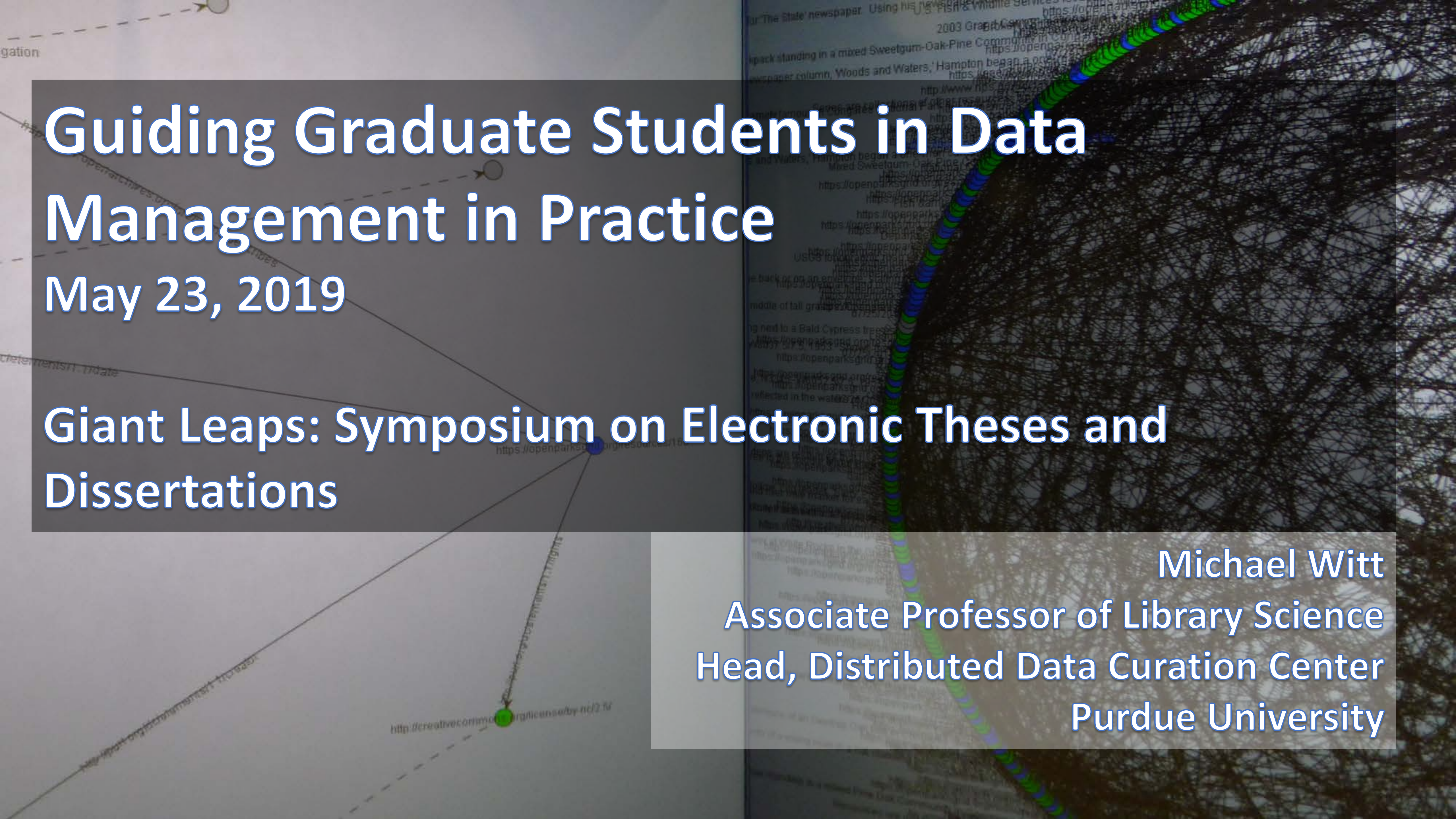
5-23-2019

Guiding Graduate Students in Data Management in Practice

Michael Witt
Purdue University, mwitt@purdue.edu



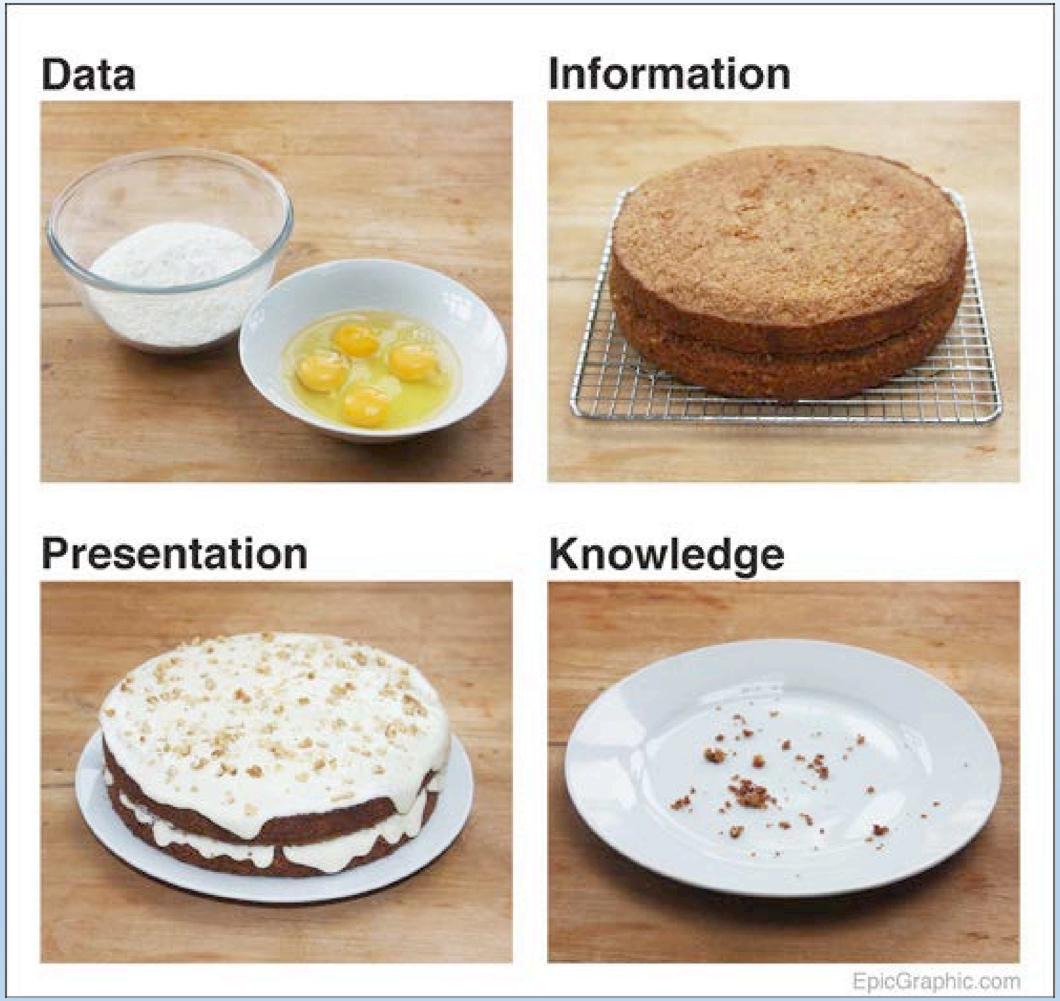
Guiding Graduate Students in Data Management in Practice

May 23, 2019

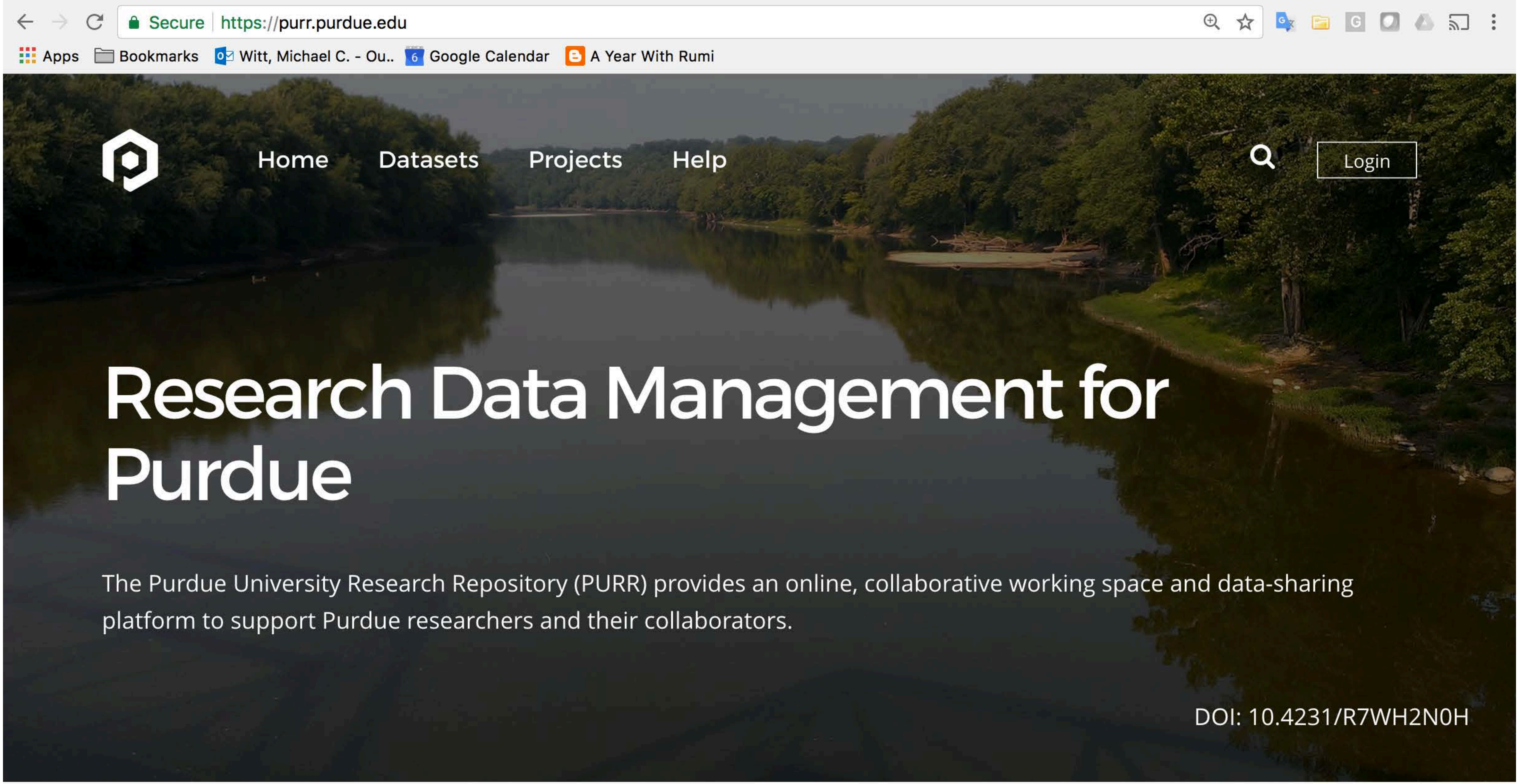
Giant Leaps: Symposium on Electronic Theses and Dissertations

Michael Witt
Associate Professor of Library Science
Head, Distributed Data Curation Center
Purdue University

Data = evidence for research



http://epicgraphic.com/data-cake



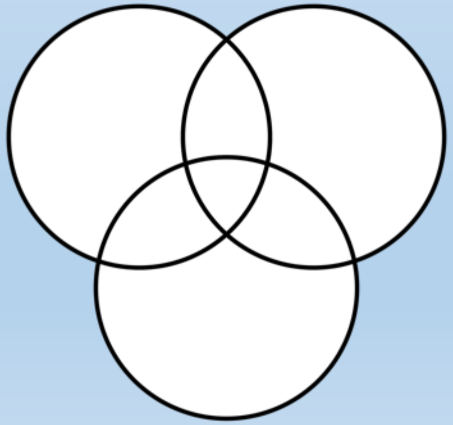
Research Data Management for Purdue

The Purdue University Research Repository (PURR) provides an online, collaborative working space and data-sharing platform to support Purdue researchers and their collaborators.

DOI: 10.4231/R7WH2N0H

Institutional Motivations for PURR

- Research office = more competitive proposals and compliance with funder requirements
- Information technology = research computing expertise, e.g., storage engineering, HPC
- Libraries = long-term stewardship and access to data as a part of the scholarly record, library and information science expertise



Purdue: 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 **student researchers**; their collaborators; and the current and future consumers of their research data.

Based on the HUBzero Platform for Scientific Collaboration open source software.

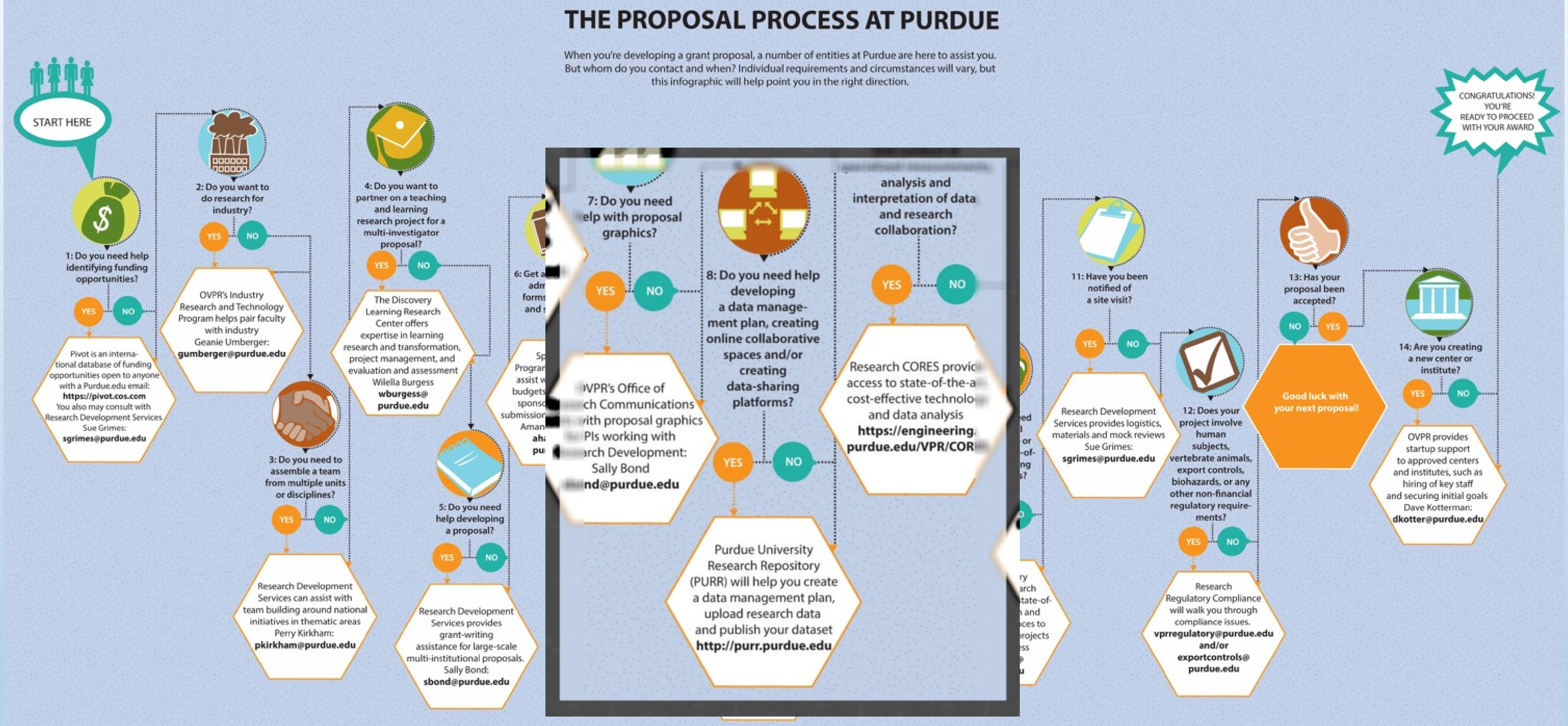
What will PURR do for you as a researcher?

1. Help you write and implement an effective Data Management Plan (DMP)
2. Provide you with private, project space to share and work with your data and collaborators
3. Publish your open data in a scholarly context with metadata and a Digital Object Identifier (DOI)
4. Archive your data for future reuse
5. Measure the impact of sharing your data

1. Write your Data Management Plan

- Boilerplate text
- Self-Assessment
- Example DMPs
- Up-to-date funder requirements
- DMPTool
- Workshops
- Tutorials (new videos coming soon)
- Reference and consultation in person 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.

START HERE

1: Do you need help identifying funding opportunities?

YES NO

Pivot is an international database of funding opportunities open to anyone with a Purdue.edu email: **https://pivot.cos.com** You also may consult with Research Development Services Sue Grimes: **sgrimes@purdue.edu**

2: Do you want to do research for industry?

YES NO

OVPR's Industry Research and Technology Program helps pair faculty with industry Geanie Umberger: **gumberger@purdue.edu**

3: Do you need to assemble a team from multiple units or disciplines?

YES NO

Research Development Services can assist with team building around national initiatives in thematic areas Perry Kirkham: **pkirkham@purdue.edu**

4: Do you want to partner on a teaching and learning research project for a multi-investigator proposal?

YES NO

The Discovery Learning Research Center offers expertise in learning research and transformation, project management, and evaluation and assessment Wilella Burgess **wburgess@ purdue.edu**

5: Do you need help developing a proposal?

YES NO

Research Development Services provides grant-writing assistance for large-scale multi-institutional proposals. Sally Bond: **sbond@purdue.edu**

6: Get a adm forms and s

Sp Progran assist w budgets sponso submission Aman **aha pu**

7: Do you need elp with proposal graphics?

YES NO

VPR's Office of ch Communications ith proposal graphics PIs working with arch Development: Sally Bond **nd@purdue.edu**

8: Do you need help developing a data management plan, creating online collaborative spaces and/or creating data-sharing platforms?

YES NO

Purdue University Research Repository (PURR) will help you create a data management plan, upload research data and publish your dataset **http://purr.purdue.edu**

analysis and interpretation of data and research collaboration?

YES NO

Research CORES provid access to state-of-the-a cost-effective technolo and data analysis **https://engineering. purdue.edu/VPR/COR**

ed l or -of- ng s?

ry arch tate-of- and ces to rojects ss @ u

11: Have you been notified of a site visit?

YES NO

Research Development Services provides logistics, materials and mock reviews Sue Grimes: **sgrimes@purdue.edu**

12: Does your project involve human subjects, vertebrate animals, export controls, biohazards, or any other non-financial regulatory requirements?

YES NO

Research Regulatory Compliance will walk you through compliance issues. **vprregulatory@purdue.edu and/or exportcontrols@ purdue.edu**

13: Has your proposal been accepted?

NO YES

Good luck with your next proposal!

14: Are you creating a new center or institute?

YES NO

OVPR provides startup support to approved centers and institutes, such as hiring of key staff and securing initial goals Dave Kotterman: **dkotter@purdue.edu**

CONGRATULATIONS! YOU'RE READY TO PROCEED WITH YOUR AWARD

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*

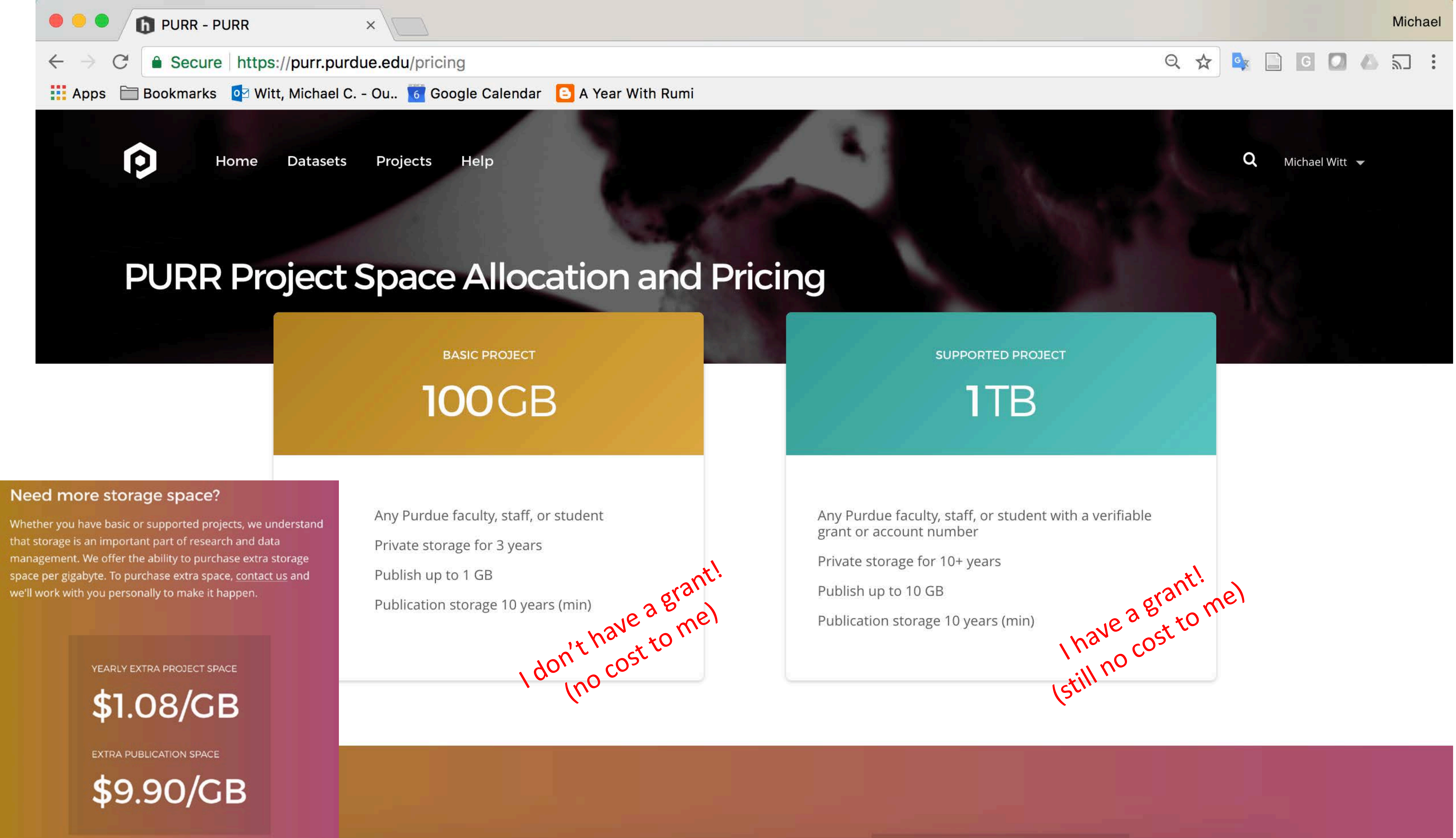
2. Create a Project and Collaborate

Create:

- any Purdue faculty, staff, or student researcher can create private projects
- describe the project and disclaim use of sensitive or restricted data
- receive a default allocation of storage
- register a grant award to increase allocation
- invite collaborators from other institutions to join as managers, collaborators, or reviewers

Collaborate:

- Private project space to upload and share files with project members (integration with sftp, Google Drive, etc.)
- Wiki, blog, to-do list management and project notes, newsfeed
- stage data publications



PURR Project Space Allocation and Pricing

BASIC PROJECT

100GB

Any Purdue faculty, staff, or student

Private storage for 3 years

Publish up to 1 GB

Publication storage 10 years (min)

SUPPORTED PROJECT

1TB

Any Purdue faculty, staff, or student with a verifiable grant or account number

Private storage for 10+ years

Publish up to 10 GB

Publication storage 10 years (min)

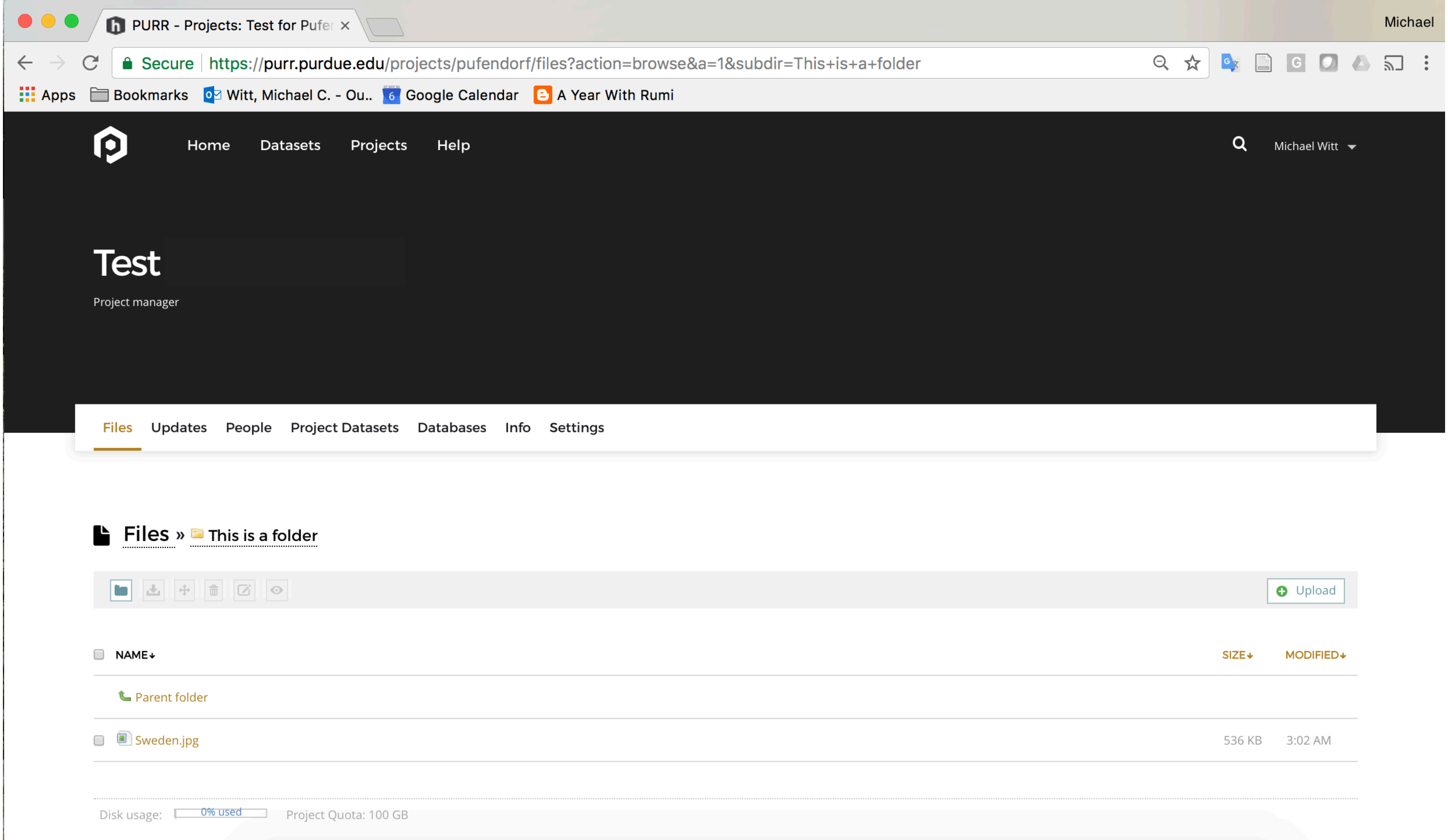
Need more storage space?

Whether you have basic or supported projects, we understand that storage is an important part of research and data management. We offer the ability to purchase extra storage space per gigabyte. To purchase extra space, contact us and we'll work with you personally to make it happen.

YEARLY EXTRA PROJECT SPACE	$1.08/GB
EXTRA PUBLICATION SPACE	$9.90/GB

I don't have a grant!
(no cost to me)

I have a grant!
(still no cost to me)



Test

Project manager

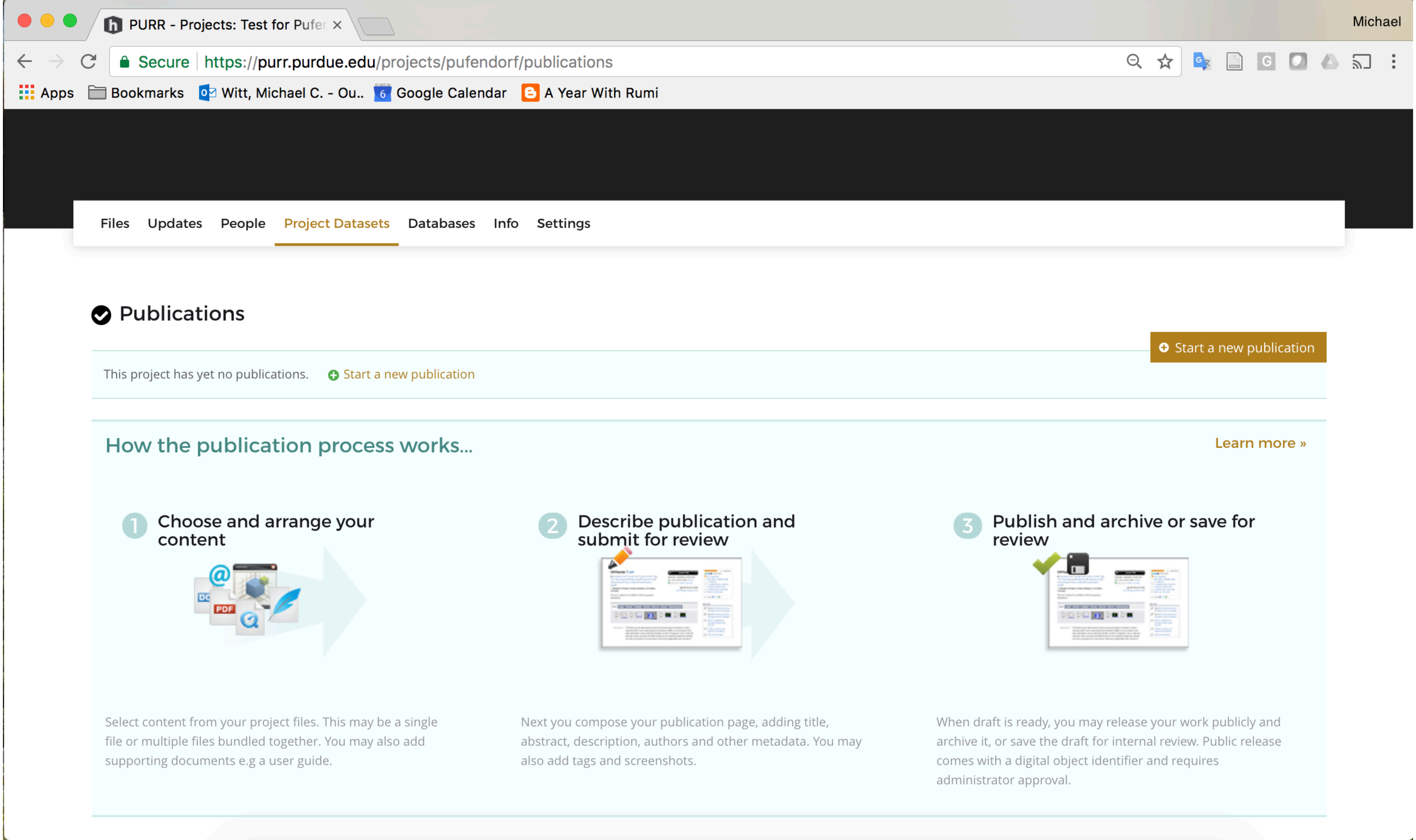
Files » This is a folder

□	NAME↓	SIZE↓	MODIFIED↓
	Parent folder		
□	Sweden.jpg	536 KB	3:02 AM

Disk usage: 0% used Project Quota: 100 GB

3. Publish your Data

- Choose the primary and supporting files that make up your dataset
- Provide metadata such as title, authors, description, keywords, citations, release notes, etc.
- Assign end-user license or suggest a new license
- Set optional release date (i.e., embargo)
- Data publication is queued for curation
  - DOI is reserved and presented
  - Review by subject liaison and repository specialist
  - Opportunity for enhancement
- Dataset published on release date (default = immediately)
- Types of publication = dataset (files), database, series



Publications

This project has yet no publications. Start a new publication

How the publication process works...

1 Choose and arrange your content

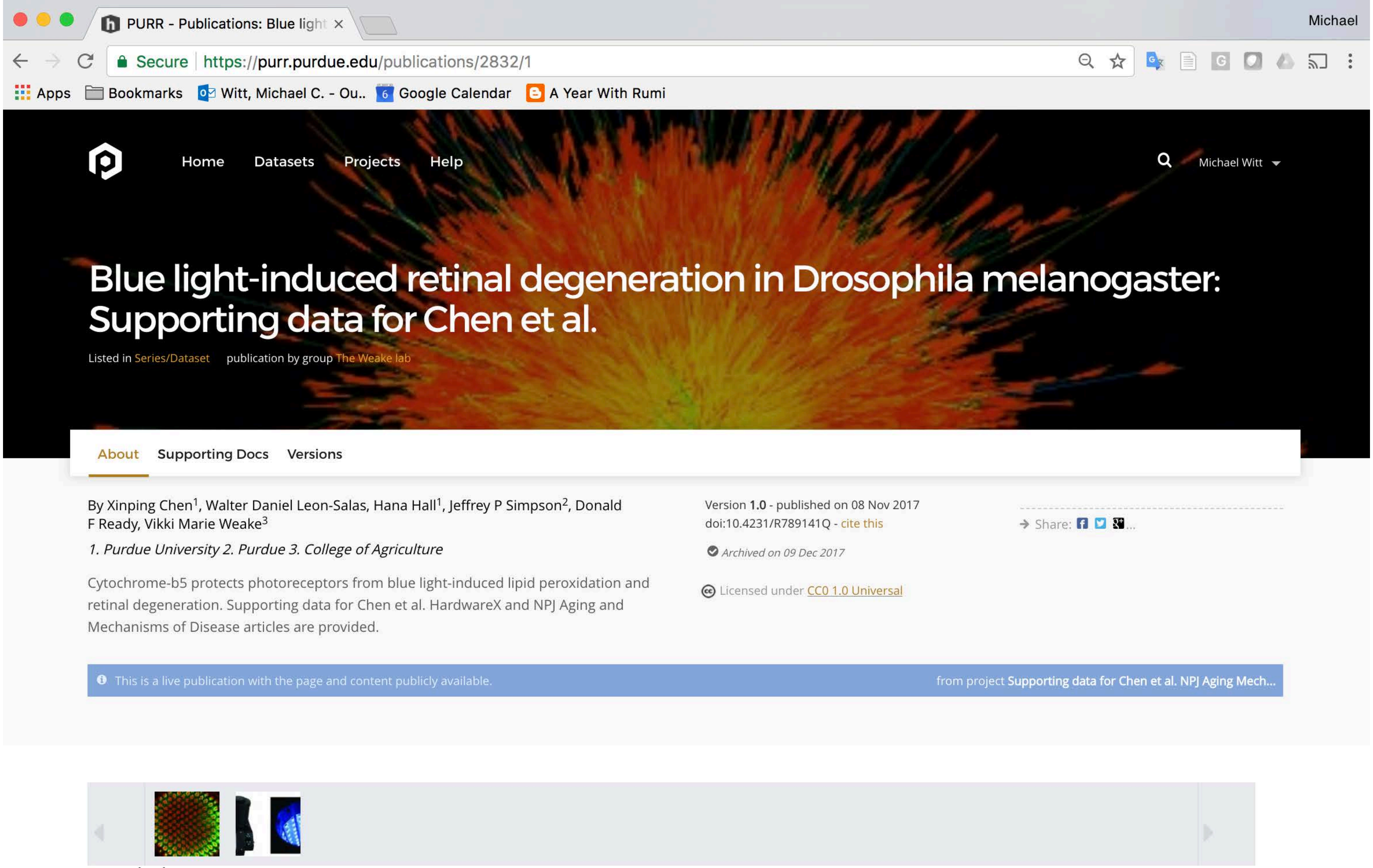
Select content from your project files. This may be a single file or multiple files bundled together. You may also add supporting documents e.g a user guide.

2 Describe publication and submit for review

Next you compose your publication page, adding title, abstract, description, authors and other metadata. You may also add tags and screenshots.

3 Publish and archive or save for review

When draft is ready, you may release your work publicly and archive it, or save the draft for internal review. Public release comes with a digital object identifier and requires administrator approval.



Blue light-induced retinal degeneration in Drosophila melanogaster: Supporting data for Chen et al.

Listed in Series/Dataset publication by group The Weake lab

By Xinping Chen[1], Walter Daniel Leon-Salas, Hana Hall[1], Jeffrey P Simpson[2], Donald F Ready, Vikki Marie Weake[3]

1. Purdue University 2. Purdue 3. College of Agriculture

Cytochrome-b5 protects photoreceptors from blue light-induced lipid peroxidation and retinal degeneration. Supporting data for Chen et al. HardwareX and NPJ Aging and Mechanisms of Disease articles are provided.

Version **1.0** - published on 08 Nov 2017
doi:10.4231/R789141Q - cite this

Archived on 09 Dec 2017

Licensed under CC0 1.0 Universal

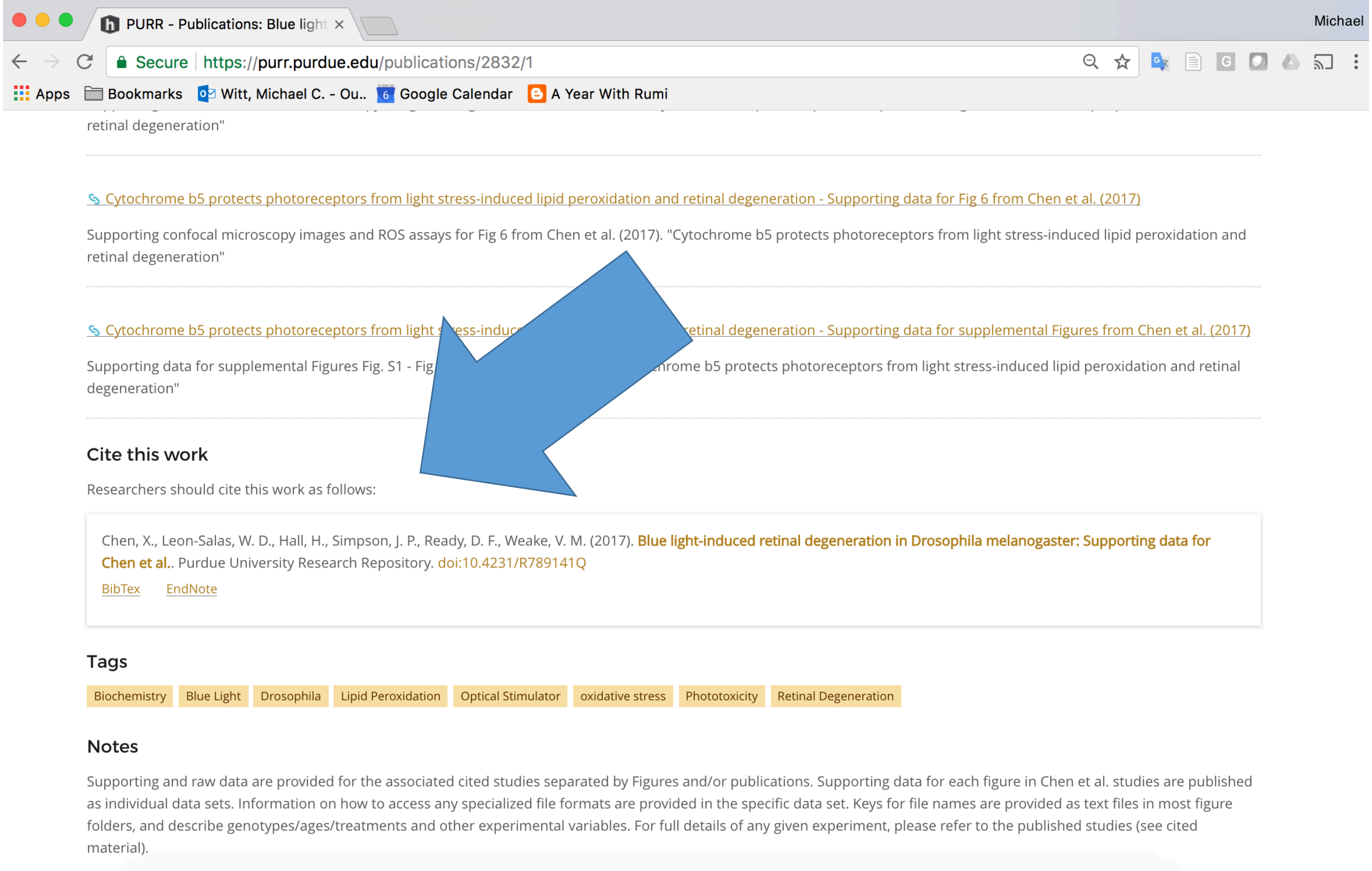
This is a live publication with the page and content publicly available. from project **Supporting data for Chen et al. NPJ Aging Mech...**

4. Archive your Data

- 30 days after data are approved for publication, they are archived
- DROID identifies files formats and records PRONOM information
- Archival Information Packages (AIPs) are serialized using BagIt with metadata (METS, MODS, PREMIS, etc.)
- AIPs are replicated to 7 distributed sites on a LOCKSS (Lots of Copies Keeps Stuff Safe) preservation network that is managed by the MetaArchive Cooperative
- Digital Preservation policies and guidance for PURR:
  - Preservation Policy
  - Preservation Strategic Plan
  - File Format Recommendations
  - Preservation Support Policy

5. Measuring Impact of Datasets

- Early days, we don’t have the same bibliometrics for data as we do for published literature
- But we have some of the pieces in place (e.g., DataCite DOIs) for a good start and as data citation practice becomes more widely adopted
- Data producers and publishers can report citations to data (but they typically don’t; Scholix in the future?)
- Once a quarter, a graduate assistant manually searches for citations of datasets and adds them to PURR
- Usage reports (views, downloads, citations) can viewed on demand and are automatically emailed to data producers once a month



retinal degeneration"

Cytochrome b5 protects photoreceptors from light stress-induced lipid peroxidation and retinal degeneration - Supporting data for Fig 6 from Chen et al. (2017)

Supporting confocal microscopy images and ROS assays for Fig 6 from Chen et al. (2017). "Cytochrome b5 protects photoreceptors from light stress-induced lipid peroxidation and retinal degeneration"

Cytochrome b5 protects photoreceptors from light s ess-induc etinal degeneration - Supporting data for supplemental Figures from Chen et al. (2017)

Supporting data for supplemental Figures Fig. S1 - Fig hrome b5 protects photoreceptors from light stress-induced lipid peroxidation and retinal degeneration"

Cite this work

Researchers should cite this work as follows:

Chen, X., Leon-Salas, W. D., Hall, H., Simpson, J. P., Ready, D. F., Weake, V. M. (2017). **Blue light-induced retinal degeneration in Drosophila melanogaster: Supporting data for Chen et al.**. Purdue University Research Repository. doi:10.4231/R789141Q

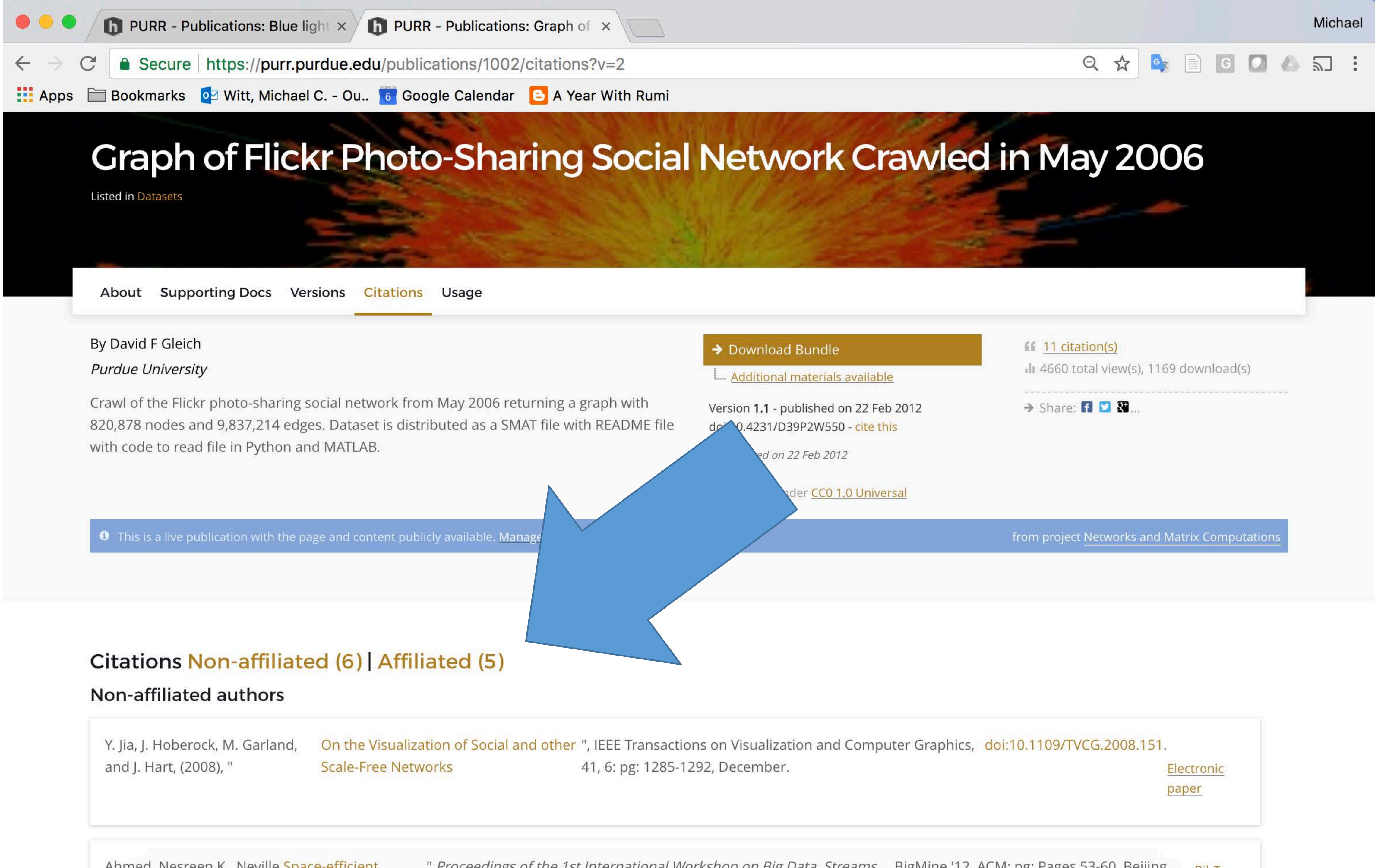
BibTex EndNote

Tags

Biochemistry | Blue Light | Drosophila | Lipid Peroxidation | Optical Stimulator | oxidative stress | Phototoxicity | Retinal Degeneration

Notes

Supporting and raw data are provided for the associated cited studies separated by Figures and/or publications. Supporting data for each figure in Chen et al. studies are published as individual data sets. Information on how to access any specialized file formats are provided in the specific data set. Keys for file names are provided as text files in most figure folders, and describe genotypes/ages/treatments and other experimental variables. For full details of any given experiment, please refer to the published studies (see cited material).


PURR - Publications: Blue light
PURR - Publications: Graph of
Michael
Secure
https://purr.purdue.edu/publications/1002/citations?v=2
Apps
Bookmarks
Witt, Michael C. - Ou..
Google Calendar
A Year With Rumi

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

Listed in Datasets

About | Supporting Docs | Versions | Citations | Usage

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 SMAT file with README file with code to read file in Python and MATLAB.

→ Download Bundle

Additional materials available

Version **1.1** - published on 22 Feb 2012
do 0.4231/D39P2W550 - cite this

ed on 22 Feb 2012

der CC0 1.0 Universal

11 citation(s)

4660 total view(s), 1169 download(s)

→ Share:

This is a live publication with the page and content publicly available. Manage

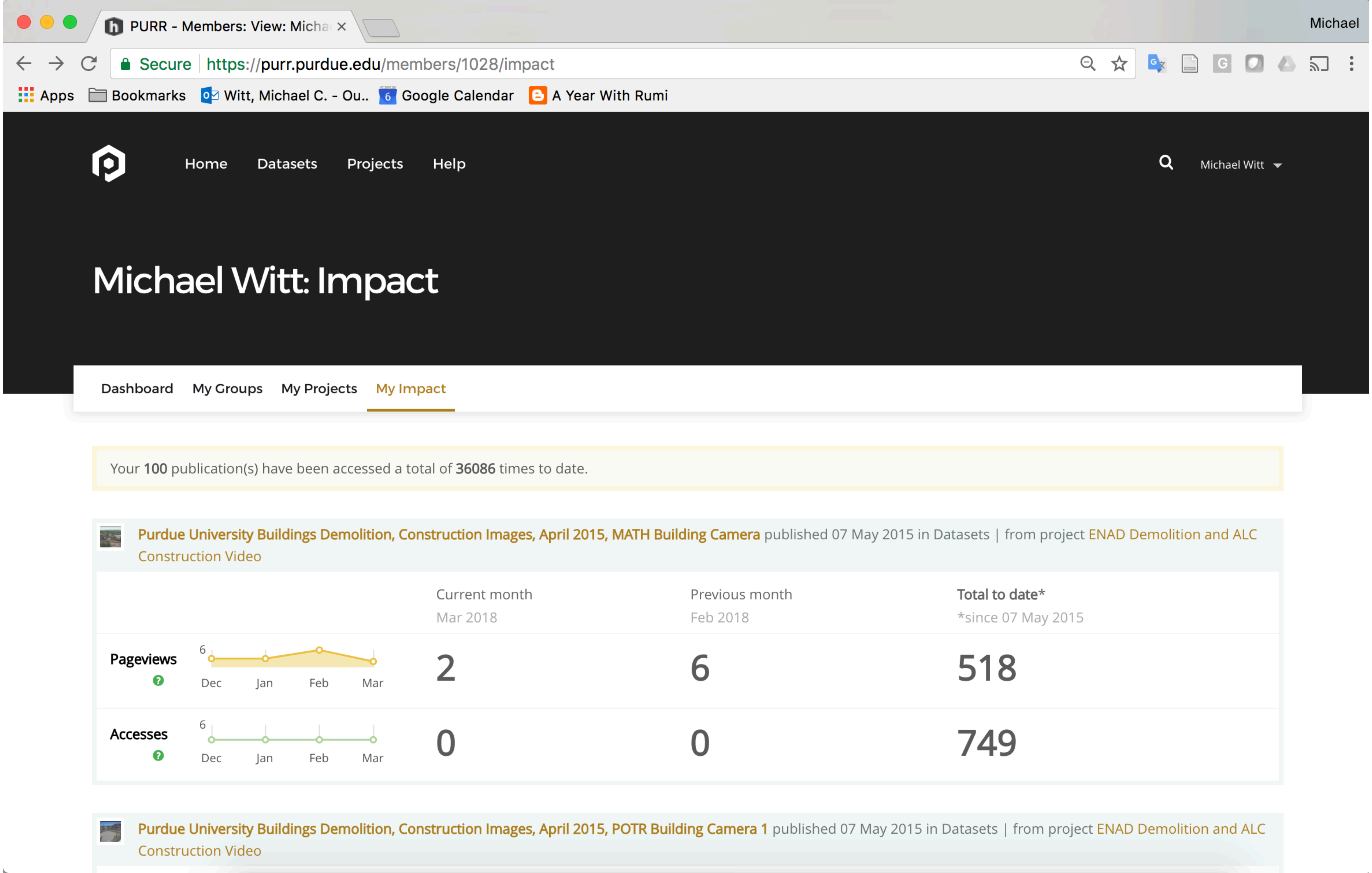
from project Networks and Matrix Computations

Citations Non-affiliated (6) | Affiliated (5)

Non-affiliated authors

Y. Jia, J. Hoberock, M. Garland, and J. Hart, (2008), "On the Visualization of Social and other Scale-Free Networks", IEEE Transactions on Visualization and Computer Graphics, 41, 6: pg: 1285-1292, December. doi:10.1109/TVCG.2008.151. Electronic paper

Ahmed, Nesreen K., Neville Space-efficient " *Proceedings of the 1st International Workshop on Big Data, Streams* ... BigMine '12, ACM: pg: Pages 53-60, Beijing



Michael Witt: Impact

Your **100** publication(s) have been accessed a total of **36086** times to date.

Purdue University Buildings Demolition, Construction Images, April 2015, MATH Building Camera published 07 May 2015 in Datasets | from project ENAD Demolition and ALC Construction Video

	Current month Mar 2018	Previous month Feb 2018	**Total to date*** *since 07 May 2015
Pageviews	2	6	518
Accesses	0	0	749

Purdue University Buildings Demolition, Construction Images, April 2015, POTR Building Camera 1 published 07 May 2015 in Datasets | from project ENAD Demolition and ALC Construction Video

PURR governance & 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

PURR by the numbers

- 3,454 data management plans (grant proposals)
- 487 grant awards
- 3,657 registered researchers
- 1,397 research projects
- 937 published datasets
- 505 data citations

Running totals to date since launch in 2012

Research data supporting theses

- Students receive email when they pass prelims with link to project in PURR that has been created for them
- Follow-up contact from data outreach specialist
- Student encouraged to invite committee to project and share data with them during their research
- Publish datasets with DOI in PURR; set embargo if desired
- Cite their datasets in their thesis; deposit thesis in Hammer (figshare)
- Citations from datasets in PURR to thesis in figshare are added quarterly so that links are bi-directional
- Pilot with 100 students
- Presentations during graduate student orientations and thesis workshops
- Documentation and support on Graduate School and PURR websites