



DATA CURATION NETWORK SUSTAINABILITY PLAN FINAL REPORT



Laurie Arp, Tom Clareson, & Carissa Egan
LYRISIS | February 2020

Table of Contents

DATA CURATION NETWORK SUSTAINABILITY PLAN OVERVIEW	2
SUMMARY PHASE RECOMMENDATIONS	3
PHASE 1: MARKET ANALYSIS	3
<i>In Person Focus Group</i>	3
<i>Summary: Report Back to Full Group</i>	3
<i>DCN Stakeholder Phone Interview: Executive Summary</i>	3
PHASE 2: SUSTAINABILITY PATH MODELS	4
<i>Summary</i>	4
PHASE 3: FRAMEWORK RECOMMENDATIONS	5
<i>Summary</i>	5
OVERALL SUMMARY RECOMMENDATIONS	6
APPENDICES	6
APPENDIX A	7
APPENDIX B	10
APPENDIX C	16
APPENDIX D	27
APPENDIX E	43

Data Curation Network Sustainability Plan Overview

This final report is intended to summarize the various phases, recommendations and provide one overall document that reflects the Data Curation Network's sustainability planning efforts.

The Data Curation Network (DCN) engaged with LYRASIS in June 2019 to assist the DCN in their sustainability planning efforts. The proposal to DCN outlined the main components as:

- Phase 1: Market Analysis
- Phase 2: Financial Model Recommendations
- Phase 3: Administrative Structure

Phase 1 focused on an assessment of the current landscape and generated ideas and feedback from the community. We were pleased to be able to conduct an in person focus group with DCN stakeholders in conjunction with the DCN all hands meeting that took place in July 2019. This gave us strong and divergent feedback that informed future activities. Specifically, it led us to adjust direction for the one-on-one interviews and to shift the original structure of the subsequent phases. We proposed and gained agreement to consolidate the original Phase 2 (financial modeling section) with Phase 3 (the organizational home section) into one section on “sustainability path models” where we could reflect on the feedback from the focus group and interviews and offer several possible avenues forward to consider depending on the consensus of the group (i.e. grow aggressively/ expand or stay small/strategically pick partners).

An expanded Phase 2 focused on development of potential Sustainability Path Models for DCN based on Phase 1 feedback. The models and spreadsheets were submitted in December 2019. Phase 3 focused on recommending a testing framework to evaluate the models presented in the previous phase in order to build on them. The framework can be used to assess what was developed in Phase 2 to shape the models into what fits best for DCN now and into the future. The output may result in the determination that the models presented should be turned into a phased approach and/or each model should be combined with another model to create a hybrid avenue toward sustainability.

Summaries and highlights of each phase are outlined below. Reports and materials for each phase are also provided as appendices.

Summary Phase Recommendations

Phase 1: Market Analysis

In Person Focus Group Summary: Report Back to Full Group

1. Success is not necessarily national (a good result could be just these 10 with strong commitments, vision and practices) or being a model for others to follow
 - a. Need to define ROI/measure success
2. Identify expertise gaps now – better to understand and fill strategically as grow
3. Refine processes – small number of data sets curated – why? Fill capacity before growing
4. Are the 10 current DCN participants committed beyond Sloan grant funding?
5. Grow strategically – match expertise to grow vs seeking different institution types (non R1, etc.)
6. Strategic partnerships – avoid landmines (Figshare causes strong concerns among some; don't have to say yes to every opportunity)
7. Define vision – identify/values
 - a. What do they want to be:
 - i. Training?
 - ii. Advocacy?
 - iii. Risk taking?
 - iv. All institutions? Similarity resourced institutions? Existing networks?
8. Strong value in tools, training, QA
9. Currently unclear commitment to data curation at the highest level (deans vs library)
10. There are many questions, but these are answerable (great work thus far!)
11. Need to create a different “story” for different audiences – ROI for leader vs library
12. Slow down – too early for super specific business plan. Don't stop sustainability planning overall, just don't expect details to be clear at this point.

DCN Stakeholder Phone Interview: Executive Summary

As part of the research into the Data Curation Network's audience needs, LYRASIS Consultant Tom Claeson conducted telephone interviews with eight leaders in the field/potential stakeholders during August 2019. Key trends from the discussions included:

- All of the interviewees were positive about the work of the DCN and its efficacy as a model for collaborative library work in the future.
- A variety of future financial models for DCN were supported by the interviewees, but most said they needed to know how the funding would be used. Seven of eight organizations interviewed expressed interest in becoming involved with DCN, but only three suggested potential membership fee levels.
- Interest in the topic of data curation is growing quickly in both the library community and the academic community overall. Data persistence and accessibility are some of the most important issues in data curation.
- Interviewees named 22 other stakeholders and leading institutions in the data curation environment; suggested 11 possible organizations for DCN to partner with and six competitors; and mentioned 12 organizations as potential organizational homes.
- Human and budgetary constraints were most often mentioned as barriers to DCN participation, particularly the current lack of capacity to contribute in-kind to the initiative at some campuses.

Phase 2: Sustainability Path Models

Summary

As a result of the LYRAsIS-led focus group at the Data Curation Network all hands meeting and subsequent phone interviews with key stakeholders, three main sustainability options rose to the surface with two alternative models also briefly mentioned.

Pathway Model Options

1. Tiered Partner Program
 - a. Institutions of various sizes would become DCN partners at a tier with an associated annual partner fee
2. Stakeholder Program (Phased)
 - a. Focus a program around the ten grant partners and, potentially, a very small group of closely aligned stakeholder institutions
3. Fee For Service
 - a. Would allow DCN to acquire revenue in a means not tied to a model exclusively focused on annual, ongoing contributions
4. Other
 - a. Merger Model
 - i. Potential integration into an existing nonprofit
 - b. Individual Contributor Program
 - i. Service where subscribers would be matched to expertise

Based on these findings, sustainability pathway models were developed to further detail how these potential avenues for sustainability *could* be implemented in order for the Data Curation Network to transition from being a grant funded project into a self-sufficient organization.

It should be noted that there was a tension uncovered in the focus groups and phone interviews in two main opposing sustainability directions/approaches that the Data Curation Network could take. These viewpoints are summarized in the following quotes:

- “Go slow, build on what's working. Don't be too ambitious. We cannot imagine not having this service organization in the future.”
- “Don't get overly insular – there is a tension between local and global issues.”

Based on the resounding number of comments falling into one of the two categories above, it is likely that a hybrid approach of the above models is the best fit for Data Curation Network going forward.

The sustainability pathways models presented above are meant to serve as starting points as the Data Curation Network transitions from grant funding. No one is model is recommended.

It was recommended that the Data Curation Network test each model with current and future stakeholders to further refine or refute details in the above sustainability pathways. Sustainability is not a one-time project that once accomplished, can be considered completed; instead, sustainability is an ever-evolving puzzle that should be reviewed and adjusted on a regular basis. The Data Curation Network will need to continue to assess and adjust whatever sustainability pathway(s) it determines is best as often as necessary in order to achieve true long-term sustainability.

[Phase 3: Framework Recommendations](#)

Summary

Phase 3 recommended a testing framework to evaluate the models presented in the previous phase in order to build on them. The framework can be used to assess what was developed in Phase 2 to shape the models into what fits best for DCN now and into the future. The output may result in the determination that the models presented should be turned into a phased approach and/or each model should be combined with another model to create a hybrid avenue toward sustainability.

Overall Summary Recommendations

We have appreciated the opportunity to work with the DCN participants in their sustainability efforts and are impressed with the strength of their achievements thus far. In our work together we have seen the dedication and commitment to a shared network of passionate professionals seeking to find innovative, scalable, long term solutions to data curation.

There is a strong sense of community and shared values around the Data Curation Network. The innovative work and collaborative relationships have served the participants in these early stages well. It will be fundamental to determine a path forward that is sustainable but that also is able to incorporate these key pillars of the DCN's early success. We strongly encourage the current participating representatives to narrow in on a few key themes to guide your work: agree on your overall sustainability goals, grow thoughtfully, pilot new ideas, and go at the pace that serves your goals and audience. As you go forward together, those thoughtful, shared goals can serve as a "north star" to guide you when you will undoubtedly need to make difficult decisions in the future.

Appendices

The various materials and reports produced during this work are included as appendices.

- A. Data Curation Network Stakeholder Focus Group Session Discussion Guide – Facilitator Version, July 17, 2019, 1pm-4pm
- B. Data Curation Network (DCN) Sustainability Focus Group Meeting Notes, Minneapolis, July 17, 2019
- C. Data Curation Network Stakeholder Phone Interview Report, August 29, 2019
- D. Data Curation Network Sustainability Pathways, November 2019
- E. Data Curation Network Sustainability Testing Framework Recommendations, January 2020

Appendix A

Data Curation Network Stakeholder Focus Group Session Discussion Guide – Facilitator Version July 17, 2019, 1pm-4pm

Introductory information:

- Background information on purpose of focus groups
- Introductions: participants and facilitators
- Agenda review
- Process Agreement

1. What is your interest in the activities of the Data Curation Network? Please describe your experience working with the group. [Facilitator note: Encourage each attendee to provide their summarized response in 2-3 minutes.]
2. What data curation activities is your organization currently involved in? Do you feel that your organization is committed to data curation?
3. What data curation activities does your organization consider to be most important? Why? What current services are the most important to sustain?
4. You are participating in this focus group as a representative of an area that could be considered as a current or future stakeholder of the DCN. Are there other organizations you would consider stakeholders or leaders in the field?
5. Talk about the interest/appetite for more and/or different services from the current DCN partner and stakeholder institutions.
6. Please give your thoughts on the interest in DCN services from the wider library/information/research community.
7. Give us your feedback on the current services and activities of DCN. How satisfied are you with the services? [Facilitator note: Ask each attendee to rank their response on a scale of 1-4 with 1 being the least satisfied and 4 being highly satisfied.]
8. Provide your feedback on this statement: DCN is currently providing the data curation services and activities that are most important to users such as researchers and current partner institutions.
9. Tell us about any overall participation constraints related to DCN services for the organizational area that you represent. [Facilitator note: Provide several examples of participation, such as financial contributions through membership dues, in-kind data contributions such as data curation expertise, other.]

10. What do you feel might be the overall participation constraints from the wider library/information/data communities?
11. What financial models make the most sense for DCN operations in the future? Can you give examples of models that work especially well in your organization area and are likely to be extensible to the DCN? Do you feel new participants would be more interested in a membership model, supporting DCN via in-kind contributions or, alternatively, with some other hybrid model with fee for services, ad hoc data curation consultancies, or something else? [Facilitator note: Briefly describe a few different models and an example of each, i.e. a membership model entails institutions paying an annual membership fee of some figure (this model is used by the ArchivesSpace project), etc.]
12. What is your opinion of DCN's current administrative structure? [Facilitator note: Ask each attendee to rank their response on a scale of 1-4 with 1 being the least satisfied and 4 being highly satisfied.]
13. In your opinion, how does the current staffing model work? What challenges could you foresee as an issue for your area? What changes are needed?
14. What administrative and operational changes do you see being needed in the next 5 years in areas such as community engagement, governance, finances, staffing, and legal support?
15. What additional partnerships do you feel make sense for DCN (administratively, financially, for service provision) in the future? Is DCN in direct competition with any other organization offering similar services (today or a few years from now)?
16. Does working with another organizational home entity make sense for DCN? Do you have suggestions for potential organizational homes? Or, is a fiscal agent or other type of contracted services model a better fit? Please keep in mind if there is a preference/ease for your institution to work with another university vs. a standalone organization. [Facilitator note: Briefly describe an organizational home as well as fiscal agent and provide examples, i.e. an organizational home is an entity that provides services such as payroll, human resources, accounting, etc. to a project in order for it to focus on day-to-day project-related activities (LYRISIS is the organizational home for ArchivesSpace, DuraSpace is the organizational home for Fedora and DSpace).
17. Talk about current DCN member community engagement. What would it take to get new members engaged?
18. What other activities/services are needed from DCN TODAY to achieve sustainability? Can any of these be fee-for-service activities? What activities/services should DCN not be doing in the FUTURE?

19. What else would you like to share with us today that you haven't already had an opportunity to say?

Data Curation Network (DCN) Sustainability Focus Group Meeting Notes

Minneapolis
July 17, 2019

Introductory info included:

- Background information on purpose of focus groups
- Introductions: participants and facilitators
- Agenda review
- Process agreement (results will be not generally individually identified, presented in aggregate)

NOTE: Questions and Question Numbers are from original discussion guide; not all questions were discussed during the session.

1. What is your interest in the activities of the Data Curation Network? Describe your experience working with the group.

- DCN participants in Circle of Friends (friendly with DCN folks) (2 votes)
- Nature of activities (curation synergies); “no brainer” to be involved
- Shared services/network – how to engage, especially small schools; digital curation
- Model (scalable) – curation, data driven, data re-use (2 votes)
- Community of practice/skill development
- Curation model – tools, output
- Potential DCN member
- Create positive change – network
- Need scale?
- Explore sustainable membership models – alternative models – level of investment – resources vs membership \$
- Feels DCN reflects 2nd wave of constituent building – can we do collective action, struggle with consortia services

2. (What data curation activities is your organization currently involved in?) Do you feel your institution is committed to digital curation?

[answers focused on second part of question]

- Different depending on level at institution; Department level commitment vs larger organizational commitment
 - o Dept – y; org, not yet

- Lib – y; org – indirect commitment (staffing); reflects in research and grants (2 votes)
- Repository (Yes); field (UL) “good” but committed? Need story for bigger level (2 votes)
 - Number of staff involved in data curation is good indicator
- Portage [confirmed okay to identify]– yes/ within new org
- Network – staffing at each
 - Curation as service; % of time
- Building abilities and relationships
- Lack of capacity to do
- Assumption of non R1 values match non R1s? Do AUL value needs?
- DCN scale of expertise – does everyone need to have these skills vs access to someone with these skills?
 - Windshield wiper model (buy wipers included installation – seller has knowledge and skills) but if you buy new ones once a year, you have lower skills

3. What data curation activities does your organization consider to be most important? Why? What current services are the most important to sustain?

- Upstream impact – make aware of issues early – documentation, organization – data migration
 - Not curation/ no IR at institution
 - Holistic discussion – awareness
- Advocacy
- Q: What is the role of institution and researcher (maybe moves)
- Set guidelines – and push details to disciplines
- Long term preservation
- Carpentry (data skills training)
- Dig preservation (cost vs expertise building – data vs digital objects)
- Weeding? Dig preservation vs. 5-year retention
- Pre-publication review
- Discoverability and use (re-use) (3 votes)
 - Auditability, Quality Assurance
- Reproducibility
 - Who on campus to influence/decide – lib/lib dean, presidents? All VERY different concerns and selling points
 - DIFFERENT value statements to target library vs president
 - Highlight different parts – target failure points
 - Tell a story
 - Values yes but → ROI important
 - What are Staff needs, \$, opportunity costs; or argument about network benefits and community

- Articulate saving plus training/community
- Larger advocacy – education, training community
- Vendor transaction – service
- Both larger advocacy and service important
 - But different message and value
- Library – local connection – hub for vs direct to vended solution (devaluing)
 - Faculty benefit – part of story
 - Funders invest in research – another audience
 - Able to recruit good faculty – better service?
 - Curation changes – funders require data migration plan

4. You are participating in this focus group as a representative of an area that could be considered as a current or future stakeholder of the DCN. Are there other organizations you would consider stakeholders or leaders in the field?

- OER (open educational resources) -- > teachers, students could work with data
- Reliable data is important
- Re-purpose – research, learning
- Different audience and level of partnership
- Curation services – referral – or rely on other
- Smaller institution – less frequency of need but part of network?

11. What financial models make the most sense for DCN operations in the future? Can you give examples of models that work especially well in your organization area and are likely to be extensible to the DCN? Do you feel new participants would be more interested in a membership model, supporting DCN via in-kind contributions or, alternatively, with some other hybrid model with fee for services, ad hoc data curation consultancies, or something else?

[the discussion didn't focus as much on models as what are the important elements of DCN and some potential ways and reasons other institutions or consortia could be involved]

- In kind services such as QA valuable
- Time vs \$ (which can you afford to give?)
- Does this program scale at 10 org – are current participants all committed beyond current grant funding?
- Prioritizes highly resourced institutions

- Prime tools valuable
- Outreach efforts valuable
- Strategic optimization - scale up as need, around core curation gaps
- Needs assessment - what is needed? Scale then.
- Core – membership and prime audience
- Training could be additional audience and revenue
- Core lab example
- Wealthier could contribute more; smaller get training and primers
- Canada – existing region, trust relationships and scale – instead of growing to incorporate them, perhaps they can use the DCN model there
- Endowment model – Gates Foundation as a possible funder

16. Does working with another organizational home entity make sense for DCN? Do you

have suggestions for potential organizational homes? Or, is a fiscal agent or other type of contracted services model a better fit? Please keep in mind if there is a preference/ease for your institution to work with another university vs. a standalone organization.

[discussion mostly continued on the previous model question]

- Core stay together once grant over?
 - o Don't have to grow – perhaps other versions such as Big 10 version
 - o Model, spin off to other groups
 - o End result doesn't have to be uber org
 - o National scale – good idea but decentralized
 - o DCN – needs to clarify purpose and identify - national? Different/other model? Big 10?
- Sense of DCN goal – want to be curation network that doesn't reinforce inequities (open to smaller resourced institutions)
- Seek specific expertise
- National survey of needs could be done
- AL – in terms of commitment prefer giving \$ over personnel
- In terms of interest in membership
 - o Need to clarify offer and see results (what would they be participating in?)
- Scale up if \$
- DPN/DPLA on minds of people – want to know won't happen again
 - o DPN – idea of “Greater good” and ability to deposit – but never deposited (limits)– if not using, why pay?
- Potential institutions might consider testing local demand and if they can't meet on campus – then utilize DCN
- Currently small numbers for network curation - No demand, no need? Lack of expertise in field– do in hope of getting funding?
- Potential for Network of networks?

- Get commitments from current group
- Grow expertise strategically
- Refine curation systems and build out expertise (areas not yet covered?)
- Survey – national too broad? Be careful about expectations
- Needs assessment to underrepresented?
- Not ready to be shared repository (too contentious now)

19. What else would you like to share with us today that you haven't already had an opportunity to say?

- Framing – commons vs vendor/market logic (libs don't do market things well)
- Who to survey/talk to ?
- Dryad – like minded, at scale, economics, full time (consider that are they already doing this service at scale)
- Need for level 1 tech checks

Summary: Report Back to Full Group

13. Success is not necessarily national (a good result could be just these 10 with strong commitments, vision and practices) or being a model for others to follow
 - a. Need to define ROI/measure success
14. Identify expertise gaps now – better to understand and fill strategically as grow
15. Refine processes – small # of data sets curated – why? Fill capacity before growing
16. Are the 10 current DCN participants committed beyond Sloan grant funding?
17. Grow strategically – match expertise to grow vs seeking different institution types (non r-1, etc.)
18. Strategic partnerships – avoid landmines (Figshare causes strong concerns among some); don't have to say yes to every opportunity)
19. Define vision – identify/ values
 - a. What do they want to be:
 - i. Training?
 - ii. Advocacy?
 - iii. Risk taking?
 - iv. All institutions? Similarity resourced institutions? Existing networks?
20. Strong value in tools, training, QA
21. Currently unclear commitment to data curation at the highest level (deans vs lib)

22. There are many questions, but these are answerable (great work thus far!)
23. Need to create a different “story” for different audiences – ROI for leader vs library
24. Slow down – too early for super specific business plan. Don’t stop sustainability planning overall, just don’t expect details to be clear at this point.

[editorial notes in square brackets]

Data Curation Network Stakeholder Phone Interview Report August 29, 2019

Executive Summary

As part of the research into the Data Curation Network's audience needs, LYRASIS Consultant Tom Clareson conducted telephone interviews with eight leaders in the field/potential stakeholders during August 2019. Key trends from the discussions included:

- All of the interviewees were positive about the work of the DCN and its efficacy as a model for collaborative library work in the future.
- A variety of future financial models for DCN were supported by the interviewees, but most said they needed to know how the funding would be used. Seven of eight organizations interviewed expressed interest in becoming involved with DCN, but only three suggested potential membership fee levels.
- Interest in the topic of data curation is growing quickly in both the library community and the academic community overall. Data persistence and accessibility are some of the most important issues in data curation.
- Interviewees named 22 other stakeholders and leading institutions in the data curation environment; suggested 11 possible organizations for DCN to partner with and six competitors; and mentioned 12 organizations as potential organizational homes.
- Human and budgetary constraints were most often mentioned as barriers to DCN participation, particularly the current lack of capacity to contribute in-kind to the initiative at some campuses.

Project Background

As part of the Data Curation Network Sustainability Study project, LYRASIS Consultant Tom Clareson developed a Phone Interview Discussion Guide based in part on the Focus Group Discussion Guide used with the DCN Advisory Focus Group on July 17, 2019. Clareson utilized the questionnaire in calls with eight stakeholders identified by DCN project leaders. The phone interviews, each 45 minutes to one hour long, were held between August 1-23, 2019. Information gained in response to each question is listed below; a summary section that reviews overall trends appears at the end of the report.

Participants in the DCN Stakeholder Interview project were (in order of interview date):

- Joe Lucia, Temple University
- Lynn Yarmey, Research Data Alliance
- Kristin Antelman, University of California Santa Barbara
- Tyler Walters, Virginia Tech University
- Nora Mattern, University of Chicago/University of Pittsburgh SLIS
- Claire Stewart, University of Nebraska

- Jamie Wittenberg, Indiana University
- Hannah Sommers, George Washington University

The calls gathered excellent information, much of which helped to amplify the findings of the July 17 in person focus group. The questions and key responses are included below.

Raw Data from Phone Interviewees

20. What is your interest in the activities of the Data Curation Network?

- Interviewees appreciated the way that DCN coordinates expertise.
- Respondents felt that community members could identify themselves as experts in a data format and ease the difficulties that others in the network are having with specific formats.
- Organizations with newer programs could learn about data curation and contribute to the DCN (if they join the initiative).
- The information “primers” are helpful ways for people to learn about data curation, both on the participant’s campus and in the overall library community.
- One interviewee sees data curation needs “hitting universities incredibly quickly;” are libraries prepared to pivot to meet demand? They felt that all libraries needing to have their own repository is a “doomed model,” (and that DCN was a more viable model).
- Interviewee is interested in “all things collaborative.”
- Another respondent said they “are interested in shared service models in libraries in general.
- Data curation activities are part of the scholarly and academic infrastructure.
- One organization said they need to build a more mature infrastructure in service to the data generated on their campus. The institution is currently “crossing a threshold where data curation is an activity we need to fund.”
- At one respondent’s campus, a number of participants in New Faculty Orientation had questions focusing on data curation services.

21. What data curation activities is your organization currently involved in? Do you feel that your organization is committed to data curation?

- One organization has a nine-person Data Services department and is very involved in data curation work on campus.
- Another organization plans to have a team of five academic professionals on this project by the end of 2019.
- Another organization has one data curation position and is adding another. The current position is working with IT and Research departments to get plans solidified, is offering data curation-related workshops, and is available to work with research teams on projects. There has been such a demand for data curation assistance from faculty on campus that the new person will be concentrating their work in this area.

- One organization said they are working with datasets that are so big they are not sure they can hold them on campus, and their I.T. department is also concerned about the size and scale of the datasets.
- Documentation of data curation is seen as important at participants' organizations.
- One caller noted the importance of DCN services to organizations that may not have a "go-to" repository.
- One organization is especially involved in education and outreach on data curation, and support for data management plans. Their data repository provides basic ingestion, DOIs, fixity checking. They are currently doing pilots with larger data centers on campus such as the Intermolecular Biocommunications and School of Natural Resources. However, capacity issues (smaller staff, data librarian budget line eliminated, overall institutional budget) have colored the resources available to do data curation work.
- Another organization has a Research Data Service on Campus and is bringing on a new repository this fall. They are based in the Library's Scholarly Communications Department and work with data curators and stewards.
- There is a need to show faculty the connections between the data-related services that are available to them, such as high-speed computing. This is part of being "good liaisons and ambassadors."

22. What data curation activities does your organization consider to be most important?

Why? What current services are the most important to sustain?

- Persistence of data and accessibility to it are most important.
- "Making sure research generated by faculty is accessible in the future."
- One organization is struggling with data being heterogenous and idiosyncratic – it does not fit well within their current collection policies.
- One participant asked "to what degree are problems (with data curation) discipline-based vs. structural or mechanical?"
- Broadening data curation expertise within their organization is most important. But, data curation activity is challenging, time-consuming, and intensive.
- Big data and how libraries can participate is a central issue.
- One participant's organization has strong humanities and music programs and is doing curation for these areas.
- While one of the institutions is not yet scaled up to accept data deposits, they try to send the materials to appropriate field repositories. DCN would have a great value to help organizations work better locally with data deposits from their institutions. The organization will never have an expert in every data format or area, but the Library would like to have wider expertise over several areas.

23. You are participating in this interview as a representative of an area that could be considered as a current or future stakeholder of the DCN. Are there other organizations you would consider stakeholders or leaders in the field?

- Purdue University Research Repository (PURR)
- Johns Hopkins University
- University of California System/California Digital Library (3 mentions)
- In the Health Sciences, leaders include New York University, the University of Maryland, and the National Library of Medicine
- ICPSR, National Library of Medicine Data Services – DCN needs to figure out a way to work with them even if the services don't overlap.
- The rapid growth of FigShare, Springer, and data curation services from commercial entities are getting organization's attention. Libraries have a "different and more compelling way to do these activities."
- National library organizations like ARL.
- Organizations with Genomics programs are leaders. University of California, San Francisco was mentioned.
- DataSite
- Data One
- Funders are important stakeholders.
- We need to think about what frameworks and incentives there are for working with stakeholders on campus such as Sponsored Research and Research Administration – it would be good if these offices could supply libraries with data management plans as the proposal is being developed and submitted (rather than after the fact).
- UNC-Chapel Hill
- Larger, state-affiliated organizations with high levels of maturity in data curation.
- One interviewee feels that DCN should be working with a wide variety of groups. Those mentioned include:
 - Publishers
 - RDA Funders Forum
 - Domains that are having conversations on data curation, including Earth Sciences
 - Societies including ACSO (clinical data)
 - Federal entities including the Department of Energy, Department of the Interior, NASA, and CDC

24. Tell us about any overall participation constraints related to DCN services for the organizational area that you represent.

- Resources are a constraint – can DCN let interested organizations know what the time commitments to participate might be?
- Interest in subject curation could constrain DCN (if DCN chooses to concentrate on only certain topics, or if competing subject curation efforts curb use of DCN in specific subject areas).
- Human resources-related constraints, such as the sharing of time of staff, could be a problem for unionized libraries
- Organizations need more information on:
 - Costs (time and money) to participate
 - Benefits

- How can we work together on a scaled solution?
- The participant's institution does not have the capacity to contribute in-kind, and may not be able to for a while at the 10% of one person level. But, they have interest and expertise, as well as strong connections to state and national legislatures.
- Interviewee said that some organizations might feel that data curation work needs to be done on their own campus.
- Are there any regulatory concerns about doing data curation between campuses?
- There have been some discussions on the appropriate role for libraries in data curation (should the researchers do it for themselves? – interviewee does not agree with this.)
- The data curation need is far bigger than what a single library or even an aggregation of libraries can provide.
- Working with both local issues and field-wide issues of data curation is “not an either/or, but a both/and.”
- Generally, people are struggling with being part of large, pay-in organizations. Concerns mentioned included membership fatigue and limited resources.
- The Dryad/DCN membership tension needs to get worked out. If DCN wants to be larger, they must do planning for organizations that are at different points of the data lifecycle and have different expectations.
- Membership cost could be a constraint
- Human Resources costs (time commitments make Library Administrators wary as it can take staff away from current activities).
- One library administrator is concerned about being “Twenty-thousand-dollar to death.” Paying at this level to participate in numerous initiatives is very challenging; this type of financial model sometimes leads organizations to start participating and withdraw later.
- The initial financial commitment can be challenging.
- Having a clear value proposition is important.
- The work of DCN requires faculty members at institutions to be creating datasets for use in curation activities.
- Bandwidth and skills are constraints, but these can be remedied with appropriate training. A discipline expert could become a good data curation contact.
- The academic community is bringing in more PhDs and resources who can work with data and guide the process; the library community needs to recognize these needs and resources and hire people to work in this area.

25. What financial models make the most sense for DCN operations in the future? Can you give examples of models that work especially well in your organization area and are likely to be extensible to the DCN? Do you feel new participants would be more interested in a membership model, supporting DCN via in-kind contributions or, alternatively, with some other hybrid model with fee for services, ad hoc data curation consultancies, or something else?

- Have both a member fee and members need to do a certain amount of curation work per month.
- The fee for service idea is good for smaller organizations, but how would it be structured? What would it look like? It needs to be affordable.
- The in-kind contribution model is “great” – many interviewees want it to continue. But, how deeply formalized is it, or should it be?
- “Membership fatigue is out there.”
- “Don’t do a standalone 501c3.”
- Could the DCN membership keep the network going, but the fee-for-service model help to build the program?
- Many libraries are doing data curation with only one staff person “covering a lot” so DCN’s services are beneficial.
- Two other multi-institutional organizations in related fields are struggling with their future business model planning as well.
- Participants would need to know “what the finances are supporting.”
- Could state university groups (such as University of California Libraries or University of Nebraska System Libraries) or other multi-institution groups come in as a consortium?
- One interviewee is interested in something hybrid. While they think membership is good, they also suggested “a premium stream on top, or pay-as-you-go,” especially for those who cannot commit to membership but want training or project-by-project help.
- If utilizing in-kind arrangement, “where is the end-point for skin in the game? Could it go away? Could curators get paid?”
- LYRISIS takes member organizations, and packages services together for a fee or offers membership at a packaged rate. This makes it easy for librarians and administrators.
- Another organization asked if DCN activities could be bundled with other services. The institution is part of LYRISIS, DuraSpace, and other open source projects, and wondered about a potential relationship for DCN. Having data curation services in an organizational home from which they already are receiving services is good. With LYRISIS, they are familiar with the management style.
- Advancing progress toward the library’s strategic role on campus is important.
- Other multi-institutional organizations are looking at tiered models based on Carnegie designation and student FTE numbers
- Smaller organizations, including other campuses or regionals in one respondent’s state would see it as a financial challenge to work in this initiative, even if there is a benefit. The respondent wants to see benefits of the program for full systems/across multiple campuses.
- When one of the respondents thinks about DCN’s value proposition, it is more about human expertise and creating a network. This leads to questions about what is sustainable – how do you keep a project like DCN going? What are paths to growth and sustainability.

- One organization wants to participate in-kind, but does not feel they are there with staffing or expertise at this point.

26. What additional partnerships do you feel make sense for DCN (administratively, financially, for service provision) in the future? Is DCN in direct competition with any other organization offering similar services (today or a few years from now)?

- Potential Partners:
 - Data Curation Centre, University of Edinburgh
 - OCLC (especially the Research Library Partnership participants)
 - Discipline-specific repositories
 - Domain Data Centers, such as Snow and Ice; ICPSR, Council of Data Facilities (earth sciences)
 - Society Archives
 - Professional Associations that researchers are connected into
 - ARL (3 mentions)
 - Groups like Figshare – “in order to be comprehensive, DCN needs to consider commercial partners,” said one participant.
 - Partnerships with library consortia such as Big Ten Academic Alliance (BTAA) that is already looking at library purchase data should be considered. Libraries also need support with curation of proprietary/licensed resources.
- Competitors:
 - Data One (although interviewee said this group is “also trying to figure themselves out.”)
 - OCLC (mentioned as both potential partner and competitor)
 - FigShare (mentioned as both potential partner and competitor)
 - Springer and Elsevier “worry” two of the participants. One said DCN needs to make a clear statement of their values on who they will partner with, because it would be off-putting to some potential members for DCN to work with commercial entities.
 - Also on the Springer/Elsevier topic, considering a comprehensive curation process vs. commercial ala carte services could put DCN into direct competition with some of the commercial providers. This could be a problem in DCN has higher costs.
 - Local data services may see DCN as a threat, even though the respondent saying this does not agree with that observation.

27. In the future, would your organization consider becoming involved with DCN (either as a member, curation contributor, contracting for specific curation services or other)? What level of fees might your organization be willing to pay?

- Seven organizations said yes, they would be interested/consider becoming a member, in particular after their specific questions were answered.
- Three interviewed organizations were willing to specify a dollar figure for a membership fee.
- One respondent said no to membership but asked if DCN would potentially consider being an affiliate member of the Research Data Alliance.

- One interviewee said they need to know what membership provides.
- Another said the fee for membership they would pay would depend on the volume of business they would transact.
- An organization said that a hybrid of membership and fee for services make sense.
- One organization was interested, but asked “what would we be paying for,” and asked about services including:
 - A common platform
 - Their name on a membership list
 - Workflow models
 - Peer-to-peer communication for support and idea generation
- One respondent asked if an annual fee would pay for an organization to be a member? Send someone to a meeting? Allow them to contribute in-kind?
- One organization said they would have to reflect on existing partnerships and memberships and do rebalancing rather than be able to use new money for this initiative.
- Another respondent had a similar question (“what does an hour of data curation cost?”) and said that \$5-10,000/year could make sense.
- One respondent said it would “be difficult to make a case if the membership fee is over \$5000/year.”
- Another said “payment of a fee of four figures is easier than five.”
- Another said the dollar level they would be willing to pay depends on the value of the services to their organization.
- Could an organization come in with other affiliated campus in their state? Or in their existing collaborations (BigTen)
- DCN needs to have one or two activities that will help take the burden off of individual institutions, such as LPC, DataSite, and SPARC do.

28. Does working with another organizational home entity make sense for DCN? Do you have suggestions for potential organizational homes? Or, is a fiscal agent or other type of contracted services model a better fit? Please keep in mind if there is a preference/ease for your institution to work with another university vs. a standalone organization.

- ACRL
- “ARL had floated a partnership idea in the past. Idea had been that DCN could be a part of “branded suite of services to ARL members.” DCN had felt that approach might alienate non-ARL libraries.
- Oberlin Group libraries.
- RDAP
- CLIR – two participants suggested this organization; one additionally suggested potential partnership with the CLIR Data Curation Fellows
- LYRASIS – three interview participants suggested this, with LYRASIS acting as a convening entity
- National Data Service (at UC-San Diego or University of Illinois)
- The library carpentries are also considering organizational homes

- “Leave it at University of Minnesota until an answer comes up.”
- DLF, Ithaka
- Educopia
- “Working with an organizational home can be important because it is difficult to negotiate MOUs and other issues, and it is not an activity that academic libraries are doing well.”
- “Could working with EDUCAUSE, Internet 2 make sense?”
- DCN does not have to live at one single institution unless it has a Hathi Trust model (separate service/standalone with own identity, incorporation, or board, but tied to a larger parent institution).

29. What else would you like to share with us today that you haven’t already had an opportunity to say?

- Some universities are concerned about their own (local) data curation cost recovery and local sustainability issues.
- “Go slow, build on what’s working. Don’t be too ambitious. We cannot imagine not having this service organization in the future.”
- Don’t get overly insular – there is a tension between local and global issues.
- Where is the community of practice located?
- The most critical thing is gathering exposure for the work of the DCN organization, and for the organization’s mission.
- Relationships with Research Offices on campus is important. Also relationships with funders (Federal and private foundations)
- DCN already has clout; sustainability and business planning could leverage that.
- One respondent sees a shift in the ways to provide services between a centralized and hub-and-spoke model. There is a lot of interest in their institution related to what services and activities lend themselves to local control vs. shared services. New initiatives need to make it easier for libraries to see the benefits and differences between local and centralized services. Work has happened on these issues in the reference and chat reference communities, but not in data curation.
- An interviewee wrapped up her interview saying they are “really interested in seeing this type of project succeed in the community.”

Summary of Trends from Phone Interviews

Throughout the interviews, key topics and trends were noted. These commonalities and important findings of the calls will be helpful to DCN in its sustainability planning and future work.

- There was support shown for a variety of future financial models for DCN, including hybrids of membership fees, fees for service, and in-kind activity, but almost all of the respondents said they need to know clearly what the budget money/funding would be going for. Several interviewed organizations expressed strong interest in allowing multi-campus/multi-organization consortial membership to DCN. The level of financial and personnel commitment for the in-kind portion of DCN work was a concern mentioned by a number of interviewees.
- When asked about their interest in the activities of DCN, coordination of expertise and developing a place for programs new to the field to learn about data curation issues were seen as key factors; this is important because interest in the library and academic community in the topic of data curation is growing quickly.
- Respondents talked about the data curation activities their organizations are currently involved in. While a few of the stakeholder organizations had large data curation staffs, most had a smaller number of people focused on these tasks but hoped to grow capacity. Documentation of data curation activities was seen as important. The library's role as a liaison/ambassador to faculty and departments on data issues was seen as very important.
- Some of the most important activities in data curation at the interviewed institutions included persistence of data and data accessibility. Several interviewees expressed concern about the challenging, time-consuming nature of data curation activities.
- When asked to name other stakeholders and leading institutions in the field, respondents named 22 organizations. Those mentioned the most include the California Digital Library (3 instances) and National Library of Medicine (2).
- Participation constraints included human and budgetary resources. As expected, costs were listed as a constraint by many. A number of the organizations interviewed do not have the staff or capacity to contribute in-kind to the project and expressed concern about that. There were also comments about differentiating between local and field-wide issues in data curation. Another common concern was that there is a need for a clear value proposition about the DCN.
- Additional partnerships and competitors included:

- Partners: ARL (3 mentions), the Data Curation Centre, discipline-specific repositories, domain data centers, society archives, and professional associations.
 - Competitors could include local data services; two respondents expressed concern about commercial services including Springer and Elsevier. OCLC and FigShare were mentioned both as potential competitors and potential partners.
- Seven of the eight organizations interviewed expressed some level of interest in becoming involved with DCN. Few of the organizations were willing to suggest a dollar figure they would pay to join, but the three that did suggest an annual membership fee level of around \$5,000 up to a high of \$10,000. Many would need to know “what membership provides” and the volume of activity they would need to undertake. It was mentioned that it may not be new money used to pay for participation, but potentially repurposed funds.
- Twelve organizations were mentioned as potential organizational homes. Three of the interviewees suggested LYRISIS as a potential home. Several people also spoke of previous discussions with ARL about some type of affiliated role.
- In the open “wrap-up” question, all of the interviewees seemed positive about the efforts of DCN. Suggestions for the future of the initiative included “go slow” in growing the activity, and gathering exposure for the work of and mission of the organization. Working with campus Research Offices and Federal and Foundation funders was also seen as important in the future activities of DCN.

APPENDIX D

DATA CURATION NETWORK SUSTAINABILITY PATHWAYS

Laurie Arp, Tom Clareson, & Carissa Egan
LYRASI December 2019

Table of Contents

- INTRODUCTION 29**
- SUSTAINABILITY PATHWAY MODEL OPTION 1: PARTNER PROGRAM 31**
 - OVERVIEW 31
 - Format 31
 - Description 31
 - DETAILS 31
 - Features, Benefits, and Advantages 31
 - Reach 32
 - Considerations 32
 - Risks 32
 - Growth Path 33
 - RESOURCES 33
 - Required Resources 33
 - Return On Investment (ROI) 33
 - Generalized Costs 33
- SUSTAINABILITY PATHWAY MODEL OPTION 2: STAKEHOLDER PROGRAM 34**
 - OVERVIEW 34
 - Format 34
 - Description 34
 - DETAILS 35
 - Features, Benefits, and Advantages 35
 - Reach 35
 - Considerations 35
 - Risks 35
 - Growth Path 35
 - RESOURCES 35
 - Required Resources 35
 - Return On Investment (ROI) 36
 - Generalized Costs 36
- SUSTAINABILITY PATHWAY MODEL OPTION 3: FEE FOR SERVICE 37**
 - OVERVIEW 37
 - Format 37
 - Description 37
 - DETAILS 38
 - Features, Benefits, and Advantages 38
 - Reach 38
 - Considerations 38
 - Risks 39
 - Growth Path 39
 - RESOURCES 39
 - Required Resources 39
 - Return On Investment (ROI) 39
 - Generalized Costs 39
- OTHER SUSTAINABILITY PATHWAY MODEL OPTIONS 40**
 - OVERVIEW 40
 - Merger Model 40
 - Individual Contributor Model 40
- SUMMARY 41**

Introduction

The following document details the second phase of LYRASIS' consulting engagement with the Data Curation Network, completed in October and November 2019, to develop sustainability pathway models. The models presented are derived from direct feedback that was received from the first phase of the LYRASIS consulting engagement that included both (1) an in-person, LYRASIS-led focus group that was conducted at the Data Curation Network All Hands Meeting in July 2019 in Minneapolis and (2) follow-up telephone interviews conducted in August 2019.

These models are meant to serve as “sustainability building blocks” for the Data Curation Network and should be used as guideposts to help successfully transition the project away from being solely grant funded. Each pathway model should be considered as a potential avenue toward sustainability as the Data Curation Network moves forward beyond the initial grant funded phase of the project. Each model contains specific details about the format, features, benefits, risks, considerations, and resources required. Additionally, an example summary of possible costs for each model is presented in a dedicated tab per model in the attached spreadsheet.

No specific sustainability pathway model is recommended at this point. The three main models presented were generated directly from focus group and phone interview feedback. Additional sustainability pathways that were not directly discussed by participants in the focus group or phone interviews are briefly included at the end of the report for further consideration.

It is recommended that further testing from those involved in the Data Curation Network should be conducted to (1) refine the sustainability pathway options and (2) determine the best short and long term sustainability strategy that aligns with the needs, interests, and values of current key partner institutions as well as existing and future stakeholders. The output of further testing may result in the determination that the models presented should be turned into a phased approach (e.g. model 1 works for conservative year 1; model 2 is a next step which is a growth year in year 2; model 3 is where the Data Curation Network plans to be in 5 years, etc.) and/or each model could be combined with another model to create a hybrid avenue toward sustainability (e.g. model 1 and model 3 could be combined into a single model that the Data Curation Network will pursue over the next 1-3 years).

The third and final phase of LYRASIS' consulting engagement with the Data Curation Network will include drafting an assessment framework for evaluating the sustainability models, which includes further refining the facets of each model, aligning the appropriate model(s) to the next stages of the Data

Curation Network's development, validating which model(s) align with the needs of the Data Curation Network participants, and quantifying the benefits and risks associated with each model.

Sustainability Pathway Model Option 1: Partner Program

Overview

Format

- Tiered, Partner Program

Description

The Data Curation Network *Tiered Partner Program* would allow institutions of various sizes to become Data Curation Network partners at a prescribed tier with an associated annual partner fee. Each tier would include a predetermined set of stipulated benefits that a partner could take advantage of throughout the year.

A tiered partner program could incorporate various distinguishing features at each tier to correlate with the program fee charged. Examples could include: participation in governance discussions at the “large” tier; ability to attend an annual partner meeting at the “medium” tier; ability to volunteer to serve in special interest capacities on Data Curation Program committees at the “small” tier; and/or a specified number of data curation hours both contributed and received in-kind at each tier. Partners could self-select which tier they wished to participate in or could be aligned with a specific tier based on certain criteria (size of institution, Carnegie classification, etc.).

Additional activities that partners could partake in that were suggested in the focus groups/interviews could relate to building a community of practice around the expert skills surrounding data curation, building training curriculum that could include suggested best practices, tools, and frameworks for data curation, etc.

Note: The term “partner” was chosen specifically over “member” due to concerns expressed in the focus groups and phone interviews over “membership fatigue”. Partnership implies joint ownership with shared rights and responsibilities which also aligned with the values expressed as being core to the Data Curation Network culture that has been built thus far.

Details

Features, Benefits, and Advantages

- + Accommodates various sized organizations:
 - Enables small to medium sized organizations to participate via a means that is accessible to them (i.e. they could self-select the tier that best fits the data curation activities underway on their campus and that would fit within their procurement structure)
 - Allows larger organizations to make an investment aligned with the state of data curation work ongoing at their institution

- + Adaptable, flexible tier structure (both the fees and associated benefits) that can be crafted to meet a wide variety of needs of partner organizations today and adjusted as needs of partners evolve over time
- + Aligns with the current state of the data curation field in that the founding grant partners of the Data Curation Network are further along in data curation activities than the majority of field (note, this is in reference to the United States only)
- + Easily understandable and commonplace model that would need fewer marketing and outreach resources to communicate more broadly
- + Addresses concerns expressed in focus group and phone interviews that specific benefits must be articulated in order for organizations to participate in the Data Curation Network

Reach

- National, broad

Considerations

- ? Focus group and phone interview participants expressed concern with membership/partnership model fatigue
- ? Potentially too early for small to medium-sized institutions to financially contribute
- ? Benefits of membership need to be clearly defined/articulated to entice partners to contribute
- ? If benefits include in-kind data curation activities, dedicated staff/systems/workflows need to be in place in order to track details/hours
- ? Potential that broad membership could dilute the depth of expertise of the Data Curation Network

Risks

- Not enough contribution to cover entire costs to administer the Data Curation Network as it was staffed during the grant phase
 - Are the current grant partners able/willing to downsize until resources are built back up?
- Level of overhead is too much of a burden in addition to actual work of conducting data curation work
 - What is the balance point?
- No institution is willing to provide administrative services
 - Will the current grant recipient institutions continue to serve as the administrative home of the project?
- Competition with/cannibalization of member dollars with other membership organizations
 - Are there collaboration opportunities and interest with other, closely related, organizations?
- Data curation work is difficult to precisely quantify and therefore align with tiered partner fees

- Broader membership could impact the current level of trust and disrupt the pace of current participants
 - Trust was a specific element highlighted in the focus group and phone interviews as a qualitative strength of the Data Curation Network.

Growth Path

- Moderate
- In the initial 1-3 years post grant funding, the goal would be to have a number of committed partners to cover costs in order to break even. Each year would require acquiring several new members to ease financial situation of program and diversify data curation expertise and activities. By year 3, program could decide to explore other sustainability path options in order to augment financial/revenue structure.

Resources

Required Resources

- Dedicated institution to provide administrative services, such as fiscal management, human resources, etc.
- Program manager
- Outreach and engagement/marketing
- (possible) Event planning
- (possible) Project management

Return On Investment (ROI)

- Initially small but could be built up and expanded, depending on the alignment of the tiers (fees and benefits) with the needs of the broader market

Generalized Costs

An example summary of possible costs for this model are provided in the attached spreadsheet in Tab #1.

Sustainability Pathway Model Option 2: Stakeholder Program

Overview

Format

- Phased, Stakeholder Program

Description

The Data Curation Network *Phased Stakeholder Program* would allow the Data Curation network to focus solely on building a program around the ten grant partners and, potentially, a very small group of closely aligned stakeholder institutions in the immediate years after the grant has been completed. This model would enable the Data Curation Network to further refine services being offered to stakeholders, build knowledge and expertise in data curation services offered to stakeholders, and could be resourced similarly to current grant work levels.

With data curation generally at a nascent stage, this approach would allow for conservative growth to align with maturity and evolution of the data curation field. The most important element of this model would be to acquire formal commitments from the existing grant partner institutions as well as a short list of highly aligned stakeholder institutions. From this commitment, a thorough list of data curation expertise provided by and to stakeholder institutions could be developed. Once assembled, a survey of the broader field could be conducted to ascertain needs that the Data Curation Network could then plan to implement to enable growth in strategically based ways based on the data surfaced from the survey (e.g. the survey might show a growing need/trend to develop expertise in astrophysics data curation and an associated stakeholder institution to approach, etc.).

Additionally, strategic growth could be enabled through developing partnerships with data repository networks already operating in the broader ecosystem (Dryad, Collaborative Archive Data Research Environment (CADRE), etc.). Longer term opportunities that partnerships such as this might enable include exploring opportunities related to revenue/program fee sharing and/or developing supplementary services for data curation and associated data deposits into a shared partner's repository (e.g. stakeholders receive ability to deposit "x" number of data sets into Dryad for additional "\$y" contribution annually, that would be shared with Dryad).

Note: This model specifically addresses the concerns expressed by participants in the focus group and phone interviews regarding the Data Curation Network trying to grow too fast and therefore, potentially diluting what has been built to date.

Details

Features, Benefits, and Advantages

- + High reward for original stakeholder institutions
- + Allows for furthering deep connections amongst individuals within stakeholder institutions
- + Aligns with current market status that larger institutions are those with dedicated staff and potentially budgets for data curation activities (this is a specific finding from focus group/interviews feedback)
- + A smaller group of involved institutions allows for the preservation of trust and commitment currently established during the grant phase of the project

Reach

- Strategic, targeted

Considerations

- ? Concerns were expressed over inclusivity, diversity, and insular nature of this approach during focus group and phone interviews
- ? Difficulty in acquiring data curation expertise in broad range of fields with targeted stakeholder approach
- ? Balance in amount of financial and personnel resources that stakeholder institutions are willing to commit, both short and long term
- ? Potentially consider developing a “Data Curation Network” framework model that could be shared with/consulted on in other regions in the United States/globally (either as another source of revenue or as a contribution to the field to bolster data curation activities worldwide)

Risks

- High level of responsibility tied to each stakeholder institution
- Potentially low financial stability/high financial volatility, as each institution would be contributing a large percentage to the sustainability of the Data Curation Network
- Personnel turnover could have significant impact on data curation expertise in one field/aspect if it is tied to only one or a few individuals at each stakeholder institution

Growth Path

- Conservative

Resources

Required Resources

- Staffing resources similar to current grant levels
- Known

Return On Investment (ROI)

- Minimal. This sustainability path allows the DCN to continue in a similar direction as what has been developed during the current grant stage of the process. It is meant to cover costs only and should be explored in the near-term only as other more long-term sustainability model options.

Generalized Costs

An example summary of possible costs for this model are provided in the attached spreadsheet in Tab #2.

Sustainability Pathway Model Option 3: Fee For Service

Overview

Format

- Consultative, expert curation

Description

The Data Curation Network *Expert Consultant Fee For Service* model would allow the Data Curation Network to acquire revenue in a means not tied to a model exclusively focused on annual, ongoing contributions. Consultative services could be the sole focus of the Data Curation Network or they could be provided as a revenue diversification stream to either of the above sustainability pathway models.

The Data Curation Network could offer consulting services regarding a wide range of data curation activities, such as:

- Making recommendations for institutions just starting to provide data curation regarding how best to begin promoting and providing curation activities on campus, upskilling existing staff, advocating for additional funding with administration, etc.
- Providing best practice assessments on data curation practices at an institution writ large and/or consulting on specific data curation activities occurring in a specific field
- Assessing specific data sets submitted for review to the Data Curation Network and providing written guidelines for how best to move forward with curation at the local institution.

The Data Curation Network could also consult with other consortia or regionally-based organizations about how to develop similar “data curation networks” within those organizations for a fee as well as consider offering in-person workshops and training sessions customized to meet the needs of the institution.

The fee for service model would require the Data Curation Network partner organizations to continue to formally commit contributing a certain percentage of data curation personnel to the project in order for services to be rendered outside of the network. A very specific memorandum of understanding would be needed between the Data Curation Network and partner institutions clearly outlining the activities and level of effort required during a specified time period.

Note: This model was specifically developed based on focus group and phone interview feedback that a fee for service offering could help “to build the program” and allow for smaller organizations to contribute in a financial way that would not be open in the partner/stakeholder models.

Details

Features, Benefits, and Advantages

- + Aligns with the current and near-term state of the market wherein small and medium institutions do not currently have dedicated data curation staff nor likely will have dedicated staff in the immediate future as volume/frequency of data curation is much lower/ad hoc, therefore need to be able to procure services is high
- + Adaptable consulting approach could be quite broad and flexible in order to apply to institutions of any size, with types of services being offered evolving over time
- + Consulting approach could apply the “It Takes A Village”¹ (ITAV) model of facets with phases to data curation lifecycle at institutions to inform consulting services and associated recommendations.
 - o For example, an institution requesting consulting services could be assessed to be at a very early “Getting Started” phase of data curation (based on the ITAV model framework). This phase would then correlate to recommendations in specific facets such as tools for discussing advocacy, as well as easy steps to get started in an economical way.
 - o Another institution may be assessed at a middle “Growing/Getting Established” stage (having just started to curate data sets) and would need guidance on how to evolve from an ad hoc data curation approach to a more fully fledged data curation workflow as the volume of data curation activities increases at that institution and it mature toward a “Stable, But Not Static” phase.
- + Aligns with market sentiment found in focus groups and phone interviews that many organizations are “crossing a threshold where data curation is an activity we need to fund”.
- + A fee for service model would address the following stakeholder comment: “If DCN wants to be larger, they must do planning for organizations that are at different points of the data lifecycle and have different expectations.” Various services could be developed and offered to closely align with current data curation needs.

Reach

- Market driven

Considerations

- ? Consulting and fee for service ventures require careful estimation and project management
- ? Volume and nature of work can be unpredictable and fluctuate

¹ LYRASIS It Takes A Village <http://lyrasis.org/itav>

- ? A workflow for meeting the consulting requests and aligning them with Data Curation Network personnel resources and skills would need to be developed along with tools to manage this process
- ? Institutions committing personnel resources to the Data Curation Network would need to receive valuable benefits in return (further work is needed to determine what those benefits would be)

Risks

- Not enough requests for data curation consulting work to cover costs
- Too many requests for data curation consulting work that needs of market are not being met
- Institutions are not willing to contribute personnel resources on which to build fee for service offerings
- Institutions may not fundamentally agree with the fee for service business model
- Institutions may feel disenchanting by DCN providing such services when they had been provided for free previously (actually or perceived)
- Volatility in requests is too difficult/complicated/time consuming to manage
- Price for fee for service offerings is not palatable for intended market (priced out of market)

Growth Path

- Unpredictable, fluctuating

Resources

Required Resources

- Dedicated institution to provide administrative services
- Project manager
- Business manager/financial oversight
- Dedicated data curation personnel (in-kind from stakeholders or dedicated staff role(s))
- Marketing
- Technical system(s) for tracking consulting details

Return On Investment (ROI)

- Unpredictable and high risk. This sustainability model could potentially have a high return if the services are priced correctly for the market and are properly staffed. However, this model could also be low return if not adequately resourced or correctly formatted.

Generalized Costs

An example summary of possible costs for this model are provided in the attached spreadsheet in Tab #3.

Other Sustainability Pathway Model Options

Overview

The following additional sustainability pathway models were not discussed in detail during the focus group or phone interviews. However, considering other paths to sustainability that may not be in consideration by partners and other participants of the Data Curation Network will be an important element to balance the feasibility of the final model(s) chosen to be pursued.

Additionally, one-time funding opportunities such as grants were not explored as part of the sustainability pathway models. Grants and similar funding streams are good ways to add one-off funding for special initiatives and can be used to supplement key initiatives but should not be considered as sustainability solutions in and of themselves. Furthermore, overall changes to how grant agencies could omit to fund data curation in the long-term were also not explored as avenues for sustainability.

Merger Model

The Data Curation Network could consider integrating the project into an existing non-profit. There are several 501(c)3 organizations that bundle various services and software programs under one umbrella entity that administers not only membership programs, but also human resources, financial and executive management, as well as other administrative functions. This sustainability pathway model was not explored in detail as the specifics of how to proceed would be dependent upon the organization into which the Data Curation Network would incorporate.

Individual Contributor Model

Another potential sustainability pathway the Data Curation Network could consider is an individual contributor program wherein PhD's, computational experts, computer scientists, etc. could subscribe to a Data Curation Network service at some level for a predetermined cost. This model could connect a researcher/data owner with an expert curator in a specific field (e.g. a match.com of data curation with a one-month subscription for \$59). This model was also not detailed above as it did not appear in the focus group nor phone interviews. However, it might be a unique avenue to consider as a supplementary sustainability model to explore.

A service offering like this could be executed in a pilot or minimal viable product means in order to confirm if there is broader interest for such a service. This service could also be pursued as part of a partnership or in close alignment with another organization in the data curation space that is also interested in exploring opportunities for individual contributor models.

Summary

As a result of the LYRASIS-led focus group at the Data Curation Network All Hands Meeting and subsequent phone interviews with key stakeholders, three main sustainability options rose to the surface: a partner program, a stakeholder program, and a fee for service program. Based on these findings, sustainability pathway models were developed to further detail how these potential avenues for sustainability *could* be implemented in order for the Data Curation Network to transition from being a grant funded project into a self-sufficient organization.

The sustainability pathways models presented above are meant to serve as starting points as the Data Curation Network transitions from grant funding. No one model is recommended; however, further assessment should be done to determine the best model or combination of models to achieve the short-term and long-term sustainability goals for the project and its participants.

It should be noted that there was a tension uncovered in the focus groups and phone interviews in two main opposing sustainability directions/approaches that the Data Curation Network could take. These viewpoints are summarized in the following quotes:

- “Go slow, build on what’s working. Don’t be too ambitious. We cannot imagine not having this service organization in the future.”
- “Don’t get overly insular – there is a tension between local and global issues.”

Based on the resounding number of comments falling into one of the two categories above, it is likely that a hybrid approach of the above models is the best fit for Data Curation Network going forward.

It is recommended that the Data Curation Network test each model with current and future stakeholders to further refine or refute details in the above sustainability pathways. Sustainability is not a one-time project that once accomplished, can be considered completed; instead, sustainability is an ever-evolving puzzle that should be reviewed and adjusted on a regular basis. The Data Curation Network will need to continue to assess and adjust whatever sustainability pathway(s) it determines is best as often as necessary in order to achieve true long-term sustainability.

DATA CURATION NETWORK SUSTAINABILITY TESTING FRAMEWORK RECOMMENDATIONS

Laurie Arp, Tom Clareson, & Carissa Egan

LYRASIS January 2020

Table of Contents

DIGITAL CURATION NETWORK (DCN) SUSTAINABILITY TESTING FRAMEWORK RECOMMENDATION REPORT	ERROR! BOOKMARK NOT DEFINED.
BACKGROUND	ERROR! BOOKMARK NOT DEFINED.
<i>Pathway Model Options</i>	<i>Error! Bookmark not defined.</i>
DCN SUSTAINABILITY TESTING FRAMEWORK RECOMMENDATIONS	ERROR! BOOKMARK NOT DEFINED.
FRAMEWORK	ERROR! BOOKMARK NOT DEFINED.
<i>Decision Points</i>	<i>Error! Bookmark not defined.</i>
<i>Session Guidelines</i>	<i>Error! Bookmark not defined.</i>
<i>Framework Steps</i>	<i>Error! Bookmark not defined.</i>
1. <i>Identify DCN's Sustainability Goal</i>	<i>Error! Bookmark not defined.</i>
2. <i>Identify Stakeholder Groups</i>	<i>Error! Bookmark not defined.</i>
3. <i>Pathway Discussion and Ranking</i>	<i>Error! Bookmark not defined.</i>
4. <i>Form Subgroup to Recommend Path Forward</i>	<i>Error! Bookmark not defined.</i>
5. <i>Discussion, Modification, Approval of Plan</i>	<i>Error! Bookmark not defined.</i>
6. <i>Outreach Plan</i>	<i>Error! Bookmark not defined.</i>
7. <i>Rollout Plan</i>	<i>Error! Bookmark not defined.</i>
8. <i>Regular Review</i>	<i>Error! Bookmark not defined.</i>

Digital Curation Network (DCN) Sustainability Testing Framework Recommendation Report

Background

The Data Curation Network (DCN) engaged with LYRASIS in June 2019 to assist the DCN in their Sustainability Planning efforts. Phase 1 focused on an assessment of the current landscape and generated ideas and feedback from the community. Phase 2 focused on development of potential Sustainability Path Models for DCN based on that feedback. The models and spreadsheets were submitted in December 2019. The goal of Phase 3 is to recommend a testing framework to evaluate the models presented in the previous phase in order to build on them. The framework can be used to assess what was developed in Phase 2 to shape the models into what fits best for DCN now and into the future. The output may result in the determination that the models presented should be turned into a phased approach and/or each model should be combined with another model to create a hybrid avenue toward sustainability. The pathway model options provided in Phase 2 are included below for easy reference.

Pathway Model Options

5. Tiered Partner Program
 - a. Institutions of various sizes would become DCN partners at a tier with an associated annual partner fee
6. Stakeholder Program (Phased)
 - a. Focus a program around the ten grant partners and, potentially, a very small group of closely aligned stakeholder institutions
7. Fee For Service
 - a. Would allow DCN to acquire revenue in a means not tied to a model exclusively focused on annual, ongoing contributions
8. Other
 - a. Merger Model
 - i. Potential integration into an existing nonprofit
 - b. Individual Contributor Program
 - i. Service where subscribers would be matched to expertise

DCN Sustainability Testing Framework Recommendations

Framework

This framework outlines the various steps that could be taken to assess the sustainability pathways and identify a path forward. The initial steps should be done with the current DCN participants. The pathways can then be refined and tested with external audiences.

Several of the steps below are best accomplished as a group. Session guidelines are provided below to help with the process.

Decision Points

- Early on in the process, the DCN should decide who will be involved in these sessions. The current DCN participant representatives meet regularly and may be the best group to go through Steps 1-3 as a group. If the group decides it is best to have a smaller group focus on Steps 1-3, it is important to loop the larger group back into the conclusions at the end of each phase to encourage buy-in and consensus building before going to the next step. It is advised that DCN not have different groups working in parallel as that process may lead to disconnects between groups such as a pathway chosen for a non-prioritized stakeholder group. For example, what if a narrow group makes a decision that the broader stakeholder /participant group included in Step 3 doesn't agree with?
- In addition, it will be important to determine, in advance, who has the final decision-making authority. Is consensus of the entire DCN participating institutions required? Would some smaller group (such as a Steering Group) decide? This needs to be clear from the beginning so authority is clearly vested.

Session Guidelines

Facilitation

- Sessions should be facilitated by someone who is not invested in a specific pathway for the program.

Format

- Sessions can be conducted online or in person. In person sessions always require a deeper commitment to the process and engenders more trust and fuller discussion, but in person meetings are not always feasible.

Notetaking

- For in person meetings, a specific notetaker (separate from the facilitator) should be assigned and should document responses on whiteboards or large pads. This helps underscore the discussion and enables participants to see if something was

misunderstood. For online meetings, the notetaker should take notes via a shared doc so all have access to the notes during the session. Notes should be distributed afterwards for review and approval.

Framework Steps

1. Identify DCN's Sustainability Goal

At the beginning, DCN participants need to gain consensus on their sustainability goal. Without working with a clear sense of a shared sustainability goal, it will be difficult to agree on a path forward. Participants should have a discussion around this. A discussion guide is given below to facilitate this process.

Draft Discussion Guide

The session should be focused around identifying and trying to gain consensus on what DCN is trying to sustain. Below are some proposed questions.

What is DCN trying to sustain?

- a. Sustain current DCN program structure and level of activity?
 - Why?
 - How could this be achieved?
- b. Affect the larger data repository environment?
 - Why?
 - How could this be achieved?
- c. Build capacity among other institutions?
 - Why?
 - How could this be achieved?
- d. Grow in certain ways (number of files curated, number of institutions, curators or subjects)?
 - Why?
 - How could this be achieved?
- e. What are other questions that resonate with current DCN participants about what the sustainability goal should be?

2. Identify Stakeholder Groups

It is important for DCN to determine who they are trying to reach and work with. Identifying and ranking stakeholder groups enables DCN to consider primary, secondary, and even tertiary groups, their goals and how to align them. Stakeholder groups may include:

- Existing DCN participating institutions
- Affiliated institutions
- Potential partners
- Potential participants

- Potential subscribers
- Advisory Group members
- DCN stakeholders that are not currently participants:
 - Funders
 - Others?
- Others

These groups should be:

- Identified
- Aligned with the goals identified previously
- Then prioritized

3. Pathway Discussion and Ranking

In advance of the session, determine which options the group is considering. Is it models 1-3 only? Will the Merger and Individual Contributor Models be considered as well? Others? Distribute background materials in advance. Ask participants to review the options in advance and come prepared to discuss them.

Facilitator Notes

1. In the session, ask each participant to give an “Instinctive rank” to each model (without thinking too much), each person should give each model a 1, 2 or 3 (1 highest/ “best”, 3 lowest). Go around the group for option 1, 2 and then 3. Do this publicly. If this is an in-person meeting, distribute paper and have each person write down their score and hold it up for all. If online, have each person verbally give their rank for each model. The notetaker should write down scores.
2. Then the facilitator should lead a discussion wherein the group talks through each sustainability pathway option, identifying pros and cons for each option. It will be helpful to understand how different options are perceived from each institutional perspective. What is a plus for some may be a negative for others. Also determine if certain pros or cons are “groupable.” For example, would small academic institutions be more comfortable with long term consistent “dues” or would larger academic institutions have more consistent ability to donate curator time?
3. Following a full discussion, facilitator should ask everyone to rank the options again to see if anything has changed. It is helpful if participants explain new scores (e.g. “I hadn’t considered that some institutions may prefer to prepay 3 years if they have end of year funds available they need to use or lose”).

4. Review overall rankings to see if there is a consensus.
 - a. Given the results, are there ways to combine or sequence options?
 - b. How do these results align with the overall sustainability goal?
 - c. How do these results map to the various stakeholder groups?
5. Identify specific data points that would confirm or clarify points. Examples include:
 - a. Is 10 or 20 the magic number of institutions participating?
 - b. Outline ways to gather that data.
6. Highlight “pros” that should be emphasized in the Model and talking points.
7. Identify “cons” that should be adapted to further adoption (i.e. 10 is too many/not enough institutions for Model 2).
8. End this session with a general summary of what was concluded or decided and a list of actions that a subgroup will be responsible for completing.

4. Form Subgroup to Recommend Path Forward

Assign a smaller group to review feedback, modify pathways as necessary, and develop a proposed model to present to the larger group. For example:

- Short term: continue with Model 2, combine with Model 3 (allocate X% of time per year in order to test demand, work with potential new organizations, etc.).
 - Potential for funding to create Model 3.
- Longer term: work towards Model 1 in X number of years – identify milestones, incremental steps and annual targets.

Remember the *DCN Sustainability Path Models Generalized Costs Examples* spreadsheet submitted in Phase 2 provides sample cost data, not absolutes; participants can and should play with and adjust all the elements to see the impact of various choices.

5. Discussion, Modification, Approval of Plan

Discuss proposed plan with the larger group and adapt as necessary. Remember this will be a living document. It represents a plan to go forward but can always be revisited as testing and other elements reveal new information.

6. Outreach Plan

Once the draft model is agreed upon, create a messaging plan that explains the goals, ROI for participants, and timelines.

- a. Test path and messaging internally (within the DCN constituency) with decision makers already affiliated with the DCN.
- b. Gather feedback
 - Develop a template to ask for feedback so you are consistent and asking for actionable feedback
 - Ask for feedback on the model and messaging. Will they commit to the model? Don't mistake "good idea" feedback for willingness to financially support it. This is not only an opportunity to ask for feedback but is a form of early outreach to encourage institutions to participate in the model.
- c. Regroup and refine the plan and its messaging
- d. Decide on the order for testing with external groups. Think about the stakeholders identified earlier. Potentially start with those familiar with DCN but not yet involved (perhaps potential consulting clients).
- e. Refine and go broader
- f. Gather feedback and further refine the pathway

7. Rollout Plan

Develop a roll out plan with timeline and milestones.

- a. Consider a pilot to test, something with a specific timeline that enables you to test critical elements and refine. For example, if piloting the consulting element, plan to test with 1-3 projects. It will enable you to determine workflows, processes, handoffs, and gather feedback. It also sets an expectation that things are being tested (internally and externally) so when there are the inevitable bumps, it reflects a need to adjust vs. a mistake.

8. Regular Review

Establish a regular review of the sustainability plan. It could be done annually or in conjunction with milestones determined by the pathways. The participants can cycle back to the first step every 3 (?) years to see how the model is working and if adjustments need to be made.