

Implementing New Data Management and Curation Services at the University of Arizona (UA)

Lessons Learned from the UA Data Management and Data Curation Pilot

<https://osf.io/fsx4t/>

Presented by Christine Kollen, University of Arizona Libraries



Outline

- Our journey – how did we get here?
 - Initiatives and projects both campus-wide and in the Libraries
- Changing environments, changing staff
 - Libraries
 - Office of Research, Development and Innovation (RDI)
 - University Information Technology Services (UITs)
- Scaling services, future directions



Research Data Management services

Development of data management services

- Researchers will be able to devote more time to research
- Supports goal of developing a RD ecosystem that facilitates reproducibility
 - Integrates data, code, workflows, other scholarly artifacts (outputs)
- Positive impact on institution's prestige and success in obtaining outside research funds

Integrates FAIR principles – Findable, Accessible, Interoperable, and Reusable



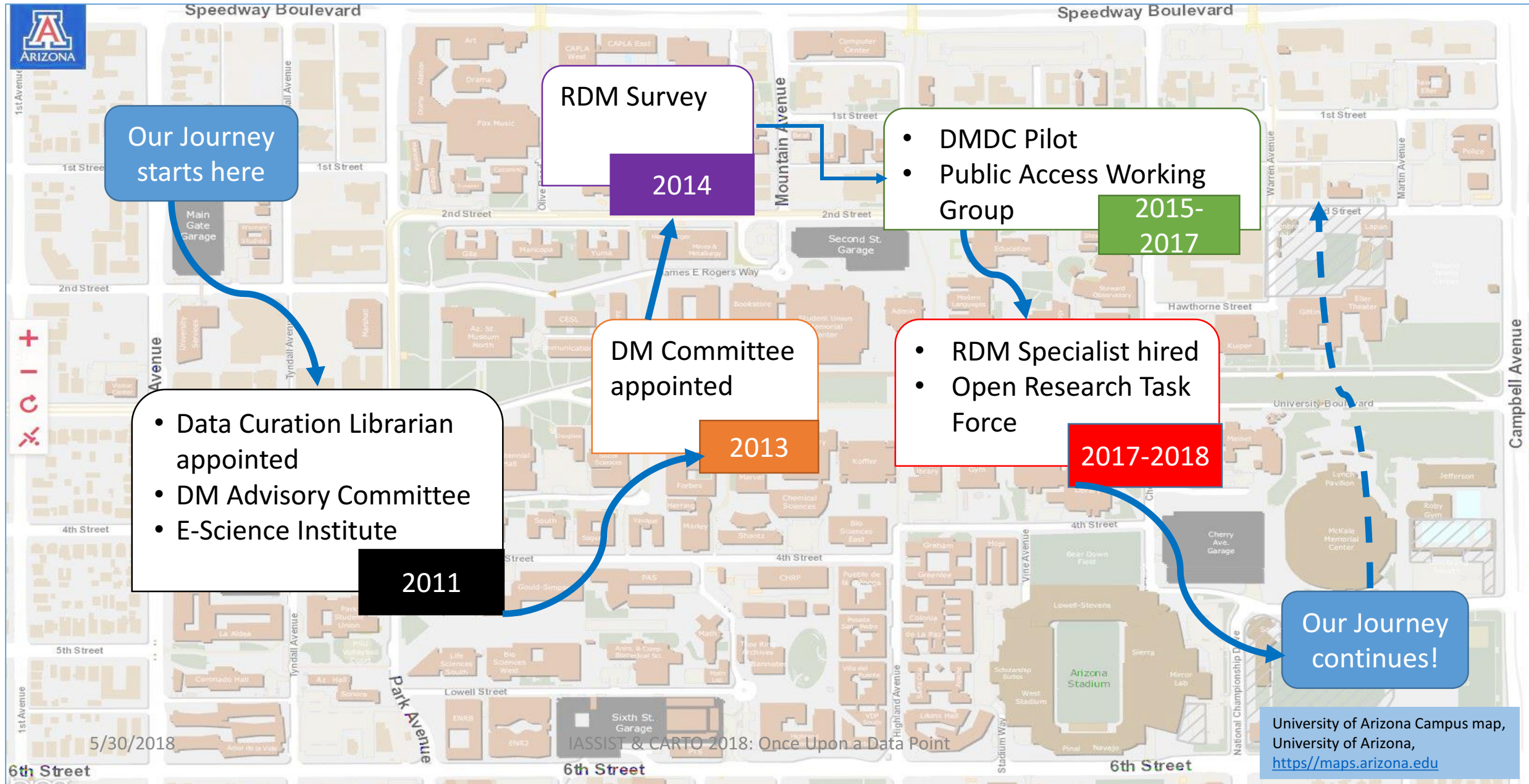
University of Arizona Quick Facts

- Land grant, doctoral granting, Carnegie RU/VH (Research 1)
- FY17 - \$606M in grant funding
- Faculty – 3,158
- Undergraduates – 34,072
- Graduate Students and Professional Students – 9,553



University of Arizona Campus, CC BY 3.0,
<https://en.wikipedia.org/wiki/File:Uofa.jpg>

Our Journey



Campus Data Management Advisory Committee

Recommendations (2011)

- Library serve as point of service for data management and creation of DMP plans
- Support for developing needed cyberinfrastructure to support data storage for both short and long term data management needs
- Libraries, UITs, and RDI develop long-term strategy for support of data storage, data access and data preservation
- Establish ongoing Campus Data Management Committee, report to Libraries, RDI, and UITs



Research Data Management Survey

Campus Data Management and Curation Subcommittee conducted RDM Survey in spring 2014 to discover RDM and curation needs of the UA community

Overall questions included:

- How is research data currently being managed?
- What is the demand for current services?
- What new services are needed?

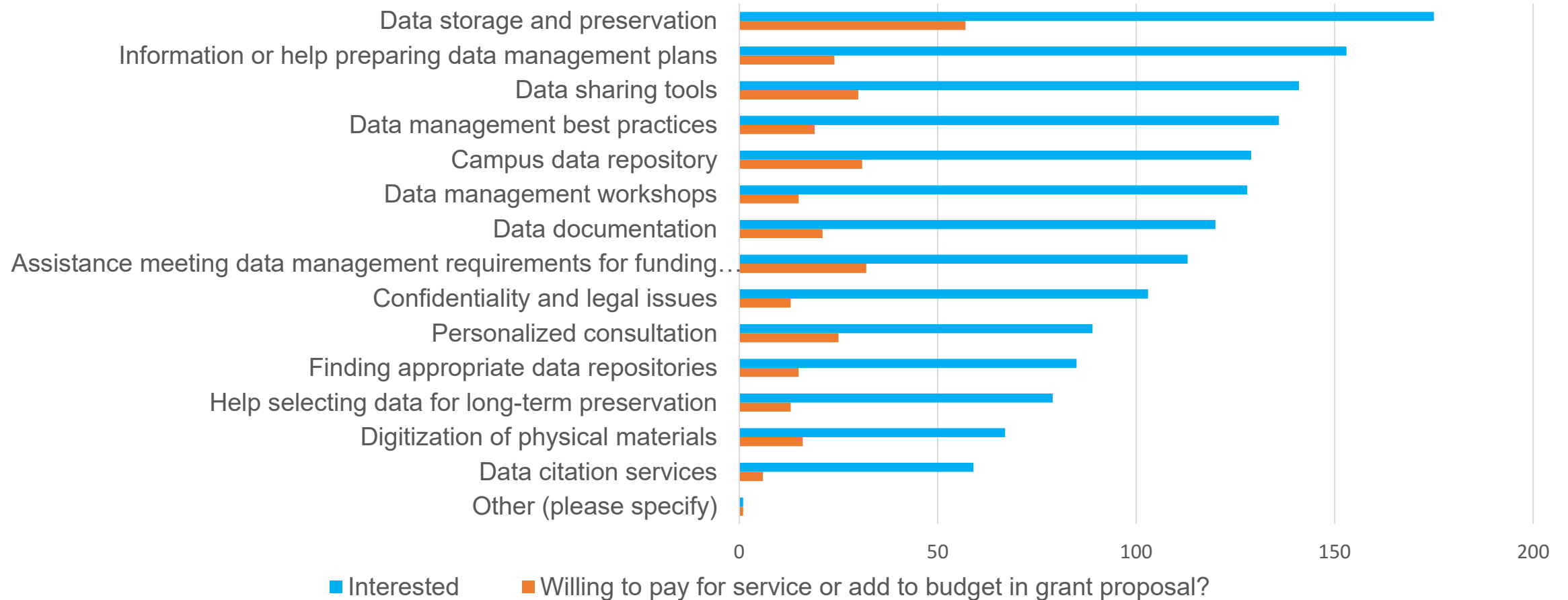
Survey sent to researchers, graduate students, post-doctoral associates

Received 375 responses across a broad range of UA departments and units



RDM Survey results

Interest in Data Management Services



RDM Survey Recommendations

Continue RDM services currently offered and explore offering RDM services related to:

- Data storage and data storage tools
- Data repository
- Data documentation and metadata
- Confidentiality and legal issues

To move forward and gain a better understanding of potential services and tools researchers need

- Developed Data Management and Data Curation Pilot
 - Funded by RCGC sponsors – Libraries, UITs, RDI, Health Sciences



DMDC Pilot

- Goals of the pilot
 - What RDM services are needed
 - What worked well, what didn't work well
 - Training needs
 - Feasibility of implementing these services campus-wide, how can we scale to campus?



Pilot participants

Project	Personnel	Data types	Est. data storage	Length of project	Sensitive or protected
Entomology	PI, 3 co-PIs, 5+ collaborators, Lab Manager	Proprietary, Excel, images	40 GB	5 years	No
Engineering Education	PI & graduate student	Excel, CSV, python scripts	250 MB	3+ years	Yes, IRB, FERPA
Wildlife Biology	2 Research Scientists, PI, 2 students, 5+ collaborators	Access, shapefiles, ArcInfo coverages, images	5 TB	25+ years	Yes, protected (endangered species)
Cancer research	Data Mgr, 3 Research Admin, 4 PIs, 5+ collaborators (3 institutions)	Images, murine models, Access, Excel, SAS, SPSS, Stata	50 GB	30+ years	Yes, HIPAA
Public Health	Data Manager, 2 PIs, 4 co-PIs, 40+ students, volunteers	Research data in RedCap	20 GB	2 years max	Yes, IRB (standard, tribal)
5/30/2018		IASSIST & CARTO 2018: Once Upon a Data Point			

Entomology

Issue – how postdocs and students collect, label, and transfer data to PI

Results

- Developed data workflow
- Recommended standard data management best practices
- Added research data to the Campus Repository

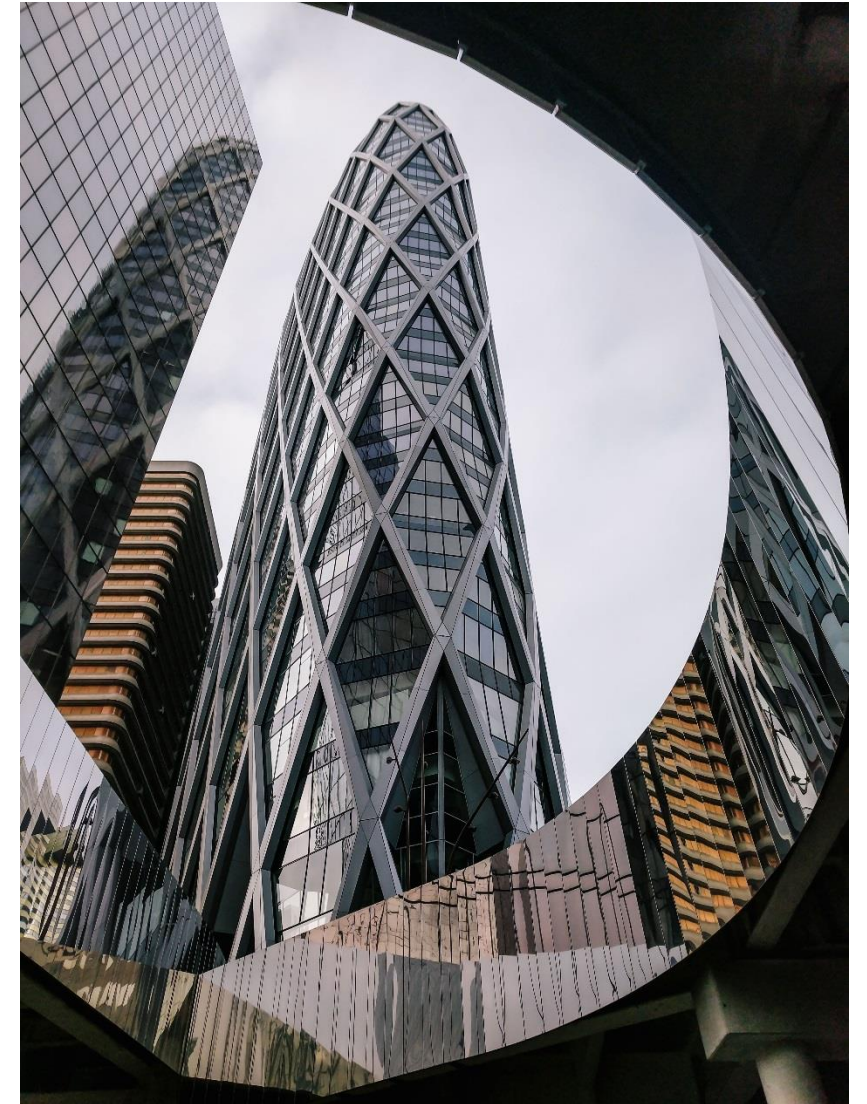


Engineering Education

Issues – moving data from Excel to SQL database, document python script, and de-identify data

Results

- Use Box to store data – anonymized data as outlined by IRB
- Documentation of python script
- Decided to keep data in in Excel and use R for analysis
- Provided information on visualization and analysis support
- Provided suggestions for journals to publish research



Wildlife Biology

Issues – file versioning, access and management protocols, managing student datasets, and sharing archived data

Results

- Document data workflow and provided feedback on documentation
- Recommended metadata schema (Ecological Metadata Language)
- Developed a procedure for student created data
- Guidance on migrating from older versions of Access



Mt. Graham Squirrel image, University of Arizona Conservation Research Laboratory,
<https://cals.arizona.edu/research/redsquirrel/mgrs-photos.html>

Cancer Research

Issues –NIH grant review, study closure and data disposition, good data management practices

Results

- Developed data workflow diagrams
- Feedback on grant proposal
- DM workshop for project team
- Recommended data repositories and developed sample check-lists for disposal or deposit of data
- Document grant and legal requirements for retaining data and physical samples

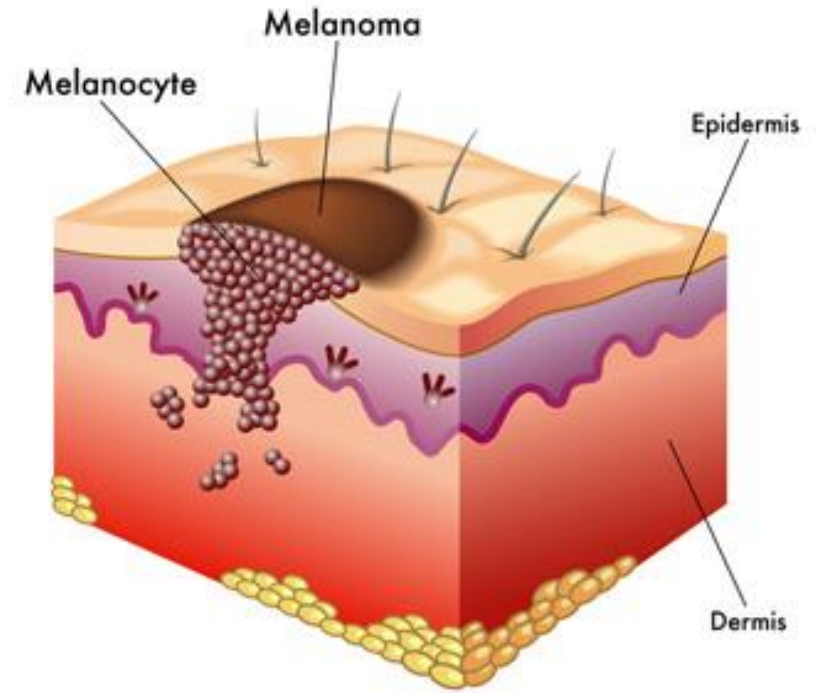


Image of Melanoma diagram. <https://curesearch.org/Skin-Cancer-in-Children>

Public Health

Issues – document and versioning control,
document data workflow and data protocols

Results

- Developed workflows for data collection
- Recommended data management best practices, access restrictions for documents
- Recommended Trello for keeping track of file and data requests
- Reviewed Tribal IRB and consent forms, legal requirements for retaining data
- Created data disposition checklist for end of project



Image of Gold King Mine Spill, Animas River, University of Arizona, Superfund Research Program,
<https://superfund.arizona.edu/info-material/gold-king-mine>



Feedback from participants

Would have liked more “hand-holding”, wanted specific templates and checklists and specific concrete tools

Consider fee-for-service for UAHS (2 projects)

Three most helpful aspects – tools we suggested to manage the project, workflow document, & awareness to be more data-savvy

“Even though I’ve been doing data management for a long time, working with the library on this data project was very useful to me. It was good to know the most recent techniques and to get an outside perspective on work that we were doing.” (Vicki Greer, UA Wildlife Biologist)

RDM practices greatly improved and skills/protocols easily transferable to other projects

RDM workshop, well-received and bolstered their credibility with the PIs

Reinforced the importance of having a DMP



General Recommendations


- Services need to be fully developed, robust, agile, and available at point of need, part of researcher's workflow and as transparent as possible
- Concrete solutions are better – checklists, rubrics, templates, protocols customized for their needs
- Offer for-credit class in RDM; offer training and support services at new faculty orientation and beginning of grants



Specific Recommendations

Service or Project	Priority Level	Responsibility
OSF for Institutions	A	Libraries (ODIS) and UITs
Coordinate services with RDI, Sponsored Projects; notifications of library's RDM services	A	Libraries (ODIS) and RDI (Sponsored Projects)
Data Repository	A	Libraries (ODIS and TeSS); RDI and UITs
Training for liaisons	A	Libraries (ODIS)
ELNs	B	Libraries (ODIS)
Training sessions for graduate students and post-docs	B	Libraries (ODIS and R & L)
Online Tutorials	B	Libraries (ODIS, R & L), RDI
Data Management website update	C	Libraries (ODIS and TeSS)

Implementation Progress

- Setup  OSF **INSTITUTIONS**
- Collaborate with other units – Graduate Studies, Center for Postdoctoral Studies
 - Held two part data literacy workshop for post-docs
 - Graduate student workshop on RDM best practices
- Developed planning proposal for Data Repository
- Exploring potential other services as part of a data repository



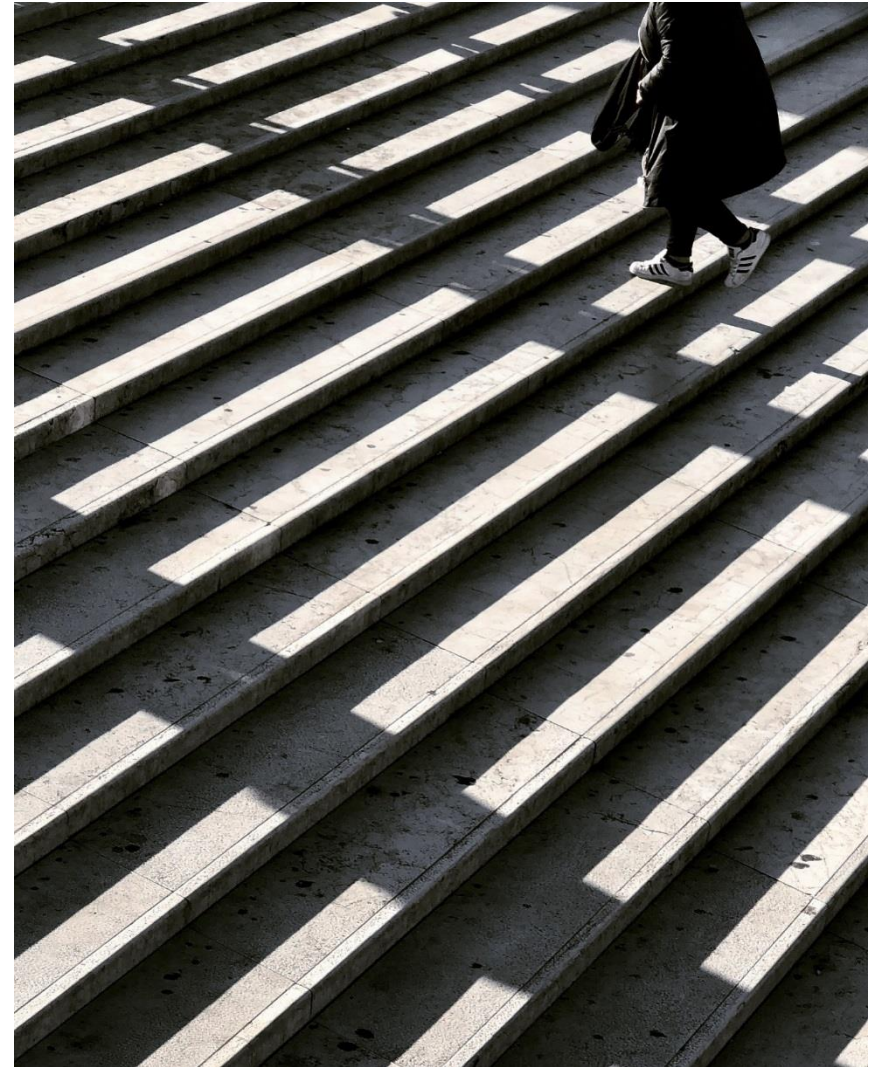
Next Steps

- Data repository
- ELN survey
- Integrate data management into graduate level courses in collaboration with liaisons
- Provide training for liaisons
- Establish a core library data management committee (meets monthly) and hold library data management meetings quarterly for all liaisons



How to scale services

- Involve liaisons
 - With only two of us, we can't do it all! We need help keeping up with disciplinary needs and changes in the landscape
- Involvement of liaisons
 - More than one-way communication
 - Liaisons provide basic RDM workshops for their departments
 - How can we integrate data literacy components into both graduate and undergraduate classes
- Support, from campus – develop services in conjunction with other units, such as RDI and UITS



Changing UA campus and Library environment

- Libraries
 - Research Data Management Specialist (Sept. 2017)
 - Data Science Specialist (Spring 2017)
 - GIS Specialist (2015)
 - Digital Preservation Librarian (Spring 2017)
 - Metadata Librarian (Fall 2016)
- Office of Research, Development and Innovation
 - VP for Research
- University Information Technology Services
 - VP for Information Strategy and University Libraries (Dean of Libraries reports to her)
 - Director of Research Technologies
 - Associate Director of Research Technologies



Open Research Task Force

Goal - develop recommendations for an Open Research ecosystem by performing a needs assessment, environmental scan, and resource assessment

- Sponsored by Faculty Senate Research Policy Committee, RDI, Libraries, and UITs
- Task force members include representatives from RDI, UITs, Libraries, and academic departments
- Final report with recommendations are currently being developed



Conclusion

- Several years providing RDM services – now able to do more outreach, workshops, develop additional services
- Hiring a RDM Specialist has made a huge difference; other positions in our department help support our services
- Setting up OSF4Institutions provided more opportunities for outreach to researchers, post-docs, and graduate students →
 - Tool that fits in well with established workflow
 - Integrates with tools researchers are already using



Questions?



Thank You!!

Chris Kollen, kollen@email.arizona.edu

