

NEIL ARMSTRONG HALL OF ENGINEERING

PURR: A RESEARCH DATA
CURATION SERVICE
MODEL USING HUBZERO

HUBBUB 2012
Courtney Earl Matthews
Digital Data Repository Specialist
Purdue University

AGENDA

PURR Background

Data Management Planning

PURR Workflow

Demonstration

Services / Support

PURR is a hub that provides a research data and collaboration solution for Purdue faculty, graduates, and staff.

Research data - spreadsheets, images, output from sensors and instruments, transcripts, surveys, software source code and tools, video, and observation logs

PURR provides:

- Collaborative Research Space
- Long term Preservation
- Support and Services
- Data Management Plan resources and consultation

DMP REQUIREMENT

GRANT FUNDING AGENCIES

National Science Foundation

National Institutes of Health

National Endowment for the Humanities

NASA Earth Sciences Program

Centre for Disease Control

GOVERNANCE

PURDUE COLLABORATORS

PUL = Purdue University Libraries

ITaP = Information Technology at Purdue

OVPR = Office of the Vice President for Research

SPS = Sponsored Program Services

Funded 3-year HUBzero hosting plan (OVPR, Libraries, ITaP)

PURR Steering/Executive Committee:

- Dean of Libraries
- Vice President for Information Technology (CIO)
- Vice President for Research

RESEARCH & DATA WORKFLOW - 1

PROJECTS & PUBLICATIONS

A **project** is a dedicated working space on PURR for you to collaborate and prepare data for publication and curation for a research project or study.

See https://research.hub.purdue.edu/kb/AboutPURR/what_is_a_project

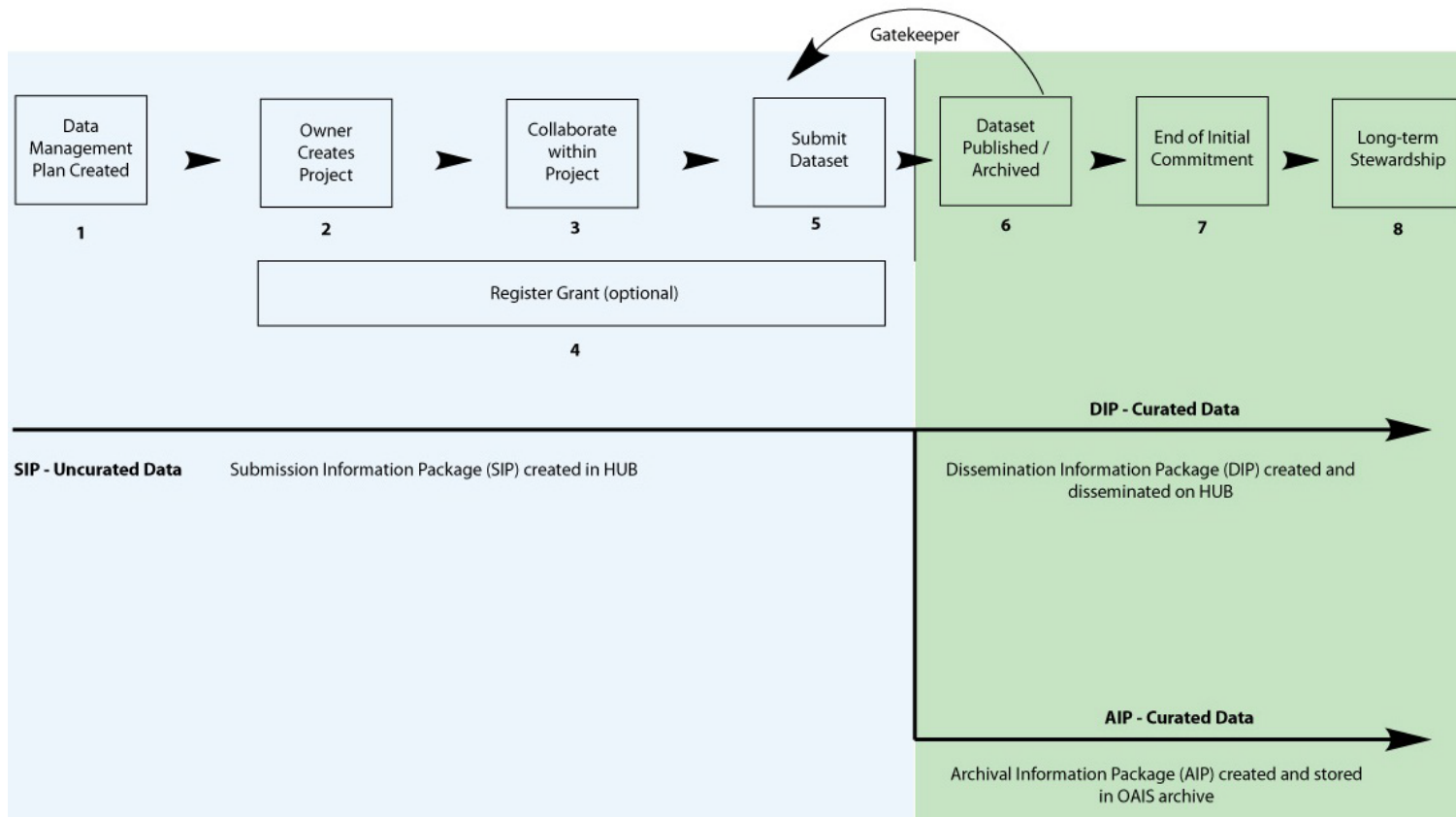
A **dataset publication** is a collection of files and metadata you choose from your project to publish.

See <https://research.hub.purdue.edu/publications?>

RESEARCH & DATA WORKFLOW - 2

PURR SERVICE WORKFLOW

PURR Workflow mapped to OAIS Reference Model



PROJECT

CREATE PROJECT - 1

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) interface. The browser's address bar shows the URL `research.hub.purdue.edu/projects/nmcomp/files`. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". Navigation links include "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the header.

The main content area displays the project details for "Networks and Matrix Computations (nmcomp)", a general project by David F Gleich. The project is marked as "private project". A sidebar on the left contains navigation options: "Updates" (2), "Info", "Team" (2), "Files" (5), "Publications", "To-do", and "Notes". Below these are "Manager options": "Edit project", "Invite people to join", and "Learn about projects".

The "Files" section shows a table of 5 files with columns for Name, Size, Modified, By, Revisions, and Associated Publication. A green notification states "File successfully uploaded". The table lists the following files:


Name	Size	Modified	By	Revisions	Associated Publication
gleich-flickr-2006(...).smat	144.32MB	02/22/2012 9:57 AM	Michael Witt	1	
Gleichflickr(...).APlusAT	63KB	02/22/2012 9:56 AM	Michael Witt	2	
README-purr.md	3KB	02/17/2012 11:58 AM	David F Gleich	2	
README_gleich-flickr(...).pdf	18KB	02/22/2012 10:08 AM	Michael Witt	1	
README_gleich-flickr(...).txt	3KB	02/22/2012 10:08 AM	Michael Witt	1	

The footer of the page includes the text "Copyright © 2011 Purdue University. All Rights Reserved." and "Powered by HUBzero, a Purdue project". The bottom right corner of the browser window shows the Zotero logo.

PROJECT

CREATE PROJECT - 2

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

My Account (matthew6) Logout  22 Report a bug

Home Browse Content **Projects** Get Started Contact Us

You are here: [Projects](#) > [Networks and Matrix Computations](#) > [Edit](#)

 **Networks and Matrix Computations** (nmcomp)
[Return to project page](#)

Edit Project










- Edit Info**
- [Edit Team](#)
- [Edit Settings](#)

Edit Info

Alias nmcomp


Title

About

B I U  **X₂ X²**      Style   

[switch to source](#)

Picture


 Upload new image (any existing image will be replaced)

No file chosen

PROJECT

CREATE PROJECT - 3

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

My Account (matthew6) Logout  22 Report a bug

Home Browse Content **Projects** Get Started Contact Us

You are here: [Projects](#) > [Networks and Matrix Computations](#) > Edit

Networks and Matrix Computations (nmcomp)

[Return to project page](#)

Edit Project

- [Edit Info](#)
- [Edit Team](#)**
- [Edit Settings](#)

Tips

About team roles

Collaborator can:

- upload and manage project files,
- edit project publications,
- use available project tools, e.g. notes and to-do's


Manager can:

- invite or remove other team members,
- change project information and global settings
- do everything that a collaborator can do


Edit Team

Add new member(s) as:

- manager collaborator

 Individual:

AND / OR

 User group:

*Find registered user by name or invite by email / type name of the group and select from drop-down

Total members: 3

Click on the user role to change. Check boxes to delete.

<input type="checkbox"/>	 Name↓	Role↓	Joined↓	 with group
<input type="checkbox"/>	 Courtney E Matthews matthew6	manager	23/09/2012	
<input type="checkbox"/>	 David F Gleich dgleich	manager	17/02/2012	
<input type="checkbox"/>	 Michael Witt mwitt	manager	21/02/2012	

[« previous](#) | [next »](#)

Project Owner:

[David F Gleich](#)

PROJECT

CREATE PROJECT - 4



Networks and Matrix Computations (nmcomp)

Return to [project page](#)

Edit Project

[Edit Info](#)

[Edit Team](#)

[Edit Settings](#)

Tips

Public projects

Public projects are listed in the public project directory, with the basic project information visible to users and visitors of the site.

Why provide grant information?

Knowing your awarded grant information allows us to allocate the appropriate amount of data storage space to your project. If you omit this information, you will receive default storage space.

Edit Settings

Access

- Project is hidden from search and listings (private project)
- Anyone can find this project in search and listings and view its basic information (public project)

Is this project supported by a research grant?

Fill in some information about your grant. If we can verify your award, we'll automatically increase the storage space available for your project.

Proposal Title:

Principal Investigator:

Funding Agency:

Total Award Budget:

Notify administrator that grant information has been updated

Save changes

Cancel

PUBLICATION

CREATE PUBLICATION - 1

The screenshot shows a web browser window displaying the PURR (Purdue University Research Repository) website. The browser's address bar shows the URL: `research.hub.purdue.edu/projects/nmcomp/publications`. The page header includes the PURDUE UNIVERSITY logo and the text "Purdue University Research Repository | PURR". Navigation links include "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the header.

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is identified as a "General project by David F Gleich". A sidebar on the left provides navigation options: "Updates", "Info", "Team" (2 members), "Files" (5 files), "Publications", "To-do", and "Notes". Under "Manager options", there are links for "Edit project", "Invite people to join", and "Learn about projects".

The "Publications" section is currently empty, with the message: "This project has yet no publications. Start a new publication". Below this, a section titled "How the publication process works..." outlines a three-step process:

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

The bottom right corner of the browser window shows the Zotero logo.

PUBLICATION

CREATE PUBLICATION - 2

The screenshot shows a web browser window with the address bar displaying `research.hub.purdue.edu/projects/nmcomp/publications/?action=start`. The page header includes the Purdue University logo and navigation links such as "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is also present.

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is identified as a "General project by David F Gleich". A sidebar on the left provides navigation options: Updates, Info, Team (2), Files (5), Publications, To-do, and Notes. Below these are "Manager options" including "Edit project", "Invite people to join", and "Learn about projects".

The central section is titled "Publications » New publication" and contains two main panels:

- Select Primary File(s):** A list of files is shown, with "gleich-flickr-2006.smat" selected. Below the list are "Browse..." and "Upload" buttons.
- Publication Primary Content:** A "Save & Continue" button is at the top right. Below, the selected file "gleich-flickr-2006.smat" is listed with the description "No description · Added at revision #1 (latest)". A "Download" button is shown with a red arrow pointing to a thumbnail of the file's download page.

At the bottom of the page, there is a footer with navigation links (Home, MyHUB, Projects, Get Started, Contact Us), copyright information (Copyright © 2011 Purdue University. All Rights Reserved.), and the text "Powered by HUBzero, a Purdue project". The Zotero logo is visible in the bottom right corner.

PUBLICATION

CREATE PUBLICATION - 3

The screenshot shows the Purdue University Research Repository (PURR) interface. The browser address bar displays the URL: `research.hub.purdue.edu/projects/nmcomp/publications/1002?section=content&step=supportingdocs&move=1`. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". The navigation menu shows "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the navigation menu.

The main content area displays the breadcrumb trail: "You are here: Projects > Networks and Matrix Computations > Publications > gleich-flickr-2006.smat". The page title is "Networks and Matrix Computations (nmcomp)" with a sub-header "General project by David F Gleich". A green notification banner states "Successfully started a new publication." and a "private project" lock icon is visible.

The left sidebar contains a navigation menu with the following items: "Updates" (1), "Info", "Team" (2), "Files" (5), "Publications" (1), "To-do", and "Notes". Below the menu, there are "Manager options": "Edit project", "Invite people to join", and "Learn about projects".

The main content area shows the "Publications" section for the dataset "gleich-flickr-2006.smat". The current version is "Version 1.0 (draft)". The "Content" tab is selected, and the "Supporting Docs" sub-tab is active. The "Select Supporting Docs" section is titled "OPTIONAL" and contains a list of files to be selected as supporting documents:

- Gleich.flickr.APlusAT.gif
- README-purr.md
- README_gleich-flickr-2006.pdf (checked)
- README_gleich-flickr-2006.txt (checked)

Below the list are "Browse..." and "Upload" buttons. A "Save & Continue" button is located in the top right corner of the "Publication Supporting Docs" panel. The "Publication Supporting Docs" panel shows the following items:

- README_gleich-flickr-2006.pdf (No description · Added at revision #1 (latest) · Edit)
- README_gleich-flickr-2006.txt (No description · Added at revision #1 (latest) · Edit)

A note at the bottom of the "Publication Supporting Docs" panel states: "Re-order items by dragging. Hit the button to save changes."

A help icon and text at the bottom of the "Select Supporting Docs" section reads: "Supporting documents can be any number of files relevant to the publication primary content. For example, here you can include a user guide for your app or course homework assignments. Supporting documents can be any number of files relevant to the publication primary content."

The footer of the page includes the Purdue University Libraries logo on the left, the navigation menu "Home MyHUB Projects Get Started Contact Us" in the center, and the copyright notice "Copyright© 2011 Purdue University. All Rights Reserved." and the Zotero logo on the right.

PUBLICATION

CREATE PUBLICATION - 3

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) interface. The browser's address bar shows the URL: `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=description&version=dev`. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". The navigation menu includes "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the navigation menu.

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is a general project by David F. Gleich. The page is currently displaying the "Publications" section for a dataset titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006". The version is "Version 1.0 (draft)".

The "Edit Description" form is visible, with the following fields:

- Title:** REQUIRED. The current title is "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006".
- Synopsis:** REQUIRED. The current synopsis is "Graph of the Flickr photo-sharing social network crawled in May 2006 returning a graph with 820,878 nodes and 9,837,214 edges. Dataset is distributed as a SMART file with README file with code to read file in Python and MATLAB."

The "Abstract / Description" field is also visible, with a rich text editor toolbar. The current text in the abstract is: "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."

The "Preview" section shows the rendered version of the abstract text.

The left sidebar contains navigation options: Updates (1), Info, Team (2), Files (5), Publications (1), To-do, and Notes. Below this, there are "Manager options": Edit project, Invite people to join, and Learn about projects.

The bottom right corner of the browser window shows the Zotero logo.

PUBLICATION

CREATE PUBLICATION - 4

The screenshot shows a web browser window with the URL `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=authors&version=dev`. The page is titled "PURR - Projects: Networks and Matrix Computations" and is part of the "Purdue University Research Repository | PURR". The navigation menu includes "Home", "myHUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". The breadcrumb trail indicates the current location: "You are here: Projects > Networks and Matrix Computations > Publications > Graph of Flickr Photo-Sharing Social Network ...".

The main content area is for the "Networks and Matrix Computations (nmcomp)" project, a general project by David F Gleich. The current publication is titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006" and is in draft status. The "Authors" tab is selected, showing the "Edit Authors" interface. The "Edit Authors" section includes a list of project team members: David F Gleich (dgleich) and Michael Witt (mwitt). The "Add" button is visible, and a note states: "Publication authors get selected from your current project team. Anyone you add as an author will also be added to your team as a project collaborator." The "Publication Authors" list on the right shows "1. David F Gleich, Purdue University" with an "Edit" button. A "Save changes" button is also present.

The footer of the page includes the Purdue University Libraries logo, navigation links, and copyright information: "Copyright© 2011 Purdue University. All Rights Reserved. Powered by HUBzero®, a Purdue project". The Zotero logo is visible in the bottom right corner.

PUBLICATION

CREATE PUBLICATION - 5

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) interface. The browser's address bar shows the URL: `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=gallery&version=dev`. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". Navigation links include "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is also present.

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is a "General project by David F Gleich". A sidebar on the left contains navigation options: "Updates" (1), "Info", "Team" (2), "Files" (5), "Publications" (1), "To-do", and "Notes". Below these are "Manager options": "Edit project", "Invite people to join", and "Learn about projects".

The main content area shows the "Publications" section for a dataset titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006". The current version is "Version 1.0 (draft)". There are tabs for "Content", "Description", "Authors", "Gallery", "Tags", "License", and "Notes". The "Gallery" tab is active.

The "Add Gallery Images" section is highlighted with a green border. It includes a "Save changes" button. Below the heading, it says "Select images from files:". A file named "Gleich flickr.APlusAT.gif" is selected and shown with a green checkmark. There are "Browse..." and "Upload" buttons. A note below states: "Screenshots in your gallery help make your publication page more discoverable and attractive to the user. Accepted image formats: PNG, JPG, JPEG, GIF, BMP."

The "Publication Gallery" section shows a preview of the selected image, "Gleich flickr.APlusAT.gif", with an "Edit" link.

The footer of the page includes navigation links: "Home", "MyHUB", "Projects", "Get Started", and "Contact Us". Copyright information reads: "Copyright© 2011 Purdue University. All Rights Reserved. Powered by HUBzero®, a Purdue project". The Zotero logo is visible in the bottom right corner.

PUBLICATION

CREATE PUBLICATION - 6

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) interface. The browser's address bar shows the URL: `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=tags&version=dev`. The page header includes the PURDUE UNIVERSITY logo and the text "Purdue University Research Repository | PURR". Navigation links include "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the header.

The main content area displays the "Networks and Matrix Computations (nmcomp)" project page. The project is described as a "General project by David F Gleich" and is marked as a "private project". The "Publications" section is active, showing a dataset titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006". The current version is "Version 1.0 (draft)".

The "Add Tags" section is the primary focus, showing a list of suggested tags and a "Picked" list. The "Picked" list includes the following tags, each with a green checkmark:

- Algorithms
- Data Mining
- Internet
- Matrices
- Representations of Graphs
- Social Network Analysis
- World Wide Web

The "Publication Tags" section shows the selected tags as buttons: "Social Network Analysis", "Algorithms", "Internet", "Data Mining", "Representations of Graphs", "World Wide Web", and "Matrices". A "Save changes" button is located at the top right of this section. The "Category" section is set to "Datasets", with a description: "A collection of research data".

On the left side of the page, there is a sidebar with navigation options: "Updates" (1), "Info", "Team" (2), "Files" (5), "Publications" (1), "To-do", and "Notes". Below these are "Manager options": "Edit project", "Invite people to join", and "Learn about projects".

The bottom of the browser window shows the Zotero extension icon.

PUBLICATION

CREATE PUBLICATION - 7

The screenshot shows a web browser window with the URL `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=license&version=dev`. The page header includes the Purdue University logo and navigation links like "My Account (mwitt)", "Logout", and "Report a bug". The breadcrumb trail reads: "You are here: Projects > Networks and Matrix Computations > Publications > Graph of Flickr Photo-Sharing Social Network ...".

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is a "General project by David F Gleich". A sidebar on the left lists navigation options: Updates (1), Info, Team (2), Files (5), Publications (1), To-do, and Notes. Below this are "Manager options" such as "Edit project", "Invite people to join", and "Learn about projects".

The central "Publications" section shows a "dataset 'Graph of Flickr Photo-Sharing Social Network Crawled in May 2006'" with tabs for Content, Description, Authors, Gallery, Tags, License, and Notes. The "License" tab is active, displaying the "Choose License" interface. It offers three options: "Standard HUB License", "CC0 - Creative Commons" (which is selected), and "Custom". A note states: "It is important that you provide a license for your publication stating your copyright and terms of use of your content."

The "Publication License:" section shows "CC0 - Creative Commons" selected in a dropdown. Below it, a green icon indicates that CC0 enables scientists, educators, artists, and other creators to waive their interests in their works. A checkbox is checked, indicating: "I have read the license terms and agree to license my work under the CC0 - Creative Commons license." A "Save changes" button is visible at the top right of this section.

The footer contains navigation links (Home, MyHUB, Projects, Get Started, Contact Us), copyright information ("Copyright© 2011 Purdue University. All Rights Reserved."), and mentions "Powered by HUBzero®, a Purdue project" and "zotero".

PUBLICATION

CREATE PUBLICATION - 8

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) interface. The browser address bar shows the URL: `research.hub.purdue.edu/projects/nmcomp/publications/1002/?section=notes&version=dev`. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". Navigation links include "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the header.

The main content area is titled "Networks and Matrix Computations (nmcomp)" and is a general project by David F. Gleich. The page shows a list of publications, with the current one being "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006". The "Add Release Notes" section is active, showing a text editor with the following content:

Version 1.0 Release Notes:

Citations:

P. Esfandiar, F. Bonchi, D. F. Gleich, C. Greif, L. V. S. Lakshmanan, and B.-W. On. "Fast Katz and Commuters: Efficient Approximation of Social Relatedness over Large Networks," in Algorithms and Models for the Web Graph, 2010.

F. Bonchi, P. Esfandiar, D. F. Gleich, C. Greif, and L. V. S. Lakshmanan. "Fast matrix computations for pair-wise and column-wise commute times and Katz scores," Internet Mathematics, vol. To appear., 2011.

Y. Jia, J. Hoberock, M. Garland, and J. Hart, "On the Visualization of Social and other Scale-Free Networks," IEEE Transactions on Visualization and Computer Graphics, vol. 41, no. 6, pp. 1285-1292, Dec. 2008.

The right side of the editor shows a preview of the citations. At the bottom of the editor, there is a note: "Version release notes usually include statements about version limitations and/or differences with previous versions, as well as any miscellaneous information that couldn't fit elsewhere."

PUBLICATION

CREATE PUBLICATION - 9

The screenshot shows a web browser window displaying the Purdue University Research Repository (PURR) website. The browser's address bar shows the URL: <https://research.hub.purdue.edu/projects/nmcomp/publications/?pid=1002>. The website header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". A navigation menu contains links for Home, my HUB, Browse Datasets, Projects, Get Started, and Contact Us. A search bar is located on the right side of the header.

The main content area displays the "Networks and Matrix Computations (nmcomp)" project page, created by David F Gleich. The page is categorized as a "private project". The "Publications" section is active, showing a draft version of a publication titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006". The draft version is labeled "Version 1.0 (draft)".

The publication details for the draft version are as follows:

Title:	Graph of Flickr Photo-Sharing Social Network Crawled in May 2006
Version label:	1.0 (default)
Version number:	1
Created:	02/22/2012 (59 minutes ago)
Created by:	Michael Witt (mwitt)
Status:	draft
Released:	11/30/1999 (1 decade ago)
URL:	http://research.hub.purdue.edu/publications/1002?v=dev

Below the publication details, there is a "What's Next?" section with three options:

- Review & submit to be published:** Publish your work for the world to see! A published version gets a Digital Object Identifier (DOI) and is included in publication listing and search.
- Review & archive:** Want to permanently store publication data without public release? You can save it as a "dark archive" with an ARK issued for reference.
- Review and share draft:** Draft is ready but you want to hold on with publication? Let others in your team know that draft is ready by finalizing and sharing it.

At the bottom of the page, there is a note: "Click on version label to edit."

PUBLICATION

CREATE PUBLICATION - 10

The screenshot shows a Mozilla Firefox browser window with the address bar displaying `https://research.hub.purdue.edu/projects/nmcomp/publications/1002/?action=post&version=dev`. The page title is "Networks and Matrix Computations (nmcomp)" and it is identified as a "General project by David F Gleich".

The main content area shows a draft of a publication titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006" (Version 1.0 Review). A yellow banner at the top of the content area asks "All looking good?" and provides buttons for "Save and share draft" and "Not now". Below this is a section for "Important notices" which states: "Access to the publication page and content is restricted to members of this project."

The page includes a sidebar with navigation options: Updates, Info, Team (2), Files (5), Publications (1), To-do, and Notes. Under "Manager options:", there are links for "Edit project", "Invite people to join", and "Learn about projects".

The publication details include the author "David F Gleich" from "Purdue University". The description reads: "Graph of the Flickr photo-sharing social network crawled in 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." A gallery shows a small image of a network graph. The "Primary Content - Files" section lists "gleich-flickr-2006.smat" (Selected file will be offered as a download) and "Supporting Docs" including "README_gleich-flickr-2006.pdf" and "README_gleich-flickr-2006.txt". The license is noted as "Licensed under Standard HUB License".

PUBLICATION

CREATE PUBLICATION - 11

PURR - Publications: Graph of Flickr Photo-Sharing Social Network Crawled in May 2006 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

PURR - Publications: Graph of Flickr P... x February 22 - databib - Purdue Librari... x +

purdue.edu https://research.hub.purdue.edu/publications/1002

My Account (mwitt) Logout Report a bug

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home my HUB Browse Datasets Projects Get Started Contact Us

You are here: Publications > Datasets > Graph of Flickr Photo-Sharing Social Network ... > About

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


By David F Gleich
Purdue University

Graph of the Flickr photo-sharing social network crawled in 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.

in Datasets

This is a final draft ready to be published. [Manage this publication.](#) from project Networks and Matrix Computations

About Supporting Docs Versions



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.

Tags [Algorithms](#) [Data Mining](#) [Internet](#) [Matrices](#) [Representations of Graphs](#) [Social Network Analysis](#) [World Wide Web](#)

Usage and feedback information is unavailable for this version of publication.

Version 1.0 (draft ready)
Draft finalized on 22 Feb 2012

Download (SMA)

Additional materials available (2)

SEE ALSO
No results found.

zotero

PUBLICATION

CREATE PUBLICATION - 12

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home my HUB Browse Datasets Projects Get Started Contact Us

You are here: Projects > Networks and Matrix Computations > Publications > Graph of Flickr Photo-Sharing Social Network ...

Networks and Matrix Computations (nmcomp)

General project by David F Gleich private project

- Updates
- Info
- Team 2
- Files 5
- Publications 1**
- To-do
- Notes

Manager options:

- Edit project
- Invite people to join
- Learn about projects

Publications » dataset "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006"

Versions » **Version 1.0 (draft ready)**

- Content
- Description
- Authors
- Gallery
- Tags
- License
- Notes

Version 1.0 (draft ready)

Title:	Graph of Flickr Photo-Sharing Social Network Crawled in May 2006
Version label:	1.0 (default)
Version number:	1 [all versions]
Created:	02/22/2012 (1 hour ago)
Created by:	Michael Witt (mwitt)
Status:	draft ready
Finalized:	02/22/2012 (54 seconds ago)
URL:	http://research.hub.purdue.edu/publications/1002

Your options

- Publish your work for the world to see! A published version gets a Digital Object Identifier (DOI) and is included in publication listing and search:
[Review & submit to be published](#)
- Want to permanently store publication data without public release? You can save it as a "dark archive" with an ARK issued for reference:
[Review & archive](#)
- Changed your mind? You can [cancel](#) and [delete](#) this draft version.

Home MyHUB Projects Get Started Contact Us

Copyright© 2011 Purdue University. All Rights Reserved.
Powered by [HUBzero](#), a [Purdue](#) project

zotero

PUBLICATION

CREATE PUBLICATION - 13

The screenshot shows a Mozilla Firefox browser window displaying the Purdue University Research Repository (PURR) website. The address bar shows the URL: <https://research.hub.purdue.edu/projects/nmcomp/publications/1002/?action=publish&version=default>. The page header includes the Purdue University logo and the text "Purdue University Research Repository | PURR". The navigation menu includes "Home", "my HUB", "Browse Datasets", "Projects", "Get Started", and "Contact Us". A search bar is located on the right side of the navigation menu.

The main content area displays the project "Networks and Matrix Computations (nmcomp)" by David F Gleich. The "Publications" section shows a draft titled "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006" (Version 1.0 Review). A yellow banner indicates that the draft is ready for publication, with the text "All looking good? Submit draft to be published" and a "Not now" button.

Below the banner, there is a section for "Important notices" with the following points:

- The publication will be reviewed by administrator before it goes live.
- A new official DOI (digital object identifier) will be issued with this version release. This means that **most version information cannot change** without publishing a new version! Please carefully review all text and chosen options on this page, and make sure that all your collaborators approve of this publication.
- This publication version is being released with **public** access: anyone can find the publication page via search and listings and view/download associated content.

The page also includes a "Manager options" sidebar with links for "Edit project", "Invite people to join", and "Learn about projects". The main content area includes a "Version 1.0" section with the title "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006" by David F Gleich, and a "Primary Content - Files" section listing the file "gleich-flickr-2006.smat" (Added at revision #1 (latest)).

The footer of the browser window shows the Purdue University Libraries logo on the left and the number "25" and the Zotero logo on the right.

PUBLICATION

CREATE PUBLICATION - 14

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home my HUB Browse Datasets Projects Get Started Contact Us

You are here: Projects > Networks and Matrix Computations > Publications > Graph of Flickr Photo-Sharing Social Network ...

Networks and Matrix Computations (nmcomp)

General project by David F Gleich

Updates 1
Info
Team 2
Files 5
Publications 1
To-do
Notes

Manager options:
Edit project
Invite people to join
Learn about projects

Publications » dataset "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006"

Versions » **Version 1.0 (pending approval)**

Content Description Authors Gallery Tags License Notes

Version 1.0 (pending approval)

Title:	Graph of Flickr Photo-Sharing Social Network Crawled in May 2006
Version label:	1.0 (default)
Version number:	1 [all versions]
Created:	02/22/2012 (1 hour ago)
Created by:	Michael Witt (mwitt)
Status:	pending approval
DOI:	10.5072/FK29888CS [->]
Submitted:	02/22/2012 (6 seconds ago)
URL:	http://research.hub.purdue.edu/publications/1002

What's Next?

Your publication has been submitted to administrator for approval to be published. You should expect to hear back from us in the next few days. Once approved, your publication will be included in the listings and search.

A DOI has been reserved for this release, and you can start citing your publication:

David F Gleich (2012), "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006." (doi:10.5072/FK29888CS).

Did you submit the draft in error and need more time? You can revert this version to draft »

26
Copyright © 2011 Purdue University. All Rights Reserved.
zotero

PUBLICATION

CREATE PUBLICATION - 15

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home Browse Content Projects Get Started Contact Us

You are here: Publications > Datasets > Graph of Flickr Photo-Sharing Social Network ... > About

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 SMAT 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 Wish List



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.

Download (SMA)

Additional materials available (2)

Version 1.1 - published on 22 Feb 2012
[doi:10.5072/FK25H7HM5](#) - cite this

0.0 RANKING

- 0 review(s) (Review this)
- 0 questions (Ask a question)
- 0 wish(es) (Add a new wish)

Share: [f](#) [t](#) [g+](#) ...

SEE ALSO

No results found.

PURDUE UNIVERSITY LIBRARIES

zotero

SERVICES & SUPPORT

DMP Overview

DMP Guides and Tutorials

DMP Q & A

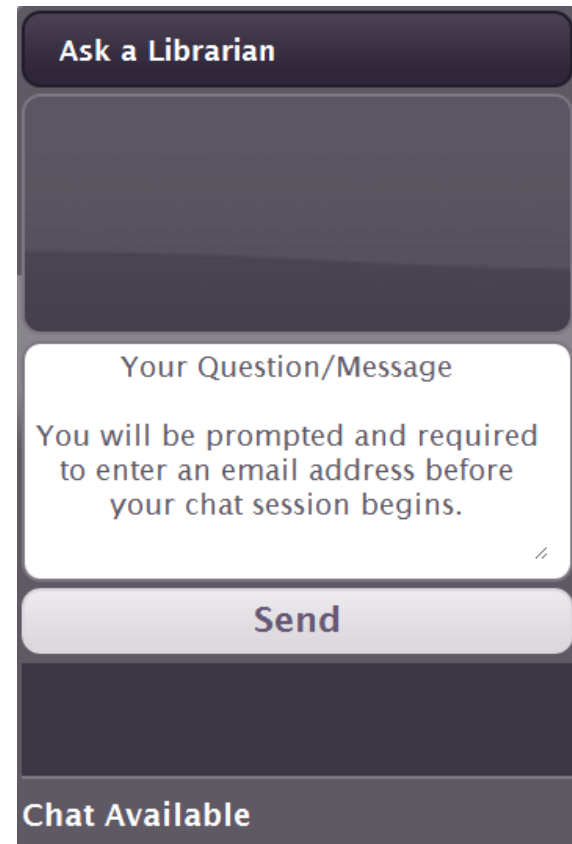
- Storage & Pricing
- Knowledge Base
- Contact Us

PURR SUPPORT / Q & A

Grant Information – SPS

Ticketing - I Tap

Data Reference – PUL



Ask a Librarian

Your Question/Message

You will be prompted and required to enter an email address before your chat session begins.

Send

Chat Available

PURR BY THE NUMBERS

430 Purdue grants include PURR in their DMPs

22 of these grants have been awarded

245 registered users

94 research projects

3 published datasets

NEIL ARMSTRONG HALL OF ENGINEERING

PURR: A RESEARCH DATA
CURATION SERVICE
MODEL USING HUBZERO

MORE INFO?

PURDUE
UNIVERSITY

LIBRARIES

Courtney Earl Matthews
Digital Data Repository Specialist

Research Department
Stewart Center, Room G50
504 W. State Street
West Lafayette, IN 47907-2058

Mailing Address:
Purdue University Libraries - RSCH
504 W. State Street
West Lafayette, IN 47907-2058

(765) 496-2770
courtneyearlmatthews@purdue.edu
<https://research.hub.purdue.edu>