

---

---

# Scientific Data Versioning at the NASA Socioeconomic Data and Applications Center

*Robert R. Downs<sup>1</sup>*

<sup>1</sup> [rdowns@ciesin.columbia.edu](mailto:rdowns@ciesin.columbia.edu)

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

SCIDATACON 2016  
Denver, CO 15-17 September 2016

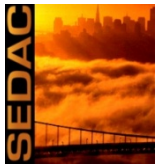
Session: Data Versioning  
Saturday 17 September 2016, 9:00 a.m.



# Versioning Scientific Data within EOSDIS, the NASA Earth Observing System Data and Information System

---

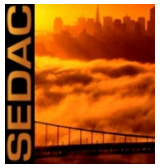
---



- EOSDIS does not use a common standard to number versions
- Diverse data producers (instrument or science teams) contribute data to the NASA Distributed Active Archive Centers (DAACs)
  - Designate versions in their own ways
  - Use various terms: Version 1, 2; Collection 1, 2; Release 1, 2; Edition 1, 2
- Number of versions vary, depending on volumes from instrument
- Absence of standard numbering scheme for versions is not problematic
  - Each data set title is identified by a particular version
  - Software version that generated data is identified, when applicable
  - Provenance is tracked independent of a standard numbering scheme
- Version information is recorded in the data metadata
- New DOIs are assigned as new versions are generated
  - Landing pages reference older versions if they exist
  - Landing pages for superseded versions will persist and refer to newer versions



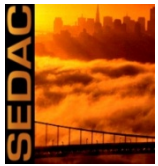
# SEDAC Practice for Versioning Data Products



- Establish titles with consistent designations for version or year
  - Uniquely identify each edition of collection, dataset, or product
  - Unique aspects can include year or range of dates for observations
- Default version number is 1; implied if not stated explicitly
  - New version number is assigned if a dataset is changed
  - Subsequent versions are Version 1.01, Version 2, or Revision 01
- New collection version reflects new stage of development
  - Collection versions are assigned as integers
  - New collection supersedes previous collection



# SEDAC Persistent Identifier Practice



- SEDAC Guide to Assigning Global Persistent Identifiers
  - Assigned to the landing page for each dataset disseminated by SEDAC
  - Included in the recommended citation for each dataset
  - Recorded and maintained to identify current location and optimize discovery
- SEDAC Procedure for Assigning Global Persistent Identifiers
  - DOIs assigned to datasets and documentation; software and services may be next
  - Assigned using EZID and the DataCite Metadata Schema
  - DOIs also recorded in the FGDC CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)
  - Related Identifier field is used to link to other data, documentation, publications
  - DOI record is modified when location of landing page changes