DMPTool2 & Alfred P. Sloan Foundation Grant Meeting Report

1. Purpose of the meeting

The primary goal of the Alfred P. Sloan Foundation funded project is to improve on the Data Management Planning Tool (DMPTool), a free, easy-to-use application that guides researchers through the process of creating data management plans. To accomplish this, we aim to build on the success of the tool to create DMPTool2, and use this improved version as a centerpiece for encouraging collaboration in data management efforts across all stakeholder groups (researchers, institutions, funders, libraries).

In support of the project goals, we convened a meeting of DMPTool partners to synchronize the project kickoff efforts and revisit our planned activities. The meeting aimed to review:

- Current DMPTool status
- Community engagement plans
- Functional development plans
- Metrics for impact and success

2. Agenda

Wednesday 20 Feb

	Block 1: Setting the stage	
8:30 – 8:45	Introductions and logistics	Andrew/Carly
8:45 – 9:30	Overview of DMPTool2 Sloan, IMLS, and USGS Projects: objectives, phases/timelines, teams, advisory groups	Andrew/Carly
9:30 – 10:30	Detailed discussion of project activities of DMPTool2 Sloan Project	Andrew
10:30 – 10:45	Break	
	Block 2: Community engagement	
10:45 – 11:15	Community engagement: proposed work plan, strategy, timeline, advisory groups	Sarah
11:15 – 12:00	Discussion and brainstorming of additional factors, opportunities to increase engagement	Amber
12:00 – 1:00	Lunch	
	Block 3: Development of new functionality	
1:00 – 1:40	Review of functional requirements	Stephen

1:40 – 2:20	Phase 1 – augment/extend current RDMS model	
2:20 – 3:00	Phase 2 – adding functionality for DMP creators	
3:00 – 3:15	Break	
3:15 – 3:30	Phase 3 – adding functionality for administrators	
3:30 – 4:00	Phase 4 – production release and documentation	
4:00 – 4:30	Summary of the day and plan for tomorrow	Andrew/Carly
5:30	Group dinner	

Thursday 21 Feb

	Block 4: Metrics for impact and success	
9:00 – 9:15	Overview of the day	Andrew
9:15 – 9:30	Recap of proposed metrics for impact and success	Andrew
9:30 – 10:45	Breakout discussion to identify additional metrics for impact and success, solutions for automation, how to add value for users	Facilitators: Sarah, Michael, Gunter
10:45 – 11:00	Break	
	Block 5: Wrap-up	
11:00 – 11:15	Summarizing	Andrew
11:15 – 12:00	Discussion, takeaways, questions	Andrew

3. Participants

Amber Budden, DataONE
Andrew Sallans, University of Virginia
Carly Strasser, California Digital Library
Daniel Phipps, California Digital Library and University of California Los Angeles
Gunter Waibel, Smithsonian Institution
Marisa Strong, California Digital Library
Michael Witt, Purdue University
Perry Willett, California Digital Library
Rosalie Lack, California Digital Library
Sarah Shreeves, University of Illinois Urbana-Champaign
Sherry Lake, University of Virginia
Stephen Abrams, California Digital Library
Stephen Eisenhauer, University of North Texas
Susan Borda, University of California Merced
Todd Grappone, University of California Los Angeles

4. Setting the Stage

A session providing an overview of where the DMPTool stands today, how the Alfred P. Sloan Foundation grant aligns with the IMLS and USGS grants, and what the meeting aims to accomplish over the next day.

- DMPTool Today
  - Free
  - Workflow for creating a DMP
  - Helps meet funder requirements
  - Institutional integration via Shibboleth
- Lots of institutions using Shibboleth now
- Many funder requirements presently supported
- Current management/operation
  - Planning/oversight
- DMPTool Steering Group
  - Operation
- Shibboleth integration (Perry)
- Manual loading of institutional text/resources (Perry)
- Manual user stats query/report construction (Perry)
- Creation of funder requirements model/loading (Perry/Sherry)
  - Outreach
- Blog
- Conference promotion is ad hoc/opportunistic
- User statistics
  - Constant growth in all aspects (a good thing)
  - Always more users than plans, meaning low repeat use (bad thing)
- Example of user statistics by view of ARL institutions
  - Users at about 77% of ARL institutions
  - Shibboleth authentication in place at 34% of ARL institutions
  - Much higher rates of plans and users at institutions using Shibboleth compared to those not, so getting Shibboleth setup suggests a certain hurdle to higher use
- Funder template statistics
  - Highest use (normalized) by NSF BIO, NSF SBE, and NIH
- Stressing that DMPTool is about data management planning made EASIER (not EASY)
  - Still complex, requires dialogue amongst many stakeholders
  - Need to understand and know about available support resources
  - Focus on simplifying and scaling common parts, building community, and providing functionality to advance services when possible
- Discussion of three common projects for 2013
  - Alfred P. Sloan Foundation grant
  - IMLS grant
  - USGS grant
- General overview of sections to be discussed later in meeting

5. Community Engagement

A session providing an overview of the community engagement activities and issues that we aim to address throughout the project, as well as a discussion section on additional ideas and opportunities. More also covered in the IMLS meeting afterwards.

- Target people: librarian, IT, admin, researchers
- Targeting particular groups for marketing such as ARL
- Webinar series
- Research development conference?
- Advisory Board
    - Administrator user advisory board
    - Researcher user advisory board
- Governance:
    - Continue prior work and roll out the governance structure as a means of formalizing how the community can participate and what roles everyone can play
- Building community
    - Why is working with researchers hard? Is it that they don't care about this, or that we just don't know them?
    - Google group?
    - Funders
        - What's the goal?
            - We want funders to contact us with new requirements
        - Focus group of faculty reviewers --> could be source of contacts
- Consider using relationship with DPN to broaden the reach of DMPTool activities and use

6. Development of New Functionality

A session focused on new functionality, requirements, and the development process. This session included extensive discussion of requirements outlined previously, but needing a revisit for currency.

Itemized review of 10-point functional requirements list, with discussion questions:
- Is a two-level template hierarchy sufficient?
- Are enumeration questions useful?
- Is there value in having strong typing for template questions?
- Are questions specific to the context of a single template?
- Template questions are associated with three types of answers (example answers, suggested answers, default answers): is the distinction between each clear and is there value in all three?
- Is there value in distinguishing between "generic" and "project" templates?
- Are these the right roles (ie. Plan owner, unit administrator, organizational administrator, DMP administrator)?
- Is a three level hierarchy for institutions sufficient (ie. Jurisdiction, organization, unit)?
- Can we assume that a given user is associated only with a single institution?
- Are these the correct actions (ie. Read unpublished plan, read published plan, create new plan, add owner to plan, edit plan, submit plan, publish plan, approve/reject plan, search/report on plan, create plan template, edit plan template)?

- Are these the correct state transitions?
- Do we need "expired" and "inactive" states?
- All collaborators on a given plan have the same set of privileges; is this correct?
- What options for institutional branding are appropriate?
- Are the necessary administrative actions restricted to: 1) create new, or edit existing, user account, or 2) create new, or edit existing, plan template?
- How sophisticated should template-editing capabilities be?
- Can we flesh out the requirements for advanced reporting (ie. Searching, matching, etc.)?
- Are there any other new requirements?
- Should we plan for interim prototype releases (perhaps end of phase 2)?
- Should we incrementally enhance the existing codebase or start from scratch?
- Who is responsible for producing updated user guide, training, and outreach materials? Where are the intersection points with the IMLS grant in this regard?

Phase 1 – preliminary
- Assemble project team (two application developers, thinking one senior and one junior)
- Refine use cases, requirements, technical specifications, UI design/workflow
- DB modification, Rails upgrade, test environment
- Projected late February through late April

Phase 2 – requirements 1-8
- Templates, users and institutions, roles, lifecycle support, search/browse, collaboration, branding
- All iterative development with user feedback and testing

Phase 3 – requirements 9-10
- Administrative interface, advanced reporting (likely a bit more basic than advanced)
- Aim to make this simple while providing enough functionality, otherwise this could be a project in itself

Phase 4 – deployment
- Production release, testing, documentation, training, outreach

7. Metrics for Impact and Success

A session focused on areas where we might better measure the value that DMPTool delivers to various users groups. The participants broke into small groups to discuss four main questions: 1) new metrics for impact and success, 2) whether the metrics could be automated or not, 3) how to add value for our four main user groups, and 4) metrics to use in meeting the needs the grant project funder. Following reports from each group, the metrics were rated in terms of difficulty of measurement, "Easy" to "Hard". We plan to prioritize these and then make plans on how to implement those which are feasible.

General
- Google analytics
- # of completed plans, how far along
- visit duration and per question
- 1st time user demographic – how to hear about tool, role

- follow-up opt-in for post testing
- What can be benchmarked now vs. new tool – particularly for Sloan

Funders
- Use by agency (**Easy**)
- Collect reviewer guides vs. plans/templates (**Hard**)
- Ability to funders to rapidly send out updates – tracking requirement changes (**Hard**)
- How to track more consistency in plans created w/ tool (**Hard**)
- Compliance (**Hard**)
- Better return on investment and data sharing – care about the big story, data management planning means better data sharing means better research means better world (**Hard**)
- Usage metrics (**Easy**)
- Lifecycle metrics (**Hard**)
- Resource usage – idea of looking at institutional preparation/investment to support data management, accreditation, ICR, rating systems (**Medium Hard**)

Researchers
- Focus on better plans, quicker creation ()
- Funder endorsements
- Success rates (**Hard**)
- How many other researchers use tool (**Easy**)
- Time to plan-quicker (TTP) (**Hard**)
- # Funders in area of research, coverage by tool (**Easyish**)
- Efficiency – in and out quickly, don't care much about data sharing (**Easy, possibly weak**)
- Support/access to services
- Compliance assured – don't know what to do, guidance in moment of uncertainty (**Easy?**)
- Usage metrics (**Easy**)
    - Abandonment – start vs. completed
    - Where in the process do they leave
    - Do they come back
- How many published/made public (**Easy**)
- Repeat users
    - # of plans (**Easy**)
    - Frequency (**Easy**)
    - # of days to complete
- # of people involved in a plan creation (via collaboration function) (**Easy**)
- Ratio – users/plans (**Easy**)
- Use of social networking/alt metrics tools (**Easy, if developed**)
- Tracking of grant names/solicitations (**Hard**)

Institutions (
- Meaning research administration, IT
- # of users (**Easy**)
- # of plans (**Easy**)
- Correlation with # of awards (**Hard**)
- Funder breakdown (**Easy**)

- Links to tool – traffic from institution (**Easy**)
- Links from tool back to institutional things (**Easy**)
- More grant success (**Hard**)
- Better return on investment
- Better use of university services – effect of DMPTool upon other services (IRs, OSP training, compliance, etc.) (**Hard**)
- Efficiencies (fewer man hours all around) (**Hard**)
- What is getting funded?
    o Does DMPTool help (**Hard**)
    o Are resources used, which? (**Hard**)
- Metrics on lifecycle compliance, etc. (**Hard**)
- Comparison of DMPs produced in tool vs. not (**Hard**)
- What peers are doing? (**Easy**)

Librarians (and other support units)
- # of contacts to library from tool and vice versa (Easy and Hard)
- Plan quality – tool use vs. not (**Hard**)
- Referrer – where from? (**Easy**)
- Mining plans for local resource mention (**Hard**)
- Relevance/ownership
- Communication/advertise services about existing services
- Data management support program straight out of box
- Business intelligence
- Sustainability of tool overall – service/tool cost versus overall usage compared with what it would cost at individual institution (**Easy to approximate**)
- Use of resources – very granular (**Easy**)
- What other institutions are doing (**Easy**)
- What's being funded
- What are the successful plans within institutions? (**Hard**)
- Distribution across domains
- Usage metrics for researchers but more segmented (**Hard**)

8. Wrap-up

Lastly, the meeting closed with final discussions of next steps, and the following immediate action plan steps for Phase 1 of the project.

Phase 1 – late February through late April
- Hiring:
- Developers hiring (Marisa),
- Community builder hiring (Sarah)
- Hiring of Tao Zhang as UI design manager (Andrew/Michael)
- Requirements specification (Stephen)
- Advisory boards setup (Sarah)
- Meeting with advisory boards (Sarah)
- Developers on- boarding (Marisa)

- Wireframe designs (Tao)
- Setup of community engagement/communications stuff (Sarah)
- Implementation of governance structure (Andrew)
- Meetings once a week for first phase