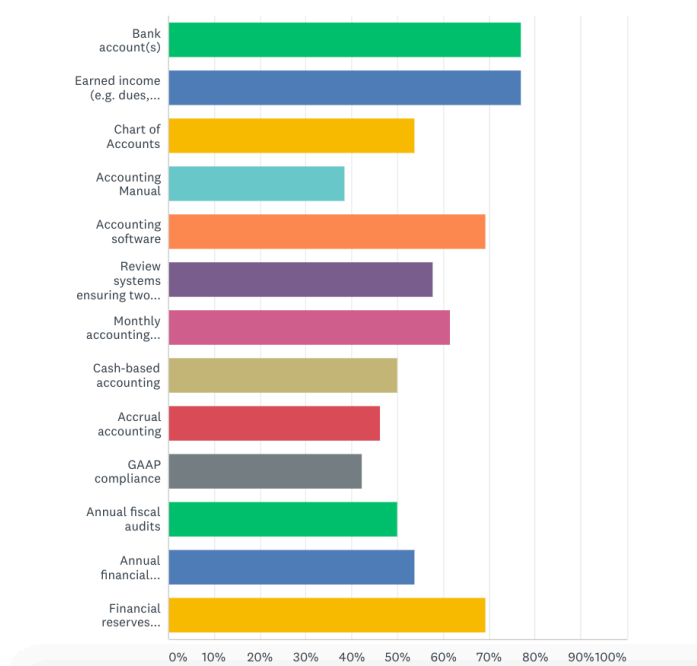


Figure 14: Accounting Systems Maintained by the SCR

Only 16 respondents rectify their books monthly, and only 15 respondents have at least two people signing off on financial transactions. Further, only 14 respondents have a Chart of Accounts, and only 14 provide annual financial reports to the public. Even the most basic checks and balances and oversight appears to be missing from nearly half of those few (26 respondents—lower than most questions in the survey) who chose to respond to this question.

Of even more concern, more than 30 percent of these respondents report they do not maintain financial reserves. And of those that do (18 respondents), eleven reported having less than a year covered by these reserves, and seven of these have less than six months.<sup>7</sup>

The Census responses provide evidence that many SCR accounting systems lack basic protections and thus leave investments in these SCRs at a greater risk of misuse, mismanagement, and failure; as described by respondents, they also make it nearly impossible for these SCRs to adequately plan toward any adaptability or resilience, let alone sustainability.

Heightening the concern already raised by these questions, nearly a third of the 40 SCRs responded “no” to a question about whether their SCR begins the fiscal year with an approved budget. Thirteen respondents wrote in with an explanation, and these open-text responses included the following:

- “It has not been necessary.”
- “Our revenue comes from grants and we discuss these budgets on an ad hoc basis upon submission.”
- “no budget.”
- “No regular source of funding at present.”

Does the SCR begin each fiscal year with an approved budget?

Answered: 40 Skipped: 15

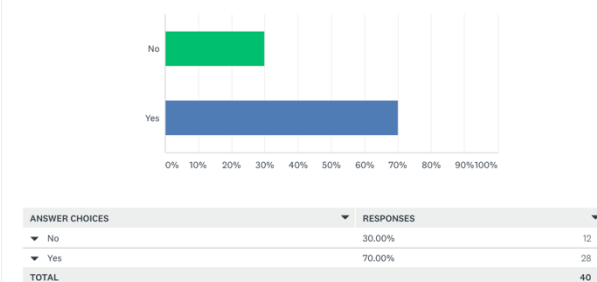


Figure 15: Does the SCR Have an Approved Budget?

<sup>7</sup> Although there is no single “safe” number, most recommendations are to maintain at least a full year of financial reserves for emergency scenarios, and an additional amount as “change capital” that can be used to help organizations through regular transition moments.

These answers point again to the insecurity undergirding so many of the SCR environments used today in the scholarly communication environment.

These responses have serious implications regarding the high-risk, low-stability scholarly communication landscape upon which we currently rely. The low number of respondents to this section, coupled with the immaturity of the systems and standards indicated by those that did respond, point to a serious need for action, not just through education, but also through fundraising for operational capital and financial reserves to ensure continuity and to make adaptability and change more possible for these SCR environments, most of which appear to be bootstrapping, agile, and very, very fragile infrastructures.

### Revenues and Expenses

The Census probed at the last three years of the SCR's fiscal reports in order to establish the SCR's *budgeted* (the estimate of revenues and expenses) and *actual* (the end-of-year tally of revenues and expenses) revenues and expenditures, and also to establish how big the gap is between these two elements. Most business environments are required to summarize their revenues and expenditures and overall worth on an annual basis, filing these in appropriate business tax documentation forms (e.g., on the US 990). We anticipated fewer SCRs would be able to report their budgeted revenues and expenditures (based on how many SCRs establish a formal budget, as covered above), and that some would still be closing their 2018 fiscal year, but that most SCRs should be able to identify and provide their revenues, expenditures, and net for each of the last three years.

... a serious need for action  
to ensure continuity and  
adaptability for these  
bootstrapping, agile, and  
very fragile infrastructures

In mature SCRs with established business practices, predictability should show relatively stable and small differences between what is budgeted and the actual revenues and expenses, particularly over a three year period. Even in nonprofit environments, which depend more heavily upon donations and grants that are sometimes hard to predict, accurate forecasting is key, and can make the difference between success and failure (especially in businesses that lack adequate financial reserves to weather a crisis). The Census also asked for an explanation of deviations or variances if the difference between the budgeted and the actuals was over 20%.

Similar to other fiscal questions, the number of SCRs reporting these numbers for 2018 (budgeted 16, and actuals 19), 2017 (12 budgeted and 20 actuals), and 2016 (11 budgeted and 16 actuals) were low. For those that shared these numbers, deviations were striking, as were overall losses reported. In 2018, revenue expectations were missed, often significantly, by seven of the 15 reporting SCRs. These predictions were off by between \$20,000.00 and \$1,250,000.00, with a total loss overall of \$(2,057,762.00) for these seven SCRs. Expenditures were also often dramatically different from budgeted, with four SCRs spending more than anticipated by between \$79,880.00 to \$248,196.00 and eight underspending their budgeted expenses by between \$16,699.00 and \$900,000.00.



Some of these variances between budgeted and actual were more than 20% of the total budgeted amount, and SCRs usually explained these variances in terms of grant-based income changes or in one case, a successful fundraising effort.

SCRs reported a wide range of revenue sources support their work. Thirty-eight respondents reported that these include: membership fees (16 of 38 respondents), private foundation grants (16 respondents), service fees (13 respondents), government grants (11 respondents), in-kind contributions (11 respondents), subcontracts on grants awarded to partner institutions (7 respondents), host subsidy (7 respondents), contracts/consulting (6 respondents), subscription fees (6 respondents), donations from corporations (5 respondents), and donations from high-net-worth individuals or family foundations (4 respondents).

What revenue streams support the SCR? Select all that apply.

Answered: 38 Skipped: 17

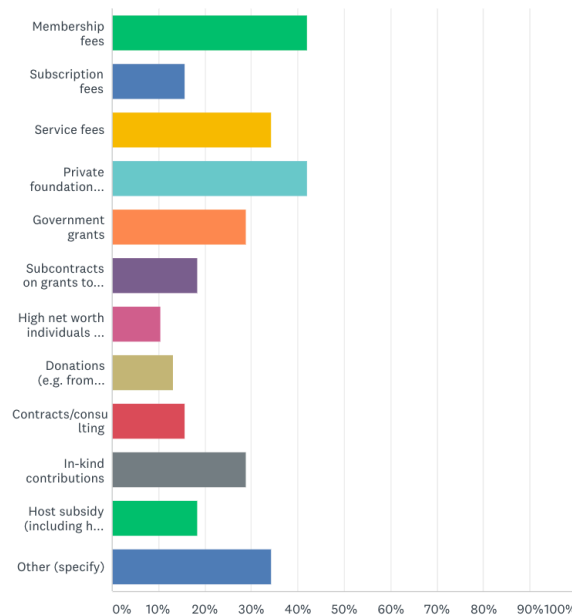


Figure 16: Revenue Streams Supporting the SCR

Additional responses with three or fewer respondents included conferences/events, registrations and sponsorships, private equity backers, licensing fees, and earned revenue from writing. The healthy diversity in revenue sources likely accounts for some of the success of these SCRs, many of which cite three or more of these sources of revenue.

Provide the average percentage for each expenditure.

Answered: 26 Skipped: 29

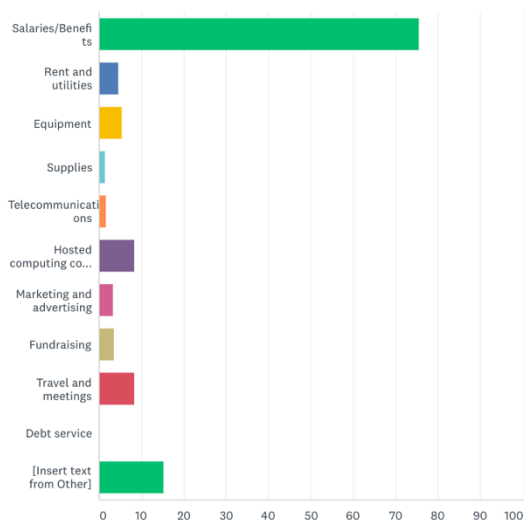


Figure 17: Expenditures of the SCR

Expenditures focused primarily on salaries and benefits. Nearly all responding SCRs (25 respondents) cited this as their top expenditure with an average of 75% of their expenditures in this category. The other expense categories cited by more than 10 respondents included “Travel and meetings” (25 respondents), “Hosted computing costs” (19 responses), “Equipment” (14 responses), and “Marketing and advertising” (14 responses). Notably, only one respondent (a for-profit organization) of the 14 that cited marketing and advertising

spent more than 3% of its total expenditures in this category; that organization reported 13% of its budget went toward this category.

Based on the responses to this Census, we believe the financial health of SCRs can be categorized as “challenged,” system-wide. High variability in the sophistication of accounting and reporting, coupled with instability in funding sources and amounts anticipated and received year-over-year, yield a fragile

**... a fragile ecosystem in which each player survives based largely on will, tenacity, and no small amount of luck.**

ecosystem in which each player survives based largely on will, tenacity, and no small amount of luck. Reinforcing the fiscal infrastructures of SCRs and ensuring proper checks and balances, month- and year-end close activities, and governance oversight are in place and functioning correctly could greatly improve the fiscal outlook for scholarly communication. Training for the Boards and Steering Committees charged with oversight in many SCRs may help to realize this opportunity to improve the likelihood of success.

## H. COMMUNITY ENGAGEMENT

The Census also evaluated the degree to which SCRs engage with their stakeholder communities, including through external outreach and internal channels of communication.

SCRs report using a wide variety of mechanisms to reach their stakeholders, and often these are highly targeted and selected based on the stakeholder groups they seek. The most frequent response was “In-person Events,” with 38 of 38 respondents citing this as an “active” area. The next most popular “active” outreach location was a website/blog (37 respondents), followed by social media (34 respondents) and live, online events like webinars and community calls (32 respondents). Real-time interaction, preferably in a live context, was unanimously prioritized by respondents.

Which of the following does the SCR use to engage stakeholders?

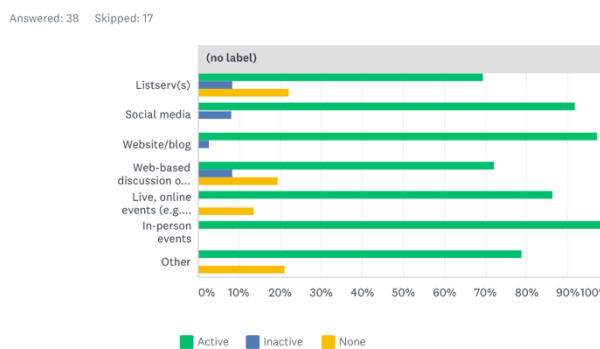


Figure 18: Outreach and Engagement Methods

Most SCRs also reported that they provide regular reports to their internal and external stakeholders on the SCR’s activities and finances. Out of 38 respondents, 33 do so annually (14 respondents), quarterly (7 respondents), monthly (5 respondents), semi-annually (2 respondents), and bi-annually (1 respondents). These go to a range of stakeholders, including community members (12 respondents), leadership (10 respondents), funders (6 respondents), staff/team (2 respondents), and the public (2 respondents). Seventeen of 31 total respondents make these reports publicly available; fourteen respondents report

that they do not. Eleven respondents produce a formal annual report, and another seven report that their SCR is included in a host organization’s annual report.

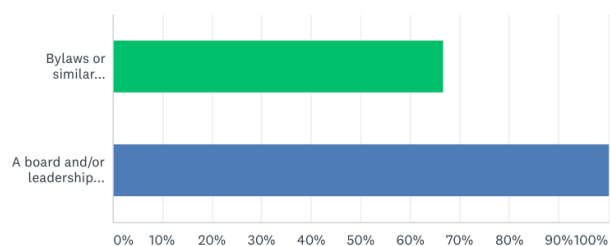
Most SCRs report recently surveying their stakeholder communities about the SCR, either within the last year (24 respondents), within the last 2 years (3 respondents), or within the last five years (2 respondents). Only four report never surveying their stakeholders. This signals a solid level of commitment to hearing information from, rather than simply conveying information to, stakeholder communities among most SCRs participating in this Census.

## I. GOVERNANCE

The Census assesses the form and function of governance for the SCRs, largely seeking to understand how decisions are made and by whom, and how checks and balances are in place to ensure accountability to stakeholders.

Which of the following does the SCR have?

Answered: 33 Skipped: 22



ANSWER CHOICES	RESPONSES
Bylaws or similar documentation of its governance structure	66.67% 22
A board and/or leadership group(s) overseeing its work	100.00% 33
Total Respondents: 33	

Figure 19: Bylaws and Boards

the work of the SCR. Two-thirds of these (22 respondents) also cited that they have bylaws or similar documentation of the SCR’s governance structure in place. Rich descriptions of these entities provided in 31 open text responses demonstrate that many SCRs expect their Boards to serve multi-year terms, and that these often provide advice and guidance to a central team or staff. Some also provide financial oversight (7 respondents) and technical guidance (10 respondents).

SCRs reported that volunteers provide much of the oversight and governance of their organizations, including 12 that cited volunteers serve as leadership of the SCR. However, only eight SCRs cited having policies in place governing its use of volunteers.

Of 33 responding SCRs, all cited that a board and/or leadership group oversee

Most of the SCRs that responded do not have a formal succession plan, contingency plans, or escrow arrangements in place to ensure continuity of operations. This signals another risk factor for current SCRs—this type of documentation, along with the fiscal reserves that often accompany it, can mark the difference between an SCR that successfully encounters and navigates crisis, transforming and maintaining continuity of operations vs. an SCR that fails in a crisis moment. Given that crisis is a predictable future state for almost every business and program, investing resources into this type of planning, education, and training for SCRs may help to make this environment more resilient to change.

Does the SCR have a formal succession plan, contingency plans, and/or escrow arrangements in place to ensure continuity of operations?

Answered: 37 Skipped: 18

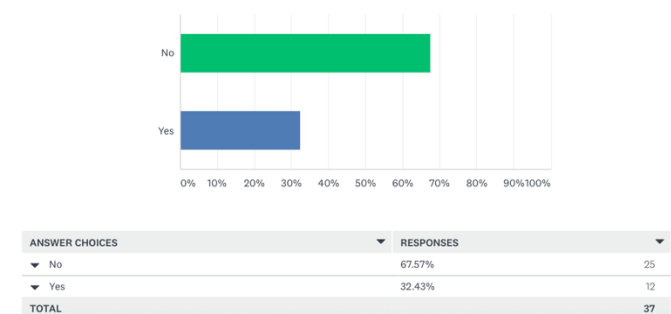


Figure 20: Succession Planning

### Section 3: Recommendations

This Census has provided a uniquely deep view across the technical, fiscal, organizational, and community frameworks used by more than 40 SCRs active today in the global scholarly communication sector. As mentioned within the report, one key finding of this research is that SCRs of various sizes, governance structures, legal structures, and fiscal levels willingly volunteered hours of their time to gather and report on their work. The reasons they cited for participating included, above all, their desire to hear similar information from others, particularly through aggregated data about their peers. Many SCRs are aware that their organizational, governance, technical, fiscal, and/or communications infrastructure needs scaffolding, and they want examples of how other projects and programs have successfully navigated the “Valley of Death” that looms between early formation and healthy acceleration of a business entity.

We are cognizant that this sample is small (we estimate it is less than a quarter of the full range of SCRs in operation today) and we are eager to see this respondent pool expand over time. We also know that the initial iteration of the Census provides a one-time snapshot of SCRs at one moment in time, and we hope to see repetitions of this Census in the future that enable comparison back against this baseline data.

To that end, our strongest recommendation is that this Census serve as an early step forward towards regular, rigorous research and assessment of SCRs’ infrastructures and models. Such assessment could help the field to quickly identify which elements of SCRs need additional attention and actively and mindfully fill these gaps. It can also inspire and influence SCRs to conduct their work according to emerging standards and norms in the industry. There are practices and operations that have demonstrated positive impacts on community and organizational health and resilience, and we can work as a field to implement more of these earlier in the process of SCR evaluation and early success. The [“Invest in Open Infrastructure”](#) initiative is one potential candidate that could coordinate this ongoing work.

We also recommend that the global intent of this survey be actualized more fully in future iterations, including through distributing invitations for participation more widely, translating the survey into multiple languages, and including business forms and descriptions that are appropriate to a broad range of different national contexts. This will require collaboration with individuals based in these national contexts, not simply producing the instrument in a US/European bias and translating that instrument into different languages.

Via the Census, we have identified a number of critical areas of SCR development that need further study and improvement. These include the following:

1. **We need a stronger taxonomy for the various functions performed by SCRs**, and one that is standardized for common use. Our team's work to develop the initial definition of an SCR and specific list of categories and sub-categories (e.g., "Publishing" as a top-level category, and sub-categories including "Submission," "Review," "Copyediting," "Design," "Layout," and "Other") provides a "straw person" version that can be improved upon by additional research and consensus building. Having a standard vocabulary will help us to quickly identify what functions exist, what tools are geared towards those functions, and what gaps and intersections we need to address. It will need to remain flexible, as scholarly communication is still quickly changing, and we can anticipate that change continuing and even increasing before we reach a long-lived, relatively stable state such as we enjoyed in the print-scholarship arena for several centuries.

As a key part of this field—its producers and users—academic institutions also need to establish what standards they want to promote and require of those providing them services. Academic stakeholders have acted as separate segments; for example, librarians, technologists, and faculty members may each engage with the same service provider in very different ways, some as purchaser, some as supplier, some as reviewer, and some as host. Coming together and consciously evaluating both the cost and the impact of different service models, and then building requirements based on these service models, is an imperative component in recalibrating the scholarly communication business model. As evidenced by some of those for-profit players that volunteered their time and information to this survey, many service providers will *welcome* this recalibration, as they, too, have a need for the system to break out of its current form to provide a more balanced marketplace in which competition and excellent service yield success.

2. **We also need to attend to the range of practices represented by our SCRs in terms of their visions, technical development and design, financial and HR models, community engagement practices, and governance frameworks.** There is no one proven pathway to success; instead, a range of processes and models produce both positive and negative results in terms of a player's adaptability, resilience, and ultimately, survival. Any assessment practice must attend to and learn from this range, not try to implement one-model-to-solve-all-challenges. Especially given the pace of change, different models may prove more or less successful depending on specific factors in the technical environment over time as well.

3. Connected to the above, **we need to recognize that not all SCRs need to or should succeed.** Sunsetting in our scholarly communication technical environment is often considered a sign of failure. Instead, we need to welcome it as a sign of a healthy overall environment. Developing and documenting methods and mechanisms for effectively sunsetting programs may help to improve understanding of this lifecycle stage and enable programs to “fail faster” (instead of being propped up in part out of a fear of failure) and thus free up resources and time more quickly for other activities. We also need to explore and document a range of other pathways to sustainability, including studying the feasibility of mergers between programs and organizations to enable sharing of some of the heftier overhead costs and expertise.<sup>8</sup>
  
4. **We need better ways to identify and prioritize solid investments at different stages of development.** In the for-profit environment (which constitutes a large portion of the scholarly communication environment), ongoing evaluation happens at several phases of potential development activity, including early formation/innovation (often with incubators and angel investors backing successful early-phase work) and acceleration (with both acceleration entities and investors backing mid-stage entities as they scale and expand). We need to introduce similar structures to support non-profit evaluation as well. Using existing tools, we are already partially equipped to evaluate the sources of our investments over time.<sup>9</sup> Matching this against real data about our library-based investment (see e.g. the 2.5% work of 2017 and the forthcoming work from the “Mapping the Scholarly Communication Infrastructure” team on evaluating library investments in SCRs) in particular could give us a quick vantage point of our current investments and our current infrastructure from which we can assess and evaluate our progress towards community-established goals.
  
5. **Almost all surveyed SCRs, and by extension, likely most SCRs, need education, mentorship, and training** in several key areas of development, including the following:
  - a. **Vision and Strategy.** The lack of understanding evidenced by the Census, including the materials submitted by SCRs as examples of their mission, vision, values documentation; their strategic plans; and their market analyses, is both striking and quickly mendable through specific, targeted investments in known business practices. These are not new practices; they extend across both for-profit entities and non-profit entities as critical foundation blocks that help to guide growth and investment over time. The lack of clear strategic documentation and measurable goals from most of the SCRs that participated in this Census marks a now-known challenge for many organizations and programs and

---

<sup>8</sup> See for example the forthcoming merger between LYRASIS and DuraSpace Foundation. Studying that merger and other hosted environments may help the field better understand how sharing infrastructure impacts the sustainability of individual programs.

<sup>9</sup> These tools include funder evaluation models (both private and federal), peer review models, the Census (Mapping the Scholarly Communication Landscape, 2019), Community Cultivation – A Field Guide (Educopia, 2018) on community and organization formation, and the It Takes a Village (LYRASIS, 2018) report and research on Open Source communities.

again, points to the need for education and training around these common practices and the value of planned development directions

- b. **Technical Infrastructure and Design.** Findings that stood out included the high variability in the number and type of coders that currently participate in SCRs and the challenges to code contribution that exist in some environments, including Open Source code spaces. The data gathered by the Census suggests there is a lack of consistent practices and standard technological design approaches in our sector. Some inconsistency may be positive, demonstrating a variety of potential roads to success; however, we question the sustainability of these code bases (particularly those that rely on only one or two software developers) over time and the lack of alignment between SCRs (e.g., as evidenced by the broad range of rarely overlapping tools listed in the “interoperability” open-text answers by SCRs). As stated above, the current tenor of development in SCRs seems to be relatively insecure, unstandardized, and *ad hoc*. This requires further study and interventions designed to move us from one-off tools to integrated environments at a swifter pace.
- c. **Financial and Staffing.** Of all of the areas of concern that have been highlighted in this report, we believe none is more compelling than the financial self-descriptions provided by respondents. Many SCRs have low-to-no financial reserves. Most of those that responded to the Census’ financial questions report that they do not rectify their books on a regular schedule, and most lack the basic checks and balances that keep businesses safe from both accidental and purposeful financial misreporting. Transparency exists for a small number of SCRs, but the vast majority do not currently provide annual financial reports to the public. Our investments in SCRs are at risk. The lack of basic accounting functions (e.g., Chart of Accounts, accrual-based accounting, adherence to GAAP, etc.) in so many SCRs today makes it improbable that they can adequately plan toward adaptability, resilience, or sustainability. Additionally, the low number of respondents to the financial questions, coupled with the immaturity of the systems and standards indicated by those that did respond, point to a serious need to evaluate and address a lack of funding for operational capital and financial reserves even in our well-established and commonly used SCR environments.
- d. **Community Engagement and Governance.** Deeper evaluation into current community engagement and governance strategies is needed at an individual SCR-level, but the collated and aggregated results from the Census show that most SCRs are engaging in a range of community-building activities (especially via in-person activities). They also show that most SCRs have some type of leadership body or board, and that the processes of selecting these, as well as the tasks these leadership groups are assigned to complete, vary wildly. A smaller number of entities reported having formal by-laws or other structures to ensure transparency and accountability beyond the founders and/or staff of an organization. Particularly now that we have witnessed several relatively spectacular demise moments for SCRs that bore very little forewarning and no chance or opportunity for crisis management, transition, or reformation (e.g., Digital Preservation Network, or DPN), we must work harder to ensure that governance bodies regularly evaluate the financial health of the organizations they are empowered to serve, and that external

structures help to train both these Boards and staff members to do functions (e.g., accounting for revenues, not just expenditures) that simply are not business-as-usual within most academic environments.



## Appendices

### APPENDIX A: CENSUS (FULL)

This census, developed in the Andrew W. Mellon-funded “Mapping the Scholarly Communication Infrastructure” project (Middlebury, 2018-2019) in partnership with the Joint Roadmap for Open Science Tools (JROST), assesses factors influencing the sustainability and “fit-for-purpose” of Scholarly Communications Resources (SCRs) - tools, services, and systems. It is intended to help to guide investments in scholarly communications infrastructures. We anticipate 75-125 respondents to this survey.

**Note:** This PDF contains ALL questions. Actual takers of the census will not see all of the questions based on how they answer particular questions. In other words, the census, while extensive, is not as long as it appears in this PDF.

Each respondent will receive a results dashboard for their own entry, as well as benchmarking against the aggregate baseline data. The first 25 respondents will also receive feedback on their current status, with guidance for strengthening and continuing to mature their SCR’s infrastructure, as well as their organization, governance, and finance. We also anticipate that this information will lead to future discoverability opportunities with a broad range of prospective funders (foundation, government, institution).

We expect this survey to take between 1.5-3 hours to complete, including the time spent researching and/or asking questions of other members of your SCR team. Before you begin, we recommend that you have available some of the documents listed below.

Please note that we know there is an American and English-speaking bias to the current census. Those answering from other geographical/national contexts, please help us improve future iterations by citing the appropriate categories for your context in the open text boxes provided.

List of documents you will be glad to have in front of you....

*(Note: most SCRs will only have some of these documents; these are not in any way required for participation)*

1. Incorporation documents, hosting contracts, and/or applications for particular business statuses, e.g., Benefit Corp or 501c3 applications
2. Mission, vision, and values for your SCR
3. Strategic plan
4. Market Analysis
5. Code of Conduct or Community Standards documentation
6. Budgets for 2015, 2016, 2017
7. Final revenues, expenditures, and net numbers for 2015, 2016, 2017 (e.g., US 990 tax return or your local equivalent)
8. Annual report (or an annual report from your host institution that includes your SCR)
9. A report you deliver to your stakeholders
10. Conflict of Interest policy

## Data Privacy

Individual and aggregate data and documents shared by respondents will only be used for the following purposes and under the following circumstances:

1. Analysis by the “Mapping the Scholarly Infrastructure” (<https://scholarlycommons.net/map-plan/>) project team and the Joint Roadmap for Open Science Tools (JROST) (<https://jrost.org/>) project team.
2. Delivery to the respondent of their own SCR’s data output (including a dashboard view of that data)
3. Aggregated data will be analyzed, reported on, and offered as an open, anonymized dataset for reuse (no individual SCR’s data will be identifiable)
4. SCR names of all responding SCRs will be collated and included with the dataset; individual respondent names will not be shared

*Data for which additional permissions are granted:*

1. Respondents can explicitly grant permissions to the research team to do the following:
2. Share their individual SCR’s response publicly
3. Share their individual SCR’s response to specific other groups conducting related surveys:  
Global Sustainability Coalition for Open Science Services (SCOSS), OSS Landscape Scan effort (MIT/SFU)

**As per our data policies, we will not share your responses with any other group without your explicit, written permission.**

\* 1. I understand and agree to participate in this survey

Yes

No

\* 2. Name of the Scholarly Communication Resource (SCR)

\* 3. Website Address(es) and codebase link(s)

1.

2.

3.

4.

\* 4. Respondent Name

\* 5. Respondent Title/Role

\* 6. Respondent Email

**Please tell us a few details about the SCR's legal structure...**

\* 7. Is the SCR standalone or hosted?

- Standalone incorporated entity
- Hosted by another entity (please name the host entity and note the SCR's relationship to this host - e.g. university, larger technology organization, fiscal sponsor or otherwise)

\* 8. What is the SCR's profit orientation? (Answer for the hosting entity if the SCR is hosted.)

- For profit
- Non-profit

\* 9. What is the SCR's legal structure? (Answer for the hosting entity if the SCR is hosted.)

- Corporation, for-profit or non-profit
- LLC, for-profit (or equivalent)
- Benefit corporation, for-profit (or CIC in the UK)
- Foundation, non-profit
- Academic institution
- Other (specify)

\* 10. What is the SCR's tax status? (Answer for the hosting entity if the SCR is hosted.)

- Non-taxable
- Taxable (for profit)

**If non-taxable...**

\* 11. Does the SCR have an international or US-based non-taxable tax status? (Answer for the hosting entity if the SCR is hosted.)

- US IRS tax determination
- International non-profit tax status

**If US-based non-taxable status...**

\* 12. What is the SCR's US IRS tax determination? Select all that apply. (Answer for the hosting entity if the SCR is hosted.)

- 501.c3
- 501.c4
- The SCR has applied for the designated status, but has not yet received this status
- Other (specify)

**If international non-profit tax status...**

\* 13. Please indicate the SCR's tax status and national context. Note whether the SCR has *applied for* or *received* this status. (Answer for the hosting entity if the SCR is hosted.)



\* 14. Is the SCR a B-corp?

No

Yes

**If B-corp...**

\* 15. Has the SCR been B-corp certified?

No

Yes

\* 16. Is the SCR fiscally-sponsored? (Answer for the hosting entity if the SCR is hosted.)

No

Yes (please indicate which organization serves as the SCR's fiscal sponsor)

**As per our data policies, we will not share your responses with any other group without your explicit, written permission.**

\* 17. Would you like to share your data in any of the following ways?

- [DEFAULT] Please do not share or report on my individual SCR data response in any public-facing report or vehicle
- Please share my individual SCR data response as open data
- Please share my individual SCR data with the Global Sustainability Coalition for Open Science Services (SCOSS) (<http://scoss.org/>) for inclusion in their dataset
- Please share my individual SCR data with the Andrew W. Mellon Foundation funded OSS Landscape Scan effort (MIT/SFU) (<https://mitpress.mit.edu/blog/mit-receives-grant-conduct-environmental-scan-open-source-publishing>)

\* 18. Provide a succinct description of this SCR (100 words or so, including the SCR's target audience).

\* 19. What part(s) of the research lifecycle does this SCR serve? Select all that apply.

- |   |  |
|---|--|
| <input type="checkbox"/> Creation - Data gathering                          | <input type="checkbox"/> Hosting and access - Preprint service   |
| <input type="checkbox"/> Creation - Data analysis                           | <input type="checkbox"/> Discovery - Persistent identifiers      |
| <input type="checkbox"/> Creation - Experimentation                         | <input type="checkbox"/> Discovery - Identity (e.g. ORCID)       |
| <input type="checkbox"/> Creation - Computation and machine learning        | <input type="checkbox"/> Discovery - Citation                    |
| <input type="checkbox"/> Publishing - Submission                            | <input type="checkbox"/> Discovery - Aggregating and indexing    |
| <input type="checkbox"/> Publishing - Review                                | <input type="checkbox"/> Discovery - Search                      |
| <input type="checkbox"/> Publishing - Copyediting                           | <input type="checkbox"/> Evaluation and comment - Review         |
| <input type="checkbox"/> Publishing - Design                                | <input type="checkbox"/> Evaluation and comment - Annotation     |
| <input type="checkbox"/> Publishing - Layout                                | <input type="checkbox"/> Evaluation and comment - User analytics |
| <input type="checkbox"/> Hosting and access - Data services                 | <input type="checkbox"/> Archiving/preservation - Curation       |
| <input type="checkbox"/> Hosting and access - Journal hosting system        | <input type="checkbox"/> Archiving/preservation - Replication    |
| <input type="checkbox"/> Hosting and access - Monographic production system | <input type="checkbox"/> Archiving/preservation - Storage        |
| <input type="checkbox"/> Other (specify)                                    |  |

\* 20. What is the SCR's primary output(s)?

\* 21. Please estimate a count of total objects the SCR has created, published, hosted, provided discovery of, evaluated, and/or archived (e.g. number of preprints, annotations, identifiers, documents crawled and indexed, etc.).

Created

Published

Hosted

Provided discovery of

Evaluated

Archived

22. Does the SCR have a documented mission, vision, and/or values?

No

Yes

**If there is a documented mission, vision, and/or values...**

23. How were these created?

24. Who participated?

25. Do you have a review process?

- No
- Yes

26. When were these last updated?

27. If available, please share.

No file chosen



28. Does the SCR have (or has it had in the past) a formal strategic plan?

No

Yes

**If there is/was a strategic plan...**

29. Does/did it include measurable evaluation criteria?

- No  
 Yes

30. Do you have a review process?

- No  
 Yes

31. If available, please share.

Choose File

No file chosen

32. When was a market analysis last conducted?

33. If it is available, please share.

Choose File

No file chosen

34. If not, please describe its form and focus.

35. With which of the following types of stakeholders does the SCR consistently engage? Select all that apply.

- |   |  |
|---|--|
| <input type="checkbox"/> End users                  | <input type="checkbox"/> Members                           |
| <input type="checkbox"/> Implementers (OS software) | <input type="checkbox"/> Advisory or governing board       |
| <input type="checkbox"/> Service providers          | <input type="checkbox"/> Code contributors/committers      |
| <input type="checkbox"/> Publishers                 | <input type="checkbox"/> Volunteers (non-development work) |
| <input type="checkbox"/> Funders                    |  |
| <input type="checkbox"/> Other (specify)            |  |

36. Does the SCR have a code of conduct or other written community standards?

No

Yes

**If there is a code of conduct or other written community standards...**

37. How was this developed?

38. Who participated?

39. Does it include enforcement guidelines?

40. Has the code of conduct ever needed to be enforced?

41. If available, please share a copy of the code of conduct.

Choose File

No file chosen

42. Please describe the SCR's technology environment (e.g., what comprises its technology stack, what coding languages does it use, what does its server infrastructure look like, etc.)

43. Do in-house staff manage the code for this SCR?

No

Yes

44. Approximately how many coders currently contribute to the code base?

Number of volunteers

Number of paid staff

45. Is the SCR's code installed by other organizations on their servers?

- No
- Yes (Approximately how many organizations host it? If you don't know, just enter "unknown.")

46. Do third-party providers support and/or host the SCR?

- No
- Yes (Approximately how many organizations support and/or host it? If you don't know, just enter "unknown.")

47. Does the SCR host an instance that others can subscribe to or use?

- No
- Yes (Approximately how many total users and monthly average users does this instance serve? If you don't know, just enter "unknown.")

48. What is the pathway for getting code included in the code base?



49. How are updates managed, and how often?

50. When were the last two releases?

1.

2.

51. What software license does the SCR use?

52. Does the SCR adhere to open standards? Select the most accurate response.

- No
- Yes, some of the time
- Yes, all of the time

53. Please list any relevant open standards that the SCR adheres to (e.g. Web Annotation, OAuth, etc.).

54. Does the SCR have a product development roadmap?

No

Yes

**If there is a product development roadmap...**

55. Is it publicly available?

No

Yes (Please include a link)

56. How it is developed and maintained?

57. With whom is it shared?

58. How are new features prioritized?

59. Does the SCR code integrate with or interoperate with other SCR systems?

- No
- Yes (please list the SCR system(s) with which the SCR code interoperates or integrates)

60. Does the SCR have mechanisms in place (e.g. APIs, export functions, etc.) to assist users in migrating off of its infrastructure?

- No
- Yes (please briefly describe)

61. Where would you characterize the SCR currently?

- Pre-prototype
- Prototype
- Alpha
- Other (specify)
- Beta
- Live
- No longer supported

62. When was the last time you formally evaluated your technical environment?

- Within the last year
- Within the last five years
- Never

63. Have you taken measures to ensure the long-term preservation of the data and/or software associated with the SCR?

- No
- Yes

**If yes...**

64. Please explain what measures were taken to ensure the long-term preservation of the data and/or software associated with the SCR and why this was done.

**If no...**

65. Please explain why measures have not been taken to ensure the long-term preservation of the data and/or software associated with the SCR.

66. Approximately how many FTE (including partial effort, e.g. "0.7 FTE" or "1.2 FTE") currently support the SCR in each of the following categories? Include developers, help desk, fundraising, administrative, etc.

Staff paid directly through SCR earnings (non-grant funded):

Staff paid through grant funding:

Staff paid through in-kind contributions:

Other:

67. Please list the job title(s) for the SCR's current employee(s). Use a semicolon (;) to separate titles in a list.

Staff paid directly through SCR earnings (non-grant funded):

Staff paid through grant funding:

Staff paid through in-kind contributions:

Other:

68. Does the SCR staff include any remote employees?

- No
- Yes (Approximately what percentage of staff are remote-only?)

69. Which of the following documents are in place for this SCR? Select all that apply.

- Organizational chart
- Employee handbook
- Job definitions for each contributing employee
- Professional development budget for paid staff members
- Regular staff evaluations

70. Does the SCR begin each fiscal year with an approved budget?

No

Yes



**If yes...**

71. Please briefly describe the SCR's budget approval process. When does it occur? Who is involved?

**If no...**

72. Please explain why the SCR does not begin each fiscal year with an approved budget.

2018 Financials

**List the SCR's total budget and total revenues/expenditures in USD (as reported on a 990 or to host your institution) for 2018.**

73. 2018 Budget

Revenue

Expenditures

Net

74. 2018 Actual

Revenue

Expenditures

Net

75. If the SCR's total revenues and expenditures differed from the total budget by more than 20% in 2018, please explain.

2017 Financials

**List the SCR's total budget and total revenues/expenditures in USD (as reported on a 990 or to host your institution) for 2017.**

76. 2017 Budget

Revenue

Expenditures

Net

77. 2017 Actual

Revenue

Expenditures

Net

78. If the SCR's total revenues and expenditures differed from the total budget by more than 20% in 2017, please explain.

2016 Financials

**List the SCR's total budget and total revenues/expenditures in USD (as reported on a 990 or to host your institution) for 2016.**

79. 2016 Budget

Revenue

Expenditures

Net

80. 2016 Actual

Revenue

Expenditures

Net

81. If the SCR's total revenues and expenditures differed from the total budget by more than 20% in 2016, please explain.

82. What revenue streams support the SCR? Select all that apply.

- |   |   |
|---|---|
| <input type="checkbox"/> Membership fees                                | <input type="checkbox"/> High net worth individuals or family foundations   |
| <input type="checkbox"/> Subscription fees                              | <input type="checkbox"/> Donations (e.g. from corporations; NOT from high net work individuals or family foundations) |
| <input type="checkbox"/> Service fees                                   | <input type="checkbox"/> Contracts/consulting   |
| <input type="checkbox"/> Private foundation grants                      | <input type="checkbox"/> In-kind contributions  |
| <input type="checkbox"/> Government grants                              | <input type="checkbox"/> Host subsidy (including host in-kind contributions)  |
| <input type="checkbox"/> Subcontracts on grants to partner institutions |   |
| <input type="checkbox"/> Other (specify)                                |   |

83. Provide the average percentage for each revenue stream.

Membership fees	<input type="text"/>
Subscription fees	<input type="text"/>
Service fees	<input type="text"/>
Private foundation grants	<input type="text"/>
Government grants	<input type="text"/>
Subcontracts on grants to partner institutions	<input type="text"/>
High net worth individuals or family foundations	<input type="text"/>
Donations (e.g. from corporations; NOT from high net work individuals or family foundations)	<input type="text"/>
Contracts/consulting	<input type="text"/>
In-kind contributions	<input type="text"/>
Host subsidy (including host in-kind contributions)	<input type="text"/>
[Insert text from Other]	<input type="text"/>

84. What expenditures support this SCR?

- |   |   |
|---|---|
| <input type="checkbox"/> Salaries/Benefits      | <input type="checkbox"/> Hosted computing costs (e.g. AWS, Azure, Google compute, etc.) |
| <input type="checkbox"/> Rent and utilities     | <input type="checkbox"/> Marketing and advertising                                      |
| <input type="checkbox"/> Equipment              | <input type="checkbox"/> Fundraising  |
| <input type="checkbox"/> Supplies               | <input type="checkbox"/> Travel and meetings  |
| <input type="checkbox"/> Telecommunications     | <input type="checkbox"/> Debt service   |
| <input type="checkbox"/> Other (please specify) |   |



85. Provide the average percentage for each expenditure.

Salaries/Benefits	<input type="text"/>
Rent and utilities	<input type="text"/>
Equipment	<input type="text"/>
Supplies	<input type="text"/>
Telecommunications	<input type="text"/>
Hosted computing costs (e.g. AWS, Azure, Google compute, etc.)	<input type="text"/>
Marketing and advertising	<input type="text"/>
Fundraising	<input type="text"/>
Travel and meetings	<input type="text"/>
Debt service	<input type="text"/>
[Insert text from Other]	<input type="text"/>

86. Which of the following does the SCR maintain? Select all that apply.

- |  |   |
|--|---|
| <input type="checkbox"/> Bank account(s)   | <input type="checkbox"/> Monthly accounting rectifications                |
| <input type="checkbox"/> Earned income (e.g. dues, fees, other revenue)  | <input type="checkbox"/> Cash-based accounting                            |
| <input type="checkbox"/> Chart of Accounts   | <input type="checkbox"/> Accrual accounting                               |
| <input type="checkbox"/> Accounting Manual   | <input type="checkbox"/> GAAP compliance                                  |
| <input type="checkbox"/> Accounting software   | <input type="checkbox"/> Annual fiscal audits                             |
| <input type="checkbox"/> Review systems ensuring two people sign off on each transaction   | <input type="checkbox"/> Annual financial reports available to the public |
| <input type="checkbox"/> Financial reserves (Approximately how many months of operations are supported by the SCR's current financial reserves?) |   |

87. Without grant funding, how long could the SCR remain viable?

- Our SCR does not rely on grant funding
- Less than a month
- Less than six months
- Less than a year
- More than a year

88. Which of the following best describes the SCR's current reality? Choose as many as apply.

- Operational deficit is covered by grants or sponsored projects
- It operates with a break-even cash flow from earned income
- It generates a surplus
- Other (specify)

89. Which of the following best describes the SCR's aspirations?

- Operational deficit is covered by grants or sponsored projects
- It operates with a break-even cash flow from earned income
- It generates a surplus
- Other (specify)

90. Which of the following does the SCR use to engage stakeholders? Choose *Active*, *Inactive*, or *None* from the dropdown menu. (Choose *Inactive* if you have the tool or platform but haven't used it in the last year.)

Listserv(s)	<input type="text"/>
Social media	<input type="text"/>
Website/blog	<input type="text"/>
Web-based discussion or support forum	<input type="text"/>
Live, online events (e.g. webinars, community calls)	<input type="text"/>
In-person events	<input type="text"/>
Other	<input type="text"/>

Actions: Engagement

91. Does the SCR regularly report to the SCR's internal and/or external stakeholders on the organization's activities and finances?

- No
- Yes

**If sending regular reports...**

92. How frequently are these produced and distributed (e.g. monthly, quarterly)?

93. Who receives these?

94. Are they publicly available?

- No  
 Yes

95. What types of information are included?

96. If available, share an example.

Choose File

No file chosen

97. Do you produce an annual report for this SCR?

- No
- Yes
- The SCR is included in an umbrella organization's annual report

**If producing an annual report...**

98. If available, share an example.

Choose File

No file chosen



99. Does the SCR make use of volunteers (including committee members, elected officials, etc.)?

No

Yes

**If using volunteers...**

100. What do volunteers do?

101. Who manages volunteers?

102. Does the SCR have policies in place governing its use of volunteers?

- No
- Yes

103. When was the last time you surveyed your stakeholder communities about your SCR?

- Within the last year
- Within the last two years
- Within the last five years
- Never

104. Please briefly describe this process.

105. Which of the following does the SCR have?

- Bylaws or similar documentation of its governance structure
- A board and/or leadership group(s) overseeing its work

106. Please provide the group(s)' name(s), size(s), and briefly describe the role(s) and checks/balances for the group(s).

107. Which stakeholder group(s) is/are represented on the board and/or governance group(s)?

- Active users/community members
- Leaders in scholarly communications
- Software developers
- Service providers to the SCR
- Funders
- Other (specify)

108. Does the SCR have a formal succession plan, contingency plans, and/or escrow arrangements in place to ensure continuity of operations?

- No
- Yes

109. Has the SCR ever had to use its succession or contingency plan?

No

Yes

110. If available, please share a copy.

Choose File

No file chosen

Closing

111. How long did it take you to complete this survey?

- Less than an hour
- 1-2 hours
- 2-3 hours
- More than 3 hours

112. If you would like to provide additional information about this SCR, or feedback regarding this survey, please do so here.



## APPENDIX B: CENSUS (ABRIDGED)

\* 1. Name of the Scholarly Communication Resource (SCR)

2. Contact's first name

3. Contact's last name

4. Contact's email

5. Website Address(es) and codebase link(s)

1.

2.

3.

4.

6. Office Location

7. Year Founded

8. Please provide a succinct description of the SCR (approx. 100 words or so, including the SCR's target audience):

9. What is the SCR's profit orientation? (Answer for the hosting entity if the SCR is hosted.)

- For profit
- Non-profit

10. What part(s) of the research lifecycle does the SCR serve? Select all that apply.

- Creation - Data gathering
- Creation - Data analysis
- Creation - Experimentation
- Creation - Computation and machine learning
- Publishing - Submission
- Publishing - Review
- Publishing - Copyediting
- Publishing - Design
- Publishing - Layout
- Hosting and Access - Data services
- Hosting and Access - Journal hosting system
- Hosting and Access - Monographic production system
- Hosting and Access - Preprint service
- Discovery - Persistent identifiers
- Discovery - Identity (e.g. ORCID)
- Discovery - Citation
- Discovery - Aggregating and indexing
- Discovery - Search
- Evaluation and Comment - Review
- Evaluation and Comment - Annotation
- Evaluation and Comment - User analytics
- Archiving/Preservation - Curation
- Archiving/Preservation - Replication
- Archiving/Preservation - Storage
- Other (specify)

11. Additional comments

## APPENDIX C: DATA MODEL

# SCIP Census: Conceptual Model

## IMPACT

- System-level:
  - Durable, scalable, and long lasting open scientific and scholarly infrastructure delivers benefits on a global scale
  - Critical groundwork that will inform ecosystem-level work to strengthen scholarly infrastructure
- SCR-level:
  - Healthy maturation through lifecycle stages
  - Users/stakeholders can trust their investment in the SCR
  - SCR contributes to the creation of a integrated, end-to-end, robust, sustainable infrastructure

## OUTCOMES

### VISION OUTCOMES

- A strong vision and roadmap for growth, including:
  - a well-defined identity
  - user community-informed growth goals
  - tracked progress against goals
  - known relationship to adjacent work underway in the broader landscape
- Stakeholders (including funders, community members, and end users) are identified
- SCR goals and activities are aligned with stakeholder interests and needs

### INFRASTRUCTURE/DESIGN OUTCOMES

- Technology is current and appropriate
- Technical debt is low
- There is a roadmap to avoid obsolescence
- Integration with other systems to allow for interoperability
  - *Optional/won't apply to all SCRs*
- Data/content produced by the SCR is secure and available for long-term use
  - *Optional/won't apply to all SCRs*

### ADMIN/FINANCE/HR OUTCOMES

- Cost of operations is known
- Fiscal and organizational models and sustainability planning are appropriate
- SCR has adequate financial and staff resources
- Financial resources are appropriately matched to function (stable funding for ongoing operations, term-limited funding for term-limited projects)
- SCR is fiscally transparent and includes appropriate checks and balances
- Paid staff are adequately supported (e.g. through training) and appropriately evaluated

### **ENGAGEMENT OUTCOMES**

- Efficient, systematic communication keeps SCR stakeholders well informed about developments, activities, and needs
- Stakeholders understand and articulate the goals and work of the SCR, its growth trajectory, and the ROI they receive from it
- Stakeholders from different groups are aligned in their understanding and activities related to the SCR
- Stakeholders have channels through which to engage with the SCR; opportunities and expectations for engagement are clear; stakeholder engagement is valued and contributions are recognized

### **GOVERNANCE OUTCOMES**

- Decision-making processes are clearly defined and well-informed by data and stakeholder needs/input
- Decision-making processes are transparent and stakeholders understand how decisions are made
- Checks and balances are in place to ensure accountability to stakeholders

## APPENDIX D: SCHOLARLY COMMUNICATION INFRASTRUCTURE PROVIDERS CENSUS PROJECT TEAM

### Mapping the Scholarly Communication Infrastructure Project:

Katherine Skinner

Melanie Schlosser

Nathan Brown

Brianna Morrow

Mike Roy

David Lewis

### Joint Roadmap for Open Science Tools:

Dan Whaley

Peg Fowler

### Invest in Open Infrastructure:

Raym Crow

Heather Joseph

Pierre Mounier

Vanessa Proudman

Kristen Ratan

Danielle Robinson