University of the Pacific

From the SelectedWorks of Michele Gibney

February 7, 2019

Data Management Planning: Two very important pages

Michele Gibney

This work is licensed under a Creative Commons CC_BY International License.

Available at: https://works.bepress.com/michele_gibney/67/

Data Management Planning:
Two very important pages

Michele Gibney

Fulbright Specialist

Outline

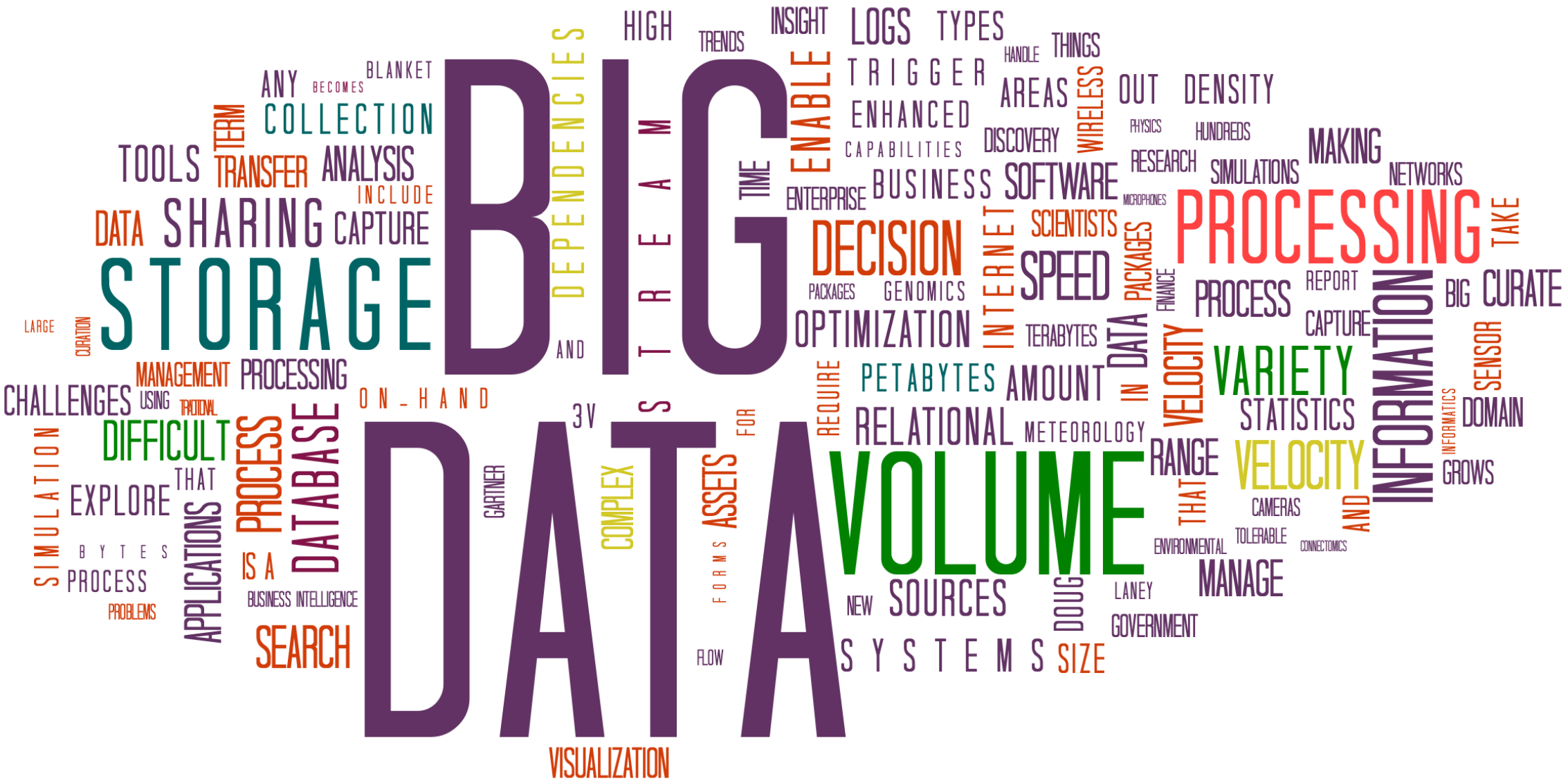
What is data?

What is a data management plan (DMP)?

Why does it matter?

DMP Resources

What is data?


BIG
DATA
STORAGE
VOLUME
PROCESSING
INFORMATION
DECISION
SPEED
VARIETY
VELOCITY
SOURCES
SYSTEMS
SEARCH
DATABASE
PROCESS
DIFFICULT
SHARING
COLLECTION
TRANSFER
ANALYSIS
TOOLS
DATA
CAPTURE
INCLUDE
ANY
BECOMES
BLANKET
TERM
LARGE
CURATION
MANAGEMENT
PROCESSING
CHALLENGES
USING
TRADITIONAL
ON-HAND
SIMULATION
EXPLORE
THAT
BYTES
PROCESS
PROBLEMS
APPLICATIONS
IS A
BUSINESS INTELLIGENCE
DEPENDENCIES
STREAM
HIGH
TRENDS
INSIGHT
LOGS
TYPES
HANDLE
THINGS
TRIGGER
ENABLE
AREAS
WIRELESS
OUT
DENSITY
ENHANCED
CAPABILITIES
DISCOVERY
PHYSICS
HUNDREDS
MAKING
RESEARCH
SIMULATIONS
NETWORKS
TIME
ENTERPRISE
BUSINESS
SOFTWARE
MICROPHONES
SCIENTISTS
PACKAGES
FINANCE
TAKE
PACKAGES
GENOMICS
INTERNET
PROCESS
REPORT
BIG
CURATE
OPTIMIZATION
TERABYTES
DATA
CAPTURE
AND
REQUIRE
PETABYTES
AMOUNT
IN
VELOCITY
SENSOR
3V
FOR
RELATIONAL
METEOROLOGY
STATISTICS
DOMAIN
INFORMATICS
GARTNER
COMPLEX
ASSETS
RANGE
GROWS
THAT
CAMERAS
AND
FORMS
ENVIRONMENTAL
TOLERABLE
CONNECTOMICS
NEW
DOUG
LANEY
MANAGE
GOVERNMENT
FLOW
SIZE
VISUALIZATION

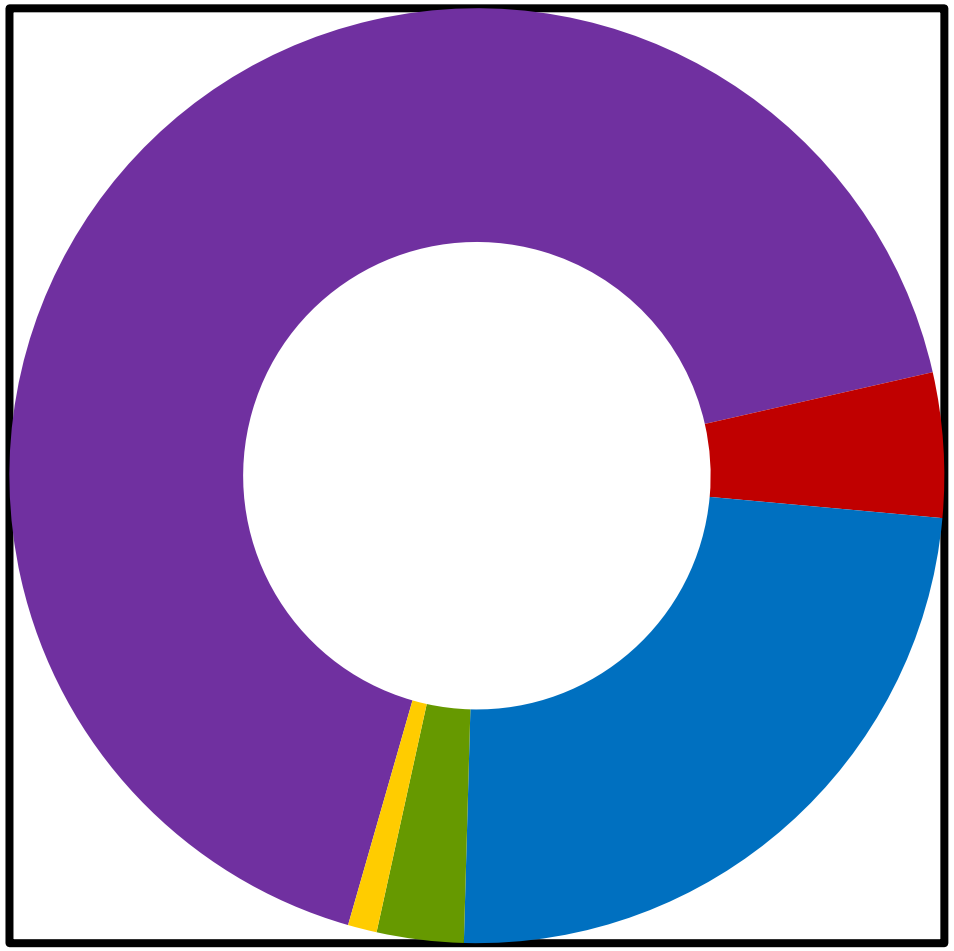
COMMON MISCONCEPTIONS: FILE SIZES

Myth: Huge

Reality: all sizes

Images free from: http://bit.ly/1MSXGnU and https://pixabay.com/p-375022/?no_redirect

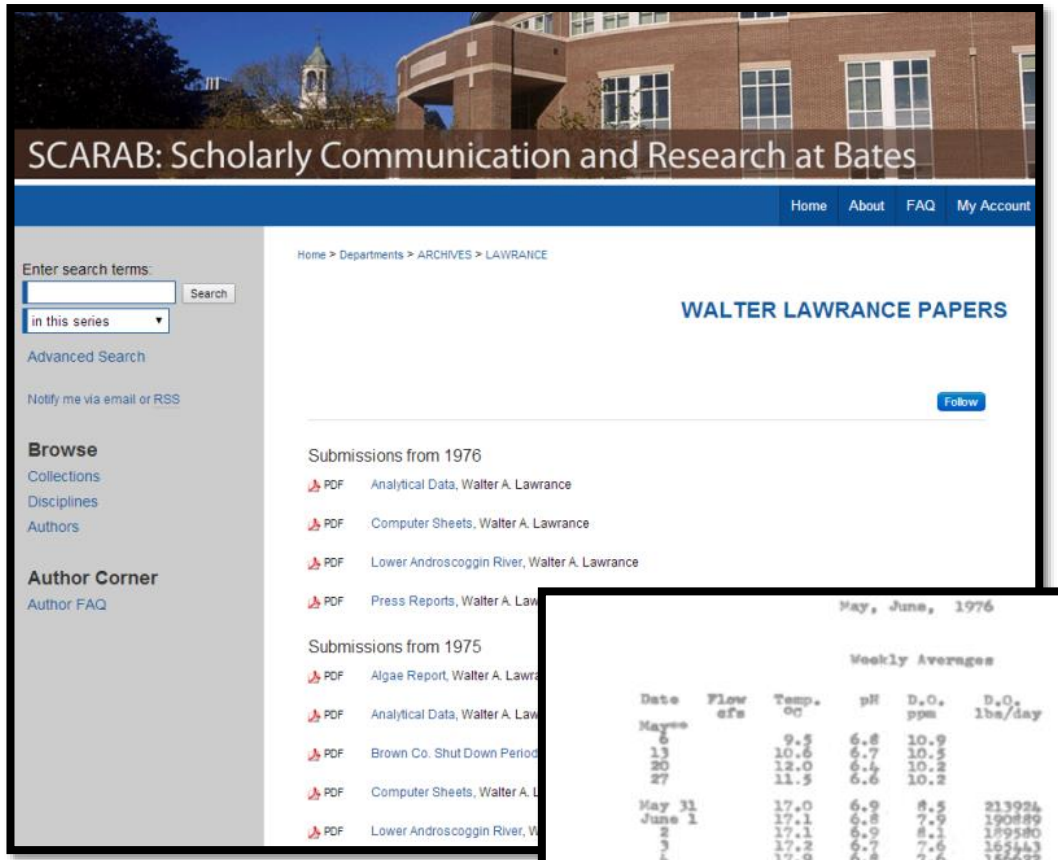
AVERAGE DATA FILE SIZES

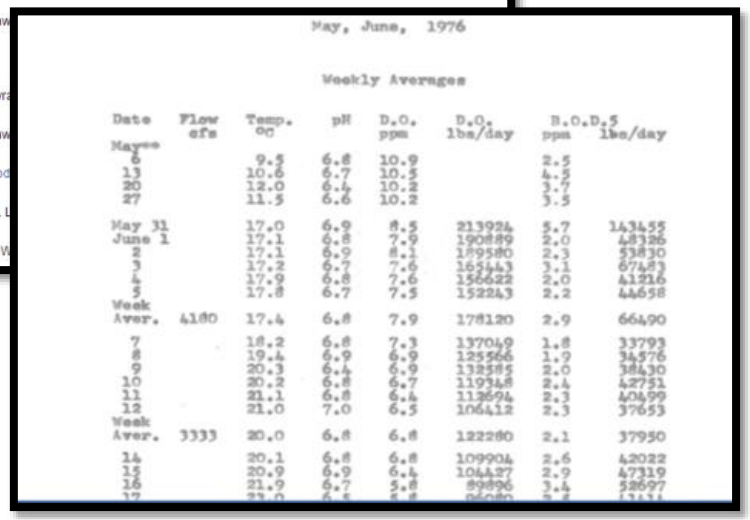


*Compiled from a random sampling of 300 data sets in Figshare, Dryad, and Harvard's Dataverse.

Data from: http://digitalcommons.bepress.com/data/1/

FILE TYPES MYTH: ALL ARE NUMERIC

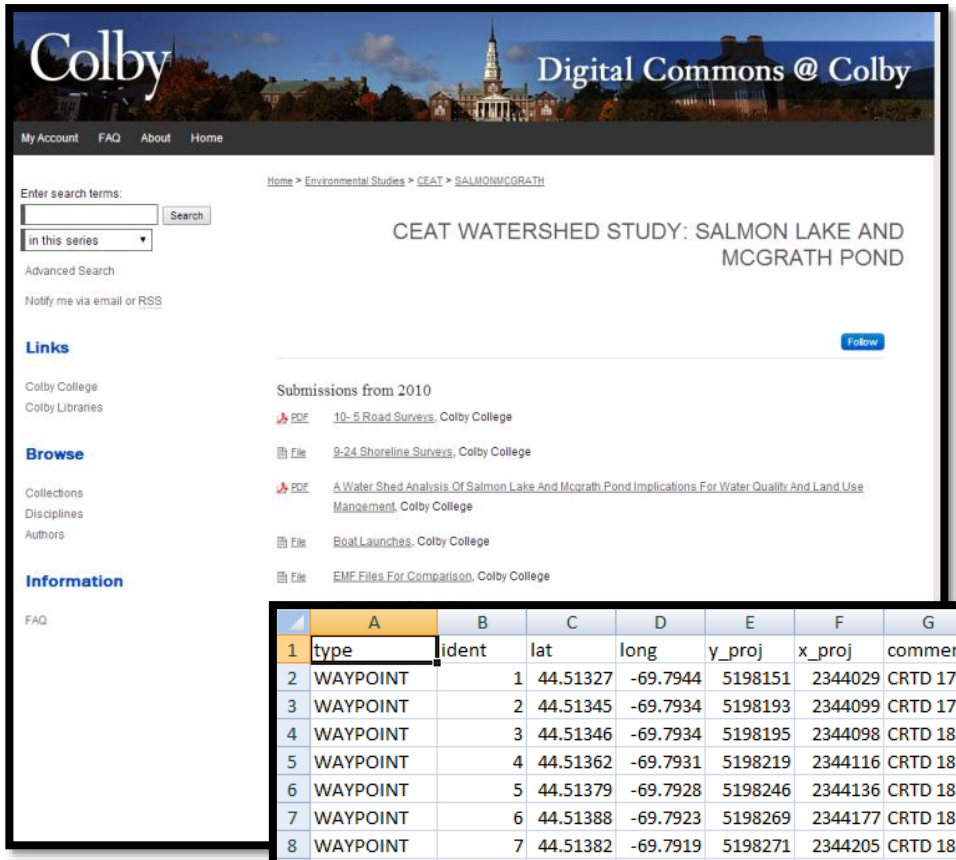


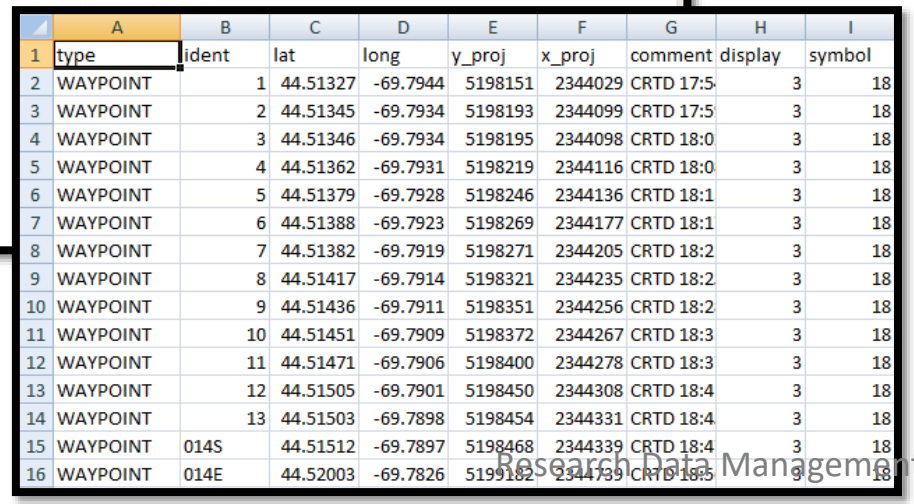


May, June, 1976

Weekly Averages

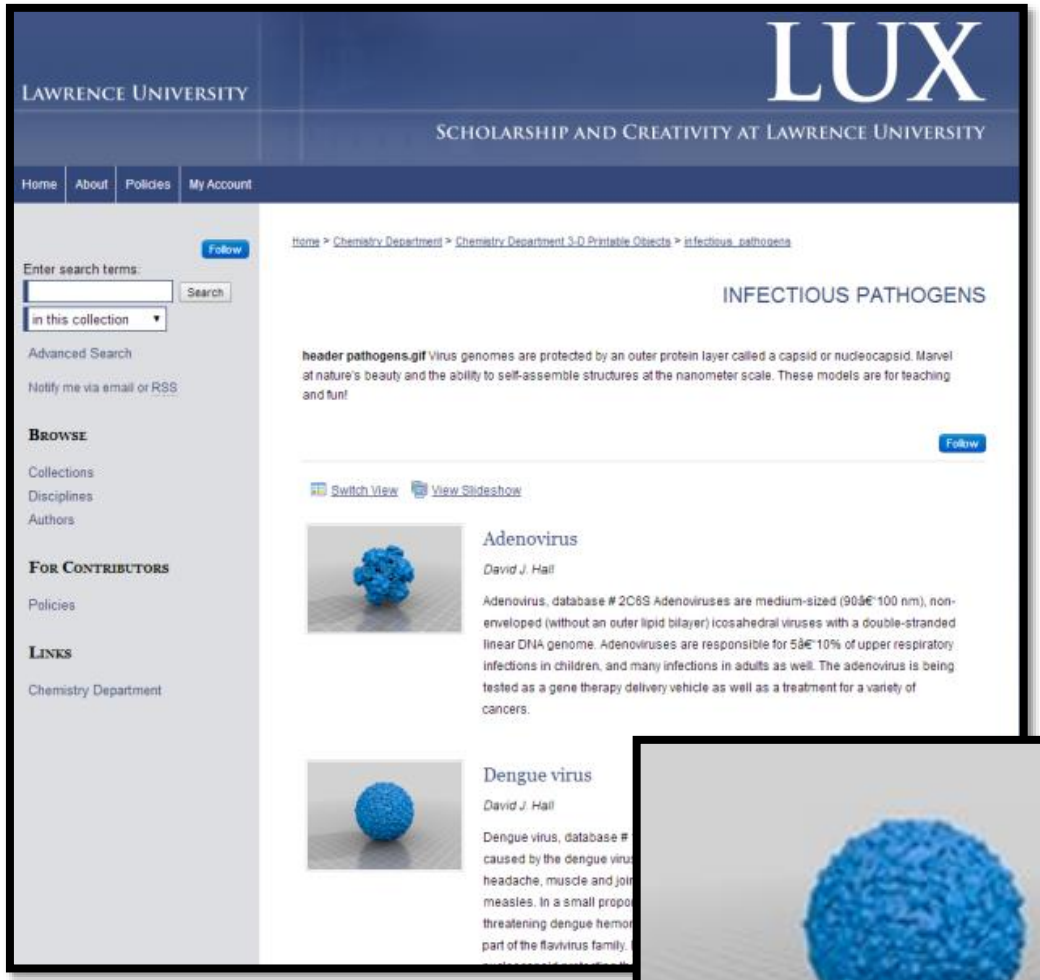
Date	Flow cfs	Temp. °C	pH	D.O. ppm	D.O. lbs/day	B.O.D.5 ppm	B.O.D.5 lbs/day
May							
6		9.5	6.8	10.9		2.5	
13		10.6	6.7	10.5		4.5	
20		12.0	6.4	10.2		3.7	
27		11.5	6.6	10.2		3.5	
May 31		17.0	6.9	8.5	213924	5.7	143455
June 1		17.1	6.8	7.9	190889	2.0	48326
2		17.1	6.9	8.1	189580	2.3	53830
3		17.2	6.7	7.6	165443	3.1	67483
4		17.9	6.8	7.6	156622	2.0	41216
5		17.8	6.7	7.5	152243	2.2	44658
Week Aver.	4180	17.4	6.8	7.9	178120	2.9	66490
7		18.2	6.8	7.3	137049	1.8	33793
8		19.4	6.9	6.9	125566	1.9	34576
9		20.3	6.4	6.9	132585	2.0	38430
10		20.2	6.8	6.7	119348	2.4	42751
11		21.1	6.8	6.4	113694	2.3	40499
12		21.0	7.0	6.5	106412	2.3	37653
Week Aver.	3333	20.0	6.8	6.8	122280	2.1	37950
14		20.1	6.8	6.8	109904	2.6	42022
15		20.9	6.9	6.4	104427	2.9	47319
16		21.9	6.7	5.8	89896	3.4	52697

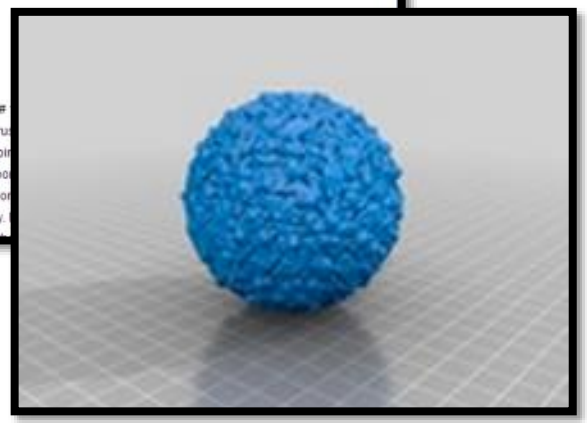


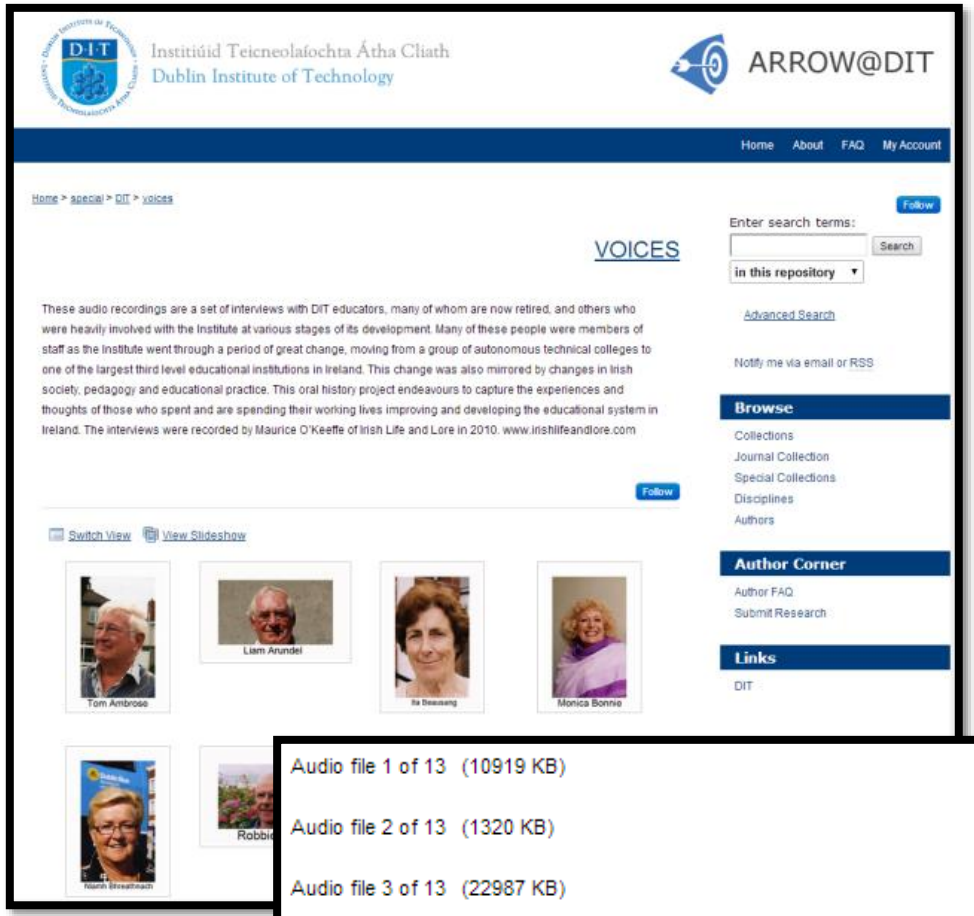


	A	B	C	D	E	F	G	H	I
1	type	ident	lat	long	y_proj	x_proj	comment	display	symbol
2	WAYPOINT	1	44.51327	-69.7944	5198151	2344029	CRTD 17:5	3	18
3	WAYPOINT	2	44.51345	-69.7934	5198193	2344099	CRTD 17:5	3	18
4	WAYPOINT	3	44.51346	-69.7934	5198195	2344098	CRTD 18:0	3	18
5	WAYPOINT	4	44.51362	-69.7931	5198219	2344116	CRTD 18:0	3	18
6	WAYPOINT	5	44.51379	-69.7928	5198246	2344136	CRTD 18:1	3	18
7	WAYPOINT	6	44.51388	-69.7923	5198269	2344177	CRTD 18:1	3	18
8	WAYPOINT	7	44.51382	-69.7919	5198271	2344205	CRTD 18:2	3	18
9	WAYPOINT	8	44.51417	-69.7914	5198321	2344235	CRTD 18:2	3	18
10	WAYPOINT	9	44.51436	-69.7911	5198351	2344256	CRTD 18:2	3	18
11	WAYPOINT	10	44.51451	-69.7909	5198372	2344267	CRTD 18:3	3	18
12	WAYPOINT	11	44.51471	-69.7906	5198400	2344278	CRTD 18:3	3	18
13	WAYPOINT	12	44.51505	-69.7901	5198450	2344308	CRTD 18:4	3	18
14	WAYPOINT	13	44.51503	-69.7898	5198454	2344331	CRTD 18:4	3	18
15	WAYPOINT	014S	44.51512	-69.7897	5198468	2344339	CRTD 18:4	3	18
16	WAYPOINT	014E	44.52003	-69.7826	5199182	2344739	CRTD 18:5	3	18

FILE TYPES REALITY: A VARIETY OF FORMATS



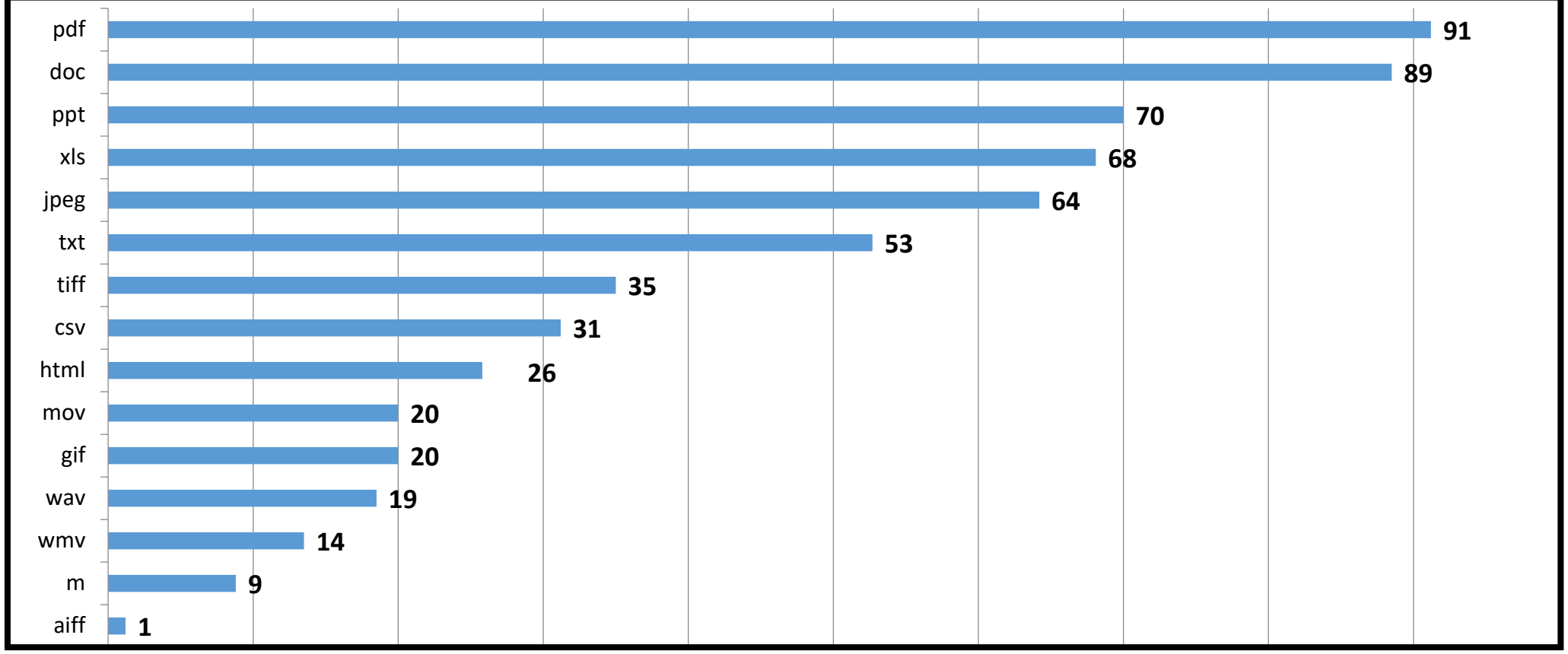




COMMON FILE EXTENSIONS

"Please select, and/or list, the file extensions you have produced, or anticipate producing, with your research."

CLEMSON UNIVERSITY



Data from: http://digitalcommons.bepress.com/data/4/

What is a data management plan (DMP)?

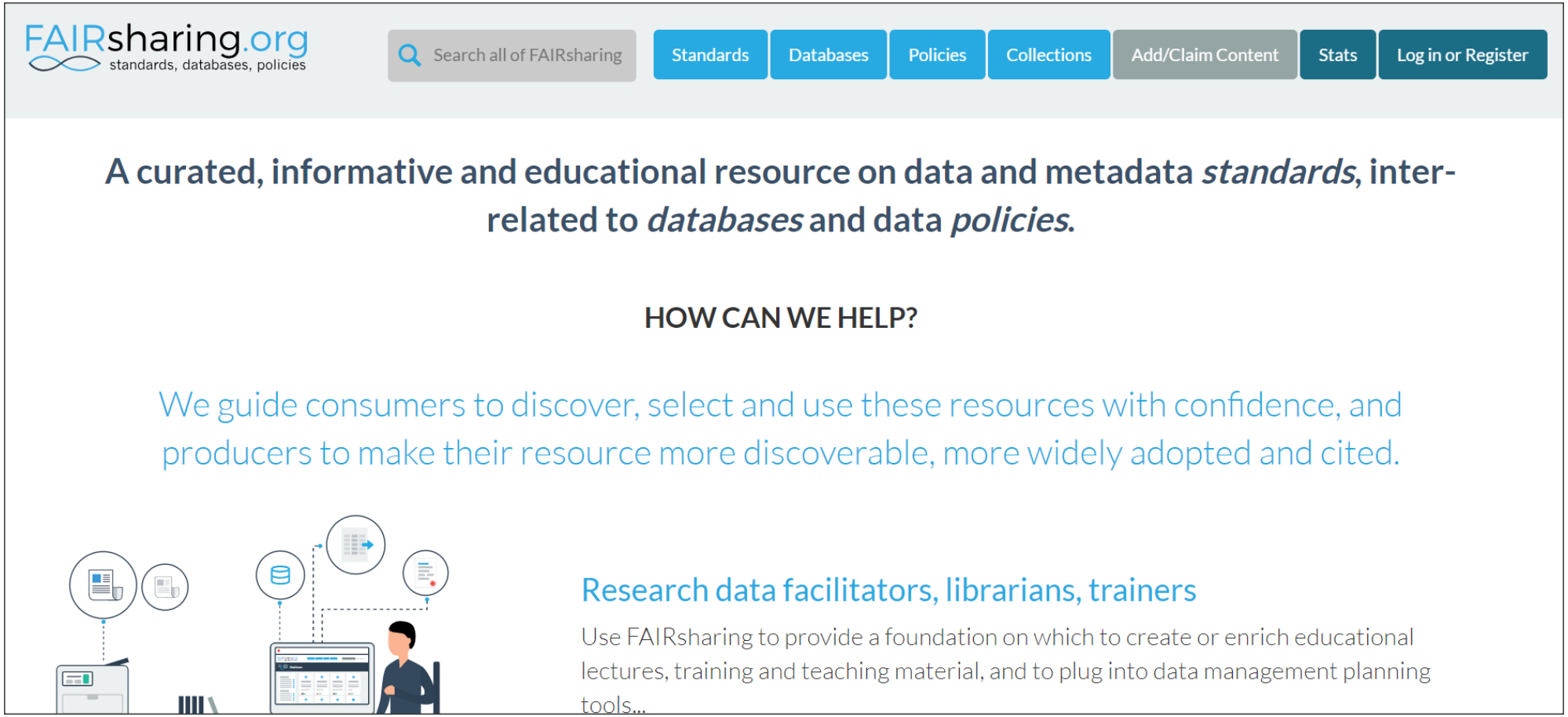
What is a DMP?

1. Description of data

What is a DMP?

2. Metadata standards

https://fairsharing.org/



What is a DMP?

3. Discussion of security and ethics

What is a DMP?

4. Plans for access, sharing, and reuse

What is a DMP?

5. Long-term management and preservation

The 3-2-1 Rule

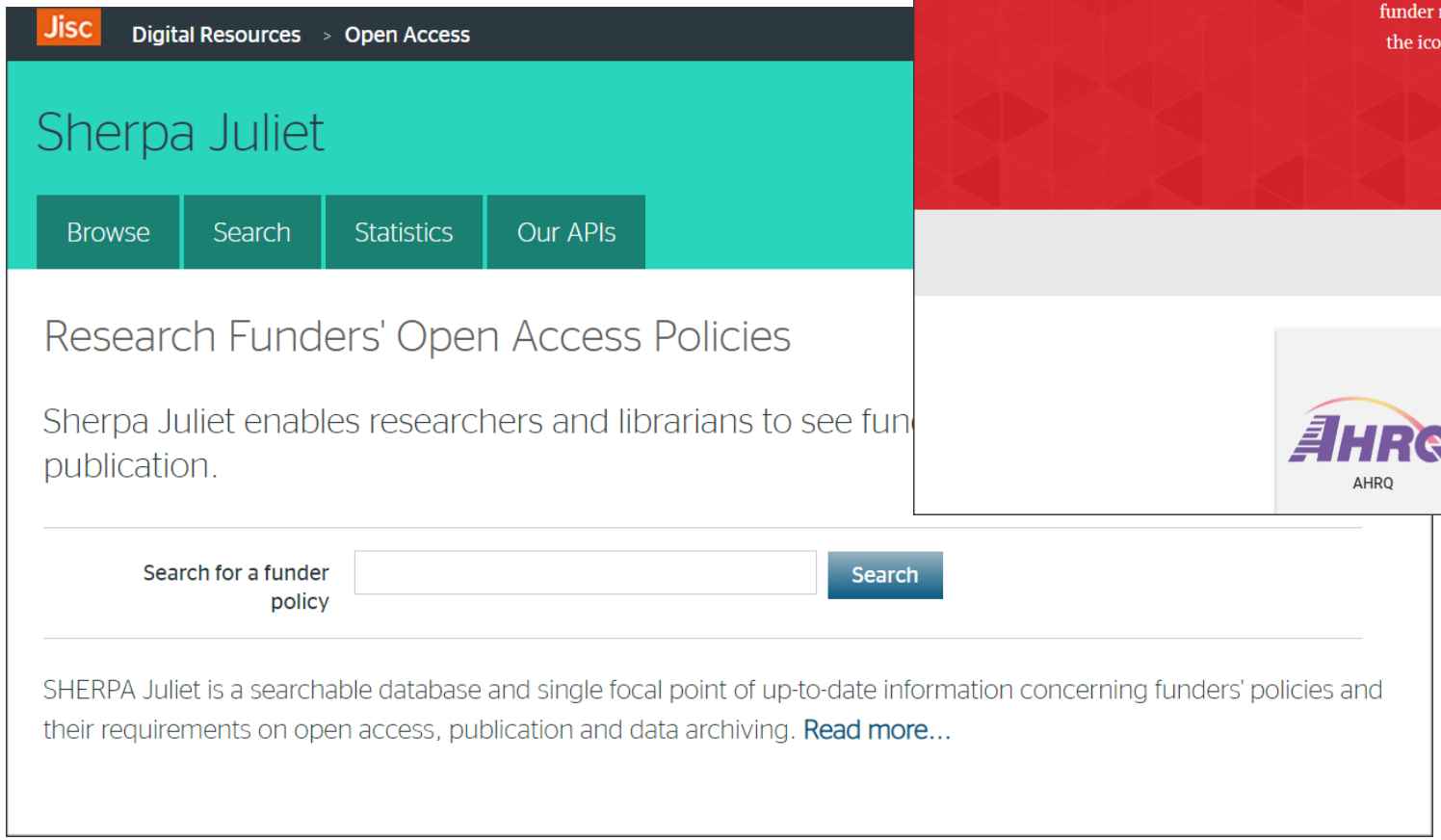
Why does it matter?

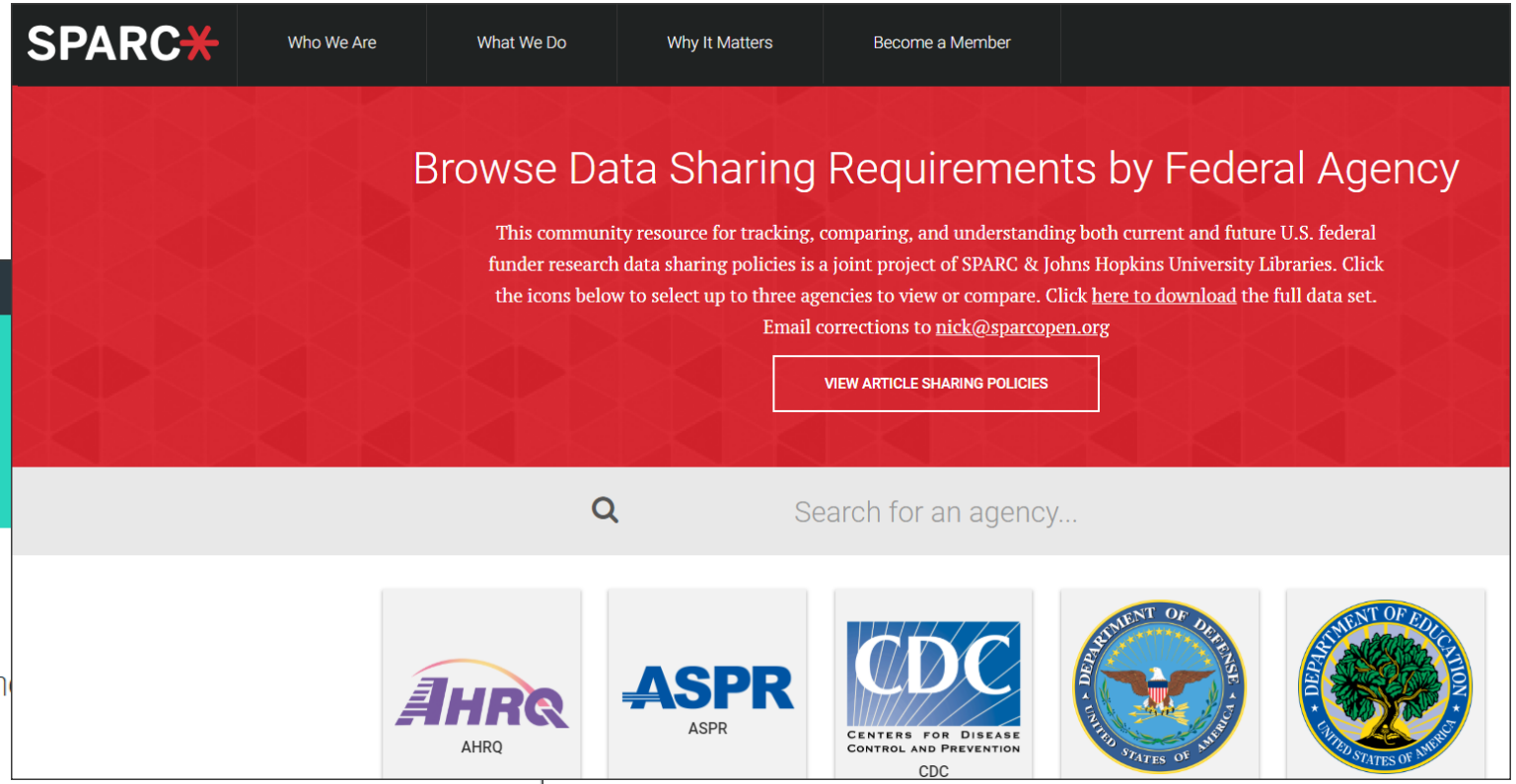
- Grant funding requirements
- Journals requiring the research data alongside the article/study
  - Increase impact and visibility
- Reproducibility
  - Maximize transparency, accountability and scrutiny of research findings
- Enable reuse and new research projects

Grant Compliance

Grant Compliance

http://v2.sherpa.ac.uk/juliet/





http://researchsharing.sparcopen.org/data

Grant Compliance

Sherpa Juliet

Browse | Search | Statistics | Our APIs

National Institutes of Health

Funder Information

Funder Name	National Institutes of Health (NIH) [English]
URL	http://www.nih.gov/ [English]
Identifiers	FundRef DOI: 10.13039/100000002
Country	United States of America

http://v2.sherpa.ac.uk/id/funder/9

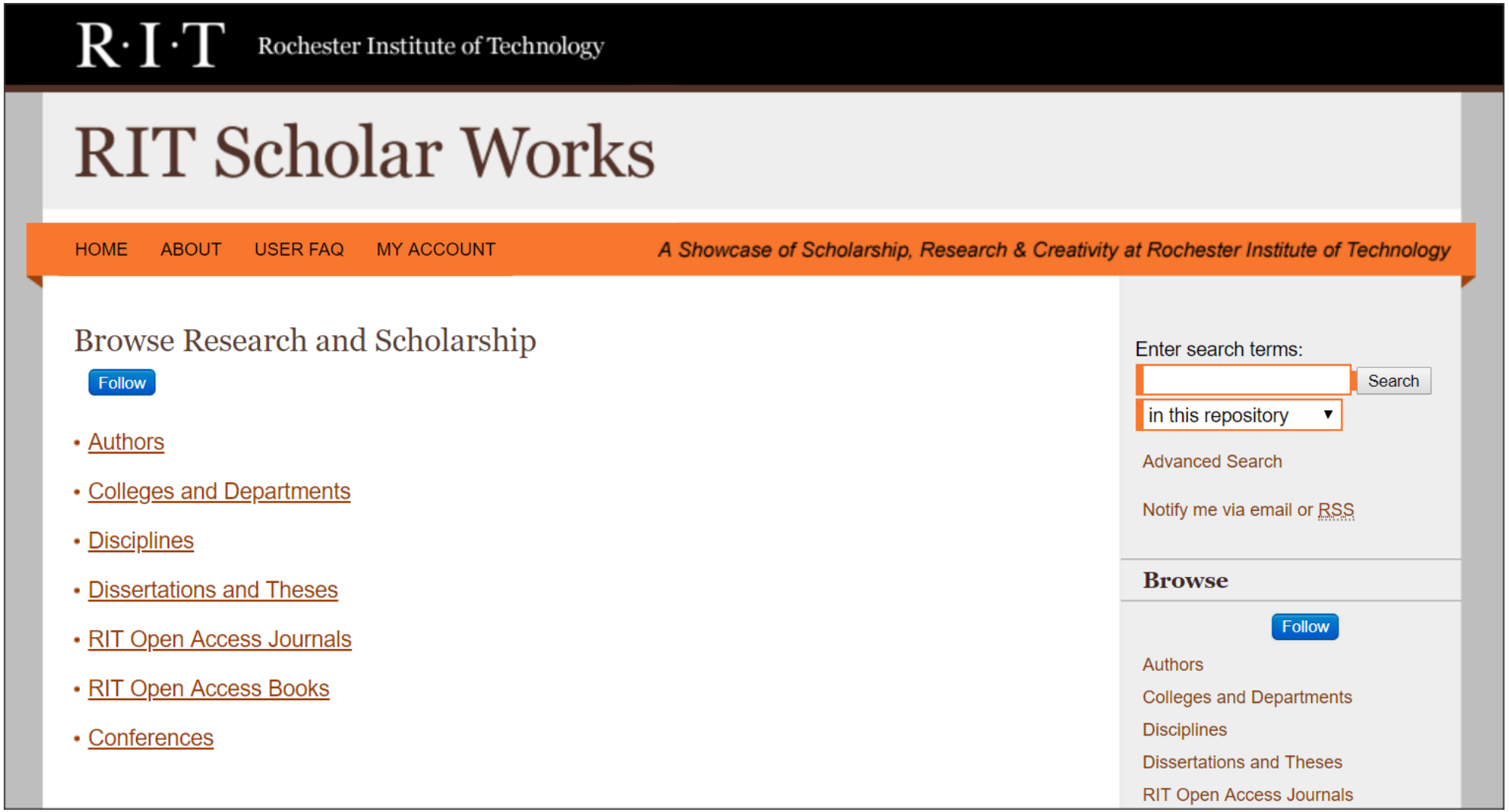
Requires Open Access Archiving

Requirement	National Institutes of Health **requires** Open Access Archiving
What to archive	Peer-reviewed publications
Publication version	Author's final version
When to archive	When accepted for publication
Permitted Embargo	12 months
Where to archive	Named repository
Named Repositories	PubMed Central http://www.ncbi.nlm.nih.gov/pmc/ (Required)

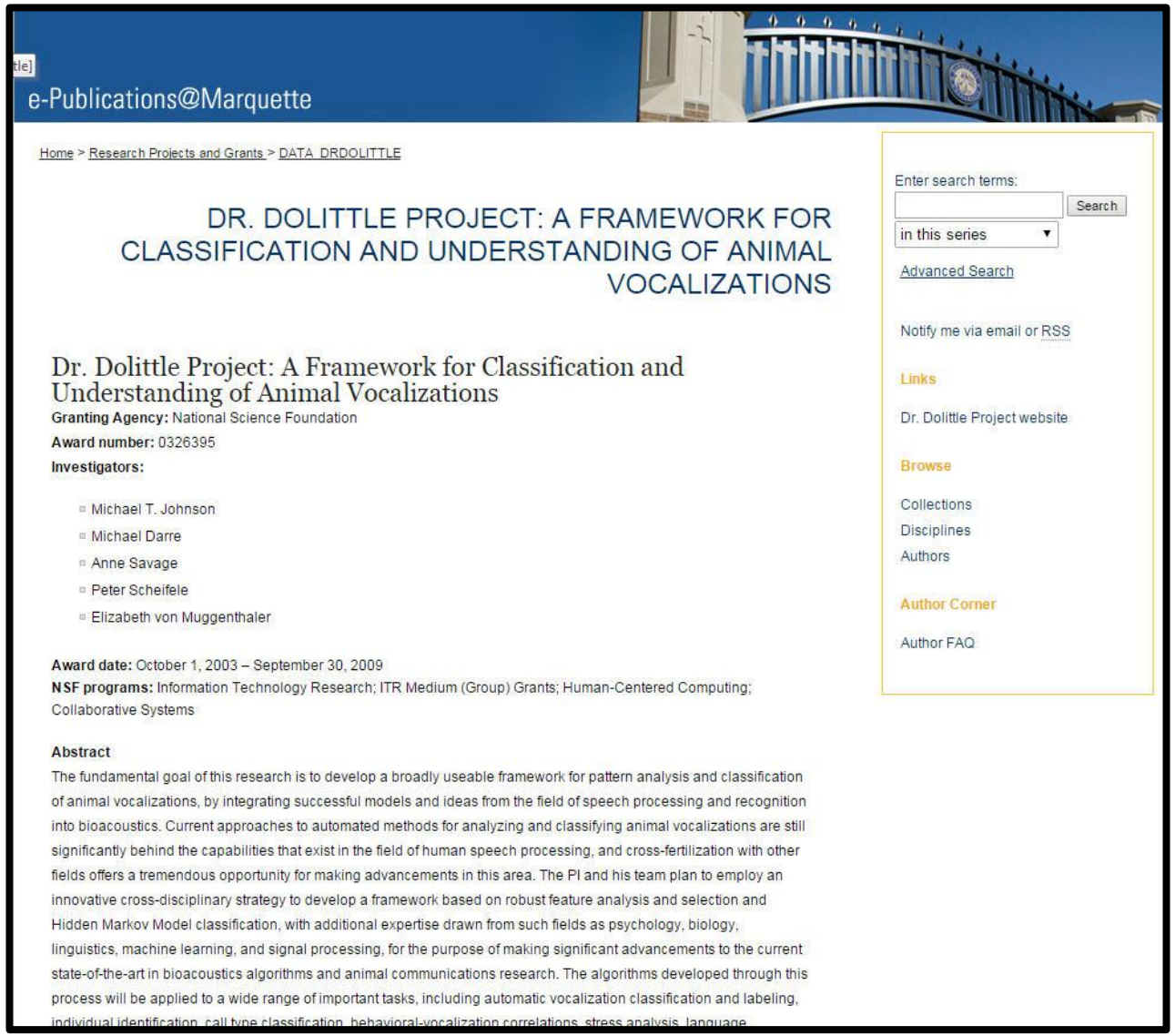
Requires Open Data Archiving

Requirement	National Institutes of Health **requires** Open Data Archiving
Types of Data	Research Data Supporting Documentation
When to archive	Within a reasonable time after completion of the work
Where to archive	Any appropriate repository (Required)
Named Repositories	NIH Databases http://www.nlm.nih.gov/databases/ (Example)
Effective for all new projects from	1 October 2003
Special Conditions	Applies to applications seeking $500,000 plus in direct costs within any year of the project period, Include a data sharing plan within application, Sharing of data no later than publication of final results
Policy links	NIH Data Sharing Policy [Policy]: http://grants.nih.gov/grants/policy/data_sharing/ Implementation Guidance [Guide/FAQ]: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Grant Compliance



GRANT-RELATED DATA



tle]

e-Publications@Marquette

Home > Research Projects and Grants > DATA_DRDOLITTLE

DR. DOLITTLE PROJECT: A FRAMEWORK FOR CLASSIFICATION AND UNDERSTANDING OF ANIMAL VOCALIZATIONS

Dr. Dolittle Project: A Framework for Classification and Understanding of Animal Vocalizations

Granting Agency: National Science Foundation

Award number: 0326395

Investigators:

- Michael T. Johnson
- Michael Darre
- Anne Savage
- Peter Scheifele
- Elizabeth von Muggenthaler

Award date: October 1, 2003 – September 30, 2009

NSF programs: Information Technology Research; ITR Medium (Group) Grants; Human-Centered Computing; Collaborative Systems

Abstract

The fundamental goal of this research is to develop a broadly useable framework for pattern analysis and classification of animal vocalizations, by integrating successful models and ideas from the field of speech processing and recognition into bioacoustics. Current approaches to automated methods for analyzing and classifying animal vocalizations are still significantly behind the capabilities that exist in the field of human speech processing, and cross-fertilization with other fields offers a tremendous opportunity for making advancements in this area. The PI and his team plan to employ an innovative cross-disciplinary strategy to develop a framework based on robust feature analysis and selection and Hidden Markov Model classification, with additional expertise drawn from such fields as psychology, biology, linguistics, machine learning, and signal processing, for the purpose of making significant advancements to the current state-of-the-art in bioacoustics algorithms and animal communications research. The algorithms developed through this process will be applied to a wide range of important tasks, including automatic vocalization classification and labeling, individual identification, call type classification, behavioral-vocalization correlations, stress analysis, language

Enter search terms:

Search

in this series

Advanced Search

Notify me via email or RSS

Links

Dr. Dolittle Project website

Browse

Collections

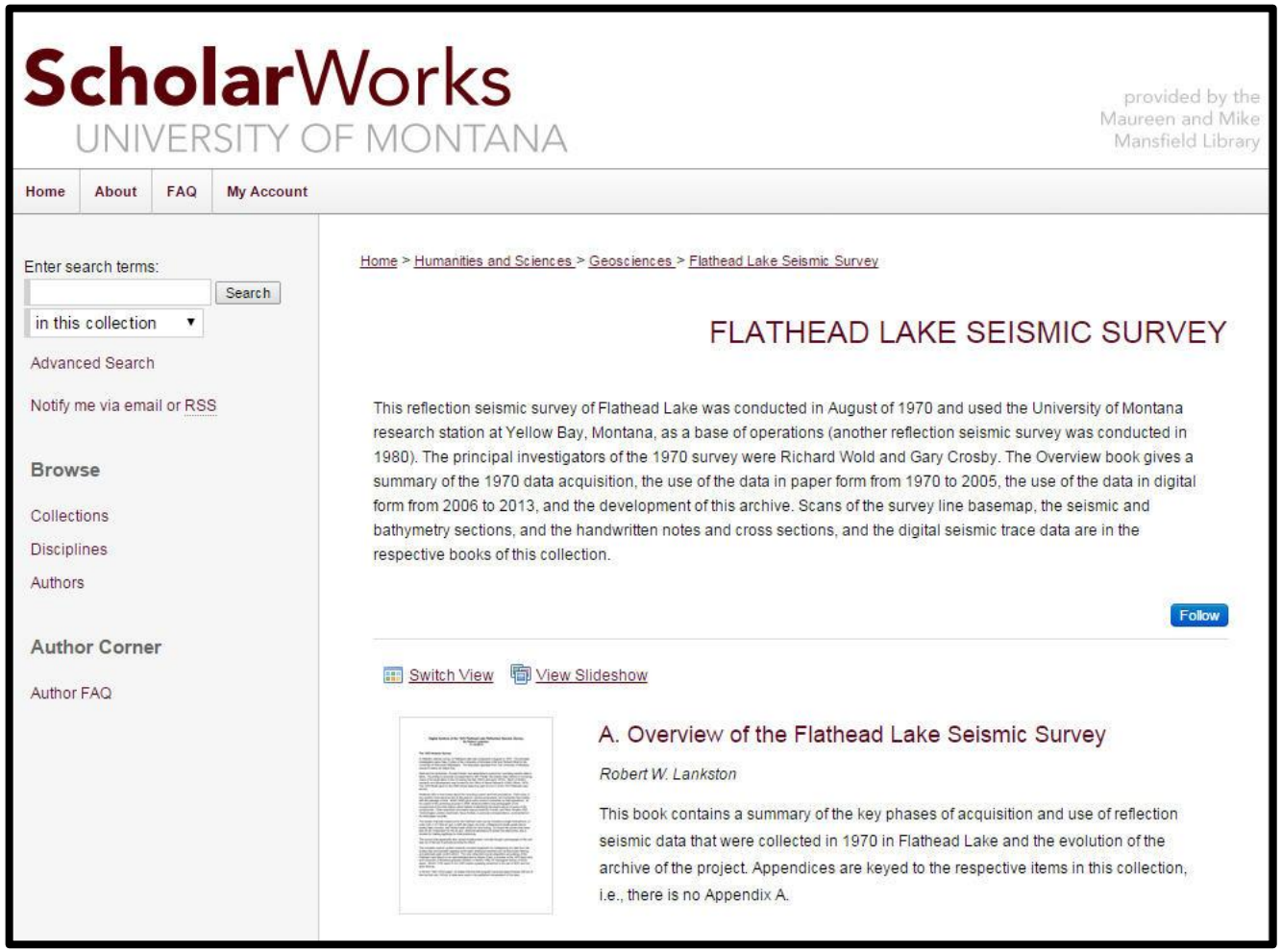
Disciplines

Authors

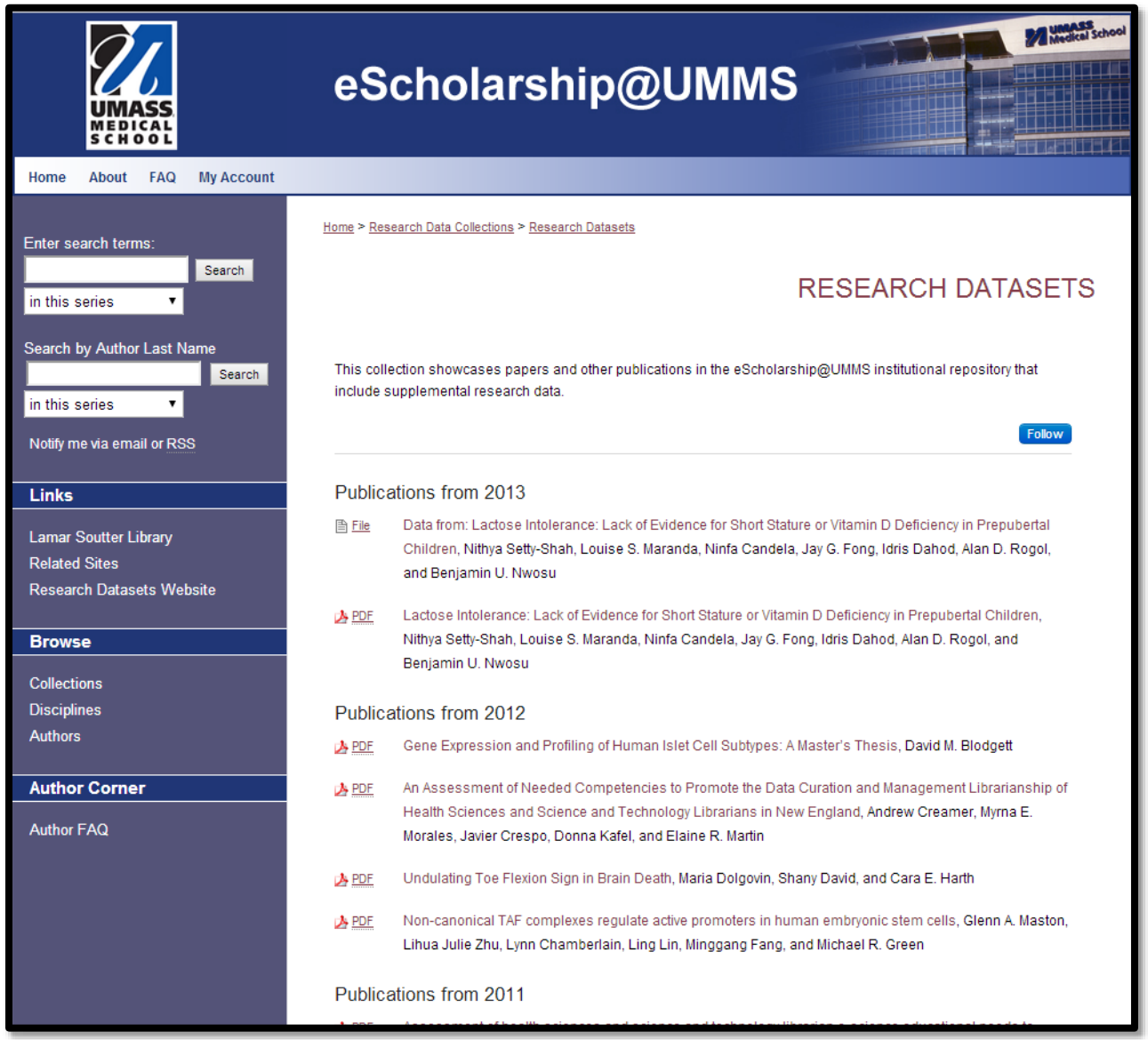
Author Corner

Author FAQ

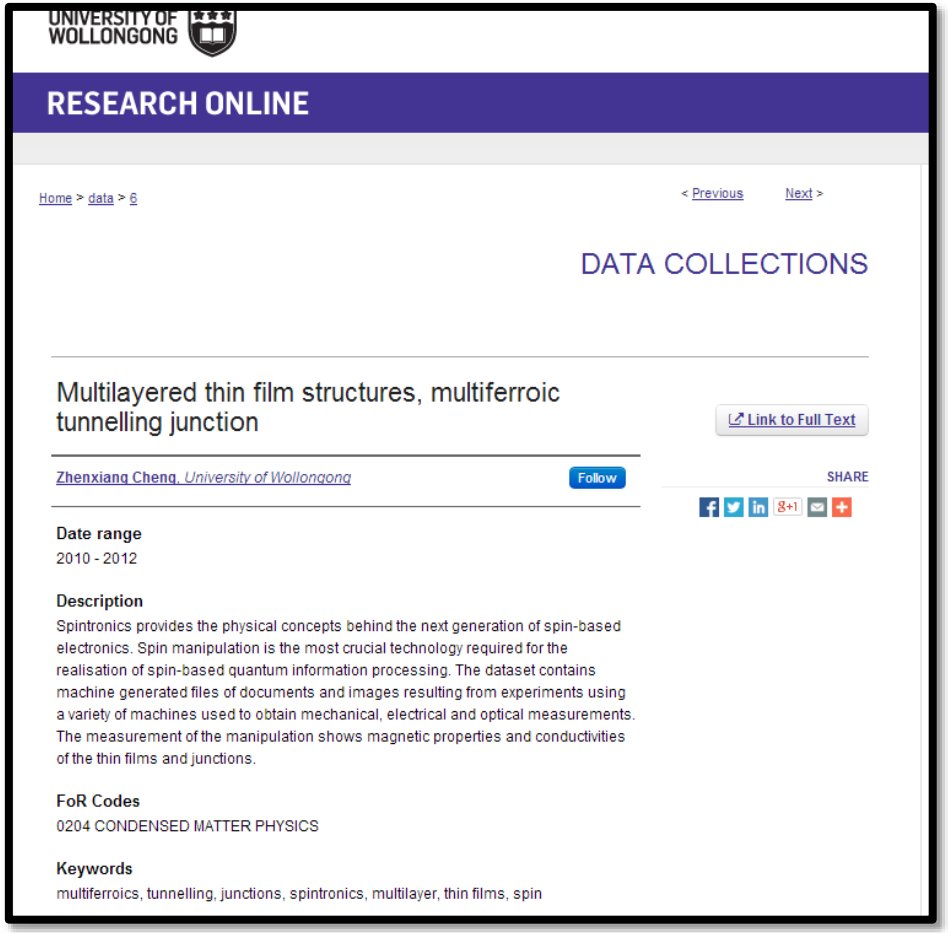
PROJECT-SPECIFIC DATA



ARTICLES WITH DATA UPLOADED AS “ADDITIONAL FILES”

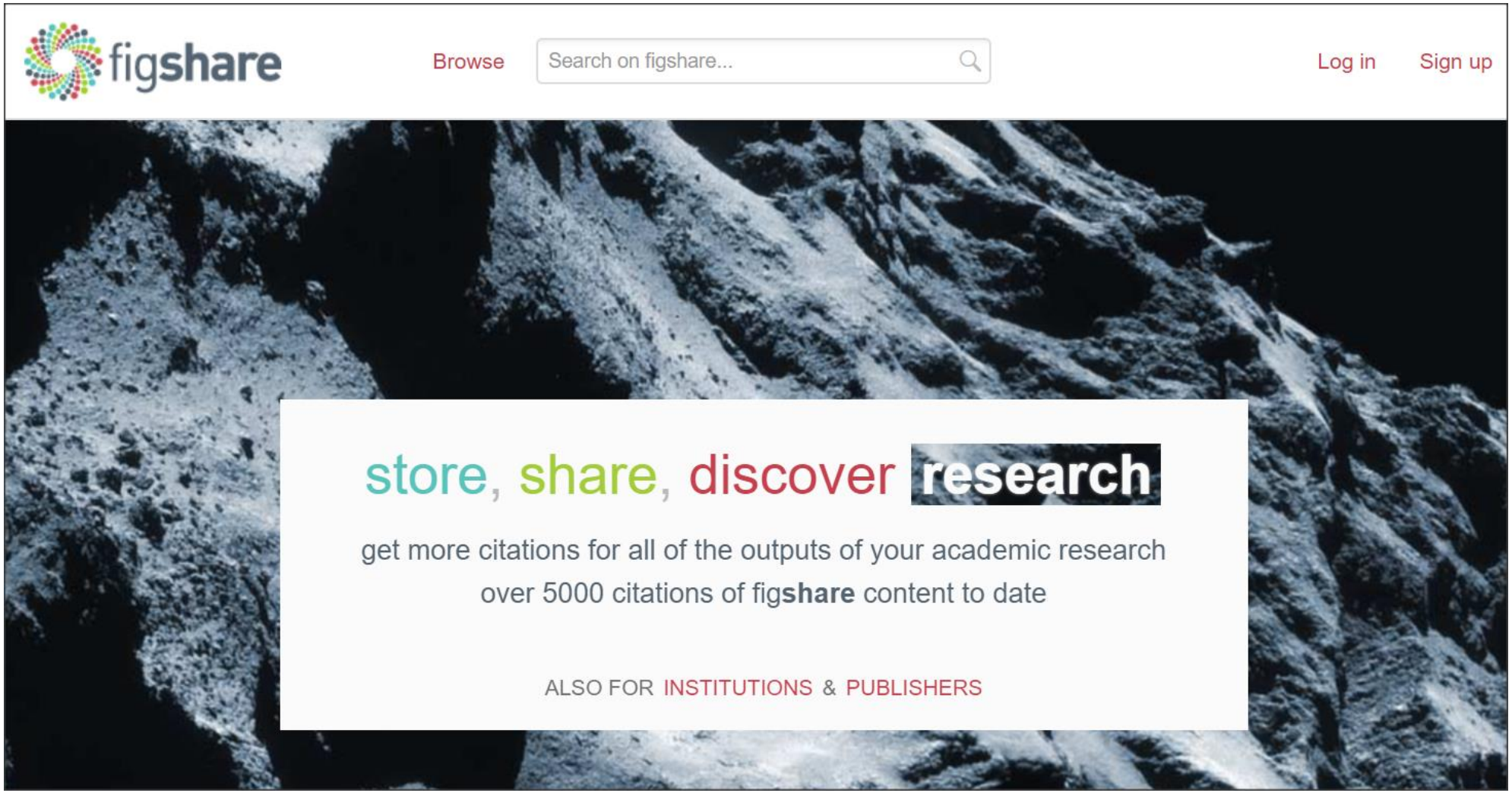


LINKING OUT TO THE DATASETS



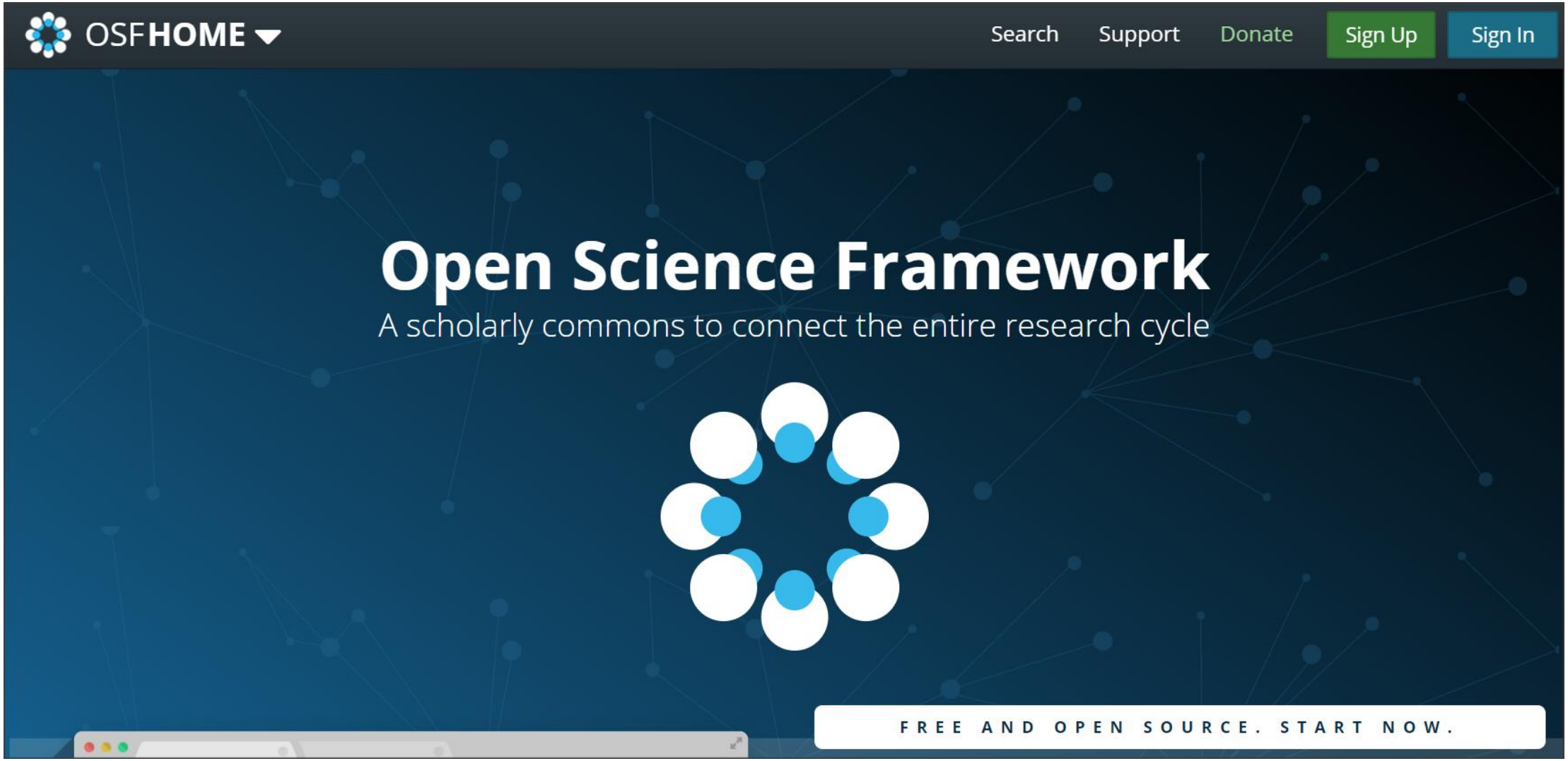
Grant Compliance

https://figshare.com/



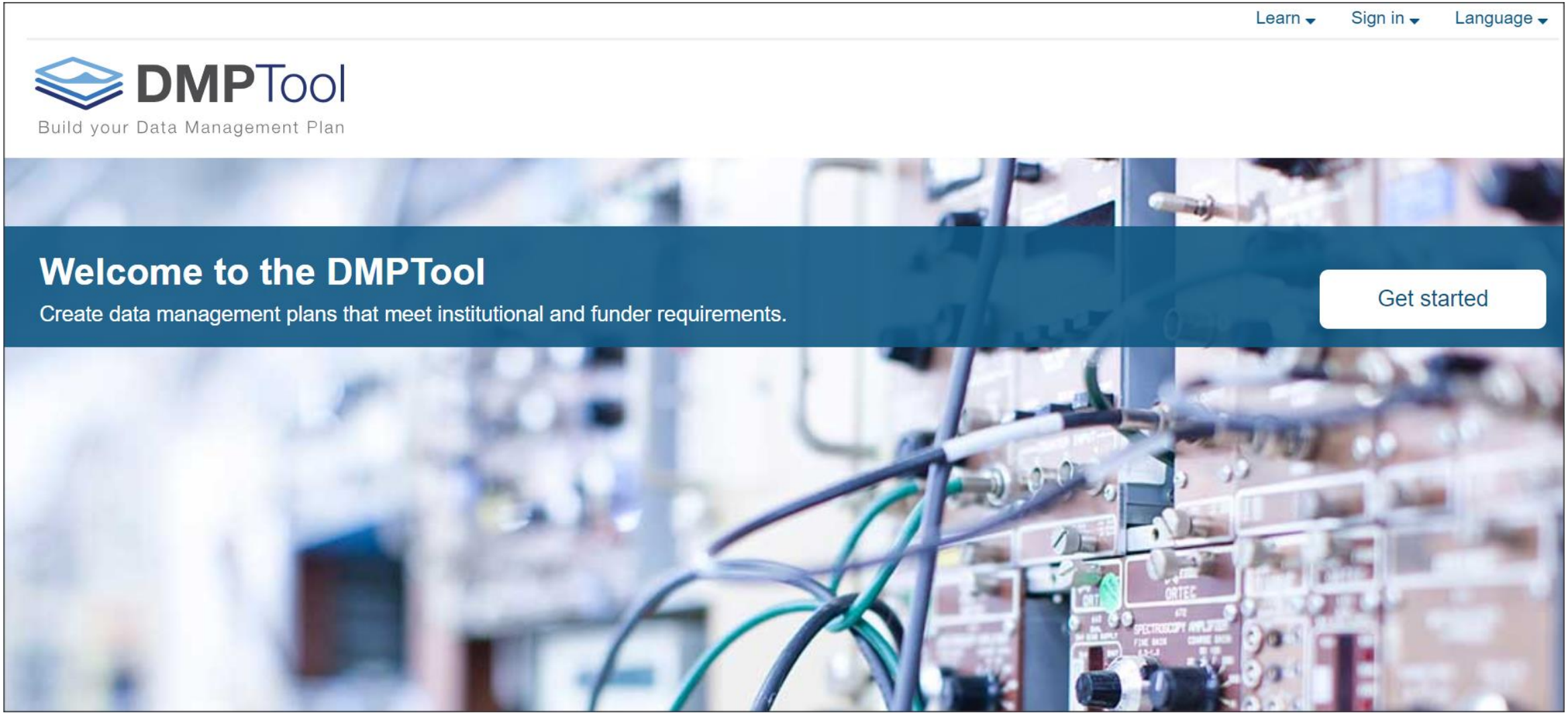
Grant Compliance

https://osf.io/



DMP Resources

https://dmptool.org/



DMP Resources

https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

DMP Resources

https://dataoneorg.github.io/Education/

DMP Resources

- Purdue Data Management Plan Self-Assessment Questionnaire
  - A document with detailed questions to guide investigators through the data management planning process.
- MIT Data Management Planning Guide
  - MIT provides a list of questions to consider when formulating a data management plan.
- Example of NIH Data Sharing Plans (1)
  - Three short examples of data-sharing plans from the NIH website.
- Example of NIH Data Sharing Plans (2)
  - Downloads a more extensive sample data sharing plan from NIH.

Questions?