Data Sharing Incentives and Metrics for Open Data

Dr. Robert R. Downs[1] and Dr. Lili Zhang[2]

[1]*rdowns@ciesin.columbia.edu*
Center for International Earth Science Information Network (CIESIN)
The Earth Institute, Columbia University

[2]*zhll@cnic.cn*
Computer Network Information Center (CNIC),
Chinese Academy of Sciences (CAS)

Presented to the
High-Level Workshop
Implementing Open Research Data Policy and Practice in China
Beijing, China,
17-18 September 2019

Session 3: Discussion of Chinese and International Data Practices across Disciplines
17 September 2019

Cultural Shift for Open Science: Transition from Data to Open Data

- Data
  - Potentially usable, useful, and used by data producers
  - Documentation may be understood by data producers
  - Metadata may not exist or be limited
  - Data conversion may be needed for use
  - Data may be unusable soon
  - Intellectual Property rights may be restricted, undefined or unknown
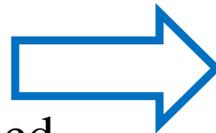
- Open Data
  - Potentially usable, useful, and used by diverse communities
  - Documentation may be understood by potential users
  - Metadata facilitates data discovery
  - Data usable in common formats, tools, or services
  - Data may be usable for decades
  - Intellectual property rights are discoverable, understandable and allow unrestricted use for any purpose free of charge

Open Data:
Responsibilities throughout the Data Lifecycle

- Create data for use by scientific community and public
  - Agree on producer contributions, responsibilities, and usage rights
- Prepare and deposit data in an open data repository
  - Grant rights for use by anyone for any purpose
- Manage, curate and preserve data for continuing use
  - Preparation to facilitate understanding and use by diverse communities
- Disseminate data for use by others
  - Enable open access and use across disciplines without restrictions
- Reuse data for new studies, products and services
  - Conduct new studies and integrate data into data products and services

Open Data – Stakeholder Opportunities

- Data producers
  - Share data referenced in published study and enable new research
- Data repositories
  - Disseminate data to support scientific progress of user community
- Hosting organizations
  - Contribute to science ecosystem and serve scientific communities
- Data users
  - Use and reference data in new studies
- Funding agencies
  - Leverage past studies and enable future studies
- Publishers and Editorial Boards
  - Provide references to data described by published articles

Incentives for Open Data: Recognition

- Data producers
  - Credit for producing and sharing data and for use of data by others
- Data repositories
  - Credit for curating, preserving, and disseminating data
- Hosting organizations
  - Credit for providing and maintaining capabilities to sustain access
- Data users
  - Credit for using data to create new knowledge
- Funding agencies
  - Credit for enabling future studies that use data from previous studies
- Publishers and Editorial Boards
  - Credit for referencing accessible data as basis for published articles

Metrics for Open Data: Are the data usable?

- Independently usable data
  - Can individuals use data without support from data producers?
- Visible and understandable intellectual property rights
  - Can potential users easily determine their rights for using data?
- Publicly accessible as open data
  - Can data be used for any purpose without fees or restrictions
- Described to facilitate use by diverse users
  - Can users with limited expertise in subject understand how to use the data?
- Analyzable with accessible tools and services
  - Can data be analyzed the with common tools or provided services?
- Citable with a recommended data citation and DOI
  - See ESIP Data Citation Guidelines*
- References to data citations available with the data
  - Can reports describing use of the data be readily found and accessed?

*ESIP Data Preservation and Stewardship Committee. 2019. Data Citation Guidelines for Earth Science Data. Ver. 2. Earth Science Information Partners. https://doi.org/10.6084/m9.figshare.8441816

6

Metrics for Open Data: Are the data being used?

- ## Scientific use
  - References to studies reported in peer-reviewed scientific literature

- ## Use across disciplines
  - References to use in studies from fields different than data creators

- ## Instructional use
  - References to data in textbooks, course materials, or learning modules

- ## Decision making
  - Evidence of use in planning, design or policy development

- ## General public
  - Evidence of use for other purposes (commercial, recreational, etc.)

Metrics as Incentives for Open Data: Recognition by Hiring, Promotion, Tenure, and Sponsorship Committees

- Contributions to open data products and services
    - Study design and data collection
    - Data preparation and submission to a data repository
    - Reviews of data products and services
    - Data curation, preservation and dissemination
    - Instructional materials for using data (course materials, user guides, FAQs)
    - Open access publications (data papers, documentation, etc.) on data and methods
- Data Use (alternative metrics)
    - Number of data landing page views and downloads
    - Mentions of data in social media
    - Monetary (lower costs/higher profits) and non-monetary societal contributions
- Data citations
    - References in peer-reviewed journals
    - References in Grey literature (reports, plans and policies, popular press)
    - References in learning resources (course materials, activities, case studies, tests)
    - References in mass media (television and radio broadcasts, webinars)

Current Community Efforts to Improve Incentives and Metrics for Open Data

- ESIP Research Object Citation Cluster
  - http://wiki.esipfed.org/index.php/Research_Object_Citation
- GEOValue
  - http://www.geovalue.org/
- RDA Data Usage Metrics Working Group
  - https://rd-alliance.org/groups/data-usage-metrics-wg
- ESIP 2019 Summer Meeting. Session: Current Approaches for Tracking and Exposing Research Object Usage Metrics
  - Webcast: https://esip.sharefile.com/share/view/s2490d8ff98048858
- CODATA Beijing Conference 2019
  - Metrics for Assessing the Use, Impact, and Value of Scientific Data
  - Expert views on evaluating data sharing efforts and practices: an international dialogue fostered by an interest group of the research data alliance

Thank you!

- We appreciate the support received from:
  - International Science Council Committee on Data (CODATA)
  - Chinese Academy of Sciences (CAS)
  - Center for International Earth Science Information Network (CIESIN)
  - Bill and Melinda Gates Foundation
  - Earth Science Information Partners (ESIP)
  - Computer Network Information Center, CAS