The 21st Century Research Infrastructure and SHARE: Challenges and Opportunities

Kelvin K. Droegemeier
Vice President for Research
Regents’ Professor of Meteorology
University of Oklahoma
The Notion of SHARE-ing

- Humans are naturally selfish, even as infants, and not sharing is an innate behavior we exhibit at a very early age!
The Notion of Sharing

• The academy encourages collaboration but still tends to evaluate and reward individual achievement – thus reinforcing at least caution in sharing
The Research Spectrum

Transdisciplinary

Interdisciplinary

Multi-Disciplinary

Disciplinary
Share will help (force) us to more quickly and effectively develop common vocabularies and build trust among researchers across disciplines.
The Value Proposition of Open Data Access

- In Research
  - Replication/reproducibility of previous results
  - New interpretation and analysis to scale for heterogeneous data sets – an open market approach
  - Opportunities for collaboration, probably numerous and multi-disciplinary
  - Follow-on studies and thus faster progress
  - Identification of unknown problems
  - Greater accountability and responsible conduct of research
  - Research experience in K through 16
  - Credit for gathering/generating/making available data
The Value Proposition of Open Data Access, continued

• In Teaching
  – Online and engaged **learning**

• In Economic Development
  – Corporate **R&D** to use in **analytics**
The Challenges of Open Data Access

- Misuse of data owing to misunderstanding → battles about results
- Burden on researchers to explain data, deal with access problems
- Applying analytics to mine vast amounts of information and make sense of it
- Misappropriation of credit
View from a Vice President for Research, Point 1

- **Key Point #1: Facilitating Research With Data**
  - A means to tackle some of the most compelling intellectual challenges at the intersections of multiple disciplines
  - It’s not only about providing access but also helping ensure **EFFECTIVE** use of data – **which is not automatic**!
  - **The institution must help:** bring people together, stimulate conversations, bridge language barriers, build trust, guide thinking, provide support
View from a Vice President for Research, Point 1, cont.

• **Key Point #1: Facilitating Research With Data**
  
  – Library has a unique role to play – a *renaissance* as the intellectual commons of our campuses
  
  – As IT has opened new doors and brought people/disciplines together, so can the “data challenge” if we handle it properly!

  • Especially critical for engagement of social sciences and the humanities
• **Key Point #2: Providing Credit and Support**
  
  – System (but importantly also a philosophy) for giving credit to faculty for generating, maintaining, and provisioning data (similar to how IP has been added to portfolio)
    
    • Provost, Deans, Tenure Committees, Senior Faculty
  
  – Building data stewardship into research metrics aka citations, impact factors, etc
  
  – Creation of persistent identifiers/tags (EZID, DataCiteConsortium) – but also WHAT credit means
  
  – Creation of an indirect cost component for data and compliance
Key Point #3: Logistics and Cost

- Coordination and consolidation of data management approaches across the institution: Provost, Library Dean, CIO, VPR
- Appropriate cyberinfrastructure, security, systems-level approach
- Integration into the broader academic ecosystem
- The strategies in which we invest today may be quite different in a short time – shifting sands
  - Division of responsibilities (local and national)
  - When will the dust settle regarding policies?
  - Unity versus diversity in approaches
Assessing Impact

- **Assessing impact** is critical if we are to understand the real benefits (and possible drawbacks) of open data access
  - We must link assessment to value proposition points described earlier and be open to new ideas of measuring value (e.g., drawing in underrepresented groups to STEM fields, creation of new domains such as digital humanities)
- Assessment also is critical for **modifying policies** to create even greater benefits and address drawbacks
Blue Sky Thoughts: Caution – A Forecast from a Meteorologist!

• Open data will generate **completely new models for collaborative interaction** – a new “social media” in research centered around **sharing**, not simply **communicating**. This will lead to **new tools and approaches** (e.g., Tableau)

• **New disciplines** likely will arise when researchers from one discipline start “playing around” with data from others, and these new perspectives will reveal new insights, **solve longstanding problems** and raise new questions. The “Fourth Paradigm.”
Blue Sky Thoughts: Caution – A Forecast from a Meteorologist! continued

- The entire notion of a “publication” will change, leading to open-ended online research release events in which new analyses and data will continually be added to previously reported outcomes, creating branches and forks in time, with other disciplines linked in a virtual intellectual commons.

- Access to data, and tools for analysis, will enhance interest and exploration by children and offer greater opportunity to increase participation by traditionally underrepresented groups.
The End Game for Research

• SHARE will **enable** NEW and **better** ways of doing old things + lots of NEW things

- Reproducibility (repeat/verify)
- Extensibility (extend/expand)
- Collaborability (across)
- Creatibility (frontier)