TODAY’S EVENT BEGINS AT 3 P.M. ET

An ECAR Working Groups Webinar • February 7, 2017

A Focus on Research Data

Kim Owen
Program Manager for Research and Education Network Resources
North Dakota State University

Ralph Zottola
CTO, Research Computing
University of Massachusetts Central Office
When faculty evaluate their institution’s support for data-intensive research, they are generally satisfied with the provision of research computing technologies (42%), but many also claimed that they do not have the IT cyberinfrastructure resources and support needed to effectively pursue their research (44%).

ECAR Study of Faculty and Information Technology, 2015
Today’s Speakers

Ralph Zottola
CTO, Research Computing
University of Massachusetts Central Office

Kim Owen
Program Manager for Research and Education
Network Resources
North Dakota State University
Considerations for Research Data Sharing

Setting the Stage for Greater Openness

Ralph Zottola, University of Massachusetts Central Office
What Is Research Data?

Just about anything, including the...
Why Share Research Data?

- Prospect of new discoveries
- Improved research transparency
- Greater interdisciplinary collaboration
- Increases the value of public funding by allowing scientists to ask new questions of extant datasets
- Enabling “big data” projects

Since most university research is publicly funded, should the public have access to the data?
Stakeholders

- Researchers
- Libraries and Repositories
- Higher Education Institutions
- Funding Agencies
- Publishers
- General Public
Barriers and Considerations

- Social and Cultural Influences
- Data Ownership
- Legal Compliance
- Funding
- Data Governance
- Information Technology Considerations
- Data Management
- Training and Education
ADNI - A Model Use Case?

- Alzheimer’s Disease Neuroimaging Initiative
- Public-private partnership with focus on collaboration and rapid data sharing.
- Multiple repositories store the data redundantly at institutions worldwide.
- Uniquely, no single researcher or institution owns the data. Both data and ideas are shared.
- The approach allowed the study to be conducted at a scale not possible by individual researchers and research organizations.
- The study has been used in over 1,200 research papers.
- To date, more than 8,500 researchers have sought access.
It Will Take a County, err, Country!

- Any path forward needs to involve all stakeholders
- Recent requests for comments from the NIH, NLM, and others
- New policies (national and local) needed?
- Different approaches?
  - Blockchain
  - Github
- Are we ready to disrupt the status quo? Or put another way, can we really stop progress? Is this a conversation at your institution yet? If no, why not?
You can have data without information, but you cannot have information without data.

Daniel Keys Moran, Programmer and Science Fiction Writer
Developing an Institutional Research Data Management Plan Service and Research Data Storage

Kim Owen
Program Manager, Research and Education Network Resources
North Dakota State University
kim.owen@ndsu.edu
February 7, 2017
Why Data Management?

1999  Public access to data from federally funded research

2013  Increasing access to the results of federally funded scientific research

Objective of this paper

Provide insights on developing data management planning services at research higher education institutions.
What the Guide Is and Isn’t

What this guide IS...

• Provides overview of common DMP service elements
• Time-sensitive – a moving target
• A broad synopsis of current practices and models
• Only one component of a larger, formalized process addressing research data management infrastructure

What this guide IS NOT...

• A comprehensive check-list intended to cover all things for all research grant proposals
• Not a guarantee for successful grant awards
The Purpose of Data Management Plans

• Communicate (findings, hypotheses, insights, replicability)
• Organize (nomenclature, terminology, disciplines)
• Build communities toward collaboration
• Document, manage, resolve controversies
• Establish precedence
• Be trustworthy
• Be reproducible
• Perturb assumptions and methods
The Benefits of Data Management Plans

• Save time
• Increase research impact
• Ensure long-term ability to preserve fragile data sets
• Organize and categorize data for efficient access, analysis, queries, etc.
• Support sharing and open-access
• Focus on data sharing as an objective of investigation
• Support data-intensive discovery across disciplines
• Promote verification and replication of research analysis and findings

UNCOMMON THINKING FOR THE COMMON GOOD
The Big Picture:
Common Data Lifecycle Stages
Developing a DMP Service

• Organizing
  o Institutional culture / Organizational makeup / Geographic dispersion

• Funding
  o Staff, resources, core service of the institution

• Service Delivery / Instructions / Guidance
  o Central IT, Faculty/Deans, Legal Counsel, Library, Research Computing, Office of Sponsored Programs, Others

• Critical skill sets: disciplines / data
Quality of the Plan

Institutional Vetting of the DMP improves quality:

• Clarifies roles, responsibilities and risks
• Storage requirements
• Curation needs
• Commitment of technology resources
• Viability of network capabilities
• Minimizes risk to the researcher and institution
Researcher Resources

• Storage options
• DMP products vs. in-house templates
• Research administration management software
• Institutional / discipline-specific repositories
• Consulting
Storage Environments Throughout the Data Lifecycle

- Central IT storage
- Research storage
- Divisional/departmental
- Cloud
In Summary

• Identify a model for local administration of research data management services
• Provide resources for convenient access to researchers during the research data lifecycle
• Designate one or more dedicated staff to serve a range of consulting needs
Research Data Storage

A Framework for Success

ECAR Working Group

July 2014
Objective

Create a framework that helps higher education institutions establish and evaluate research and data storage efforts on their campus.
Framework

- Governance, Policy and Oversight
- Data Life Cycle
- Use of Third-Party (Cloud) Providers of Storage
- Communication and Collaboration
- Infrastructure
- Operation and Management
All founded on communication ...

Conversations about research data storage planning and management are critical across departments.
Anticipated outcomes

- Starting the conversation at your institution
  - Come to a common understanding of what research data storage maturity means across your institution.
  - Address accountability issues with stakeholders.

- Strategizing
  - Use or modify matrix components to identify your institution in the continuum of higher ed institutions.
  - Identify specific areas that, when addressed, will help your institution move up to the next level.
  - Use as leverage when justifying needs in planning and budgeting.
Understanding the Maturity Classifications

Leading Edge
Ongoing communication, collaborative planning and development across the institution: review and piloting of new tools and resources is a joint venture, especially among those in academics, research, the libraries and information technology.

Industry Standard
A general awareness exists across the institution but developing critical practices may be viewed as unnecessary to address randomly recurring issues; minimal or sporadic communication occurs when needed. ‘Hot spots’ exist across campus where leading edge activities are more frequent.

Lagging
Issues are addressed in isolation and with little consideration to the bigger picture of the institution’s goals; communication is minimal.
Example: Compliance with Research Grant Data Storage Requirements

<table>
<thead>
<tr>
<th>Leading Edge</th>
<th>Industry Standard</th>
<th>Lagging</th>
</tr>
</thead>
<tbody>
<tr>
<td>A formal process exists to understand and comply with storage requirements based on pending grant proposals and awards, as does a formal mechanism for the institution to respond to those requirements effectively and efficiently.</td>
<td>Compliance for research grant data storage is considered and addressed when new grants are awarded or new agency requirements are announced.</td>
<td>Compliance is addressed independently by grant PIs, often with little communication or engagement with centralized IT and other departments charged with maintaining the institution’s practices and policies.</td>
</tr>
</tbody>
</table>
In Summary

• Research data growth and needs continue to expand at a rapid pace
• Use framework to initiate and expand discussion surrounding this topic at your institution
• Categories and maturity classifications invite engagement from diverse disciplines and support entities
Questions?
Resources

• ECAR Research Data Sharing Working Group - Project Description
• Research Data Storage: A Framework For Success
• Developing An Institutional Research Data Management Plan Service
• ECAR Study of Faculty and Information Technology, 2015
• EDUCAUSE Data Governance and Chief Data Officers Constituent Group
• EDUCAUSE Research Computing Constituent Group
• ECAR Working Groups Home
Give Us Your Feedback!
Please take the time to fill out this brief evaluation

Save the Date!
Tuesday, March 7 • 3-4 p.m. ET
Focus on Administrative Data

Tuesday, April 4 • 3-4 p.m. ET
Focus on Academic Data

Find out more at:
www.educause.edu/ecar/ecar-working-groups