



Research Data Strategies to Support Studies of Human Interactions in the Environment

Robert R. Downs, PhD

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

THE EVOLUTION OF RESEARCH DATA:
STRATEGIES FOR CURATION AND DATA MANAGEMENT

Charlottesville, Virginia
Thursday May 16, 2013





Some Example Data Needs Questions of the Research Community

- How can the data be found?
- Have the data been reviewed?
- Have the data been preserved for future use?
- Are the data managed throughout the data lifecycle?
- How have the data been used?



Scientific Data at SEDAC

Center for International Earth
Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

- Data on human interactions in the environment
 - Organized online within 30+ collections
 - SEDAC website: <http://sedac.ciesin.columbia.edu/>
- Key SEDAC data collections
 - Environmental Treaties and Resource Indicators (ENTRI)
 - Gridded Population of the World (GPW), v3
 - Global Rural-Urban Mapping Project (GRUMP), v1
 - Last of the Wild, v2



Finding SEDAC Data by Themes

NASA Earth Data Data Discovery Data Centers Community Science Disciplines Search EOSDIS

SEDAC

DATA MAPS THEMES RESOURCES COMMUNITIES ABOUT HELP

NASA SOCIOECONOMIC AGRICULTURE AGRICULTURE RESEARCH CENTER (SEDAC) – HOSTED BY CIESIN AT COLUMBIA UNIVERSITY

CLIMATE
CONSERVATION
GOVERNANCE
HAZARDS
HEALTH
INFRASTRUCTURE
LAND USE
MARINE AND COASTAL
POPULATION
POVERTY
REMOTE SENSING
SUSTAINABILITY
URBAN
WATER

Map Layer » Population Count 19

Follow Us: | Share:

Gridded Population of the World, Version 3

To serve a wide user community by providing the latest data on human population distribution that can be used in interdisciplinary studies of the environment.

- Data Collection
- Data Sets (10)
- Map Gallery

Featured Uses of Data

Global Forecasts of Urban Expansion to 2030
Proceedings of the National Academy of Sciences – February 23, 2013

More Featured Uses...

- Global Forecasts of Urban Expansion to 2030
- Sea Level Rise Impacts on Wetlands
- Prosperity Shining
- Water Balance of Global Aquifers Revealed by Groundwater Footprint

News

More News...

- New Online Resource Supports Learning About Scientific Data Management Practices
- CIESIN Represented at Annual Population and Geography Conferences
- Students from France Begin Spring Internships at CIESIN
- Environmental Indicators Examined for Policy and Management Uses



SEDAC Data are Prepared for Use by Diverse Communities

Center for International Earth
Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

- Research
 - Scientists and scientific research staff
 - Data developers and managers, system developers and managers
 - Data librarians, data laboratory staff, and other intermediaries
- Education
 - University faculty and students
 - PreK-12 teachers and students
- Policy and Planning
 - International, national, state, and local government planners, decision-makers, and policy-makers,
 - Government contractors
 - Commercial and non-profit organizations
- General Public
 - Lifelong learners
 - Individuals and groups interested in professional and personal development



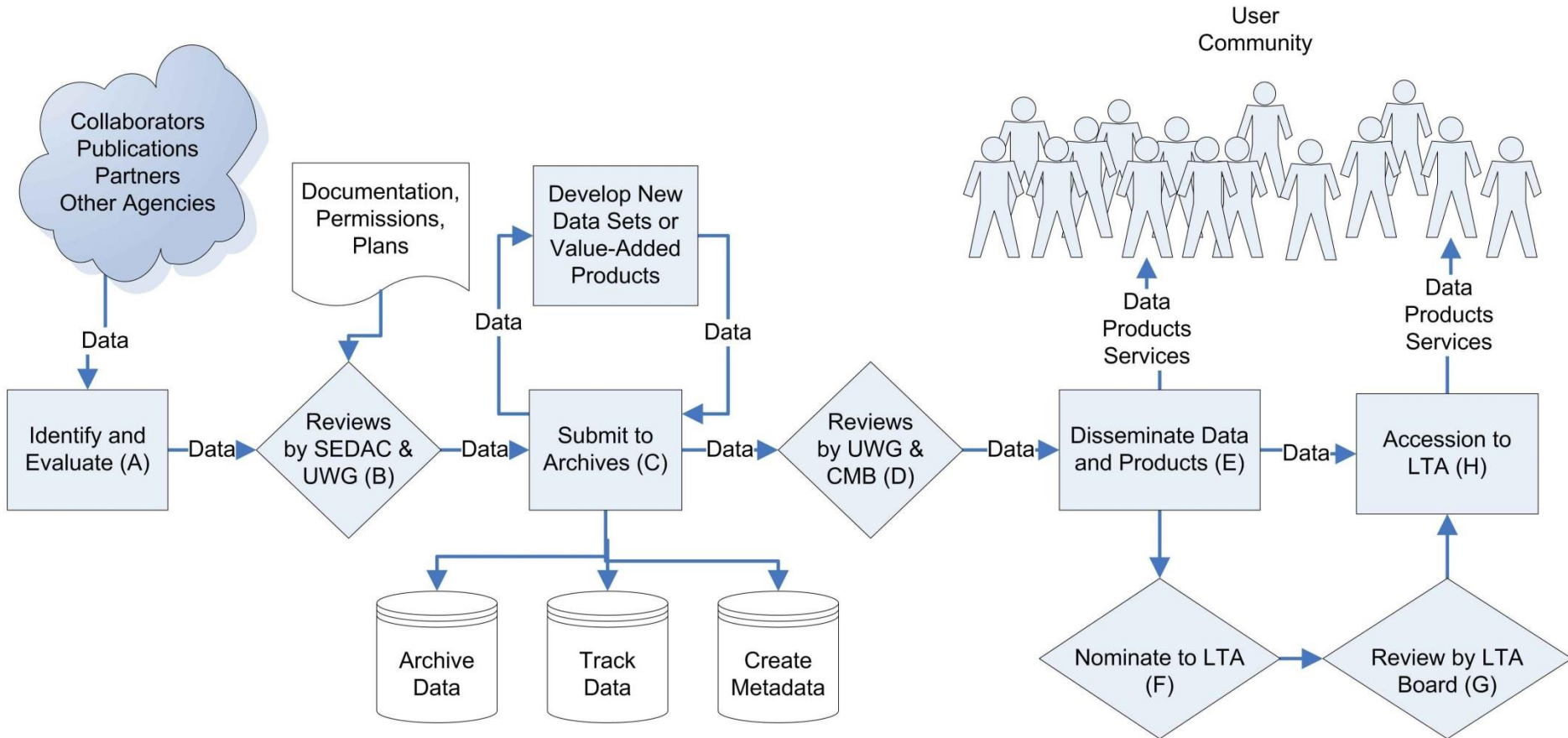
- **Community Voices**
 - Publishing articles and reports describing research data needs
 - Participating in workshops and meetings that identify research data needs
- **Data Producers**
 - Collecting data for dissemination as products or services
 - Integrating data into new products or services planned for dissemination
- **Reviewers**
 - Recommending data products and services for dissemination
 - Recommending features for products or services planned for dissemination
 - Testing data products and services prior to release
- **End-Users**
 - Conducting and publishing studies that use SEDAC data
 - Teaching and learning about human interactions in the environment
 - Recommending decisions based on data and other information resources
 - Reproducing or combining data with other data to inform decisions



- SEDAC User Working Group members serve in multiple roles
 - Represent the community of data users
 - Provide advice and recommendations for activities, plans, and priorities
- Community needs drive strategic planning
 - Strategic plan provides a 5-year outlook for data acquisition, development, and dissemination
- Strategic plan informs data management during data lifecycle
 - Acquisition and Creation
 - Appraisal and Selection
 - Ingest and Preservation
 - Dissemination
 - Access and Use
 - Long-Term Archiving



Workflow for SEDAC Data





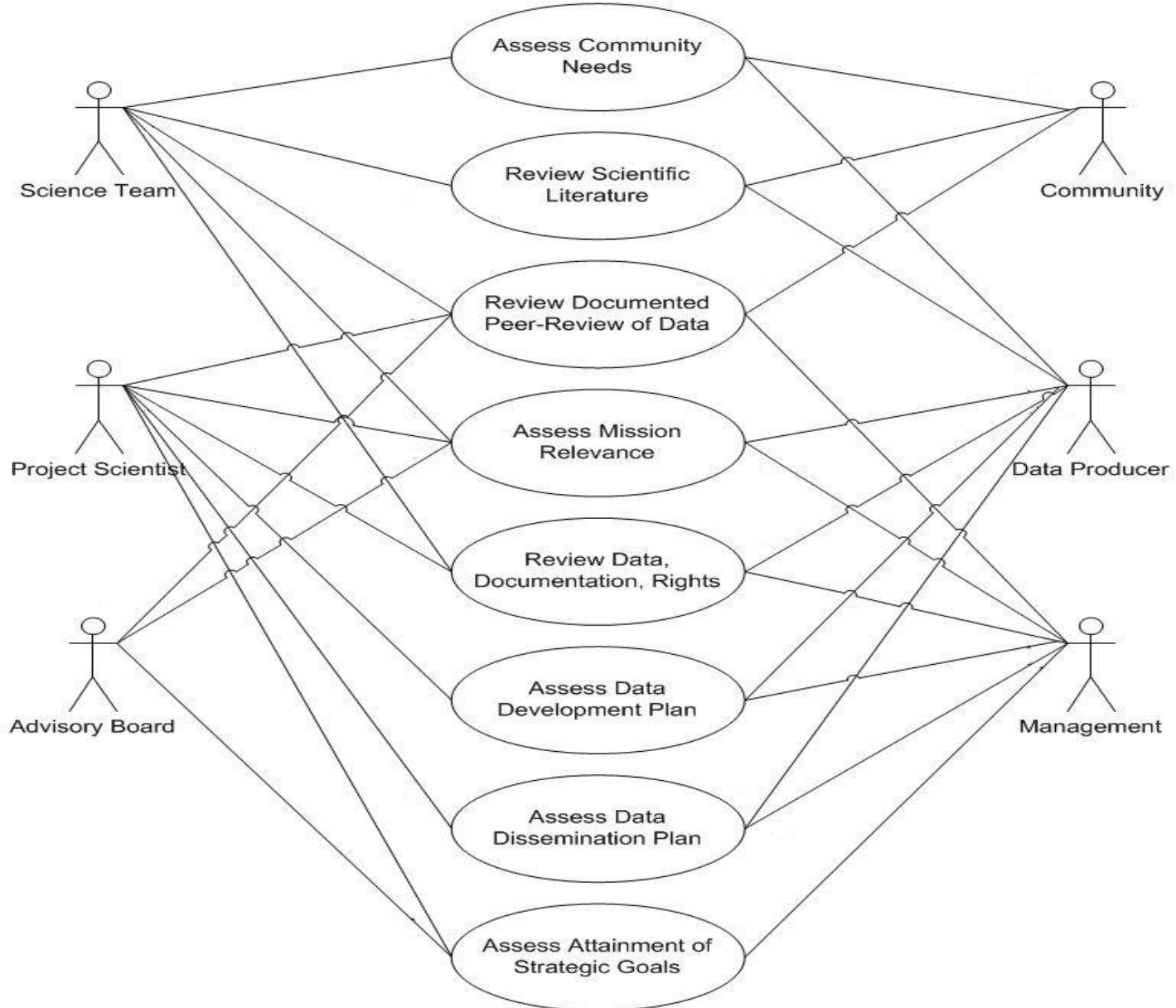
- Identifying candidate data for dissemination
 - Scientific value
 - Potential usefulness to user community and groups
 - Relevance to strategic plan and current priorities
 - Feasibility of data development and dissemination
- Options for candidate data products and services
 - Develop data products or services
 - Add value to third-party data
 - Distribute third-party data



- Internal data review and nomination
 - Meets identified community needs
 - Rights and restrictions
 - Available scientific expertise and resource requirements
 - Timetable and cost estimates for development and release
- User Working Group Review
 - Justification (science and community needs)
 - Data development and dissemination plan
 - Applicability to current strategic plan, goals, and priorities
 - Previous reviews and citations of data or methodology
 - Plan for review and quality control



Use Case for Data Appraisal and Selection





Ingest and Preservation

- Receive
 - Data, description, documentation, and rights
- Inspect
 - Complete, correct, and virus-free
- Package
 - Inventory, profile, and serialize
- Store
 - Preservation, security, and distribution copies
- Track
 - ID, volume, status, rights, contacts, etc.

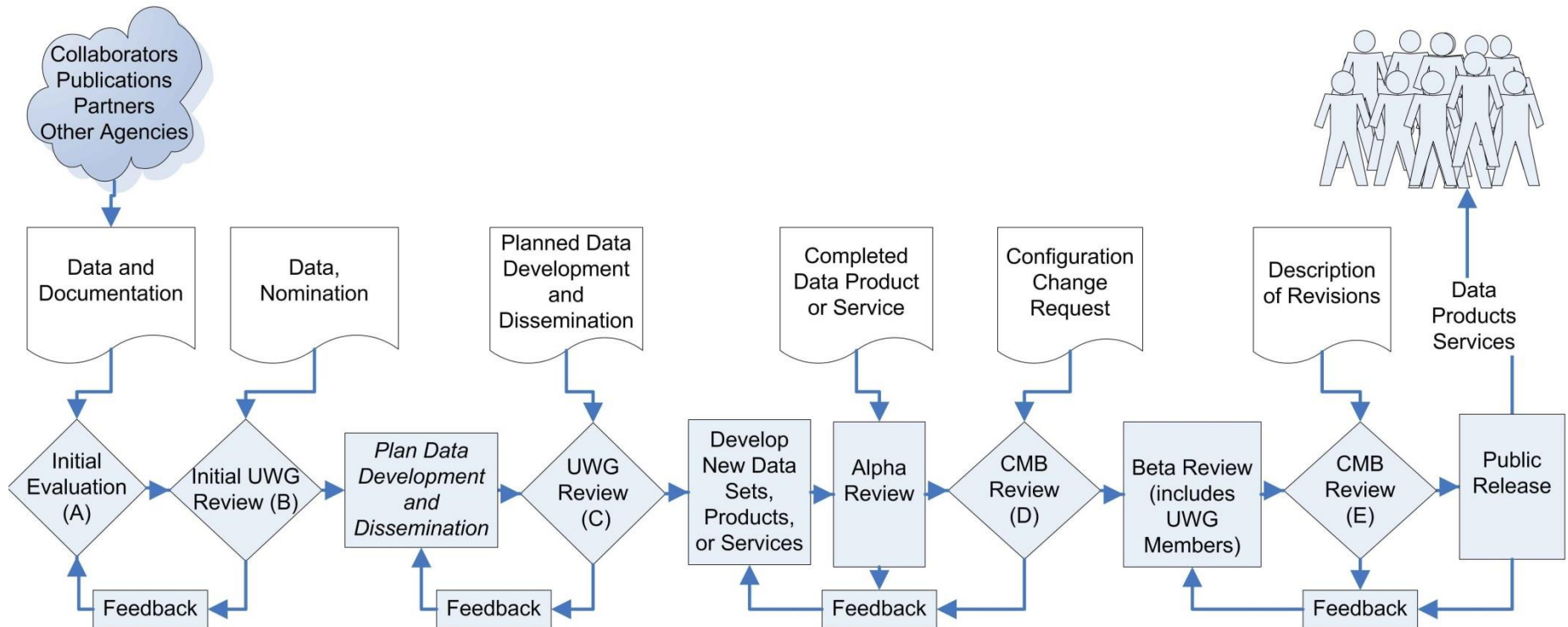


Dissemination

- Alpha review completed and suggestions addressed
- Pre-Beta Configuration Management Board (CMB) review and required changes completed
- Beta review completed and suggestions addressed
- Pre-Production CMB review and required changes completed
- Dissemination Information Package obtained from archive
- Registration on Earth Observing System Data Information System (ESDIS) Metrics System (EMS)



Reviews for Data Development



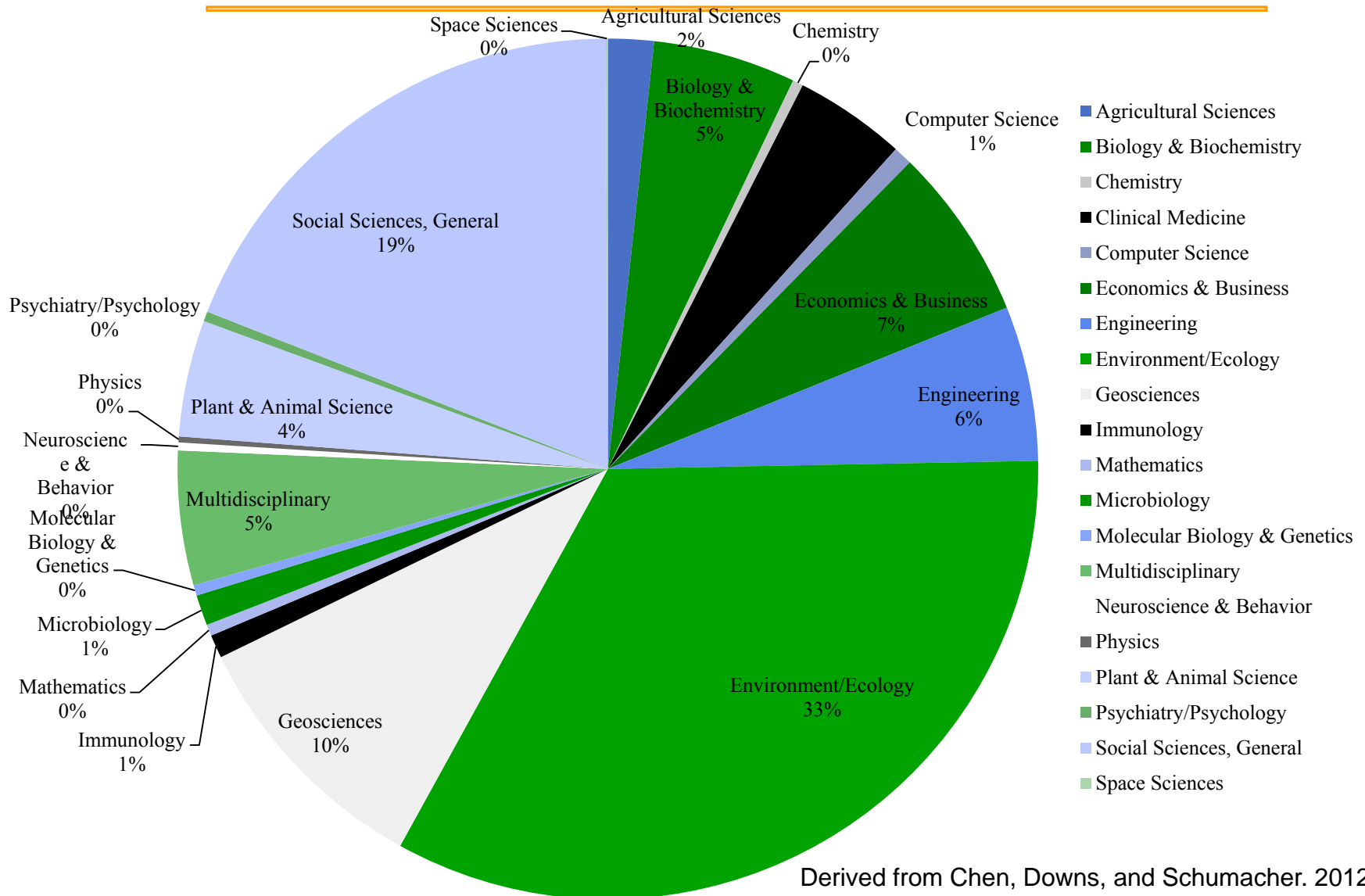


Access and Use

- Data set available on standard pages
 - Landing page and tabbed download and metadata pages
 - Landing page includes summary and recommended citation
- Discovery enabled via multiple paths
 - Organized and searchable by collection, theme, or free-text query
 - Searchable via the NASA Global Change Master Directory (GCMD)
- Use of data cited in journals, books, and popular media
 - Data citations in 695 articles from 2007-2011 in 331 journals
 - Data citations in 130 books or chapters from 2007-2011 in 110 book titles
 - Collection and metadata records in Thomson Reuters Data Citation Index



Disciplines of 331 Journals that Published 695 Articles From 2007-2011 Citing SEDAC Data



Derived from Chen, Downs, and Schumacher. 2012.
<http://hdl.handle.net/10022/AC:P:19170>



- Recommended for nomination to LTA
 - Data no longer used or has been superseded
 - Request approval by User Working Group
- Nominated for accession to LTA
 - Appraisal and selection by LTA Board
 - Recommended levels of service for dissemination and preservation
- Sustainable governance of LTA
 - LTA Board: Columbia University Libraries, Earth Institute, and SEDAC
 - Contingency plan for reorganization in the event of funding lapse