

eScholarship@UMassChan

Purdue University Research Repository

Item Type	Presentation
Authors	Witt, Michael
DOI	10.13028/a6c1-zc46
Rights	Copyright the Author(s)
Download date	2026-04-10 15:54:56
Item License	http://creativecommons.org/licenses/by-nc-sa/4.0/
Link to Item	https://hdl.handle.net/20.500.14038/28660

2015 E-SCIENCE SYMPOSIUM

APRIL 9, 2015 – UNIVERSITY OF MASSACHUSETTS

Purdue University Research Repository



Michael Witt

Head, Distributed Data Curation Center

Associate Professor of Library Science

<http://www.lib.purdue.edu/research/witt>

E-mail: mwitt@purdue.edu

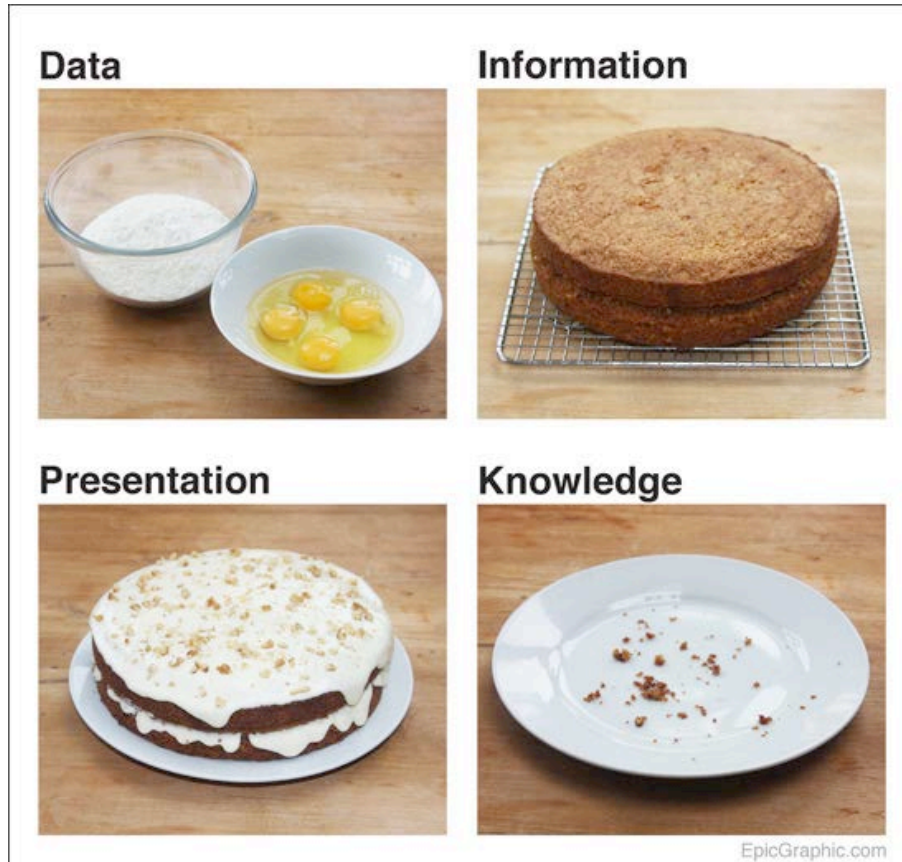
PURDUE
UNIVERSITY
LIBRARIES



D2C2

DISTRIBUTED DATA
CURATION CENTER

DATA = EVIDENCE



<http://epicgraphic.com/data-cake>

SOME EXAMPLES OF DATA

Tables and spreadsheets

Sensor and instrument output

Software source code and simulations

Images, video, and audio

Observation logs

Workflows

Interview transcripts

Survey instruments and results

Etc.

CAMPUS COLLABORATION

Purdue University Research Repository (PURR)

The PURR service is a collaborative effort of the *Purdue University Libraries, Executive Vice President for Research and Partnerships, and Information Technology at Purdue*. PURR is a designated university core research facility.

Designated community:

Purdue University faculty, staff, and graduate student researchers; their collaborators; and the current and future consumers of their data.

LIBRARY STRATEGIC PLAN

Data is written into the three pillars of our strategic plan:

- Learning: “...*information literacy defined broadly to include digital information literacy, science literacy, data literacy, health literacy, etc...*”
- Scholarly Communication: “Lead in data-related scholarship and initiatives”
- Global Challenges: “We will lead in international initiatives in information literacy and e-science and ... contribute to international information literacy, learning spaces, data management, and scholarly communication initiatives.”

<https://www.lib.purdue.edu/sites/default/files/admin/plan2016.pdf>

CURATION LIFECYCLE SERVICE MODEL



Witt, M. (2012). Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. *Journal of Library Administration*, 52(2). DOI:10.1080/01930826.2012.655607. http://docs.lib.purdue.edu/lib_fsdocs/6/

Digital Curation Centre's Curation Lifecycle Model: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

PURR

PURDUE UNIVERSITY RESEARCH REPOSITORY

FACT:

MANY FUNDING AGENCIES
REQUIRE
DATA MANAGEMENT PLANS
WITH GRANT PROPOSALS.

purrr.purdue.edu

PURDUE
UNIVERSITY

PURR IS YOUR SOLUTION FOR:

DATA MANAGEMENT PLANS

READY-MADE BOILERPLATE TEXT TO PUT IN YOUR PROPOSALS, TUTORIALS AND BEST PRACTICES, SUPPORT FOR DEVELOPING A GOOD DATA MANAGEMENT PLAN

COLLABORATION

PURR PROVIDES A FREE HUBZERO™ PROJECT SPACE FOR PURDUE RESEARCHERS AND THEIR COLLABORATORS TO WORK TOGETHER ON RESEARCH AND SHARE DATA AND TOOLS ONLINE

PUBLISHING YOUR DATA

PUBLISH YOUR RESEARCH DATA WITH DIGITAL OBJECT IDENTIFIERS THAT OTHER SCHOLARS CAN USE TO EASILY FIND AND CITE YOUR DATASETS

ARCHIVING YOUR DATA

YOUR DATA WILL BE PRESERVED AND MADE ACCESSIBLE FOR LONG-TERM ACCESS IN A SECURE, TRUSTWORTHY DIGITAL REPOSITORY

OVER **500** GRANT PROPOSALS FROM PURDUE HAVE INCLUDED
PURR IN THEIR DATA MANAGEMENT PLANS SINCE 2011

PURR IS A PURDUE RESEARCH CORE FACILITY DEVELOPED BY
THE OFFICE OF THE VICE PRESIDENT FOR RESEARCH, PURDUE
LIBRARIES, AND INFORMATION TECHNOLOGY AT PURDUE

To learn more, visit
PURR at purrr.purdue.edu
or contact **Courtney Earl Matthews**
purrr@purdue.edu
765-496-2770

DATA MANAGEMENT PLANS

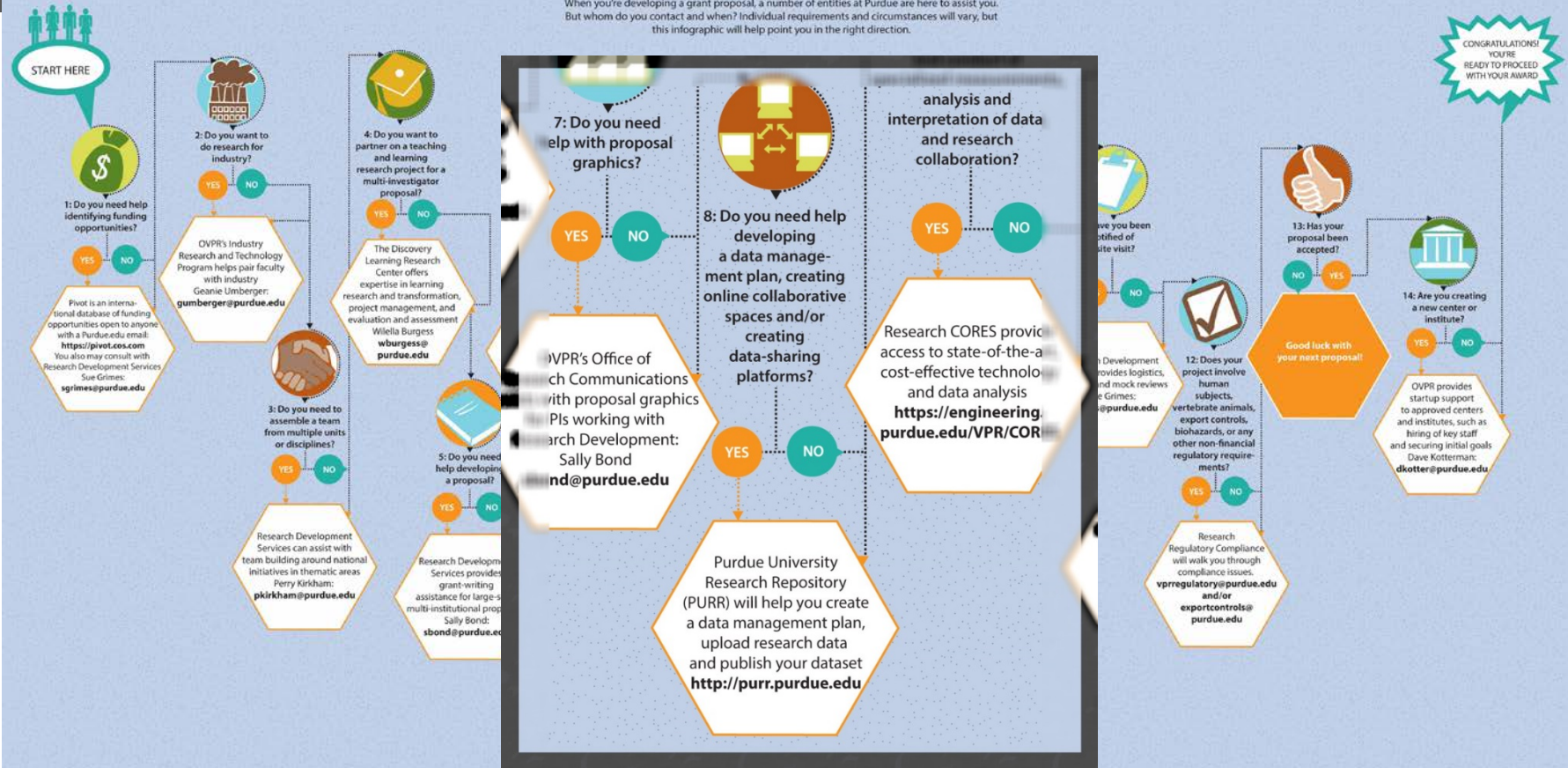
- Boilerplate text
- Example DMPs
- DMP Self-Assessment
- DMPTool
- Workshops
- Tutorials
- Reference and consultation with subject-specialist librarian and/or data services specialist



<https://purr.purdue.edu/dmp>

THE PROPOSAL PROCESS AT PURDUE

When you're developing a grant proposal, a number of entities at Purdue are here to assist you. But whom do you contact and when? Individual requirements and circumstances will vary, but this infographic will help point you in the right direction.



Dimensions of Discovery (Winter 2013). Office of the Vice President for Research, Purdue University, <http://www.purdue.edu/research/vpr/publications/docs/dimensions/Winter2013.pdf>

CREATE A PROJECT

The screenshot displays the Purdue University Research Repository (PURR) interface. At the top, the Purdue University logo and 'Purdue University Research Repository' are visible. Navigation links include 'Home', 'Publications', 'Projects', 'Get Started', and 'Contact Us'. A search bar is located on the right. The breadcrumb trail reads: 'You are here: Projects > Social Network Graph Analysis > Files'. The main content area shows the project title 'Social Network Graph Analysis (socnet)' with a 'Project manager' button. Below this, the 'Files' section lists two files: 'README_socnet-flickr-2006.pdf' (18KB, Jan 30, .me) and 'socnet-flickr.smat' (3KB, Jan 30, .me). The 'socnet-flickr.smat' file has a green checkmark and the text 'Social Network ...'. At the bottom, a 'Disk usage' section shows '0.4% used' and 'Project Quota: 10GB'. A 'Git Status' link is also present.

Purdue University | **Purdue University Research Repository** | PURR

My Account (mwitt) | Logout | 660 | Report a bug

Home | Publications | **Projects** | Get Started | Contact Us | Search

You are here: Projects > Social Network Graph Analysis > Files

Social Network Graph Analysis (socnet) | Project manager

Private project by Michael Witt

Files | Max upload size: 1.5GB

Upload | [Icons]

<input type="checkbox"/>	Name +	Size +	Modified +	By	Publication
<input type="checkbox"/>	README_socnet-flickr-2006.pdf	18KB	Jan 30	me	
<input type="checkbox"/>	socnet-flickr.smat	3KB	Jan 30	me	<input checked="" type="checkbox"/> Social Network ...

Disk usage: Project Quota: 10GB | Git Status

PURR project tutorial video:

http://www.youtube.com/watch?v=q5xGO_oF9uQ

USE PROJECT TO COLLABORATE



Create:

- any Purdue faculty, staff, or graduate student researcher can create projects
- describe the project
- disclaim use of sensitive or restricted data
- receive a default allocation of storage
- register a grant award to increase allocation
- invite collaborators to join project

Collaborate:

- git repository to share and version files (sftp & Google Drive integration)
- virtual machine/s
- wiki
- blog
- to-do list management and project notes
- newsfeed
- stage data publications

STORAGE ALLOCATION

Storage space for PURR projects currently covered by central resources				
Option:	For what:	Space Available:	For how long:	For who:
 Default/trial projects	Just trying things out, or don't need much space	10 GB	3 years	Any Purdue faculty, staff, and graduate student
 Default data publications	Great for small publications	1 GB	10 years	Any Purdue faculty, staff, and graduate student
 Supported projects	Funded projects with PIs from Purdue	100 GB	Life of grant	Purdue faculty with a verifiable grant or account number
 Supported data publications	Publishing work done on a funded project	10 GB	10 years	Purdue faculty with a verifiable grant or account number
Estimated cost of additional space *				
Option:	Space Available:	For how long:	For who:	Cost:
Extra project space	As needed	Per year	Any Purdue faculty, staff, and graduate student	\$2.10 per GB per year **
Extra publication space	As needed	10 years	Any Purdue faculty, staff, and graduate student	\$14.30 per GB

<https://purr.purdue.edu/about/pricing>

DATA PUBLICATION & ARCHIVING

The screenshot shows the Purdue University Research Repository (PURR) interface. At the top, the PURR logo and navigation menu are visible. The main content area features the title "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006" by David F. Gleich. A "Download (SMA)" button is present, along with a version information box stating "Version 1.1 - published on Feb 22, 2012" and a DOI link. A sidebar on the right contains statistics: 0 reviews, 0 questions, 0 citations, and 131 users with 220 downloads. Below the main content, there are tabs for "About", "Supporting Docs", "Versions", "Reviews", "Questions", and "Citations". The "About" tab is active, showing an abstract and a citation template. The abstract describes the graph's construction from a Flickr crawl in May 2006. The citation template provides the author, year, title, and DOI.

Graph of Flickr Photo-Sharing Social Network Crawled in May 2006

By David F Gleich
Purdue University

Crawl of the Flickr photo-sharing social network from May 2006 returning a graph with 820,878 nodes and 9,837,214 edges. Dataset is distributed as a SMA file with README file with code to read file in Python and MATLAB.

Listed in Datasets

This is a live publication with the page and content publicly available. [Manage this publication.](#) from project [Networks and Matrix Computations](#)

About Supporting Docs Versions Reviews Questions Citations



Abstract Flickr is a popular online-community for sharing photos, with millions of users. This graph is representative of its social network, in which the node set V represents users, and the edge set E is such that (u, v) is in E if and only if a user u has added user v as his/her contact. We start with a crawl extracted from Flickr in May 2006. This crawl began with a single user and continued until the total personalized PageRank on the set of uncrawled nodes was less than 0.0001. The result of the crawl was a graph with 820,878 nodes and 9,837,214 edges.

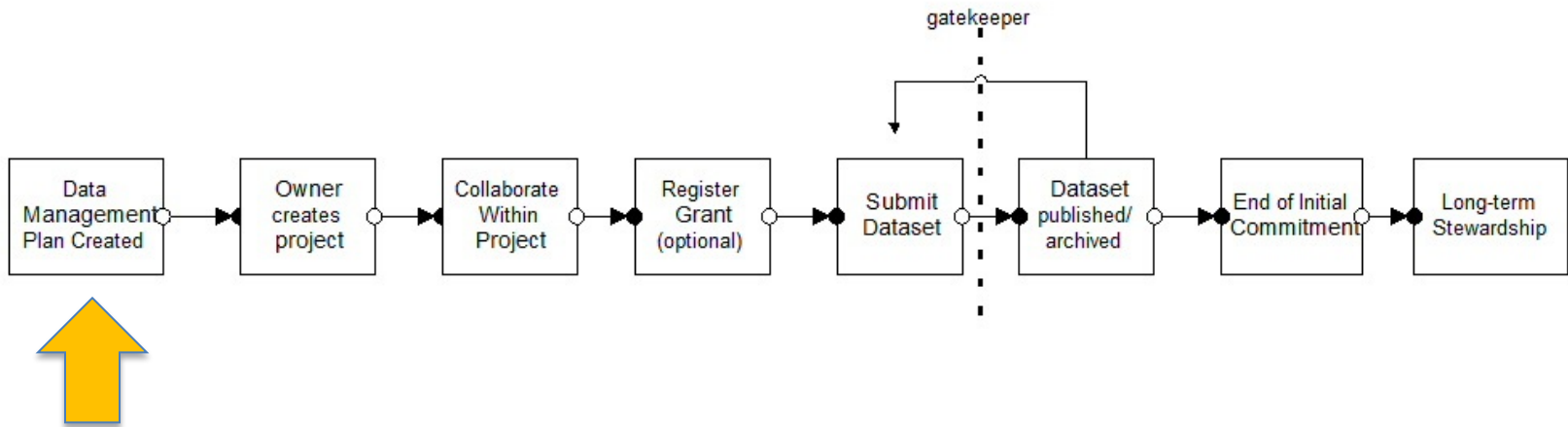
Cite this work Researchers should cite this work as follows:

David F Gleich, (2012), "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006": (DOI: 10.4231/D39P2W550)
[BibTex](#) [EndNote](#)

PURR publication tutorial video:
<http://www.youtube.com/watch?v=jYBcsfiRhio>

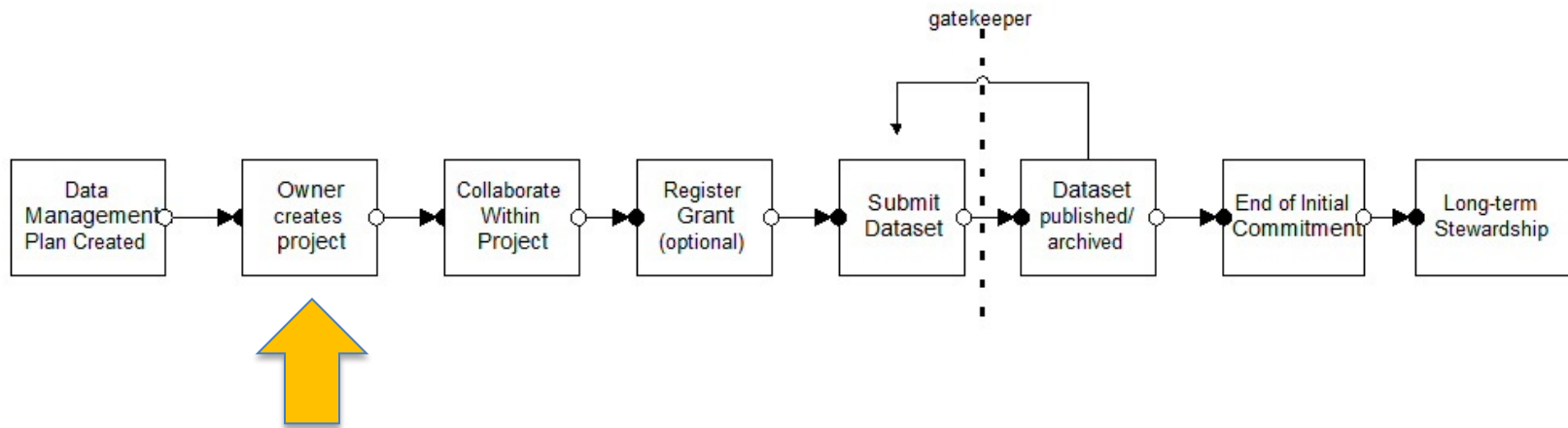
ORGANIZATION AND STAFFING

- **Executive Committee:** Dean of Libraries, Vice President for Research, Chief Information Officer
- **Steering Committee:** 2 from libraries, 2 from IT, 2 from research office and sponsored programs, 3 domain faculty researchers
- **Personnel:** Project Director (.50), Technologists (3.85), HUBzero Liaison (.35), Metadata Specialist (.20), Digital Archivist (.25), Repository Outreach Specialist (1.0), Data Curator (1.0)
- **Key players:** Subject-specialist librarians & data services specialists



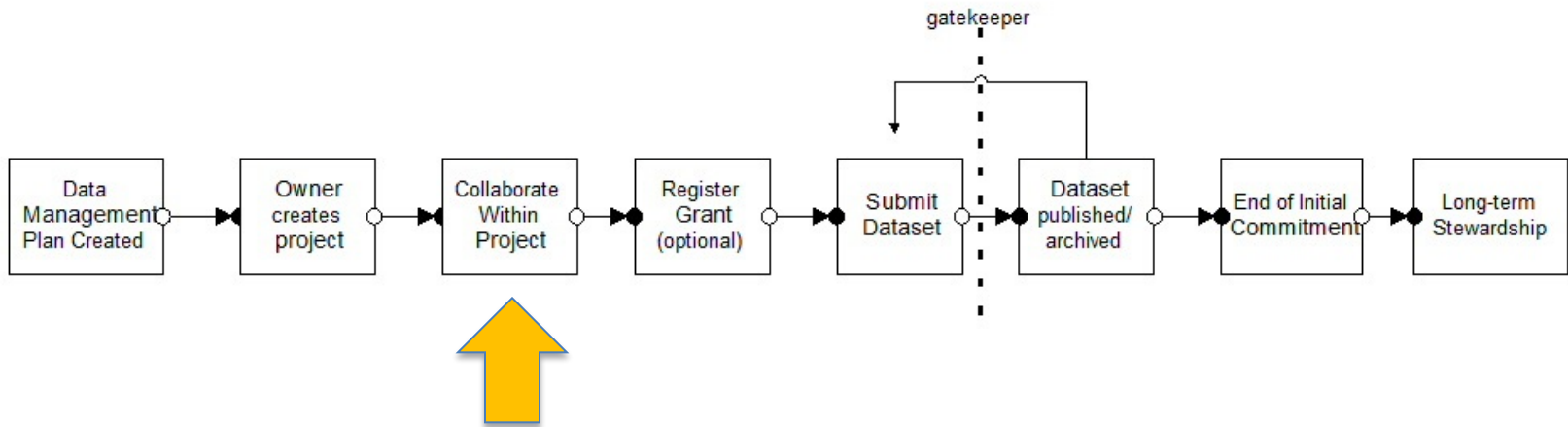
Librarians consult on data management plans in their subject areas.

Creating opportunities for librarians to interact with researchers about data



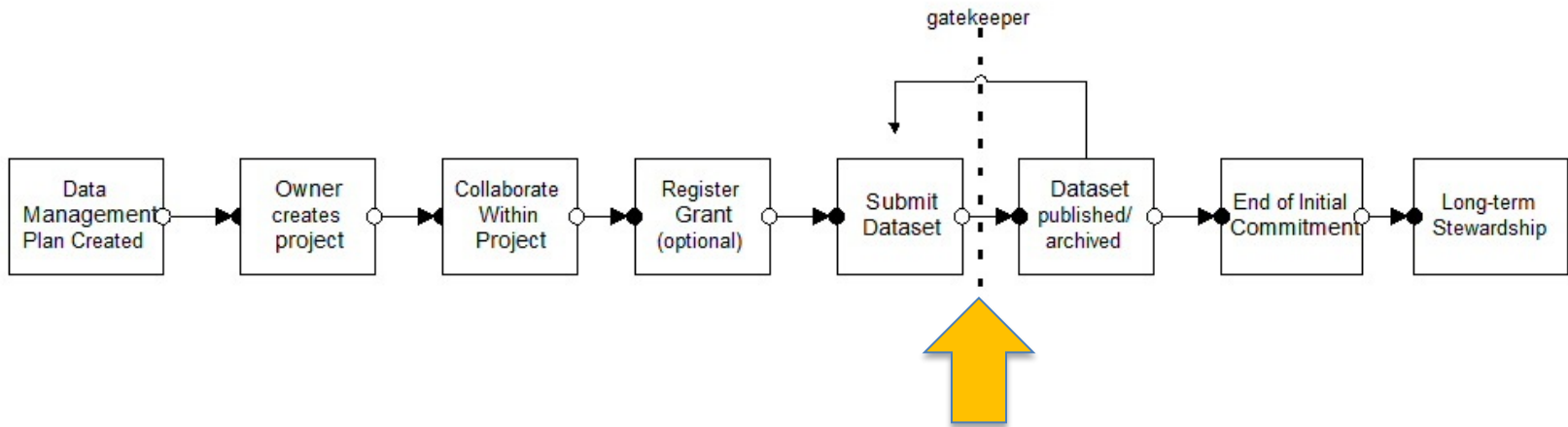
Librarian is notified by e-mail when a new project is created or a grant is awarded, based on department affiliation of Purdue project owner.

Creating opportunities for librarians to interact with researchers about data



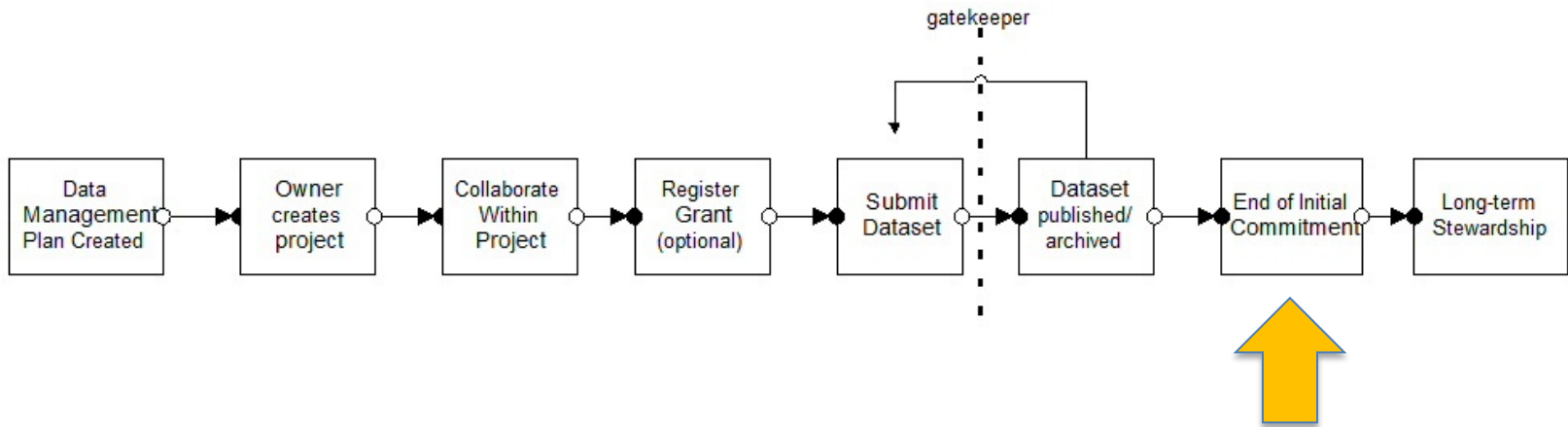
Librarian may consult or collaborate on project if needed.

Creating opportunities for librarians to interact with researchers about data



Librarians review and post submitted datasets.

Creating opportunities for librarians to interact with researchers about data



At the end of initial commitment (10 years), archived and published datasets are remanded to the Libraries' collection. A librarian working with the digital archivist selects (or not) the dataset for the collection.

Creating opportunities for librarians to interact with researchers about data

ASSESSMENT

2013 was first full year of PURR in operation, to date:

- 1,405 data management plans
- 163 grant awards
- 1,295 registered researchers
- 477 research projects (455 active in last 30 days)
- 120 citations

DMP analysis (n=111 NSF proposals from Purdue, Jan-Jun 2013)

- 49% PURR
- 29% Local computer or server
- 14% Disciplinary repository (e.g., ICPSR, Protein Data Bank, nanoHUB, NEES)
- 8% No data or not applicable

THANK YOU

<http://bit.ly/1HUfKJI>



PURR: <http://purrr.purdue.edu>

PURDUE
UNIVERSITY
LIBRARIES



D2C2

DISTRIBUTED DATA
CURATION CENTER

Michael Witt

Head, Distributed Data Curation Center

Associate Professor of Library Science

<http://www.lib.purdue.edu/research/witt>

E-mail: mwitt@purdue.edu