Data Lifecycle
Review of Research 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

RESEARCH DATA ALLIANCE 5TH PLENARY
San Diego, California 8-11 March 2015

Joint Session: IG Active Data Management Plans, IG Preservation e-Infrastructure, and IG RDA/WDS Certification of Digital Repositories
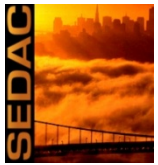Tuesday 10 March 2015, 9:00 a.m.

Selected Reasons for Conducting Reviews of Research Data During the Data Lifecycle
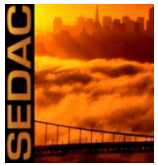
- Confirmation of **values collected from observations**
- Verification of **data compilation and completeness**
- Inspection and correction of **errors**
- Validation of **combined or computed variable values**
- De-identification of **restricted, sensitive, or personal information**
- Submission Information Package (**SIP**) Preparation
- Appraisal for archival **acquisition**
- Selection for domain **collections**
- Archival Information Package (**AIP**) Preparation
- Identification or verification for potential **usability**
- Dissemination Information Package (**DIP**) Preparation
- **Verification and reproduction** of findings in study reports and publications
- **Comparison** with data from other studies
- Applicability to study new **hypotheses or research questions**
- **Integration** with other research data
- Selection for **retention, deaccession, or transfer**

- Each stage of data lifecycle affects data
  - Different stakeholders involved throughout data lifecycle
  - New purposes for using data emerge
  - Data products and services evolve
- Foster current and future users and uses of research data
  - Limited initial usage: data collection team
  - Subsequent usage: within domain and beyond
  - Expanded usage: (interdisciplinary, range of expertise)

Examples of Drivers for Data Reviews During the Data Lifecycle
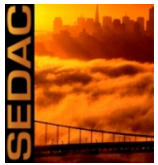
Published Studies

Data analysis and results reporting reviews
Reproducibility and data comparison reviews

Data Creation

Study design and data collection reviews
Preparation for submission reviews
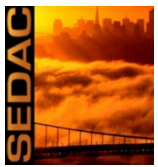
Active Archiving and Dissemination

Pre and post-release reviews
Product and service development

**Post-Active and Long-Term Archiving**

Data retirement and de-accession
New product development

# Scientific Data Center Reviews for Public Release of Data

- ## User Working Group and NASA reviews
  – Data product or service concepts
  – Data development plans
  – Beta review of data product or service prior to production release

- ## Configuration Management Board reviews
  – Alpha review of data product or service
  – Beta review prior to production release

SEDAC Active Archive Data Review. Derived from Downs, 2014
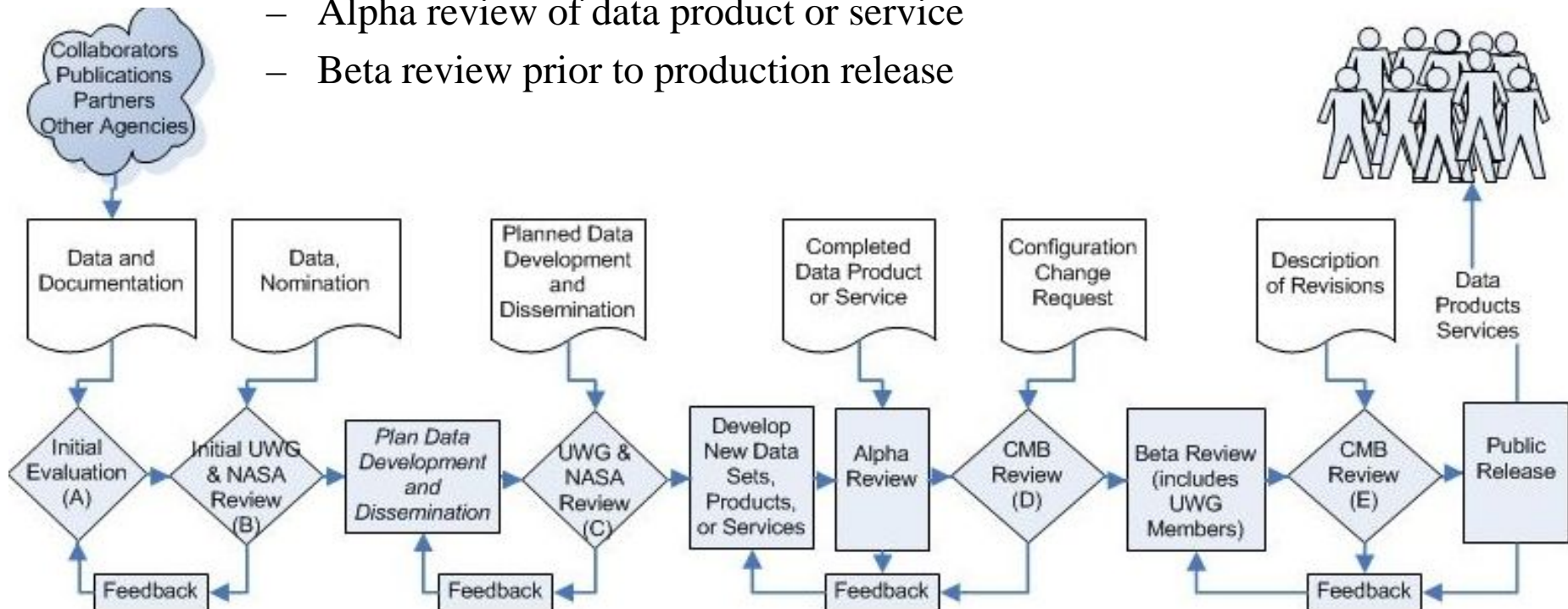
Additional Opportunities to Benefit from Reviewing Research Data throughout the Data Lifecycle
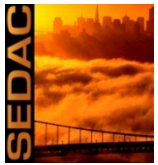
- Usability
  - New tools available for use of data (analysis, integration, etc.)
  - Cumulative learning (and documentation) about the data and its use
  - Capturing provenance from each review event

- New data products and services
  - New or enhanced data products by integrating data sets
  - Services to discover, access, integrate, and analyze data

- Strategy and Policy development
  - Changes in needs and technology leading to new strategic plans
  - Gaps in policies that could benefit from enhancements

- Process improvement
  - Quality control for data curation processes
  - Efficiency and automation