Repositories, Resources, and Relationships

- Context: Unexpressed relationships exist between different forms of research output held in different repositories
  - Data → Publications, Software ← Data, Publications ← Software
- Approach: Demonstrate how repository resource interconnections can be established and maintained
- Scope: Proof-of-concept implementation of connection between repositories provided and managed by NCAR
  - OpenSky repository – peer-reviewed & grey literature
  - Earth System Grid (ESG) – scientific data, models, and software
- Goals:
  - Enable researchers who deposit resources in one system to initiate deposit of related resources in the other
  - Links between the related resources held in the two systems will be exchanged and made visible in the respective systems

Stakeholders and Key Tasks

- Cross-linking implementations need to support multiple stakeholder tasks
  - Data set providers can upload publications into OpenSky when depositing data sets in ESG
  - Publication authors can create new datasets in ESG when depositing documents into OpenSky
  - Repository content curators are notified when new submissions are initiated from external systems (e.g. OpenSky curators are notified when a publication submission comes from the ESG, and vice versa)
  - Research content creator is notified of completed process and may share deposited resources with colleagues via persistent URIs
  - Research output consumers (either human or machine) can use the repositories to find resources and their relations

Technical Components

- Technical components under investigation to support repository cross-linking
  - Notification service
    - Automated emails
    - File deposits via SWORD (Simple Web-service Offering Repository Deposit)
  - Data model for metadata exchange
    - Metadata inheritance and relationship types
    - Information necessary for maintaining relationships between resources over time as updates occur
  - Effective UI displays that make the cross-repository relationships transparent to users
  - Coordinated security and authentication to smooth cross-repository notification and deposit workflows

Discussion points

- Datasets and publications have varying creator and consumer norms: how might repository cross-linking bridge between different communities of research practice?
- Data and publication repositories have widely different metadata and other technical requirements – this a challenge for system interoperability and cross-platform discovery. Are there common or third system tools that might be leveraged to facilitate exchange?
- We have discovered that the interconnection between publications and data is more like a web than a series of one to one relationships. What terms might be useful to include in a typology for describing relationships between datasets and publications?

Acknowledgments: Thanks to our project collaborators: Don Middleton, Eric Nienhouse, and Jonathan Ostwald. This project is supported by NSF grant # 1449668. NCAR is funded by the U.S. National Science Foundation.