

Marquette University

**e-Publications@Marquette**

---

Library Faculty Research and Publications

Library (Raynor Memorial Libraries)

---

5-1-2014

## Data Management for Research Grants: A Marquette Pilot Project

Rose Fortier

*Marquette University*, [rose.fortier@marquette.edu](mailto:rose.fortier@marquette.edu)

Lynn K. Whittenberger

*Marquette University*, [lynn.whittenberger@marquette.edu](mailto:lynn.whittenberger@marquette.edu)

Follow this and additional works at: [https://epublications.marquette.edu/lib\\_fac](https://epublications.marquette.edu/lib_fac)



Part of the [Cataloging and Metadata Commons](#)

---

### Recommended Citation

Fortier, Rose and Whittenberger, Lynn K., "Data Management for Research Grants: A Marquette Pilot Project" (2014). *Library Faculty Research and Publications*. 64.

[https://epublications.marquette.edu/lib\\_fac/64](https://epublications.marquette.edu/lib_fac/64)

# Data Management for Research Grants:

A Marquette Pilot Project

Rose Fortier – Coordinator of Digital Programs  
Lynn Whittenberger – Metadata Librarian





A Data Inventory Survey was sent out February 2012 to discover MU faculty's data preservation/access needs and desires.

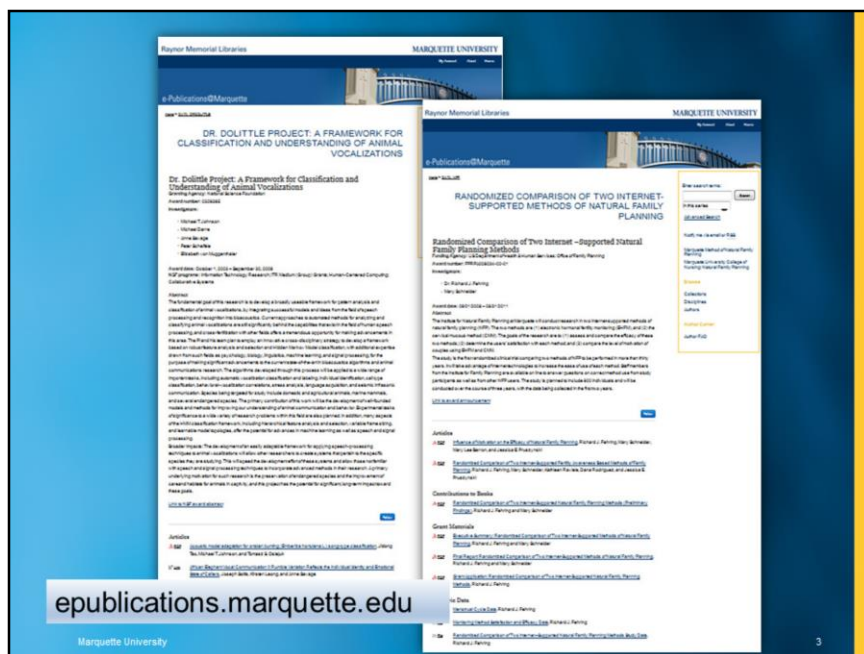
Questions were crafted to determine faculty research data habits, etc...

There was a 16% response rate from faculty.

Responses indicated an unfilled need for aid in managing research data on campus.

Responses also indicated a wide range of data management practices, output formats, and time ranges for data retention.

Decision was made to move forward and attempt to fulfill what the libraries could with current software and staff, hence the data pilot project was born!



## Success! Downloads as of 4/23/14

Title	File type	Downloads
Randomized Comparison of Two Internet-Supported Natural Family Planning Methods Study Data	Fulltext	65
Grant Application: Randomized Comparison of Two Internet-Supported Natural Family Planning Methods	Fulltext	36
Comparison of Abstinence and Coital Frequency between Two Natural Methods of Family Planning	Fulltext	25
Final Report: Randomized Comparison of Two Internet-Supported Methods of Natural Family Planning	Fulltext	14
Executive Summary: Randomized Comparison of Two Internet-Supported Methods of Natural Family Planning	Fulltext	13
Randomized Comparison of Two Internet-Supported Fertility Awareness Based Methods of Family Planning	Fulltext	11
Randomized Comparison of Two Internet-Supported Natural Family Planning Methods (Preliminary Findings)	Fulltext	4
Menstrual Cycle Data	Fulltext	4
Randomized Comparison of Two Internet-Supported Natural Family Planning Methods (Final Efficacy Results)	Fulltext	4
Monitoring Method Satisfaction and Efficacy Data	Fulltext	3
Randomized Comparison of Two Internet-Supported Fertility Awareness Based Methods of Family Planning	Additional	3
Monitoring Method Satisfaction and Efficacy Data	Additional	2
Monitoring Method Satisfaction and Efficacy Data	Additional	2
Randomized Comparison of Two Internet-Supported Fertility Awareness Based Methods of Family Planning	Additional	2
Menstrual Cycle Data	Additional	2
Influence of Motivation on the Efficacy of Natural Family Planning	Fulltext	1
Monitoring Method Satisfaction and Efficacy Data	Additional	1
Randomized Comparison of Two Internet-Supported Fertility Awareness Based Methods of Family Planning	Additional	1
Randomized Comparison of Two Internet-Supported Fertility Awareness Based Methods of Family Planning	Additional	1

Natural Family Planning pilot went live in December 2013. Downloads are from almost a five-month period from December 2013-April 2014. Dr. Dolittle numbers aren't as good as the Natural Family Planning numbers, but Dr. Dolittle isn't as far along as NFP. Most of the materials in Dr. Dolittle have only been available for a couple of months.

## From survey to pilot

- Consider how to structure the repository to include research data
- Populate the structure

These processes happened concurrently. As we were trying to contact faculty to invite them to participate in the pilot, at the same time we were reviewing National Science Foundation data management plans to help us determine how we would structure the data series.

Even though we talk about the two processes separately, they happened at the same time.

## Setting up e-Publications

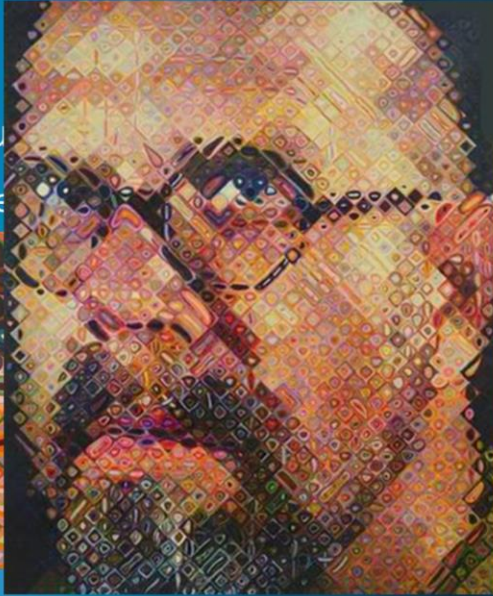
- Known issues with dataset preservation
  - Difficulties in describing data for use *and* discoverability
  - Data is less useful out of context
- Variety of data formats
  - Especially proprietary data formats

There are two sets of needs for describing data. The data needs to be described so people can find it, i.e. what the data is about. It also needs to be described so people can use it. The details about the files will be important so that users will be able to access the data and make sense of it and reuse it.

The sheer volume and variety of data formats is a big hurdle, especially when proprietary data formats are thrown into the mix.

## Setting up e-Pu

- Data is less use





## Creating a context for data

- Created a collection series based around a single research project or grant
  - Includes:
    - Project/grant proposal
    - Data
    - Supporting/interpretive data materials
    - Research output

In creating this context, we defined research output very broadly: publications, presentations, dissertations/theses, software, project websites, etc... Supporting data materials were anything that might help interpret the data, but that had no intrinsic research value, such as codebooks, data dictionaries, etc...

## What we came up with

- [http://epublications.marquette.edu/data\\_nfp/](http://epublications.marquette.edu/data_nfp/)
- [http://epublications.marquette.edu/data\\_drdolittle/](http://epublications.marquette.edu/data_drdolittle/)

## RANDOMIZED COMPARISON OF TWO INTERNET-SUPPORTED METHODS OF NATURAL FAMILY PLANNING

### Randomized Comparison of Two Internet-Supported Natural Family Planning Methods

**Funding Agency:** US Department of Health & Human Services: Office of Family Planning  
**Award number:** FPRPA00034-02-01

**Investigators:**

- Dr. Richard J. Fehring
- Mary Schneider

**Award date:** 09/01/2008 – 08/31/2011

**Abstract**

The Institute for Natural Family Planning at Marquette will conduct research in two internet-supported methods of natural family planning (NFP). The two methods are (1) electronic hormonal fertility monitoring (EHFM), and (2) the cervical mucus method (CMM). The goals of the research are to: (1) assess and compare the efficacy of these two methods; (2) determine the users' satisfaction with each method; and (3) compare the level of motivation of couples using EHFM and CMM.

The study is the first randomized clinical trial comparing two methods of NFP to be performed in more than thirty years. It will take advantage of internet technologies to increase the ease of use of each method. Staff members from the Institute for Family Planning are available on line to answer questions on correct method use from study participants as well as from other NFP users. The study is planned to include 600 individuals and will be conducted over the course of three years, with the data being collected in the first two years.

[Link to award announcement](#)

[Follow](#)

Articles

Enter search terms:

in this series

[Advanced Search](#)

Notify me via email or RSS

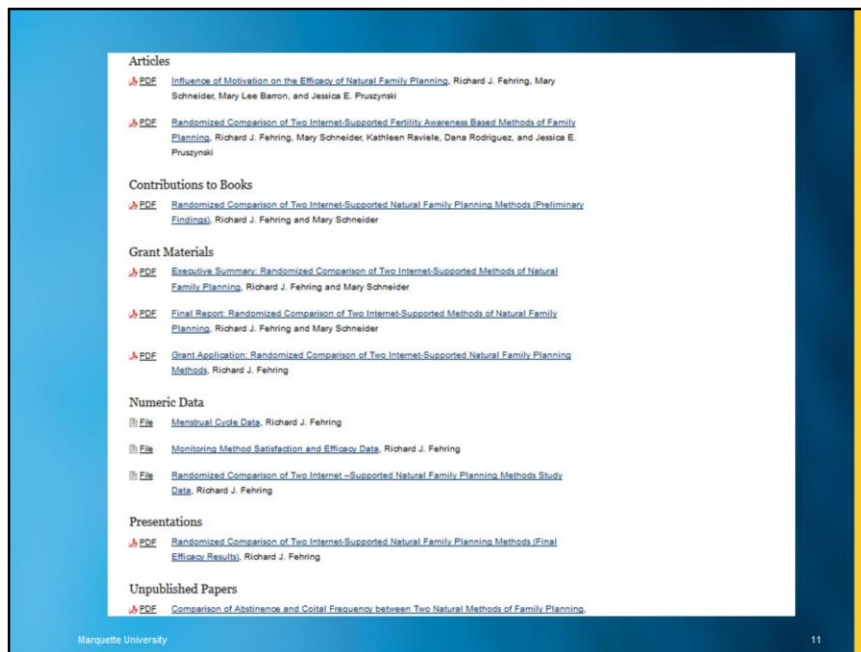
Marquette Method of Natural Family Planning  
Marquette University College of Nursing: Natural Family Planning

[Browse](#)

Collections  
Disciplines  
Authors

[Author Corner](#)

[Author FAQ](#)



Explanation of grant materials and numeric data:

Grant materials are the materials that were used in the grant application. They offer more context on the aims of the project, and its expected as well as actual outcomes. Numeric data contains the actual datasets. In Dr. Dolittle project, this is called Audio Data, to reflect the difference in data type.

## Digital Commons customizations for datasets

- Series description
  - Digital Commons allows customizable introductory text which we further customized using HTML



We included all of the information that we felt would be important for describing the project. Although it may seem repetitive, it's important to include this information at the series and document level to ensure that context is preserved throughout, no matter how a user navigates to the project.

This is how the series description looks when viewed on the site's back end.

e-Publications@Marquette


[Home](#) > [DATA\\_NFP](#)

## RANDOMIZED COMPARISON OF TWO INTERNET-SUPPORTED METHODS OF NATURAL FAMILY PLANNING

### Randomized Comparison of Two Internet-Supported Natural Family Planning Methods

**Funding Agency:** US Department of Health & Human Services: Office of Family Planning  
**Award number:** FPRPA00034-02-01  
**Investigators:**

- Dr Richard J. Fehring
- Mary Schneider

**Award date:** 09/01/2008 – 08/31/2011

**Abstract**  
The Institute for Natural Family Planning at Marquette will conduct research in two internet-supported methods of natural family planning (NFP). The two methods are (1) electronic hormonal fertility monitoring (EHFM), and (2) the cervical mucus method (CMM). The goals of the research are to: (1) assess and compare the efficacy of these two methods; (2) determine the users' satisfaction with each method; and (3) compare the level of motivation of couples using EHFM and CMM.  
The study is the first randomized clinical trial comparing two methods of NFP to be performed in more than thirty years. It will take advantage of internet technologies to increase the ease of use of each method. Staff members from the Institute for Family Planning are available on line to answer questions on correct method use from study participants as well as from other NFP users. The study is planned to include 600 individuals and will be conducted over the course of three years, with the data being collected in the first two years.

[Link to award announcement](#)

[Follow](#)

Articles

Enter search terms:

in this series

[Advanced Search](#)

Notify me via email or RSS

Marquette Method of Natural Family Planning  
Marquette University College of Nursing: Natural Family Planning

[Browse](#)

Collections  
Disiplines  
Authors

[Author Corner](#)

[Author FAQ](#)

Marquette University
13

This is how the series description looks on the public view of the site.

## Digital Commons customizations for datasets

- Object description
  - Each individual object within a series has some repeated metadata
  - Some metadata fields are suppressed to allow the objects to be collected together using Digital Commons' promised search feature

Custom field searching in Digital Commons. Bepress is working on making custom fields available for dedicated searching. At the moment, the fields are searchable as part of the whole, but not on their own.

# Influence of Motivation on the Efficacy of Natural Family Planning

[Download](#)

Available for download on  
Thursday, November 20, 2014

[Find in your library](#)

Included in  
[Maternal, Child Health and  
Neonatal Nursing Commons](#)

SHARE

[f](#)
[t](#)
[in](#)
[s+1](#)
[ma](#)
[+](#)

COinS

[Richard J. Fehring](#), [Marquette University](#) [Follow](#)

[Mary Schneider](#), [Marquette University](#) [Follow](#)

[Mary Lee Barron](#), [Saint Louis University](#)

[Jessica E. Pruszyński](#), [Medical College of Wisconsin](#)

**Grant Title**  
Randomized Comparison of Two Internet-Supported Methods of Natural Family Planning

**Document Type**  
Article

**Publication Date**  
11-2013

**Source Publication**  
MCN, The American Journal of Maternal/Child Nursing

**Source ISSN**  
0361-929X

**Abstract**  
Purpose: Mutual motivation is recognized as essential for effective behavioral methods of family planning. Few studies have studied this factor in family planning efficacy. The purpose of this study was to determine the influence of mutual motivation on unintended pregnancy rates of couples who used natural family planning (NFP) methods to avoid pregnancy.  
Study Design and Methods: Using an online taught NFP method, 358 women and (their male partners) indicated "how much" and "how hard" they wished to avoid

This is document level description as viewed on the public side of the site. The grant title is the only publicly visible piece of repeated metadata.

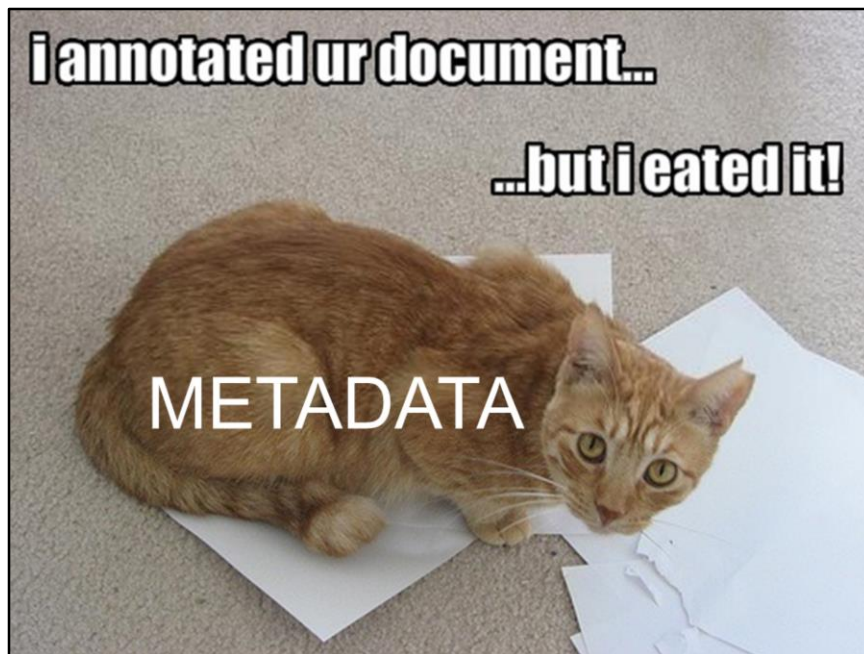
15



Metadata Field	Value
Title	Influence of Motivation on the Efficacy of Natural Family Planning
Authors	Richard J. Fehring, Marquette University; Mary Schneider, Marquette University; Mary Lee Barron, Saint Louis University; and Jessica E. Pruszyński, Medical College of Wisconsin
Streaming Media	- empty -
Media Format	flash_audio
Grant Title	Randomized Comparison of Two Internet-Supported Methods of Natural Family Planning
Principal Investigator	Richard Fehring
Funding Agency	US Department of Health & Human Services: Office of Population Affairs
Award Number	FPRPA006034-02-01
Document Type	Article
Format Type	- empty -
Access Instructions	- empty -
Publication Date	11-2013
Data file last updated	- empty -
Data collection date range	- empty -
Source Publication	MCN, The American Journal of Maternal/Child Nursing
Source ISSN	0361-929X

Marquette University 16

This is document level description as viewed on the back end of the site. Repeated metadata is grant title, principal investigator, funding agency, and the award number.



## Default fields

- Title, Author, Document Type, Publication Date, Source Publication, Source ISSN, Keywords, Embargo Date, Disciplines, Abstract, Document Rights and Citation of Original, Recommended Citation

Out of the box, Digital Commons comes with a certain set of default fields. We've added additional defaults to that list for e-Publications. We typically use most of these fields for general IR purposes; they're most often used with faculty publications. Each series is further customizable from this default set, which is what we had to do for the data series.

## Document Type

- Clustering of objects within the series is driven by this field
  - Developed list of types likely to appear most frequently
    - Can be modified if necessary

Document type maps to the Dublin Core Type field.

The field can be modified and we anticipate that further modifications will be necessary as time goes on and the project expands. Because there's such a variety of different output types from research grants, initially, it was important for us to identify commonly recurring types, so the series public display would be logically grouped.

## Document Type

- Default types
  - Article, Book, Contribution to book, Conference proceedings, Educational resource, Gallery guide, Popular press, Unpublished paper, Presentation, Technical report, Trade publication, Video

Currently, ePubs is populated by a variety of document types. It wasn't remotely complete enough for data types, and made no mention of any kind of data at all, except as "other."

## Document Type

- Added types
  - Audio data, Dissertation/Thesis, Grant material, Image data, Numeric data, Project website, Software, Supporting data documentation, Textual data, Video data, Other research output

Marquette University

21

We added the following document types.

To clarify what we mean by certain types:

Grant material: materials relating to the management of the grant, i.e. grant application materials, progress on grant, final reports.

Supporting data documentation: codebooks, lab notebooks, data dictionaries, etc...

Project website: Special Collections and Archives are archiving project websites so we are assured of stable and long-term access to the site, even after it no longer exists.

## Document description fields we added

- Grant information fields
  - Grant Title, Principal Investigator, Funding Agency, Award Number
- Data descriptor fields
  - Format Type; Access Instructions; Data file last updated; Data collection date range; Data collection location; Data rights, permissions, & privacy

Marquette University

22

Came to many of these fields through analysis of research DMPs, provided to us by ORSP. The DMPs were analyzed for different data and file formats, research output, and current data storage practices.

Grant information: added for context. Also, to add access points in searching (when bepress gets its act together) so that like information can be retrieved together.

Data descriptors: Format type-file extension of the data file. Access instructions-if the data files require any specific software or anything other than “click and play.” Data file last updated. Data collection date range. Data collection location.

Fields that aren’t populated aren’t displayed, so many of these fields aren’t active in our two pilot series, but we want to include this information if known/applicable.

## Data rights, permissions, & privacy

- Human subjects data
  - Statement explaining why data reuse is allowable (or not)
- Licensed data

Human subjects: not a problem for either of our two pilots, but something we had to consider for the Natural Family Planning grant. We didn't end up having a problem because the data was sufficiently anonymized, and the reuse of electronic files of data was spelled out in the consent form for the research.

Licensed data: The Dr. Dolittle grant made use of audio/video files that were obtained through licensing. Using those files would require gaining permission from the license holders. Were that permission granted, that information (along with any restrictions) would need to be recorded here.





## Identifying potential faculty partners

- Partnered with Office of Research and Sponsored Programs
  - Viewed a sampling of NSF grant applications, with particular attention to their data management plans (DMP)
  - Reviewed DMPs from a variety of other federally-funded grant applications

Marquette University

25

Used the grant applications and DMPs to determine likely candidates for the pilot. Particularly interested in NSF grants at the outset because those grants now have a mandate to provide access to the data generated by these grants.

Data management plans gave us an idea of the breadth of data types being produced, and their current plans for storing and managing the data.

Likely candidates for the pilot were identified and contacted, however the response rate was minimal (only 1 of the initial 3 contacted showed any interest).

To broaden the pool participants in the data pilot, we contacted Librarian liaisons asking if they knew of faculty with strong research production that might be willing to participate.

## Acquisition of Data

- Challenges in faculty participation for the pilot
  - Had to modify our original plans based on difficulties in faculty recruitment
- Possible reasons
  - Not all faculty are sold on the benefits of an IR
  - Faculty don't necessarily have easy access to their own data files

Marquette University

26

A possible solution to the challenge of faculty participation is the library's insertion in the data management process. This will/may make it easier to access data and accompanying files for inclusion in ePubs. The NSF requires a Data Management Plan, and with the federal government's open data policy, the day will come when federally-funded research will likely have a requirement for providing access to the data being produced by the research.

Both reasons for non-participation are common refrains in IRs.

## Faculty collaboration

- Digital Programs librarian, Librarian liaisons and Metadata librarian met faculty volunteers, to assess their data management needs.



Marquette University

27

For the faculty that agreed to participate, we met with them and asked a set of questions about their data.

For the pilot, we gathered together as much information on the grant beforehand, prior to contacting the faculty. We did the leg-work to identify and ingest the research output, for Dr. Johnson's project, based upon information on the NSF grants website and the project DMP. For Dr. Fehring's project, we had a project website to go from, but it included much less information than the NSF grant. By talking to Dr. Fehring, we were able to discern what materials we needed to procure from him.

## Data interview questions

- Project background
- What is the data's form and format?
- How is it stored?
- Who is the potential audience for the data?
- Who owns the data?
- Does it contain any sensitive information?
- Are there restrictions on the data?
- What publications, etc. resulted from the data?

Marquette University

28

Needs assessment includes: what and how much data faculty currently had, in which formats, its retention policy, if there were any privacy/confidentiality issues with the data, which associated files there might be (codebooks, data dictionaries). We also asked about associated research output such as journal articles, software, dissertations/theses from students working on the project, presentations, project webpages, etc.

Data restrictions are likely to include: Should it be embargoed for a specific period of time? Is the data more appropriate for access than for sharing?

## Ongoing concerns: What is data???



Marquette University

29

What do we consider as data for the purposes of data management?

What is the research output/input? Data means more than just datasets. It includes audio, video, images, observational data, interview data, surveys, instrumental readings, etc.. etc... etc...

## What is appropriate to include for dissemination?

- Raw data
- Processed/aggregated data
- Final data
- Proprietary vs. non-proprietary file formats

It depends!

Non-proprietary file formats are our preference for dissemination and preservation reasons. Preferred practice is to offer both proprietary formats with a non-proprietary version, however, this may not always be possible. Non-proprietary formats also allow for continued format migration down the road.



Partners for the pilot.

Partners for the project at a larger, on-going stage will likely include all of these, but will also need to include campus IT, and possibly Special Collections and Archives.



## Office of Research and Sponsored Programs

- One of the biggest campus stakeholders
- Worked closely with ORSP for pilot
  - They supplied us with grant proposals and Data Management Plans
  - Provided a list of potential faculty partners

ORSP will likely be the most important partner in recruiting data for the IR. They shepherd faculty through the grant-writing process, and they have the most reason to be interested with open data becoming a more prevalent requirement in research grants.

## Office of Research and Sponsored Programs

- Future plans:
  - Involve them in referring faculty to the library as a place to disseminate their research data
  - Include DMP template as part of NSF grant package materials

DMP template forces researchers to consider ways they could/should be managing their research data and serves as another pointer to the IR. It also gets them (hopefully) to organize their data in a way that makes it easier to ingest into the IR, should they choose to do so.

## Office of Research Compliance

- Need to make sure faculty has appropriate permissions to share data
  - IRB approval letter, consent form, waiver of consent (secondary analysis), IRB protocol
- Future plans: work with ORC to insert boilerplate language to human subject consent forms to allow for future data reuse

Partnership is mostly on a consultation basis to make sure that human subject data and privacy issues are being appropriately handled.

## Liaison Librarians

- Provider of background information about faculty
- Participated in the data interview

Benefits to liaisons: gets them in front of their faculty and increases their presence in their departments. Gives them a broader knowledge of what their faculty are involved in.

Liaisons are involved in more than one to one contact, will also be involved with planned training on DMPs, their requirements, and the implications.

## Liaison Librarians

- Future plans:
  - Point of continued contact with faculty
  - Advocate to faculty that ePubs is one place they can satisfy the data requirements of their granting agency
  - Use subject knowledge to help contextualize data



Even with the pilot under our belts, there are still issues and questions that we need to resolve.

## Access vs. storage vs. archiving

- Librarians' perceptions of these terms can be \*very\* different than others on campus
- True data archiving is out of our reach at the moment
  - Our capabilities will allow short to medium-term access
  - Long-term access/archiving is beyond our capabilities

## Data sharing vs. data access

- Not all data is appropriate to have as an immediate download
- How to handle data sharing when a potential user must contact the data creator?

We haven't had to deal with this question yet, but it's something that will come up. If the data is only available upon request to the researcher, how do we handle continued access? What happens when a researcher goes on sabbatical, leaves the university, retires, or dies? Here is an opportunity to involve Special Collections and Archives... maybe.



## Technical limitations

- Digital Commons was not created for research data
  - Workflow complications
  - A lot of work goes into creating the context and framework for these series
    - How much of the process can be standardized/systematized?

Each data series will likely require more customization than a typical faculty publication series in the IR. The small size of the pilot means that while we are now more aware of the issues and some of their solutions, it wasn't big enough to get a handle on what a workflow might look like on a larger scale.



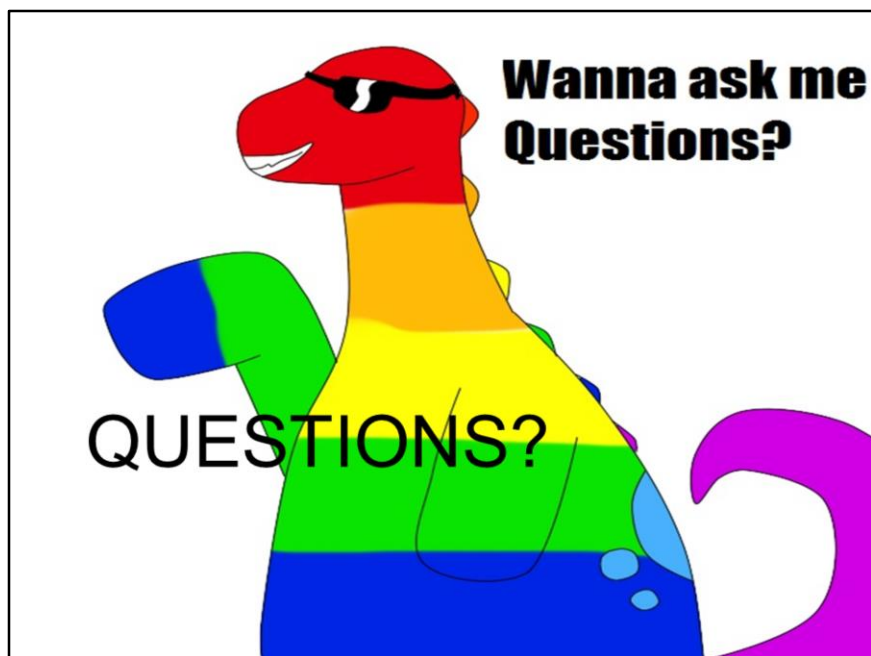
# FUTURE PLANS

- Expand the pilot
  - Recruit faculty from humanities and/or social sciences disciplines
- Data management and DMP training for faculty, in collaboration with ORSP and library liaisons

All of this takes time and in order to take on a systematic approach, serious investment will be required. What form that investment takes is unclear, but to continue in anything other than an ad hoc fashion will be extremely difficult without that investment.

We need to be ready to respond to federal government's open data initiative. The work we've done puts us in a better position to do so, but we are by no means ready to take on large-scale data preservation tasks.

- We need to create internal documentation based on what we've learned
  - Data dictionary
  - Workflows
  - Procedures



- Rose Fortier
  - [rose.fortier@marquette.edu](mailto:rose.fortier@marquette.edu)
- Lynn Whittenberger
  - [lynn.whittenberger@marquette.edu](mailto:lynn.whittenberger@marquette.edu)