UCAR/NCAR Data Citation Initiative

Matthew Mayernik
Research Data Service Specialist
NCAR Library/Integrated Information Services (IIS)
National Center for Atmospheric Research (NCAR)
University Corporation for Atmospheric Research (UCAR)

Bridging Data Lifecycles Workshop, UCAR, April 5, 2012
Interest in Data Citations

• Federal agency emphasis
• Federation of Earth Science Information Partners (ESIP)
  – NCAR is a member
  – ESIP has data citation guidelines
• 2009 AGU position statement:
  “The scientific community should recognize the professional value of such [data] activities by endorsing the concept of publication of data, to be credited and cited like the products of any other scientific activity, and encouraging peer-review of such publications.”
• AMS Board on Data Stewardship
NCAR/UCAR Data

Climate model output data

Longitudinal time-series data

Observational data from field studies

All images: copyright University Corporation for Atmospheric Research
Motivation for Data Citations

• Understand use and impact of data
  – Measurements of data use
  – Give scientists and data centers credit for producing, managing, and curating data
  – Metrics requirements as an FFRDC

• Connecting data and scholarship

• Increase transparency of data and science
What to Cite?

NOTE: PLEASE CITE ORIGINAL REFERENCE WHEN USING THIS DATA!!!!!
NCAR Data Citation Initiatives

1. Technical

2. Policy/procedural

Image copyright University Corporation for Atmospheric Research
1. Technical: DOIs for Data

- Working to assign Digital Object Identifiers (DOIs) via EZID
- Integrating DOIs into existing data systems
  - Building scripts that use EZID API to create/update DOIs
  - Deciding where to include DOI assignment in data processing workflows
  - Updating databases, displays, etc.
2. Procedures for Data Citations

- Data archives
  - Who should be assigning DOIs
  - How to assign DOIs to resources
  - How to register DOI metadata
  - How to ensure that DOIs will be maintained
  - How to provide recommended citations to users

- Outreach to data users
  - Make citation information easy to find/use
  - Proactively inform users about citations and DOIs
Progress

• Established membership with EZID
• White paper draft - DOI and citation recomm.
• Community building
• Identifying and organizing pilot projects
  – Data sets
  – Software
  – NCAR publications
Citation Challenges

1. Diversity

2. Granularity

3. Version Control

4. Maintenance Over Time
Diversity

- Diversity in:
  - Data sizes and formats
  - Technical architectures
  - Data uses and users
  - DOI metadata needs
    - Desire for standardization
    - Need for flexibility
Citation Challenges

1. Diversity
2. Granularity
3. Version Control
4. Maintenance Over Time
Granularity

• What is the right granularity at which DOIs should be assigned?
  – How do you manage resources internally?
  – How do you display resources to users?
  – What do you want users to cite?
  – What burden are you putting on users?

• Tension between ease for users and desire for transparent referencing
Citation Challenges

1. Diversity

2. Granularity

3. Version Control

4. Maintenance Over Time
Version Control

• How to cite and assign DOIs to different versions of a resource?
  – Continuously growing data sets
    1. Assign DOIs to discrete time chunks - OR
    2. Assign one DOI to the whole data set and request that users indicate in the citation that data are continuously updated
  – Bounded data sets that are updated
    • Assign DOIs to major versions, not minor versions
    • Request that users include version indicator in the citation
Citation Challenges

1. Diversity

2. Granularity

3. Version Control

4. Maintenance Over Time
Maintenance Over Time

• Data repositories as publishers?
  – DOI is part of the publishing world
  – There is an understanding that a resource, once assigned a DOI, will be available
  – Implies that the institutions assigning DOIs will take on the responsibility for maintaining them, and thus for taking on long-term curation of resources.

• Maintenance of DOIs
  – Keeping links & metadata updated
  – Providing information if a resource is removed from an archive
Future Efforts

• Develop methods for counting/measuring citations
  – Necessary to assess progress and success
  – Necessary to promote data citations as an indicator of scientific impact

• Fill UCAR/NCAR data curation gaps
  – Many scientists do not formally archive their data
  – NCAR data centers archive data of specific types or from specific projects
  – How to provide DOI and citation services to PIs who do not formally archive their data?
Thank You

UCAR/NCAR data citation white paper draft available at: https://wiki.ucar.edu/display/DatacitePublic/

Email: mayernik@ucar.edu
END