Open Data
Open Access

Pennsylvania State University
What is Open Data?

• Data that IS, MUST BE, or WANTS TO BE publicly accessible
• Possibly *used by* your researchers
  • May come with simple compliance terms — e.g. “must cite”
• Possibly *created by* your researchers
  • Publishers may mandate that data be made publicly available
  • Researchers may believe in Open Data and want to make it available

• Tools for Open Data, Issues around Open Data...
Open Data Repositories (1)

- Institutional Open Data Repository (Optional)
- Penn State has Data Commons
  - Focuses on geospatial data and Penn State researcher data
- Many domain-specific Open Data repositories
- Many generalist Open Data repositories
- Many Open Data repositories identified at [www.re3data.org](http://www.re3data.org)
Open Data Repositories (2)

- Why not just use others’ repositories?
- Institutional awareness of institutional research data
  - Where did all your research data go? No one knows
- Support for the Open Data movement
  - Consider providing a domain-specific repository for others
- Convenience
  - When in doubt, your researchers have a place to put their data
Issues Around Open Data

- Faculty may have competing priorities for their limited time
  - Data preparation and curation takes time out of next research project
  - Data publication not a contributor to P&T evaluation

- Compliance when using “must cite” Open Data
  - Compliance with terms around Open Data almost certainly in the hands of the PI – likely no institutional oversight

- Technical and Administrative Issues
  - May need massive storage spaces
  - May need high-speed data transfer mechanisms
  - May need data curation staff
What is Open Access?

• The idea that a university’s research output should be accessible to the public
  • Especially at an APLU institution
  • Especially the output of publicly-funded research

• Tools for Open Access, Policy for Open Access, Issues around Open Access...
Research Repositories (1)

• Institutional Research Repository (Optional)
• Penn State has ScholarSphere
  • Based on open source Sufia and Samvera projects
• For Penn State publisher-accepted research
  • Pre- and Post-peer-review versions
  • Where no copyright restrictions are in place
• For the data that goes with the published research
  • But has severe restrictions on the amount of data that can be uploaded
  • Not really an Open Data solution, per se
• For working papers, technical reports, conference papers, presentations
Research Repositories (2)

- Many public research repositories available
  - ROAR – Registry of Open Access Repositories
  - OpenDOAR – Directory of Open Access Repositories
- Why not just use others’ repositories?
- Institutional awareness of institutional research
  - Where did all your research go? No one knows
  - Your own repository gathers “all” your research in one place
- Convenience
  - When in doubt, your researchers have a place to publish their research
Open Access Policy

• A means of increasing participation among the faculty
• A means of ensuring your research is accessible to the public
• A means of ensuring publicly funded research serves the public

• Penn State policy formation in progress
  • Faculty Senate resolution passed in April 2015
  • Policy based on 2008 Harvard policy
    • Institution retains nonexclusive, irrevocable right to distribute their scholarly articles for any non-commercial purpose
    • Policy waived automatically when publishers object
Issues around Open Access (1)

- Faculty may have competing priorities for their limited time
  - Preparation and curation takes time out of next research project
  - Open Access compliance not a contributor to P&T evaluation

- Publishers may prefer not to participate
  - Publishers may decline to waive exclusive copyright
  - Publishers may require a fee to waive exclusive copyright
    - Fee-reimbursement program for researchers should be instituted
Issues around Open Access (2)

• Policy Compliance
  • Compliance with Open Access policy (if there is one) almost certainly in the hands of the PI – likely no institutional oversight

• Technical Issues
  • Beyond self-reporting, need an administrative means of knowing when research becomes available for Open Access
  • If you have no institutional repository, you should at least have an institutional database that links to your university’s research online
  • Must have a robust back-end to the institutional repository
  • Should integrate with research administrative information systems, P&T evaluation systems, public-facing research activity reporting systems
Questions?

• Greg Madden

Senior Advisor for Research Computing and Cyberinfrastructure
Office of the Vice President for Research
Pennsylvania State University
gem19@psu.edu