

Notes From the 2016-Jan-08 Discussion at ESIP

Present: on webex: Philip Tarrant, Greg Janee, Sherry Lake, Steve Daley-Laursen, Margaret O'Brien, Corinna Gries; 2 on the phone unidentified (Kerstin Lehnert). in person: Nancy Hoebelheinrich, Ruth Duerr, Dave Vieglaiss, Peter McCartney, Abby Benson, Rebecca Koskela, Matt Mayernik, Anne Wilson, Shannon Rauch (BCO-DMO), Cyndy Parr

Background information:

- Wiki Page: http://wiki.esipfed.org/index.php/Sustainable_Data_Management
- References Link: <http://wiki.esipfed.org/index.php/References>
- List of action items from Tempe meeting:
[http://wiki.esipfed.org/index.php/Action_Items_from_Collaborative_Strategies_for_Sustainable_Environmental_Data_Management_workshop_\(Tempe,_AZ_Nov_2015\)](http://wiki.esipfed.org/index.php/Action_Items_from_Collaborative_Strategies_for_Sustainable_Environmental_Data_Management_workshop_(Tempe,_AZ_Nov_2015))

Summary from Anne's raw notes, at back (please amend, correct):

- In order to think about how to define Return on Investment (ROI), would be helpful to think about describing in terms of those who see value in repositories **and fund them**
- How about thinking of two main types of funding sources (stakeholders) who see value in repositories
 - **Members.** Organizations that provide a block of money for you to build and maintain the repository. Examples: a federal agency or foundation that shares your passion for the repository and wants you to provide services, for them and others, or a university administration providing institutional commitment....such FTE / staff for the repository. The member organizations trust that you're doing the "good" work.
 - **Transactional-Contractual.** Researchers who are responding to a law or regulation that requires them to take care of data. They want you to provide a specific service, like storage and curation, and they'll pay a fee for that particular service. They're interested in getting a specific product for a certain amount of money; uses a Service Center metaphor.
- Would have to figure out a level of abstraction that allows us to see similarities that enable comparison of apples to apples, such as:
 - Basic infrastructure models
 - Overall purpose (the advancement of science), but expressed as proxies for this
 - Stakeholders groups (users, funding agencies, professional societies, staff)
- But also, allow for differences that would make a difference in terms of still wanting to measure values that are supported / returned to stakeholders (e.g., level of granularity of items or collections within the repository?)
- The spectrum of what we could come up with may range from "simply" providing guidelines for building a good assessment plan for their stakeholders to identifying and describing the metrics that can be used to measure values returned to similar types of stakeholders.

- In either case, our work could describe a framework for measuring value that should include:
 - The identification of stakeholders and their values
 - A characterization of the repository's data holdings and level of usage
- An example way of thinking about how to stratify populations of stakeholders that could further characterize their values (the value proposition between each stakeholder type and repositories), and the metrics that measure impact (whether the value relationship is actually working and generating satisfaction for both parties) is, as follows:
 - Host Universities who house researchers (measure # of users? How?)
 - Agencies of Government with mandates and orders
 - Researchers
 - Public users (citizen science)
 - Funding agencies who fund researchers that use the repositories/data...e.g. NSF funding to repository advances a particular brand of research. What's the proof that NSF-funded researchers are advanced by having repositories available.
- Depending upon the level of similarity in systems, it might be possible to write tools to help measure, and thereby simplify how metrics are collected.
- Should look to see what others are doing / have done on this topic including the resources on the ESIP wiki list at: <http://wiki.esipfed.org/index.php/References>
- To continue work on these topics, it might be advisable to engage a broader segment of data repositories in the sciences (and/or beyond), so should think about who / how to engage
 - The broader group of people who attended the Tempe meeting
 - Other ESIP members via esip_all
 - Perhaps by establishing an EarthCube RCN (and requesting funding to do so)
 - Contacting other organizations that have engaged in similar efforts including JISC, ANDS (Australian National Data Service, others?)
 - Pull together another face to face meeting in the months before ESIP Summer and/or at ESIP summer

Original & Updated Action Items for Return on Investment (ROI) group:

1. Review JISC paper to understand what is covered and how we can use (All)
2. Identify quantitative and qualitative methods of study across reports for applicability
3. Discuss & summarize at ESIP Winter mtg [DONE]
4. Identify values per stakeholder type & metrics for measuring the impacts
5. What metrics are generic across the stakeholders?

6. Look at supplementary papers for info about costs, and other details of doing the studies
7. Ruth will contact Josh Greenberg at Sloan re the funding of a grant to do a similar study
8. Anne will talk with Peter very informally about the possibility of an EAGER grant for planning and also look at the formal solicitation and report back at ESIP
9. In the planning grant, Include a literature / environment scan to see what has been already done in this area
10. Ruth to disseminate lecture info on this topic to group
11. Talk with Erin about setting up a cluster on this topic (Nancy) [DONE]
12. Investigate bringing issue to RDA (Ruth talk to Mark Parsons) [DONE]
13. Develop a 1 - 2 page statement about goals, (timeline) for the overall plan, i.e, the outcome of the longer term goal (At ESIP Winter) based on discussions from yesterday (see above) (TBD)
14. Define what we mean by ROI for this area
15. Anne to ask Matt Jones for relevant material [DONE]
16. Nancy will send email to esip_sustainableDM to see who else is interested among the other interest groups
17. Nancy: email to esip_all ditto
18. Plan sustainableDM Workshop #2 (Corinna, Margaret, Philip), at summer ESIP? include stakeholders? depends on what's accomplished wtr/spr

Three groups from Tempe were:

Question: should future work of these be branched? do they need different funding mechanisms (rcn vs other), timelines, etc?

The below lists of people from:

[http://wiki.esipfed.org/index.php/Notes_from_Collaborative_Strategies_for_Sustained_Environmental_Data_Management_workshop_\(Tempe,_AZ_Nov_2015\)](http://wiki.esipfed.org/index.php/Notes_from_Collaborative_Strategies_for_Sustained_Environmental_Data_Management_workshop_(Tempe,_AZ_Nov_2015))

Landscape Analysis and Gaps for Environmental Data Repositories

<https://docs.google.com/document/d/1rYs3nOku5oCZCYcemNn-1dwJ4pahP8fCpHFD0YVFIHg/edit#> Link:
<http://bit.ly/env-repo-landscape>

- Paul McCartney
- Cyndy Chandler
- Sherry Lake
- Yaxing Wei
- Inna Kouper
- Erin Clary

- Kerstin Lehnert
- Matt Jones
- Corinna Gries

Return on Investment of Data Repositories for Society

https://docs.google.com/document/d/10KKUjKnO3LbzCB3Qu5_JhVBVaNiBn67nmqfRASDVeg/edit

URL: <http://bit.ly/1HZa3ci> for group

- Anne Wilson
- Ruth Duerr
- Peter McCartney
- Steve Daley-Laursen
- Nancy Hoebelheinrich
- Carolina Murcia
- Erin Robinson
- Dave Rugg

https://docs.google.com/document/d/1CoAT_ZIZSzNPb0cZHC3bfxbc8xO4ZLrShy_8Ow9CD0/edit#heading=h.k09oe6hl9os

URL: <http://bit.ly/1HH5Yz7>

- Cyndy Parr
- Dave Vieglass
- Ken Casey
- Margaret Hedstrom
- Greg Janee
- Mark Sevilla
- Kevin Browne

Note: there was a fourth group, called “entrain stakeholders”, but it was folded into the other 3, because all have stakeholders.

Anne’s (?) Summary notes:

What are possible outcomes ?

spectrum: one end, repos might simply get guidelines for how to build a better assessment plan for their stakeholders. other end: our users are so similar that we can identify a set of metrics, and if systems are similar enough maybe we can write tools, and simplify how everyone does it. depends on how easily you can characterize the repos, and how similar they are.

similarities:

all have users who download data.

all belong to societies
all have funding agencies (members)
model for infrastructure is quite similar, although implementations are different.

Differences:

level of granularity of data holdings
? do any have transactors? example: ___

examine from stakeholder PoV (see below)

Question to NSF: what criteria do you use to characterize value? Peter: don't have much to go on: manual searches, personal questions, anecdotes.

Ultimate value: science has been advanced. [asking the proposer to justify puts a large burden on them, eg keeping track of project numbers of users, etc] what are proxies?

Anne's raw, real time meeting notes:

[please feel free to correct or amend these notes! AW]

r: ruth, p: peter, m: matt, ...

goal:

r: eco data repo sustainability

dev framework to be used by repos to assess their ROI in apples to apples to way

eg diff levels of service that repos provide

2 other topics came out, but today only talking about ROI

how to proceed

planning grant funds, eg sloan

ruth has candidate that if data were lost would be painful

p: 3 parties, provider, user, 3rd: party that might give resources based on the value

but latter may not be same people

so ROI is dependent on to whom advancing science? tenure? incentives vary

sustainability will translate in \$ at some point

a step:

identify stakeholders and their relevant values (they may vary)

which may vary based on repo type

which IS part of the framework

dev framework vs just study

m: framework includes how to identify stakeholders

id repo, it's users, and other potential stakeholder

steve: eg funders tied to specific transactional needs, versus general "member" needs (eg foundations)

eg nkk usgs and foundations were very interested in being members, gave \$, want to partner, supports their efforts, general, believe in mission

vs. faculty members, that keep up with rule, mandate, etc. care about specific cost

link to paper is on wiki carol ember, hanisch

another step:

id characterize the holdings

kind, granularity, SLAs, distinctions between SLAs, volume

and usage and SLAs

p: values may be expressed not in terms of the db

so, need to advance in their value terms

eg a researcher wants to get tenure

r: it's a partitioning issue

codata study assessed funding mechs and strats currently in use

found very few repos have only 1 funder, typically address multiple communities,

a: jisc report asked stakeholders directly what would happen if data went away

r: NSIDC demonstrated an outraged user community when Nasa tried to kill it

'direct investment and use value'

'contingent value', willingness to pay

'efficiency impact

'wider community benefits'

mapping values of stakeholders to impacts

methods for establishing value and impacts

M; typically people gather easy stats, # papers published, downloads, citations to pubs,

diff about jisc is trying to apply economic measures, directly talked to users

a framework, but repos would need \$ to do the study

characterize relationships into groups would ultimately be tailored to repo

p: helpful to characterize relations: specific return for specific exchange, whereas other is diffuse, based on trust

n: add that dist doesn't nec consider funders eg academic where a PI or repo manager has to argue to university admin that resource is providing value, more targeted than an agency perspective, 'institutional commitment' eg grows research at the U, more focused than 'we trust you'

meeting goal: go over action items

should we pursue a planning grant to ...

in order to est framework need a

by planning how to get more info about various data repos, and the associated values for them in order to determine ROI

id methods by which to get info

who we want to survey

framework - a doc advising a repo that needs a strategic plan,

so we read doc that says

estab set of stakeholder communities and their values to determine which methods to use to establish ROI. map things to that, measure this that,

so framework is instruction book, like a pm plan

much work done, which needs to be assessed: JISC, Australia,

m: tons of bibliometric studies: download metrics, papers, citation

but these are very system specific, which impacts metrics eg can't login in

also few have controls or comparisons, no reference point

nasa acsi report, which does user surveys

p: if this group was to be successful:

generic outcome: equipped to better figure out metrics, stakeholders,

eg guidelines vs

other extreme: stakeholders are all so similar that we can identify metrics and a more concrete product, like possibly tools that simplifies operations

how easy can a variety of stakeholder needs be met with small number of metrics

how much commonly is available to commonize around

s: many repos have similar models but impl differs

also at workshop, identify similarities accross repos

maybe this is place to focus at first?

meeting of tech people and user reps to dev tech plan and white paper to make recommends

could have sig impact

r: but SLAs vary so much

eg dryads provides same SLA for all, but is very low

-> repo selection matters

but if you find something in common, that is prob a metric

r: can't find that. R was on dryad archiving board, adv committee, SLA is so different that can't see how comparison would be done

p: familiar w/ dryad as a funder, recogns that while centers are wildly diff, but from his perspective the value is the same: how to they advance research?

if find that commonality

s: take diff stakeholder groups, what info do they need

eg proposals who mention dryad in proposal

% of archive downloaded every years

p: wants to know that science has been advanced

will need proxies

steve: added to doc categories of stakeholders, with interests in repos

need categorize prop in each category

(updating the list of action items on ESIP wiki)

review papers, id methods

id stakeholder type and metrics

p: what other tempe working groups are going to persist, maybe a

eg tool development

along line of RCN

other groups had other action items

landscape analysis and gaps across repos

common technical vision there is a similar RDA group

p: good news/bad news is: we have multiple pieces with gaps, but all are funded independently to meet own values on own cycles

be sure to include their input

apply for RCN? with other groups as a whole , funded by several nsf programs, a community builder, eg dev of env bio,

how to characterize target community

Note that funding remains in the EAGER grant that could cover another meeting for workshop participants.

Next steps, action items

action item: nancy to draft 2 emails, send to group for review, get results, discuss them on phone session at summer ESIP meeting