Tools + DMPTool2

IMLS Grant Meeting

Table of Contents

Logistics

Meeting goals

1. Identify software tools that might be integrated into future versions of the DMPTool.
2. Determine the level of effort needed to implement integration of various tools.
3. Prioritize future development efforts with an eye towards tools that may work with the DMPTool2.

IMLS grant text: [Improving Data Stewardship with the DMPTool: Empowering Libraries to Seize Data Management Education](#) (downloadable PDF)

Relevant section: The number of software tools available to help researchers with their data management and analysis is rapidly growing. Many of these tools are good candidates for working in conjunction with the DMPTool to streamline researchers' workflows. We propose to explore these potential collaborations more fully in the last quarter of this grant's funding period. We will solicit suggestions from the community at large, including researchers, librarians, and other information professionals, for a list of tools and their utility. We will generate discussion about these tools using an online forum, which will allow us to prioritize future development efforts. One deliverable from this grant will therefore be a summary of existing software tools ripe for integration, and how those tools might best serve the DMPTool community.

Agenda

8:30 - 9:00	Introductions & meeting goals
9:00 - 9:30	Overview of DMPTool2 features for plan creators & administrators
9:30 - 10:00	Technical overview of DMPTool2
10:00 - 10:30	Q&A; discussion
10:30 - 11:00	Break
11:00 - 11:15	Overview of breakout exercise & tasks
11:15 - 12:00	Breakout groups - brainstorm tools
12:00 – 1:00	Lunch (provided)
1:00 - 2:00	Report back and consolidate
2:00 - 2:15	Break
2:15 – 3:15	Top 3 tools/concepts from each attendee; dot voting
6:30	Group Dinner at [The Forge](#)

Attendees

Bhavi Vedula	CDL	Bhavi.Vedula@ucop.edu
Carly Strasser	CDL	carly.strasser@ucop.edu
Shirin Faenza	CDL	Shirin.Faenza@ucop.edu
Marisa Strong	CDL	marisa.strong@ucop.edu
Stephen Abrams	CDL	stephen.abrams@ucop.edu
Trisha Cruse	CDL	patricia.cruse@ucop.edu
Perry Willett	CDL	perry.willett@ucop.edu
Andrew Sallans	COS	andrew.sallans@gmail.com
Dan Phipps	CDL/UCLA	dphipps85@gmail.com
Michael Witt	Purdue - Databib	mwitt@purdue.edu
Susan Borda	UC Merced - DataUp	sborda@ucmerced.edu
Sherry Lake	UVA	sah@virginia.edu
Sarah Shreeves	UIUC	sshreeve@illinois.edu
Amber Budden	DataONE	aebudden@dataone.unm.edu
Scott Fisher	CDL	scott.fisher@ucop.edu
Puneet Kishor	CC	punkish@creativecommons.org
Noah Whitman	ResearchHub	wittman@berkeley.edu
Karthik Ram	rOpenSci	karthik.ram@gmail.com
Aaron Culich	UC Berkeley	aculich@eecs.berkeley.edu
David Minor	UCSD	dminor@ucsd.edu
David Vieglais	DataONE	dave.vieglais@gmail.com
Jonathan Crabtree	UNC	jonathan_crabtree@unc.edu
Chris Hoffman	UC Berkeley	chris_h@berkeley.edu
John Kratz	CDL	john.kratz@ucop.edu

Part 1: Overviews

Overview of DMPTool2: Issues & Ideas

- Public plans should have a cc license of some kind
- Sample plans and the plan library
  - Think about distinguishing plans that got funded to limit the field of sample DMPs
  - Opportunities to work with institutions to identify good plans? Funded plans?
  - Can we construct an argument that they (institutions) should build up a roster of sample plans to start from, when they have last minute requests?
  - Who is in a position to judge plans as good examples? CDL can't do it, researchers aren't knowledgable in this area.
- Reviewing plans
  - Feature idea: allowing researchers to select repository, automatically communicate with

repository for review?

- ○ Opening up review to repositories, curators, etc.
- ○ The "reject" term is too strong - consider "revise" in a context where it has no teeth
- What people ask of the tool: use stats and who's using it at their institution. Make this available via API, but also consider an admin dashboard with stats displayed. Not all libraries have tech people to work with APIs.
- Interface with Databib to make recommendations for repositories for data
- Consider connection between DMPTool and ERA system
- Can the DMPTool be used to measure compliance? Verify that the users do what they say? Check for deposits in repositories?

It would be great if the DMPTool Could...

- Provide service providers notification that they are mentioned in a plan.
- Connect to a project management platform, useful throughout the lifecycle for researchers.
    - ○ transfer information generated by DMPTool to the project software as an outline or a checklist
- Recommend repositories – maybe with keyword indexing for data types?
- Connect to Call For Proposals – auto-fill as many fields as possible.
- Allow us to perform meta-anlysis of trends in DMPs- allow data mining
- Create a checklist for researchers to aid project management.
- Help track compliance
- Estimate cost of data management- inform researchers of the cost of data management using a calculator of some kind.
- Use API to guide repository capacity planning, support funding bids with projected use.
- Allow researchers to see institution-specific stats. This would allow collaboration between researchers using similar resources to coordinate and save money/effort
- Tie into researcher profiles
- Connect to ERA for post-award management.
- Mint a unique identifier that could be submitted with proposal.
- Provide useful analysis for grant officers. Grant officers might be motivated to return where researchers aren't.

Part 2: Tools brainstorm

Prompt: What software tools or projects would complement DMPTool? What integration would benefit DMPTool users? What would that integration look like?

Group 1	Group 2	Group 3
Carly Strasser	Trisha Cruse	Marisa Strong
Shirin Faenza	John Kratz	Dan Phipps
Sherry Lake	Bhavi Vedula	Amber Budden
Chris Hoffman	Michael Witt	Karthik Ram
David Minor	Puneet Kishor	Scott Fisher
Andrew Sallans	Noah Whitman	Susan Borda
David Vieglais	Jonathan Crabtree	Sarah Shreeves
Aaron Culich	Stephen Abrams	Perry Willett

Breakout Group 1

1. SHARE project
2. VIVO or Harvard profiles catalyst (UCSF uses this system "profiles")
3. CalNET (UCB identity refinement) campus id management
4. UCB API management portal (management system, others might might be using it, browsing and looking for apis)
5. cost estimate for doing data storage - calculator (budget) for budget - market place
6. License decision tree (dataverse?)
7. Funder system (Fastlane, grants.org, etc.)
8. ORCID
9. campus grant management systems (cayuse); admin management systems - Kuali
10. Amazon web service public data sets, w/ "requestor pays subset"
11. EZID
12. re3data.org/ BioSharing
13. other lifecycle management (OSF) ELM -
14. LIMS RedCap
15. local repo verification – e.g., campus storage to see if data is registered.
16. interactive local resources, not just links of resources
17. gitLab, Gitannex assistant (management data same way as source code) can use w/ github
18. Fundref
19. Crossref
20. metadata export to metadata creation tools, part of summary (e.g. morpho)
21. integrated training - more support of training at time of need - not pointing to a link, but a pop up video (integrated into workflow)

22. on-boarding and off boarding of faculty integrate DMPTool as part of storage -

23. Hydra development

24. Solicitation/level DMP requirements. would need solicitation scraping.

25. collab software - SharePoint, alfresco, Sakai, etc.

26. analytics software, Reporting software - piwik - information value to admins and institutions

27. Standard way to refer to something - like Facebook "@" - autofill, auto completion, have an index of the choices. data is normalized as people are clicking

28. Project management tools - Pivotal, ms project (template generated widget)

29. integration w/ latex - markdown; write using form of rest of proposal

30. CV creator - integrate w/ google scholar - ORCID connection???

31. cross-linking proposals - get the proposal that goes w/ the DMPTool (upload the proposal)

32. Safe archive @ UNC

Breakout Group 2

1. Identity/CV: consider IDs for plan, user, agency, institution
   - ORCID
   - DataCite
   - Profile enrichment – Shibboleth
   - Vivo
   - Fundref
   - Social networks
   - Campus LDAP directory
   - EZID – for plans (when to use ARKs, DOIs, etc?)
   - Profiles

2. Metrics and credit
   - Usage stats- how often shared/used
   - Compliance metric – are they doing what they said they would do
   - Something that allows users to rank good plans
   - symplectic – digital measures
   - Altmetrics/DCI

3. Repository selection
   - Match-making wizard
   - "getting started" email from service provider
   - ID for each repository
   - Databib hook

- Look at what funders recommend and enforce
- Broaden definition of repositories – google drive, dropbox

4. Metadata creation
   1. consider metadata for the plan versus the research object
   2. recommend metadata schemas – hook into repository for this
   3. metadata for DMP (provided via API?)
      - identifier
      - repository
      - institution
      - agency
      - orcid
      - etc
      - licensing and use rights

5. Storage and preservation
   - Different needs throughout life cycle. Do we want to encourage use of tools that are hindrance to long-term preservation?
   - Set up storage space at the beginning of the project
   - SafeArchive
   - Integrate with citation managers – zotero etc.
   - do commercial entities want plans?
   - Integrate with grant management systems
   - Hive – suggests ontologies. Automated metadata.
   - Social networking to find relatoinships to people, data to people...

Breakout Group 3

1. Google Analytics
   - Track which resources are being used
   - Track how people are navigating around the tool, where they are coming from
   - Other analytics tools exist
2. Minting ARKs: automatically generate an identifier?
3. FundRef/CrossRef: Track publications funded by specific funders
4. ORCID
   - Potential user error
   - Part of single sign-on? (ideal)
   - Orcid has an API for fetch?

5. Metrics/Alt Metrics

   ○ ImpactStory tracks Figshare

   ○ We can punt this, unless DMPTool wants to be known as a repository of data management plans

   ○ Can we do this on the institutional level rather than the individual level? Institution more likely to care

   ○ Becomes a way to support people who need to generate reports, encourage top-down use from Office of Sponsored Projects (and similar named sub-orgs)

6. Institutional Integration Stuff

   ○ Ask a librarian would be neat, but are there librarians on ask a librarian who know about data issues

   ○ The widget itself is easy to implement

   ○ How can we support ask the librarian staff?

7. Storage and Preservation

   ○ One-button deposit into figshare? Merritt?

   ○ Merritt is keeping it in the UC system

   ○ Proposals that are funded are made public. Where are they? How do we get them?

   ○ Electronic Research Administration tools

Report Back & Consolidate

Identifiers for plans themselves, users of DMPTool, funders, and institutions
1. EZID for plans
2. ORCID for individuals – capability to look up existing ORCID or request creation of new ORCID
3. Fundref for agencies/funders. Would be part of DMP template creation phase. Widget for selecting agency
4. social network profiles – Twitter, LinkedIn, Facebook
5. Profiles – e.g., UCSF profiles, Harvard profiles, NC Link
   would allow integration of data into larger bodies of profile and funding data.
6. Vivo
7. Symplectic
8. SciVal

Altmetrics at the institution level to generate reports for OSP etc. Consider using figshare (see below)

Figshare – plans deposited there, and they would generate altmetrics for us (via ImpactStory etc). But then our DMPs would be in figshare instead of under our control.

SHARE project – notification system for research results of all types (data, papers, etc). Would be captured and metadata would be shared. Used by funders, institutions etc. to track outputs.

Metrics & Credit

1. Usage statistics – how often things are viewed, downloaded
2. Compliance metrics – did you do what you said you'd do
3. Evaluation metrics for plans themselves. Apply a ranking (in-house? Crowd-sourced?) so people can know quality plans. Up/down vote.

Tools: ImpactStory, Altmetric, DCI / Scopus

Google or other analytics (Piwik) – how people are using the tool. Where they clicking, where from, where they are going.

Storage / budget calculator. Help estimate costs while you are creating your DMP. Could have a marketplace concept where storage companies (Dell, amazon, …) could bid.

Repository Selection - match-making wizard. Supply description of data. On basis of responses the tool suggests a list of repositories that might fit needs. Need expansive definition of "repository". Commercial alternatives included (but moral dilemma?)
- re3data.org, databib, biosharing – dropdown menu.
- wizard takes into account things like sensitivity, IRB, privacy and personal info, country, funder, etc. weighted.
- should have IR pop up at the top of that list.

Metadata creation – metadata for what?
- for plan – structured summary
- post-award activity; can integrate with metadata tools for research outputs (like Morpho)
- wizard for metadata standard selection
- DCC metadata registry; RDA registry
- export DMP metadata so available for import into metadata creation tools. Then DMP and data are linked.

Other stuff
- License decision tree – look at Dataverse. Both License for plan and for data: (1) in the plan about the data produced by the project, described in the DMP (2) for the plan itself. Nice to decide this on your profile as default license for all plans you create.
- Post-award: gather information on funded proposals. Scrape info on funded proposals and update DMP database.
- Tie call for proposals with DMP. Some info automatically filled in. Does NSF has API to show current funded proposals?
- Scrape info on solicitations from funder websites to inform DMP templates for specific solicitations. Crabtree knows about tools that can search CFPs for data-generating proposals that have been funded.
- LaTeX / markdown creation of DMPs
- Library chat widgets – chat with your librarian about DMPs while creating one
- Lookup/autocomplete service. Pre-populate list and as you type it fills in the term (repository name, metadata standard, etc.). Keeps track of mentions. Then other services can tie into DMPTool – look for mentions of "Databib" for example. Or if it detects words it can provide custom help text.
- Hive – controlled vocabulary tool. Uses standard vocabs but can also add custom vocabs. Ontology library.
- Safe Archive tool at UNC – determines compliance. Measures what you said versus what you did.

- Information management systems for lifecycle management. LIMS (redcap), OSF
- Integrated training popups – for example video pops up. Pushes training into DMPTool use since that's when they need it.
- New faculty onboarding – grant officers or research office teaches researchers how to do DMPs using DMPTool
- Project management tools – MS Projects, Pivotal Tracker. Spit out a template for the tool that has the timeline etc related to data management. Project management institute – look at parts DMP can fulfill
- CV creators – google scholar. Related to IDs, profiles.
- Intelligence for pointing to local data management resources. Not just big repositories. Storage options appropriate for their kind of data. Would replace multiple templates for different groups, disciplines etc.
- data bridge – NSF grant in for it. Uses social network analysis to look at relationships among data, people, repositories etc. and recommends answers for DMP prompts. Looking at scalability.
- checklist

Part 3: Hopes & Dreams

Highlighted = "won" for dot voting

1. Tie to larger proposal writing so not just DMPs. More seamless. Minor (4)

2. DMPTool information export-able for downstream use. Kratz (API) (3)

3. Licensing of research products. Legal status of content created by DMPTool. Widget. Puneet (6)

4. Find out funded grants. Scott.

5. Checklist for researchers. More extensive. Chris (4)

6. Link to submission process – funder buy-in. Amber. (8)

7. Identifiers for people, funders, institutions, plans – Dave. (27)

8. Recommender for repositories – databib + local options. Stephen. (16)

9. Metadata standard recommender – Susan

10. *(repeat) ORCID – Carly*

11. Integrate with other services that do things well. Don't try and do everything. Make a good API. Karthik. (15)

12. Demand reporting to service providers (e.g., repository service) Noah (2)

13. Public display of data management practices – gold stars, badges for institutions. Dan. (3)

14. Funders help put content in for DMPTool. Sherry (4)

15. OSF - Connect to things that support life cycle management. Sallans.(6)

16. Library chat widget. Perry. (3)

17. Pre-populate DMP with call for proposals. Michael. (1)

18. Export mechanism with quantitative measures for audit compliance. SafeArchive. Jonathan. (3)

19. Having a write-back API. Especially institution-generated parts of plan. Aaron (3)

20. databridge of social networking. Bhavi (2)

21. Analysis tools for institutional administrators. What's happening overall. Sarah. (5)

22. Autocomplete/vocabulary/ontology David M. (2)

23. Button for human subject data – affects a lot of components for DMP. Kratz.

24. Altmetrics/metrics for creating plans. Can help motivate plan creation. Scott. (1)

25. Integration with campus grants management systems. Chris. (10)

26. Become a DMP repository with a quality search tool. Amber. (2)

27. Import existing plans – deposit your plan in the DMPTool. Marisa

28. Highly structured internal data. Stephen. (3)

29. Providing meta-analytics for how people are using the tool at your institution. Dan (7)

30. *(repeat) Fundref. Michael.*

31. Metadata for the plan itself- well-formatted file that follows the plan. Jonathan.

32. DMP diff. Look for patterns. What differs or doesn't. Aaron.

33. Users can come back and update the plan – what they did. Shows up in a view for audit. Puneet. (4)

Part 4: Final List of Ideas

Summarized from all breakout groups, report backs, and brainstorms.

Idea category	Description	Software/application
Cost	Provide cost estimate for data management/storage - calculator for budget	
	"Market place" of storage options – allow storage providers to "bid" for your project	
Funder & funding	Integrate with funder systems – submit directly to funder from tool	Fastlane, grants.org
	Tie call for proposals directly to data management plan. Some info automatically filled in. Scrape info on solicitations from funder websites to inform DMP templates for specific solicitations.	
	Cross-linking proposals – upload or link to the proposal that goes w/ the DMP	
	Gather list of funded proposals to help identify successful dmps	
Lifecycle management	Checklist, outputs for post-award tasks	Project management tools (MS projects, pivotal tracker), Collaborative tools (Sharepoint, alfresco, sakai), workflow systems (OSF)
Outreach	Integrate DMPTool education with faculty orientation and exit interviews	
	training at time of need – integrated in DMPTool as pop-up video, for example	
Identity	ORCID for individuals – capability to look up existing ORCID or request creation of new ORCID	ORCID
	Integrate with CV creators to list DMPTool as a research output	google scholar, academia.edu (likely via ORCID)
	Campus profile systems - would allow integration of data into larger bodies of profile and funding data.	Profiles, Vivo, Symplectic
	Widget for selecting funding agency (Would be part of DMP template creation phase)	Fundref

Idea category	Description	Software/application
	Track publications funded by specific funders	CrossRef (with Fundref link)
	Use social network profiles as login	Twitter, LinkedIn, Facebook
	Integrate with campus identity management	LDAP Directory
	Identifiers for plans (ARKs, DOIs, etc)	EZID
	Identities for each repository	
Metrics & analytics	Plan-level: Plans deposited and altmetrics would be generated for each plan	Figshare (advantage – they already have altmetrics; disadvantage – for-profit). Merritt
	Institution-level: which resources are being used; where they are coming from	google analytics, piwik
	Tool-level: how people are navigating to/from the tool	google analytics, piwik
	Analysis tools for institutional administrators- repositories mentioned, number of plans, list of plan creators, templates used etc.	
Repository & storage	match-making wizard - supply description of data, and tool suggests a list of repositories that might fit needs. Need to define repository- Google drive? Dropbox? Do we want to encourage use of tools that are hindrance to long-term preservation? Wizard takes into account things like sensitivity, IRB, privacy and personal info, country, funder, etc. weighted. Should have IR pop up at the top of that list.	
	Metrics/notification to IR staff about data. E.g., capability for "getting started" email from service provider if particular repository is chosen	
	One-button deposit of plans into a repository	figshare, Merritt, escholarship
	Interactive local resources based on responses. Would replace multiple templates for different groups, disciplines etc.	LIMS RedCap, Hydra
	dropdown menu that lets you select your repository from a list	re3data.org, databib, biosharing
Metadata	Metadata for DMP: identifier, repository, institution, agency, orcid, licensing and use rights, etc	API?
	Recommend metadata schemas (perhaps via repository	DCC metadata

Idea category	Description	Software/application
	selection?)	registry; RDA registry
	Plan metadata export to metadata creation tools to start populating project-level metadata fields	morpho
	Suggests vocabulary / autofill terms. One implementation would be Facebook "@" - autofill, auto completion, have an index of the choices. data is normalized as people are clicking	Hive
Institution admin	Grant management systems - Electronic Research Administration tools	Cayuse
	Administration management systems	Kuali
	"Ask a librarian" widget	
	Write-back API. Especially institution-generated parts of plan.	
Compliance	Export mechanism with quantitative measures for audit compliance.	
	Users can come back and update the plan – what they did. Shows up in a view for audit.	SafeArchive - determines compliance.
	Button for human subject data – affects a lot of components for DMP.	
Governance	License decision tree for data discussed in dmp	Dataverse
	License for plan Nice to decide this on your profile as default license for all plans you create.	CC widget
Misc	Provide ability to write DMP with LaTeX and/or markdown; write using form of rest of proposal	WriteLaTeX
	Become a DMP repository with a quality search tool	
	Import existing plans – deposit your plan in the DMPTool	