1-9-2016

Where's the Data?

Andrea Payant
Utah State University

Betty Rozum
Utah State University

Liz Woolcott
Utah State University

Where's the Data?

Andrea Payant
andrea.payant@usu.edu
@rusros25

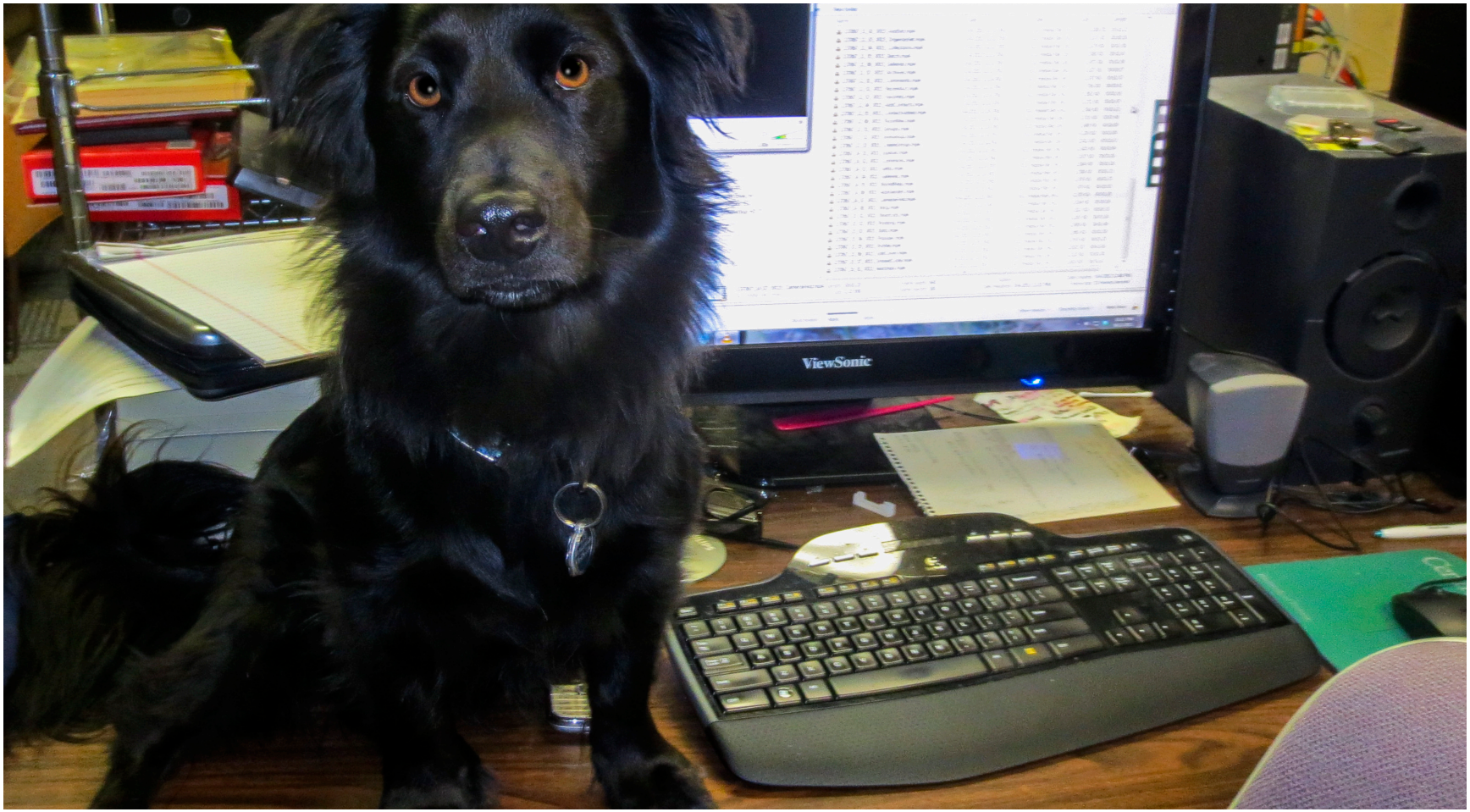
Betty Rozum
betty.rozum@usu.edu
@brozum

Liz Woolcott
liz.woolcott@usu.edu
@lizwoolcott

UtahStateUniversity

Presented to the American Library Association, ALCTS Cataloging Norms Interest Group
January 9, 2016

The Problem



Lots of Data

- Researchers generate Lots of Data
- Not all researchers organize and manage their data
- Faculty *may* be required to deposit data to make it publicly accessible
- University is ultimately responsible for complying with terms of grants

Can you find where your institution's data is hiding?

Why Bother?

- Federal Policy
  - *digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.*
    - OSTP Memorandum, February 22, 2013
- Will we be able to locate this data after the grant is completed? What will happen in an audit?
- Where do your faculty store data for NIH and NSF, which currently require Data Management Plans?

Lead People to Your Data

Data Management Plans

- Short documents that accompany grant applications
  - Type of data collected (format, how much)
  - How will it be described?
  - Will it be shared? When? How?
  - Where will it be made publicly accessible? Preserved?
  - IP, copyright, sensitive data, any other issues?
- *To make data publicly accessible to search retrieve and analyze*

Indicate How It Can Be Shared

Bone 2 photo by jb CC-by-2.0

DMPs from Successful Grants

- Some are Stellar! Morgan Ernest, now at University of Florida:
  - Data access will take place primarily through a project website hosted by the White Lab at USU as well as through the Sevilleta LTER website (http://sev.lternet.edu) . . . . In addition, all datasets will be discoverable and downloadable via the LTER NIS data portal, as well as through the Sevilleta LTER and White Lab websites. . . . Sequence data . . . will be deposited in the publicly available NCBI Genbank database. Simulation data will be archived in Dryad (http://datadryad.org/) . . .

DMPs from Successful Grants

- Some have a pretty solid idea of sharing and access
  - USU's IR, Digital Commons identified as storage and access point
    - A few problems with metadata, file formats, etc.

DMPs from Successful Grants

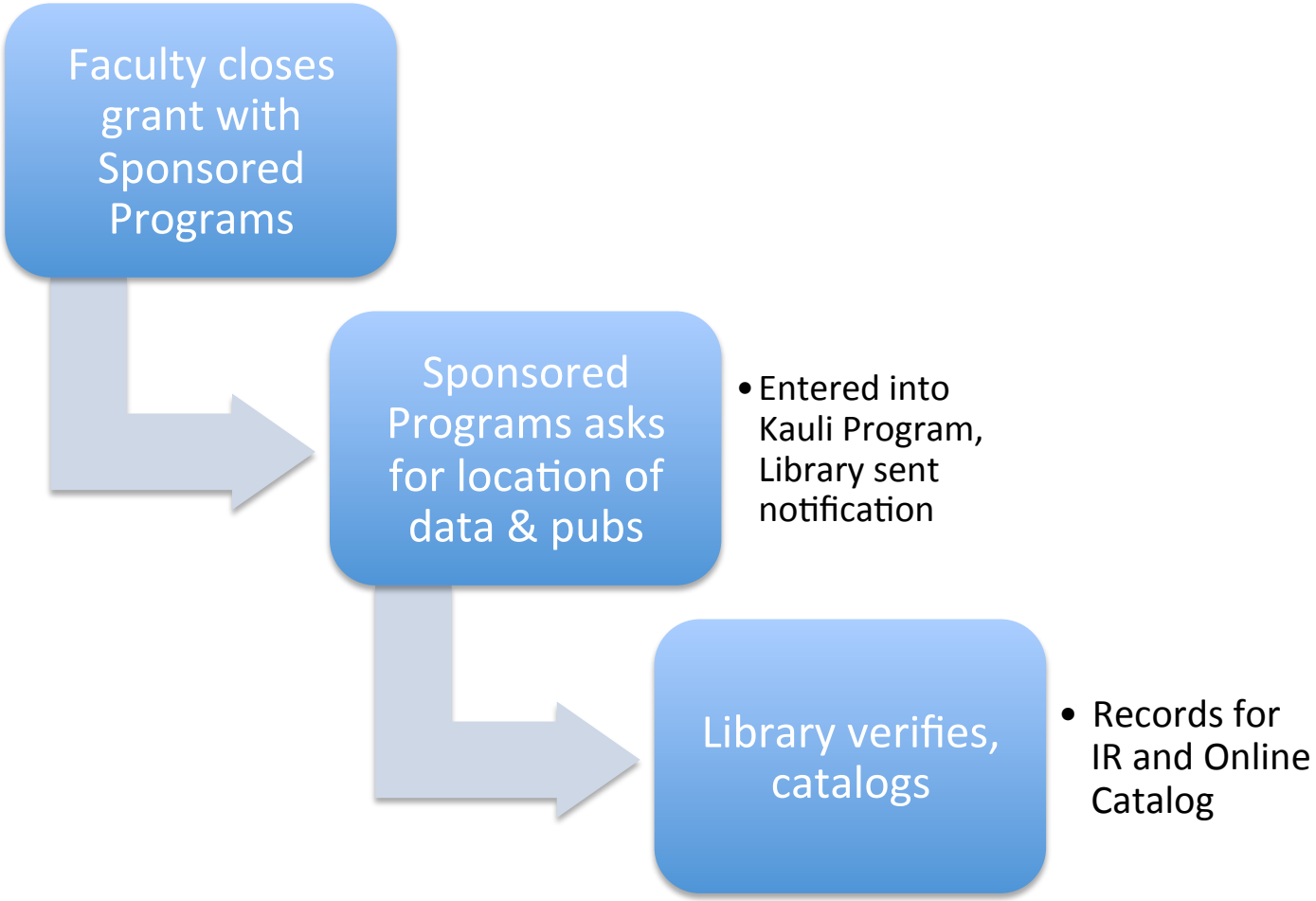
- Some share, but don't currently deposit
  - "The PI of this project will provide access to the data and information . . . to other scientists who are interested…This information will be available through a web portal of USU."
  - Share data upon request and through publication

It’s Here!

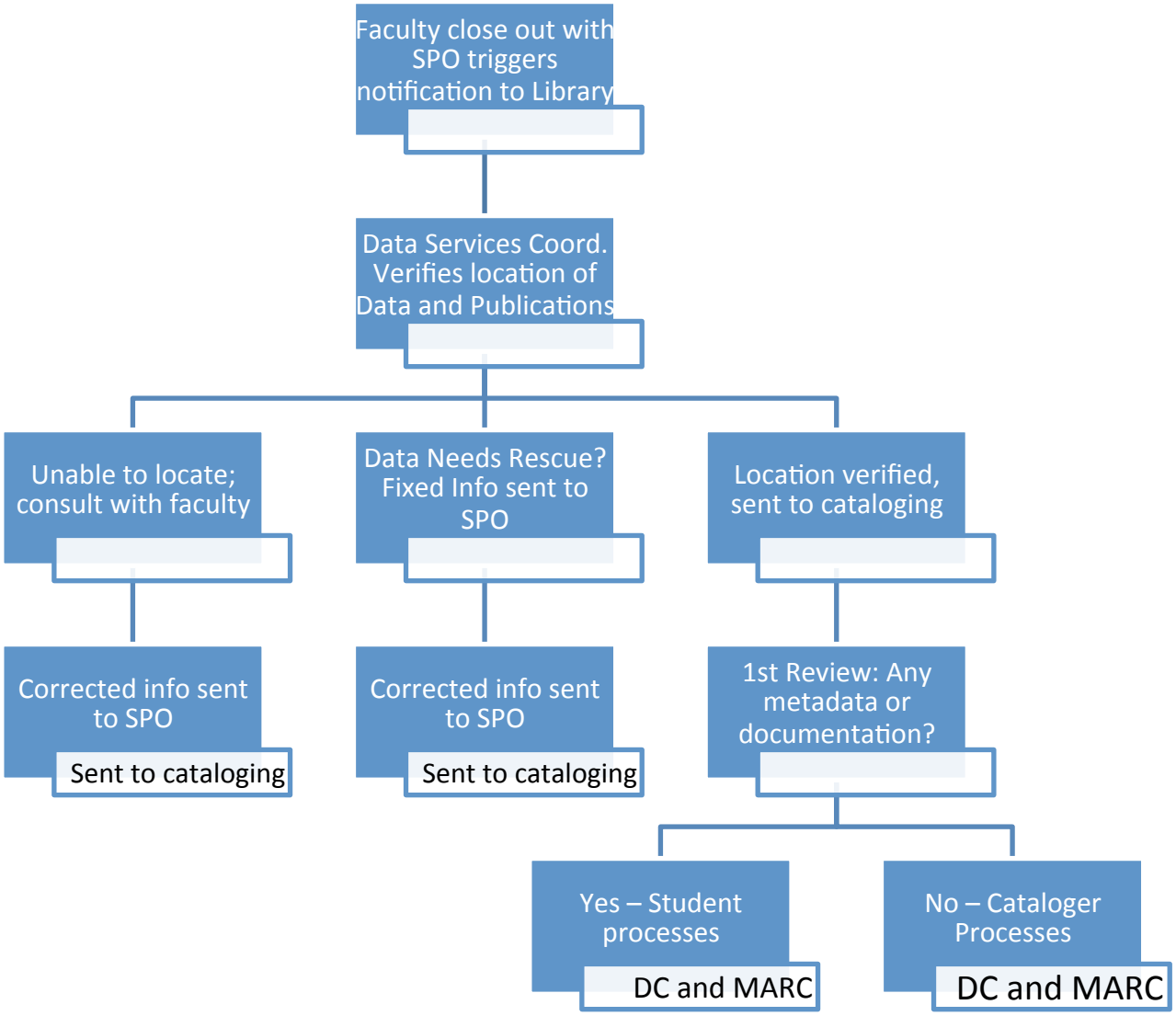
How Can Libraries Help?

- We know how to Catalog and make things discoverable!

The Project (so simple)



The Project (A few More Details)



Cataloging Data

OCLC NEW

Save File 1

Computer Files | Rec stat n | Entered 20160104 | Replaced 20160104

Type	m	ELvl	K	Srce	d	Audn		Ctrl		Lang	eng
BLvl	m	Form	o	GPub				MRec		Ctry	utu
Desc	i	File	u			DtSt	s	Dates	year	,	

007			c ǂb z ǂh m
040			UUS ǂb eng ǂe rda ǂc UUS
090			ǂb
049			UUSA
100	1		[Name of Researcher], ǂe [relator term].
245	1	0	[Name/Title assigned to data] / ǂc [Researcher(s)/Authors of the data].
264		1	[Place where data originated] : ǂb [Institution name], ǂc [year of data publication]
300			1 dataset.
336			computer dataset ǂ2 rdacontent
337			computer ǂ2 rdamedia
338			online resource ǂ2 rdacarrier
347			[Digital file characteristics = [file type] ǂb [encoding format] ǂc [file size]
500			[Granting Agency/Grant Award # Information]
500			[Any additional information pertinent to the data/dataset that is not otherwise related in the MARC record - use readme file, if available, as a reference]
500			[Add Citation/Reference Information for associated article]
520			[Summary/Abstract]
538			[System details note = include information about the characteristics of computer files for example: mode of access, software programming language, and computer requirements - use readme file, if available, as a reference]
590			USU Data Deposits.
590			[University Department]
650		0	[Subject of research data]
700	1		[Name of Additional Researcher], ǂe [relator term].
856	4	0	Dataset deposited here: ǂu [insert url for dataset]
856	4	0	Dataset DOI: [insert doi for dataset]
856	4	2	Link to the associated journal article here: ǂu [insert link for associated article]
856	4	2	Article DOI: ǂu [insert doi for associated article - must include "http:" or the record will not validate]

Required

Recommended

Cataloging Data

OCLC 933762486 Held by UUS - no other holdings

Computer Files | Rec stat n | Entered 20160106 | Replaced 20160106135207.3

Type	m	ELvl	K	Srce	d	Audn		Ctrl		Lang	eng
BLvl	m	Form	o	GPub	s			MRec		Ctry	utu
Desc	i	File	u			DtSt	s	Dates	2015	,	

007			c ǂb z ǂh m
040			UUS ǂb eng ǂe rda ǂc UUS
090			ǂb
049			UUSA
100	1		David, Michael, ǂe researcher.
245	1	0	How uncertainty in the neutral wind limits the accuracy of ionospheric modeling and forecasting.
264		1	Logan, Utah : ǂb Utah State University, ǂc 2015.
300			1 dataset.
336			computer dataset ǂ2 rdacontent
337			computer ǂ2 rdamedia
338			online resource ǂ2 rdacarrier
347			data file ǂb zip ǂc 133172 KB
347			data file ǂb zip ǂc 154616 KB
347			data file ǂb zip ǂc 136181 KB
347			data file ǂb zip ǂc 135839 KB
347			data file ǂb txt ǂc 3.6 KB
347			data file ǂb txt ǂc 3.2 KB
500			National Science Foundation Grant AGS-0962544.
500			"TDIM output files are zipped. There are about 700 files in all."
520			"The output consists of altitude profiles of ion densities, neutral densities (from MSIS), and ion, electron, and neutral temperatures. All densities are in units of cm-3; temperatures are degrees Kelvin. In addition to the altitude profiles, there are a number of other items that are returned via common blocks." --explanation-of-lastpage-files.txt
538			"The .lastpage and .eachpage files contain the output from runs of the Utah State University TDIM ionospheric model. They are ascii files, and can be read with the subroutine READ_NEXT_STEP_EACHPAGE(IUNIT,ALTS,PROFS,ISTATUS)." --explanation-of-lastpage-files.txt
590			USU Data Deposits.
650		0	Ionosphere.
650		0	Ionospheric forecasting.
700	1		Sojka, Jan, ǂe researcher.
700	1		Schunk, R. W. ǂq (Robert W.), ǂe researcher.
856	4	0	Dataset deposited here: ǂu http://digitalcommons.usu.edu/all_datasets/9/
856	4	2	Link to the associated journal article here: ǂu http://onlinelibrary.wiley.com/doi/10.1002/2015JA021544/full
856	4	2	Article DOI ǂu http://doi:10.1002/2015JA021544

Throw Us a Bone!

MARC/Dublin Core Mappings

MARC Mapping	DC Mapping	Field Description
100	Creator	1st Author/Researcher listed
245 $a	Title	Title/Name assigned to data set
245 $c		All authors/researchers listed
264 $a		Place where data originated
264 $b		Primary institution name
264 $c	Date	Year of publication/deposit
347 $a		Digital characteristics - file type, refer to the file extension
347 $b		Digital file characteristics - encoding format
347$c	Format.Extent	Digital file characteristics - file size
500	Description	Granting Agency, grant award number
500	Description	Any additional information pertinent to the data/dataset that is not otherwise reflected in the record. Refer to the readme file, if available
500	Relation.IsReferencedBy	Citation for original publication based on this data set
538	Relation.Requires	Include information about the characteristic of the files, noting mode of access, software or computer access. Refer to the readme file, if available.
520	Description.Abstract	Include any summary information about the content of the dataset, such as an abstract.
650	Subject	Subject of research data
700	Creator	Name(s) of additional researcher
856	Description	URL for location of dataset
856	Identifier	Dataset DOI
856	Relation.IsReferencedBy	Link to associated Journal Article
856		Associated Journal Article DOI
	Type	Indicate the DCMI type (typically "Dataset")
	Format	Indicate the file format of the dataset, refer to the MIME types
	Publisher	Indicate where the data set is housed
	Coverage.Spatial	If reported, include the spatial coverage for the dataset
	Coverage.Temporal	If reported, include the date coverage for the dataset

Sometimes Data Doesn't Behave According to the Cataloging Rules

Problems encountered

- Incomplete information
  - Data sets are not always named as a group (even when individual files may be)
  - Lack of DOIs
  - Unclear about authors/contributors and pertinent roles
  - Unspecified/uncommon data formats
  - Lack of readme files
- Locating specific data sets on funding agencies websites/data depositories can be tricky

Rewards (Benefits)

Benefits

- Capture the location of data while it's (relatively) fresh in the mind of researchers
- Create permanent records of data
- Increases discoverability of data
- Opportunity to "rescue" data insecurely stored
- Compliance is verified

Benefits - Partnerships

Benefits (Rewards)

- Library increases is value on campus, strengthens partnership with Research Office
- Library increases interactions with faculty and demonstrates value by securing data and helping faculty comply with DMP
- Library gains opportunity to help faculty learn about better options for data deposit
- Reporting functions allow for University to understand where data is housed
- Cross campus synergistic relationships

What could go wrong?

It's a pilot project for a reason…

- Data deposited with GenBank or NCAR and no other clues provided
- Data deposited "on PI website"
- Data has not been properly de-identified
- Data is embargoed
- Researcher does not remember
- Time to identify location of data is a huge concern - sustainability
- Faculty reluctant to release data
- Actually cataloging data – many formats

Questions?

Throw Us a Bone!

Contacts

Andrea Payant
Data Management Metadata Specialist
andrea.payant@usu.edu
@rusros25

Betty Rozum
Data Services Coordinator and Undergraduate Research Librarian
betty.rozum@usu.edu
@brozum

Liz Woolcott
Head of Cataloging and Metadata Services
liz.woolcott@usu.edu
@lizwoolcott

UtahStateUniversity