

May 29, 2013

GPN/GWLA/KU Big Data Summit

Audience participation activity (Summary to be available online later.)

Instructions: Turn to two or three people seated near you, either from your same institution or another.

Work with them to respond to the questions below.

1. Name some things you would envision as elements of a good solution to issues of research data management at your institution:

180 comments were received from 38 forms handed in after the audience participation exercise. Responses are summarized on following pages.

2. Name some stakeholders (people, positions or roles) that must be involved for efforts to succeed.

3. Please share any ideas you have on how institutions could cooperate on data management and curation.

Question #1: Name some things you would envision as elements of a good solution to issues of research data management at your institution.

Comment Categories	DATA SUMMIT PARTICPANT INPUT May 29, 2013	# of Received Comments
Skills / Expertise / Personnel	<p>1. Institutional Solution Elements</p> <ul style="list-style-type: none"> *Several skill sets; *Dedicated person to help faculty in a [?] manage amd organize data; *Profssional lib positions should no require the MLS; *Maximize roles/expertise that different people bring to the problem->solution; *Qualified staff; *Multiple people w/ different bactrounds Lib-research fac. in multiple areas together-IT; *Personnel Organization Mostly. Structuring support services to facilitate data management. *Address data services/librarian role - this person can help researchers find data [and] help figure out where their data should go (function); *Skill sets needed for scientists & librarians & workforce development, e.g., data curricula; *Knowlegeable staff; 	10
Service or Policy Definition / Best Practice & Procedures	<ul style="list-style-type: none"> *Clear idea for library; **Identify levels of service (or areas you intend to support) focus on those areas; Consistency; *taking responsibility; ****policies-uniform (make easy across the board); Policies & procedures;Policies [x2];Policies in place so everyone deposits datasets; *development common standards / standards for the way the data is collected / comprehensive standards; *finding out what the limitations are to achieving best practices; *guidelines about what should be included in datasets for long - term storage; *Establish stakeholders & governance; *Standards for interoperability / sharing /cross-institutional; **Scalability [x2]; *A great process for defining and archiving the data and associated programs. **Categorization of research data -- lab books, equipment, supplies, software;Definitions of data; *How linkage of data sets affect life cycle of each data set; *Make appropriate data management a criteria for tenure? *IRB project review element? *Start with convincing faculty that data mgmt is an appropriate function for the library/institution to perform for the faculty work; *OA & Copyright Policy Baseline[?]; **Establish a framework: policies, procedures, lifecycle, training for data management for creators and users; Lifecycle; 	29

Question #2: Name some stakeholders (people, positions or roles) that must be involved for efforts to succeed.

DATA SUMMIT PARTICPANT INPUT May 29, 2013

of Received
Comments

2. Stakeholders Needed For Success

LIBRARIES: library; librarians; librarians/archivists;Library IR; library IT; library personnel;Dean behind plan; Dean of Library; Dean of Libraries;[Library-liaison librarians, metadata librarians];	28
RESEARCH ADMINISTRATION: research office; research admin; office of sponsored research; VP Research;admin for research/sponsored programs' Vice Provost Research-directors;research administration; VC of Research; university research councils;VPR;research support/office staff;the researchers of tomorrow	26
INFORMATION TECHNOLOGY: IT; IT services; IT - data storage; Central IT;IT units; IT folks; campus computing;IT staff;system administrators; [CiO/Director/DBA, System Admin/Manager/Network Admins];Mike and Andy - IT Guys;CIO;CIO/director;CIO, etc; systems/IT;	31
RESEARCHERS: faculty; researchers; faculty (faculty council); research faculty & Pis;research stakeholders/federated groups of researchers; researchers disposed to work with us;researchers/faculty;	21
ACADEMIC AFFAIRS: Acad. Affairs/VPAA;	1
PUBLIC: Public Users/Funders of Data Collection; Citizens/taxpayers;	2
RESEARCH COMPUTING: Supercomputing Center; IT Supercomputing' HPC Unit; High performance or resarch computing;high performance / academic computing; research computing	8
ADMINISTRATION: President/Chancellor behind plan; Provost/Univ. Admin; Admin (need some motivation from the top); Administration; Administrators; Chairs; Provost; DMP for big data needs to come down from the top-b/c it will require institutional support;University preidents;Deans;Sr. Administration;[Provosts, Deans, dmin in general];University admin.	18
Auditors;	1
Grad Students; GPSGA; Postdocs; Maybe the Grad college	7
University Finance/VP Econ Development/CFO	4
Funding agencies; Sponsors; Financial Sponsors; Sponsors/Financial;	4
General Council for Compliance; IP/copyright and lawyers; IPI People	3
Undergrad researchers	2
Legislators; legislative support for some institutions;	2
Governing Board (Regents?); Board of Regents/Governing Bodies; State governing body -- regents; Board members/state gov. body;	4
Grant writers; College grant coordinators;	2
Grant reviewers	1

Question #3. Please share any ideas you have on how institutions could cooperate on data management and curation.

Comment
Categories

DATA SUMMIT PARTICIPANT INPUT May 29, 2013

of Received
Comments

Skills / Expertise /
Personnel

3. Institutional Cooperation

- *expertise (e.g., HathiTrust);
- *sharing staff & resources -- look at data domain expertise -- pulling in people to contribute

2

Service or Policy
Definition / Best
Practice & Procedures

- *best practices re:policy development;
- *codevelop standards;
- *provide standard formats;
- *create access to templates or framework;
- *Guidelines on ownership of data -- institutions - Researchers?
- *Know what's already available and working at other institutions;
- *Do not reinvent the wheel. There are already many building blocks for layered services in place in Europe and Australia;
- *Institution-level or state-level DMP or DMP development process;
- *establish a statewide task force to derive a data management process that can be used as a general template for public institutions located with the state/region;
- **Establish a framework: policies, procedures, lifecycle, training for data management for creators and users;work together and establish a framework;
- *Heirarchy of data management to establish flow and methodologies;[x2];
- *Establish methodologies;
- *Lifecycle;
- *Policies & procedures;
- *Develop a repeatable framework that we can all use!

17

Assessment

Involvement / Partnership/ Connection/ Communication

- *Sharing the load;
- *Scientists already have. Ask NSF recipients to show what they put in their grants for data sharing;
- *In Oklahoma there is a memorandum 1OK Cyberinfrastructure and there is a lot of cooperation;
- *resource sharing;
- *data sharing networks;
- *Libraries have an existing interoperable structure that could be enhanced;
- *Keep Librarians and technology people talking about standards that allow wide-spread collaboration;
- *KALCDD - Kansas Acad. Library Consortium could add this to their mission/talks;
- *KALCDD - Statewide library consortium could help with the issues for smaller schools;
- *Communication;
- *Consortiums?
- *Share, pool resources & expertise via large consortia;

11

Three questions given to participants at the Big Data Summit as an audience participation exercise.
 Instructions: Turn to two or three people seated near you, either from your same institution or another. Work with them to respond to the questions below. (Questions are noted on the following pages along with the responses.)

Big Data Summit Participant Comments, Interest by Area for Institutions and for GWLA/GPN Partners
 (highest percentages highlighted)

180 Comments
 Total / 38 Individuals

Needs for institutions, opportunities for partners.

Solutions Identified by Participants

	<i>Institution</i>	<i>Partners</i>		<i>Institution</i>	<i>Partners</i>		<i>Totals</i>	
Human Resource - Skills / Expertise / Personnel	10	2		8.13%	3.51%		12	6.67%
Service or Policy Definition / Best Practice & Procedures	29	17		23.58%	29.82%		46	25.56%
Assessment	5	0		4.07%	0.00%		5	2.78%
Involvement / Partnership/ Connection/ Communication	22	11		17.89%	19.30%		33	18.33%
Metadata	4	0		3.25%	0.00%		4	2.22%
Preservation	4	0		3.25%	0.00%		4	2.22%
User-focused Services	12	3		9.76%	5.26%		15	8.33%
Training / Professional Development / Education	13	6		10.57%	10.53%		19	10.56%
Tools / Infrastructure	17	17		13.82%	29.82%		34	18.89%
Funding / Sustainability	6	1		4.88%	1.75%		7	3.89%
Other	1	0		0.81%	0.00%		1	0.56%
TOTALS:	123	57		100%	100%		180	100%

Metadata 0

Preservation 0

User-focused Services Explore shared service through GWL for data services; 3

Talking points to sell to faculty, work with them to make their work more productive.

RECOGNIZE DISCIPLINARY DIFFERENCES

Training / Professional Development / Education *training; training; 6
*GWLA could promote a program in data science--provide a credential
MOOC-Instruction;
*Regional training workshops;
*Worksohs to frame data management;
*Train and grow people who can re-train their institutions. Service and Training.
*Training > users & [?];

Tools/Infrastructure *****collaborative digital repositories & consortium level 17
repositories;Community-based repository; repository; trans-institutional
repositories that share infrastructure; regional repositories;
***tools; Make tools available for use;Expand efforts like GLOBUS
**Researcher Navigator; Research Navigator, configure what we saw
today;
*sharing infrastructure; common infrastructure;
***Federated search & discovery;Discovery; one stop search for
institutions;
*Storage;
*Backup of data from other instituions;

Funding / Sustainability "Economies of scale"; 1

Other 0

Domain specialists that operate as an interface between [other stakeholders];	1
Lab Manager	1
Departmental IT	1
IRB	3
Accreditors	1
Knowledgeable staff	1
Faculty Senate	1

Assessment	<ul style="list-style-type: none"> *Environmental scan as a starting point: what are the needs of researchers / what services are currently available; *Determine from researchers the most needed system elements; *Requirements gathering for data volume & activity; *Survey-where are we now, what is needed? *Survey faculty & GRAs to determine the state & data on campus. Where is it being generated? By whom? How much? Formats? 	5
Involvement / Partnership/ Connection/ Communication	<ul style="list-style-type: none"> *At table with researchers/being at table early; *promoting a culture that emphasizes the importance of data management; ***Better outreach & collaboratio with grant writers / seekers / admin. Regard[ing] services offered in data management planning; Outreach;Publicize & Outreach; ***Develop partnerships with central IT, Research & Graduate studies at all "levels" (e.g. top admin & mid-management researcers etc); Partnerships/collaboration; Partnership with campus stakeholders to craft services--VP Research, college grant coordinators, IT; *Establish governance btwn: IT, Research & Libraries. Determine where/how data will reside on campus; *Team approach-people from different areas at the table-lib, IT, researchers, areas of [?]; *Removal of beauracratic obstacles; *More awareness at all levels (fac,admin); *Awareness building - awareness of the issues of data management; *Awareness building; *Establish protocols & common language; **National Incentives; [x2] for researchers and librarians; *Convincing faculty of data neds. Why do they need a system; **Work together; Work together and establish a framework; *Communication; *Collaborations & partnerships (Office of VP for Research, Consortium, 	22
Metadata	<ul style="list-style-type: none"> *Metadata; *Library creation of metadata; *Metadata application/tools; *Standard metadata practices 	4
Preservation	<ul style="list-style-type: none"> *Preservation; *Preservation Plan for data [software, migration, preservation, emulcation]; *Forever? Really? Decisions about how long stuff should be kept. *Long tail of data not supported [?] has grant monies; 	4

User-focused Services	<ul style="list-style-type: none"> **Curation; Easy access to standardized tools and people who know how to use them; *Knowing when to connect them [researchers] with others; *KISS (Keep it Simple Stupid) for the researcher; *EASY EASY EASY *Outreach to faculty and student researchers; *Publicity; *A Data Mgmt plan + Implementation that is "easy" to use with guidance from research librarians & informatics folks; *Coordinating / Management / Frameworks /Methodology; **Ease ->; ease of depositing datasets; *Safe; *Gather requirements from researchers that will meet 80% of their technical requirements. 	12
Training / Professional Development / Education	<ul style="list-style-type: none"> *I.T. should offer courses in Postgres; *LIB schools should work on topic modelling for metadata; *There should be several cross-college courses in data science; *Education--data info. Literacy; *What about a panel discussion who on campus - what are you doing for NSF grant application; *Training - bring in someone to provide training; **Education; Training; *Training researchers in best practices of data management; *Program/instruction for new researchers on where/how to deal with big data; *Developing skill set for scientists and librarians; *Training > users & [?]; *Heightened awareness via workshops, publicity, etc... 	13
Tools / Infrastructure	<ul style="list-style-type: none"> **Discovery; ability to discover/access data; *Infrastructure; *put it somewhere for starters; *Use the same tools? *Searchable catalog; *Notebook information automated? **Interoperability [x2]; *Storage; storage & computing capacity **Local & national clearinghouses to distribute resources & applications [mentioned 2x]; *Effectively archiving & representing software that is itself the research product; ***Cyberinfrastructure w large, low-cost storage; Storage; Tons of storage capability; *Data management system to store/organize/provide secure authenticated access. 	17

Funding / Sustainability	<ul style="list-style-type: none"> *Funding; *Think through the funding model to think about data that falls through the funder gaps; **Sustainability - local, national [x2]; *A plan to keep the resource in place after any initial implementation; *Scalability; 	6
Other	<ul style="list-style-type: none"> *Make Kepler avail/scientif [sic] 	1