Envisioning and Enacting a Coherent Organization-Wide View of Data

Karen S. Baker¹, Matthew S. Mayernik², Cheryl A. Thompson¹, Eric Nienhouse³, Steve Williams⁴, Steve Worley³

¹University of Illinois Urbana-Champaign, Graduate School of Library & Information Science
²National Center for Atmospheric Research, ³NCAR Library, ⁴Computational & Information Systems, ⁵Earth Observing Laboratory

### Background

Amidst discussions of open data and national coordination, there is also need for organization-centric discussions about scientific data and data management. At the National Center for Atmospheric Research (NCAR) and its parent organization the University Corporation for Atmospheric Research (UCAR), data collections began in the mid-1960’s with individual labs and divisions providing their own data tools and services.

### The Challenges

- Aligning independent data management efforts

<table>
<thead>
<tr>
<th>Differing Data Assemblies</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Domain Repository – Data sets</td>
</tr>
<tr>
<td>✔ Domain Repository – Projects</td>
</tr>
<tr>
<td>✔ Institutional Repository</td>
</tr>
<tr>
<td>✔ Instrument Collections</td>
</tr>
<tr>
<td>✔ Code Repository</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Initial state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intended state</td>
</tr>
</tbody>
</table>

### Four Collective Endeavors

The NCAR/UCAR case provides an example of envisioning and enacting community-building activities and organization-wide communication over the past 20 years to increase coordination and develop more coherence of data services across NCAR/UCAR. These data activities are embedded within an evolving IT infrastructure with each coordination effort constrained by very different technologies.

#### 1995 IITA

- **Information Infrastructure Technology Applications**, a new organizational unit
  - **Scope**: Enterprise IT organizational infrastructure
  - **Outcome**: IT strategic plan for organization

#### 2001 DMWG

- **Data Management Working Group**, a cross-cutting group
  - **Scope**: Scientific computing interoperability of existing systems
  - **Outcome**: Community Data Portal

#### 2010 CDC

- **Committee on Data Citation**, an ad hoc group
  - **Scope**: Scientific computing research digital services
  - **Outcome**: NCAR-wide DOI Services coordinated with documentation via technical reports

#### 2014 DSET

- **Data Stewardship & Engineering Team**, a formally recognized, cross-cutting group
  - **Scope**: Scientific data stewardship
  - **Intended Outcome**: Common discovery of resources

### Conclusion

This study informs current data sharing efforts by focusing on internal alignment and coordination within organizations. The evolution of an intra-organizational understanding of data – beyond the individual research lab or project – is a critical enabler of increased data integration and discovery across organizations.

This work demonstrates alignment not as a one time event but as an iterative, ongoing process of envisioning & enacting.

### Lessons Learned to Date

- Balance top-down managerial input with bottom-up interests and capabilities
- Recognize and build on peer developments
- Establish goals based on user needs
- Leverage standards and existing IT systems
- Achieve success through well-ordered work plans that emphasize sustainability from the beginning
- Document and communicate the process & the outcomes

### Acknowledgements

This work is supported by the Data Curation Education in Research Centers (DCERC) IMLS Award #RE-02-10-0004-10.

---

January 2015