

A Scientific Data Center Perspective on Publishing Data

Robert R. Downs

rdowns@ciesin.columbia.edu

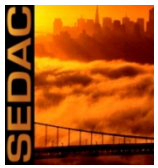
NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network
(CIESIN)
The Earth Institute, Columbia University

2015 WINTER ESIP FEDERATION MEETING
Washington, DC, 6-8 January 2015

Breakout Session: What does it mean to publish data?
Tuesday 6 January 2015, 3:30 p.m.



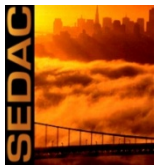
Organization and Role



- Program: SEDAC
 - NASA Socioeconomic Data and Applications Center¹
 - Focusing on human interactions in the environment
 - Integrating socioeconomic and Earth science data
 - Serving as an information gateway between the Earth and social sciences
 - Operated for NASA by Columbia University
 - One of 12 Distributed Active Archive Centers (DAACs) in the Earth Observing System Data and Information System (EOSDIS)
- Department: CIESIN is a research and data center
 - Center for International Earth Science Information Network²
 - Interdisciplinary department (social, natural and information sciences)
 - Information management, online data, and spatial data integration
 - Research on human interactions in the environment
 - Part of the Earth Institute of Columbia University
- My Roles
 - Senior Digital Archivist and Senior Staff Associate Officer of Research
 - Acting Head of Cyberinfrastructure and Informatics R&D



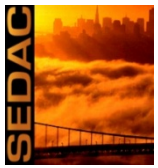
Definition of “Publishing Data”



The act of publishing data includes the provision of online access capabilities to enable the discoverability, accessibility, usability and citability of quality scientific data products and services for diverse communities, today, and in the future.



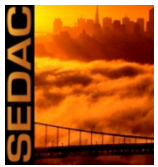
Scientific Data Center Role in “Publishing Data”



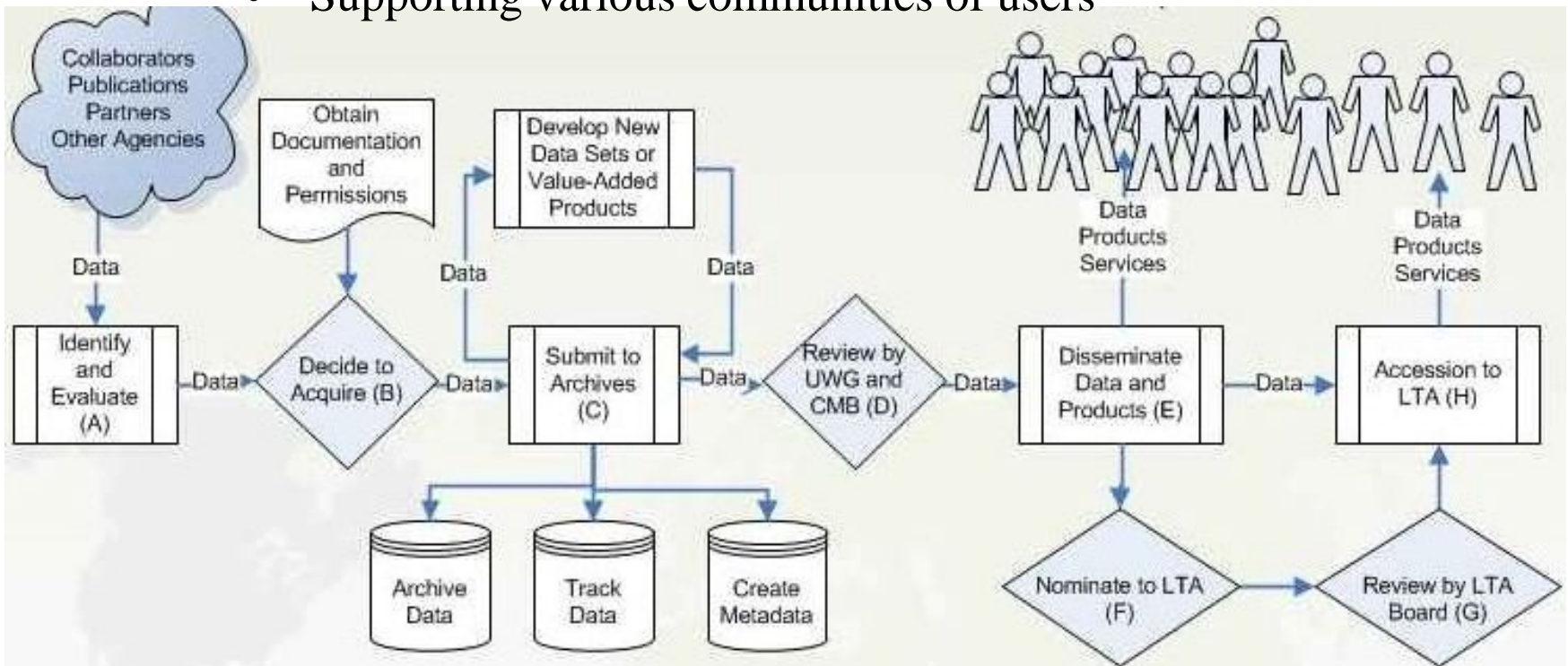
- Lifecycle approach to publishing data
 - Establishing, maintaining, and improving process
- Verifying quality throughout entire process
 - Continuous review, assessment, quality control, and inspection
- Ensuring developed data meet community needs
 - Community involvement from beginning
- Providing capabilities for diverse uses
 - Improving the use of data by potential users
- Supporting ongoing data usage
 - Enabling continuing and future use of the data



How Data are Published at SEDAC

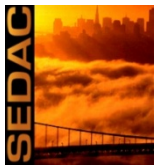


- Identifying, appraising, and selecting data
- Developing data products and services
- Disseminating data products and services
- Enabling data discovery, exploration, analysis
- Preserving the data for future users and uses
- Supporting various communities of users

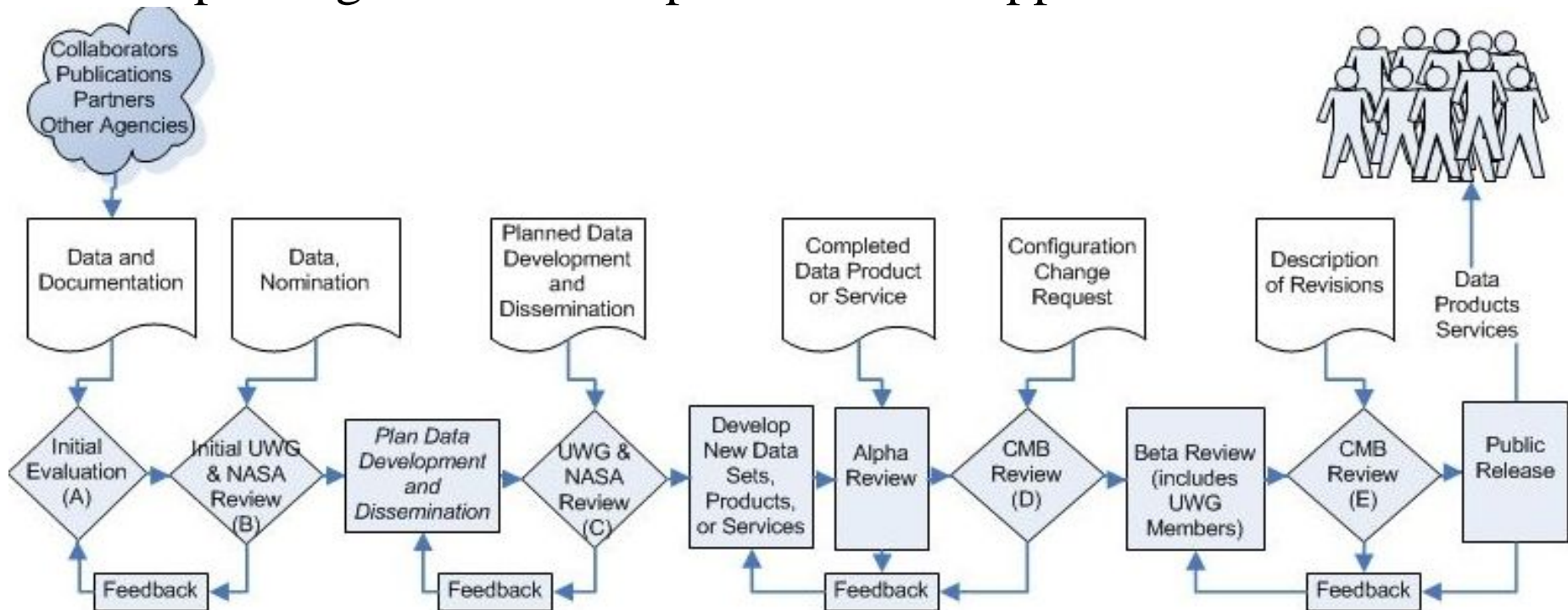




Challenges with “Publishing Data”

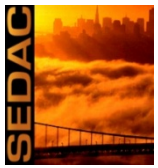


- Ensuring that competing demands do not compromise data products and services offered
- Providing capabilities to enable a variety of users to efficiently use data products and services
- Improving both internal processes and opportunities for users





Minimum Requirements for Publishing Data



- Plan
 - Description of product or service and how it will be created
- Review
 - Scientific and management; User Working Group and NASA
- Acquire
 - Data and rights obtained to enable development and use
- Develop
 - Creation of products or services and related information to meet user needs
- Quality Control
 - Product release in Alpha and Beta Configuration Management Board reviews
- Archive
 - Completed inventory, inspections, validation, packaging, and redundant storage
- Disseminate
 - Data discoverable and accessible online as complete usable products or services