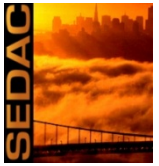
Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY



Facilitating Integration of Socioeconomic and Remote Sensing Data to Support Interdisciplinary Research and Applications

Robert R. Downs[1] , Robert S. Chen[2]

[1] rdowns@ciesin.columbia.edu

[2] bchen@ciesin.Columbia.edu

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

IASSIST 2017
Lawrence, Kansas, 23-26 May 2017

G4: Data Sharing and Reuse Across Boundaries
Friday, 26 May 2017; 9:00 a.m. - 11:00 a.m.

Interdisciplinary Research

- Need for Interdisciplinary Research
  - Study issues outside each individual discipline
  - Employ methods & tools from other disciplines
  - Overcome limitations of data collected by one discipline
- Value of Interdisciplinary Research
  - Enabling studies of issues that fall between or cut across disciplinary boundaries
  - Facilitating problem-focused research that addresses real-world needs and challenges
  - Utilization of perspectives and data from multiple disciplines and analytic frameworks
  - Ability to integrate different spatial and temporal scales and units of analysis (e.g., people vs. pixels, watersheds vs. provinces)

Value in Interdisciplinary Use (or Reuse) of Data

- Data gathered for one purpose may have value for other purposes in ways not understood by original data collectors
- Combination of data from multiple disciplines may yield potential insights and knowledge valuable for both disciplinary and interdisciplinary research
- Well integrated interdisciplinary databases and value-added data products derived from interdisciplinary research may enable new applications both in research and practice

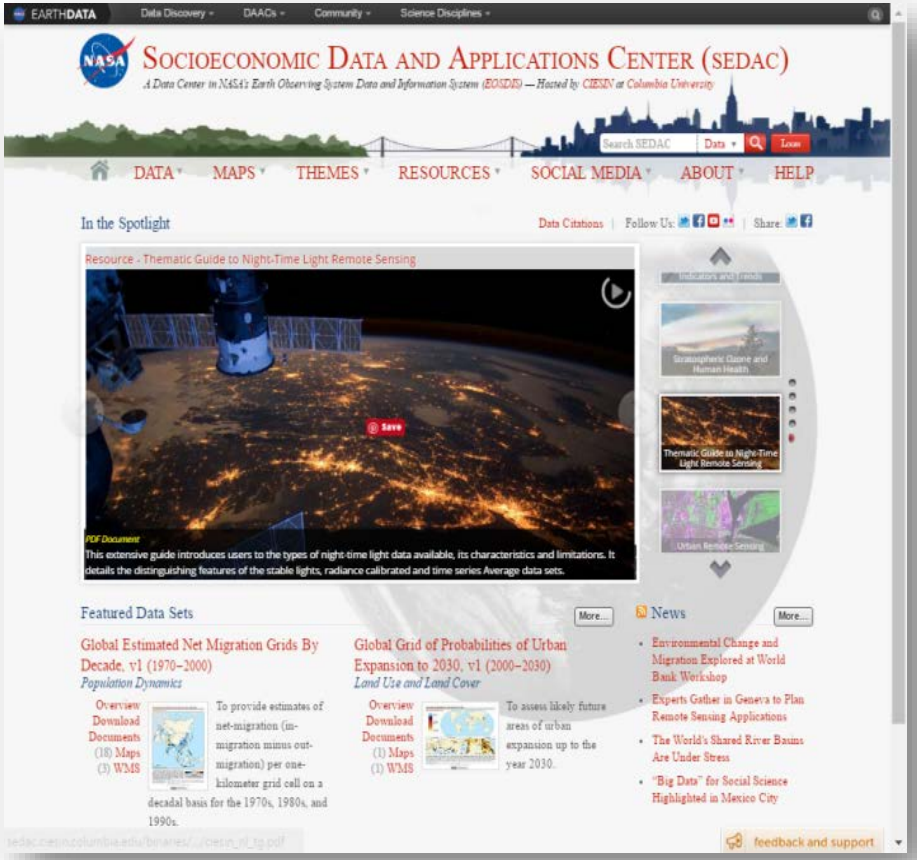
Based on Downs, 2017. Measuring the Interdisciplinary Value of SEDAC Data

Integrated Remote Sensing and Socioeconomic Data: Examples of Value for Practice

- Policy and Decision Making
  - Land management, population growth, infrastructure planning
  - Populations and property risk reduction, safety and security
- Disease epidemic prevention, preparation, and response
  - Remote sensing identification of locations of disease vectors
  - Socioeconomic data identification of populations at risk
  - Public health data for understanding vulnerability, developing and implementing response strategies
- Disaster resilience: preparation, response, and recovery
  - Prediction: Forecasts of natural, technological, and societal hazards
  - Exposure and Vulnerability: Identification of human populations and resources at risk
  - Resilience: Damage assessment, emergency response, and recovery/reconstruction planning

Interdisciplinary Aspects of SEDAC Mission

- Mission: “develop and operate applications that support the integration of socioeconomic and Earth science data and to serve as an "Information Gateway" between the Earth and social sciences”



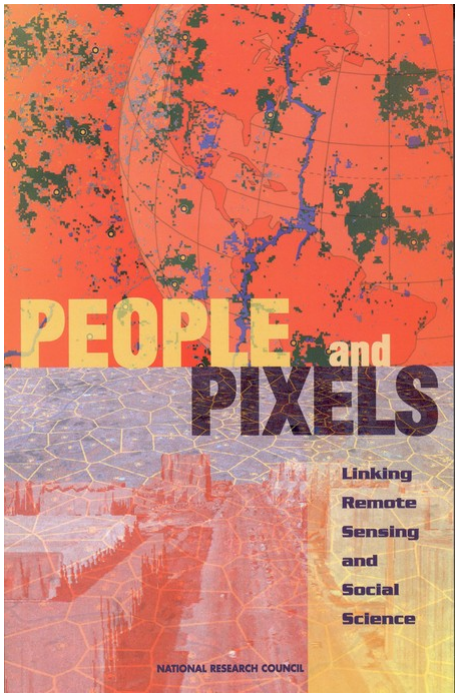
Source: http://sedac.ciesin.columbia.edu/about

Challenge of Accessing Data from Diverse Disciplines

- Data discovery and access
  - Data may or may not be available from domain-specific or institutional repositories; often buried in places only known and available to their own communities
  - Described with discipline-focused terminology
  - Different practices with respect to documenting data, assumptions, methods, inputs
  - Different cultures with respect to data sharing, including timeliness, contribution of data to shared community databases and archives, openness in sharing metadata

Some Challenges of Integrating Data from Diverse Disciplines

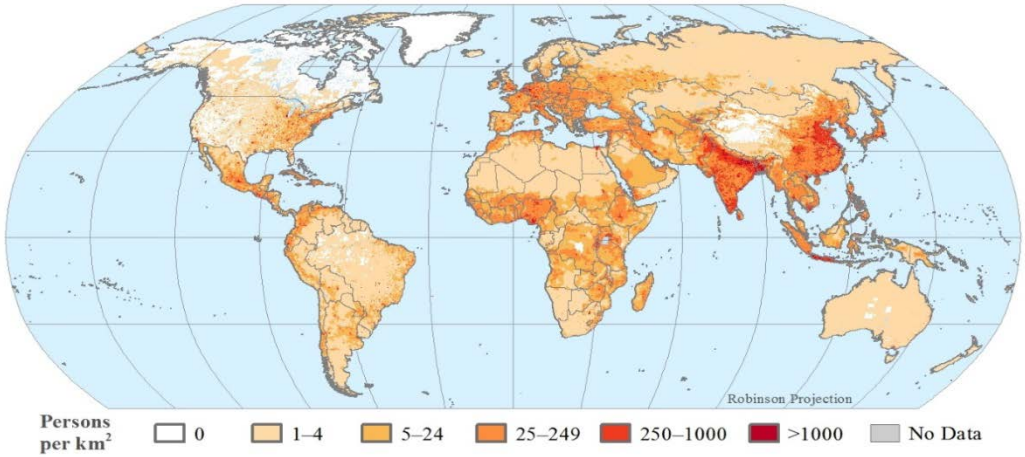
- Different units of analysis
  - e.g., “people vs. pixels”
- Different spatial scales and frameworks
  - e.g., fixed geographies vs. evolving administrative areas
- Expanding array of similar data products
  - makes it hard for those outside of originating discipline to understand and use appropriately
- Confidentiality and privacy issues for social/health data
  - e.g., human subjects and institutional review board requirements
- Intellectual property/access differences between disciplines, sources
  - e.g., use of administrative data in research; different science cultures regarding data sharing and access



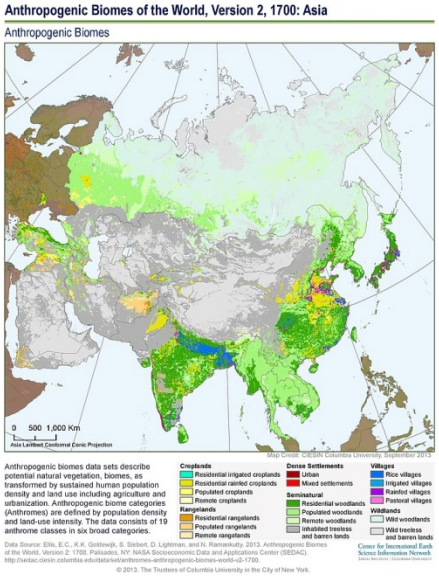
NRC 1998

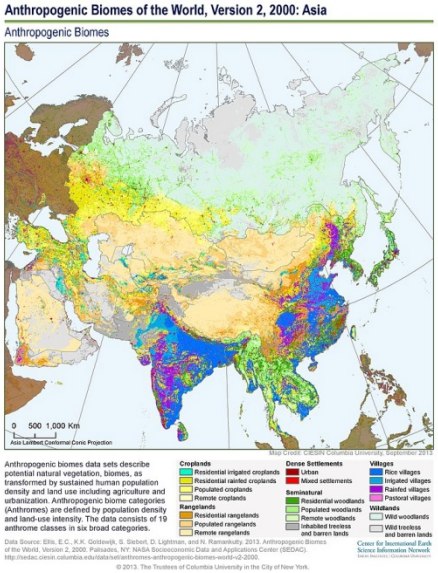
Enabling Interdisciplinary Use of Data - 1

- Convert data to more "integratable" forms
  - e.g., grid population data
- Support and disseminate integrated data products
  - e.g., create integrated data and policy-relevant indicators; work with external partners to disseminate integrated data



Gridded Population of the World, Version 4
http://sedac.ciesin.columbia.edu/data/collection/gpw-v4

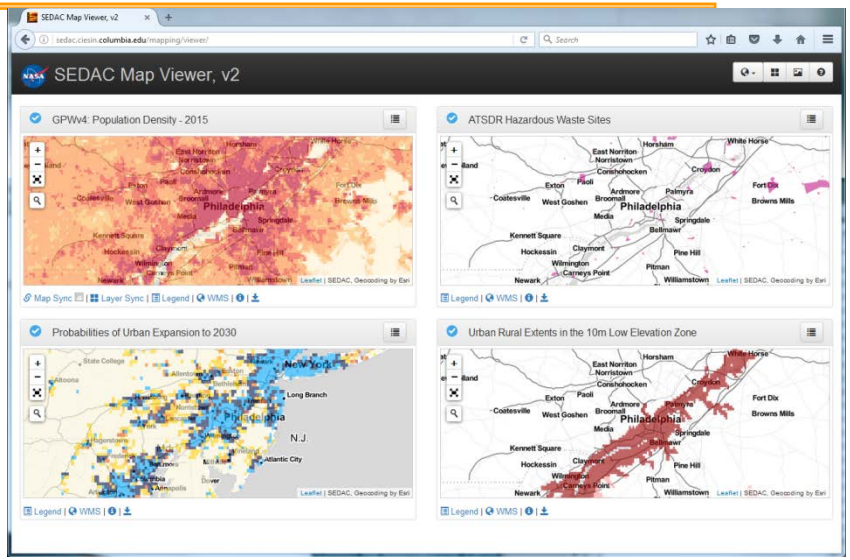




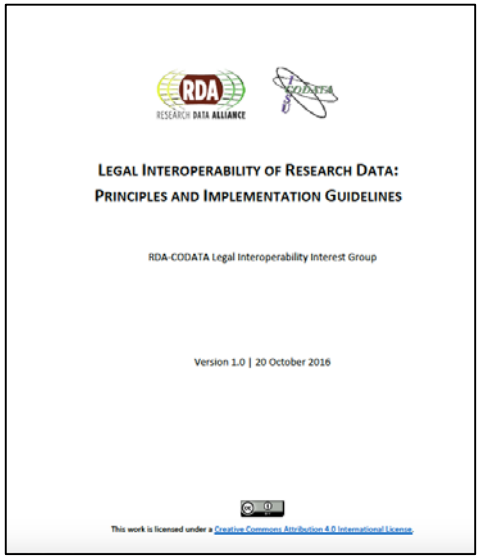
Anthropogenic Biomes, 1700-2000
http://sedac.ciesin.columbia.edu/data/collection/anthromes

Enabling Interdisciplinary Use of Data - 2

- Provide visualization and analysis tools to facilitate data integration
  - e.g., mapping clients and tools; spatial query services
- Address IP and privacy/confidentiality concerns
  - e.g., promote open access licensing; support anonymization methods

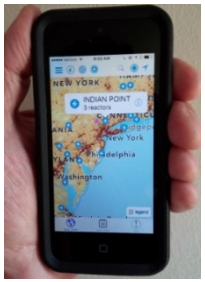


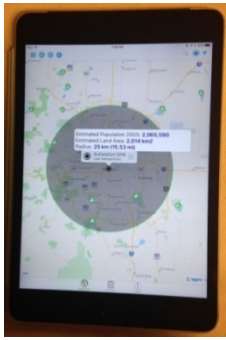
SEDAC Map Viewer, V2
http://sedac.ciesin.columbia.edu/mapping/viewer/



Legal Interoperability Principles
https://zenodo.org/record/162241#.WER7MX1vDxU

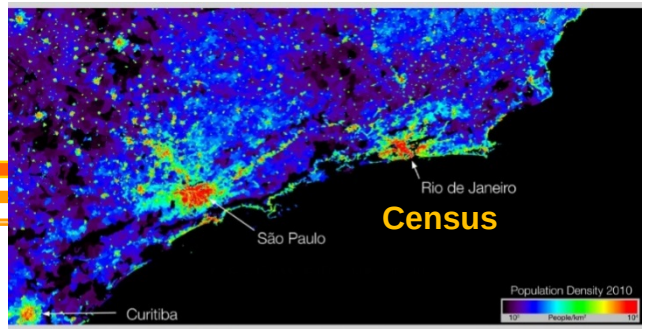
HazPop Mobile App for iOS
https://itunes.apple.com/us/app/hazards-population-mapper/id1092168898?mt=8

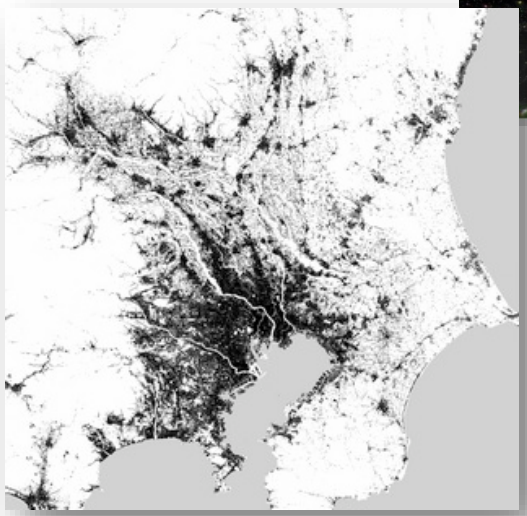




Enabling Interdisciplinary Use of Data - 3

- Example: How to make sense of growing number of gridded population and human settlement data products?
  - Night-time lights (DMSP >1 km → VIIRS ~750m)
  - Landsat (~30 m)
  - Radar (~12 m)
  - High resolution imagery (< ~3m resolution)





DLR

internet.org by facebook

http://ciesin.columbia.edu/data/hrsl/

https://urban-tep.eo.esa.int/#

http://ghslsys.jrc.ec.europa.eu/index.php

JRC EUROPEAN COMMISSION

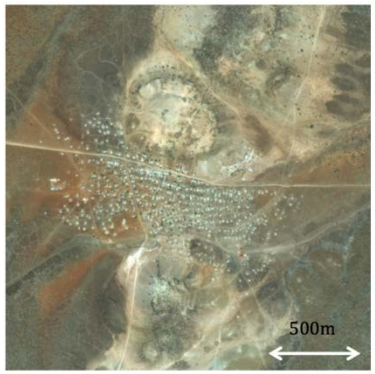
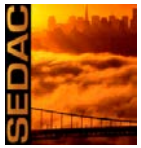


Figure 2A: Dense settlement where a short-range wireless hotspot would be efficient. Imagery: DigitalGlobe

Figure 2B: Sparse, scattered settlement that would benefit from long-range cellular technology. Imagery: DigitalGlobe

Solution: A "Data Collaborative" for Settlement, Infrastructure, and Population Data

- Public-private data partnership involving intergovernmental organizations, national & academic research institutions, large and small companies, NGOs, foundations, universities, data stewards, etc.

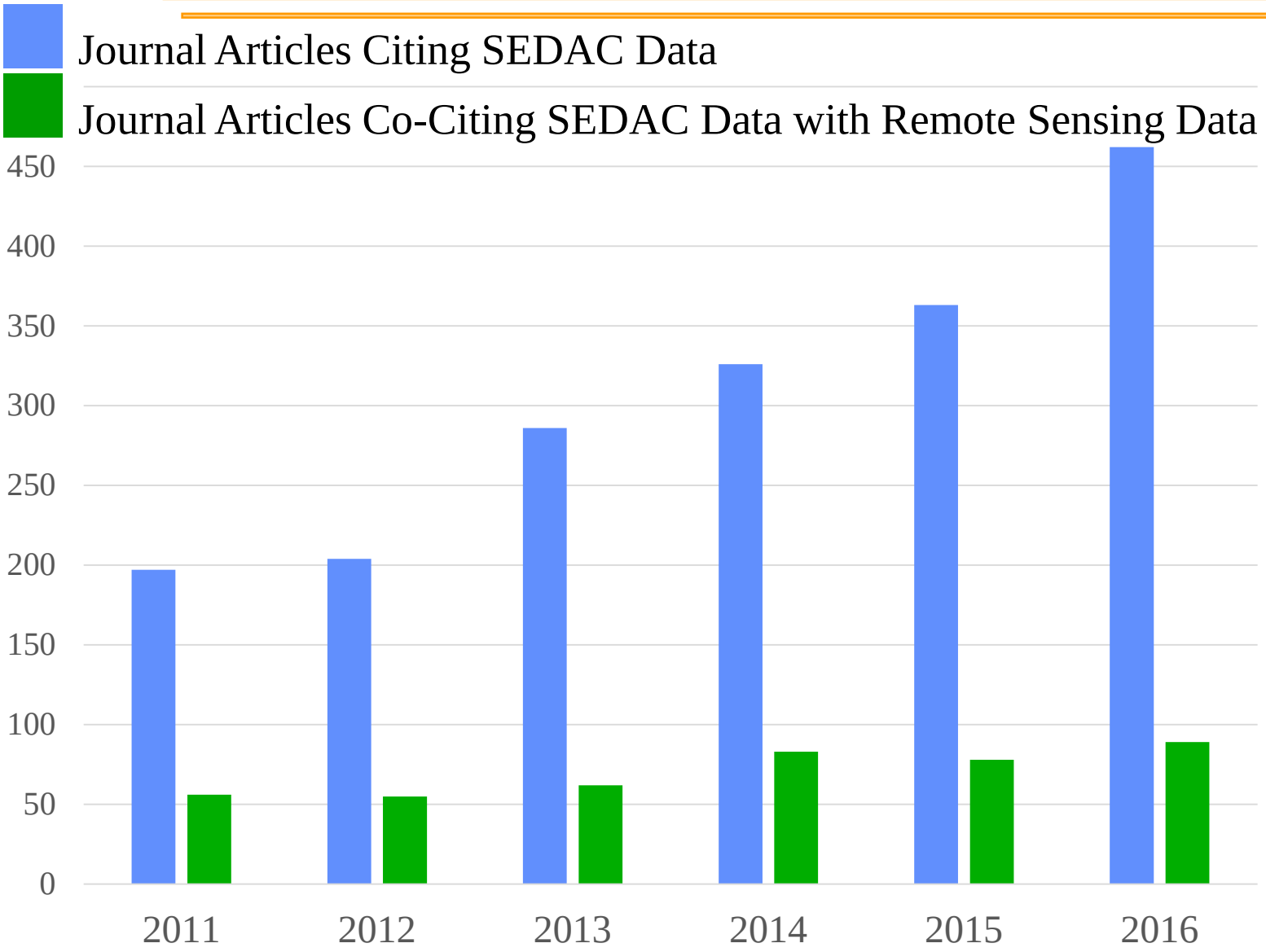


Methodology for Assessing Interdisciplinary Use in the Scientific Literature

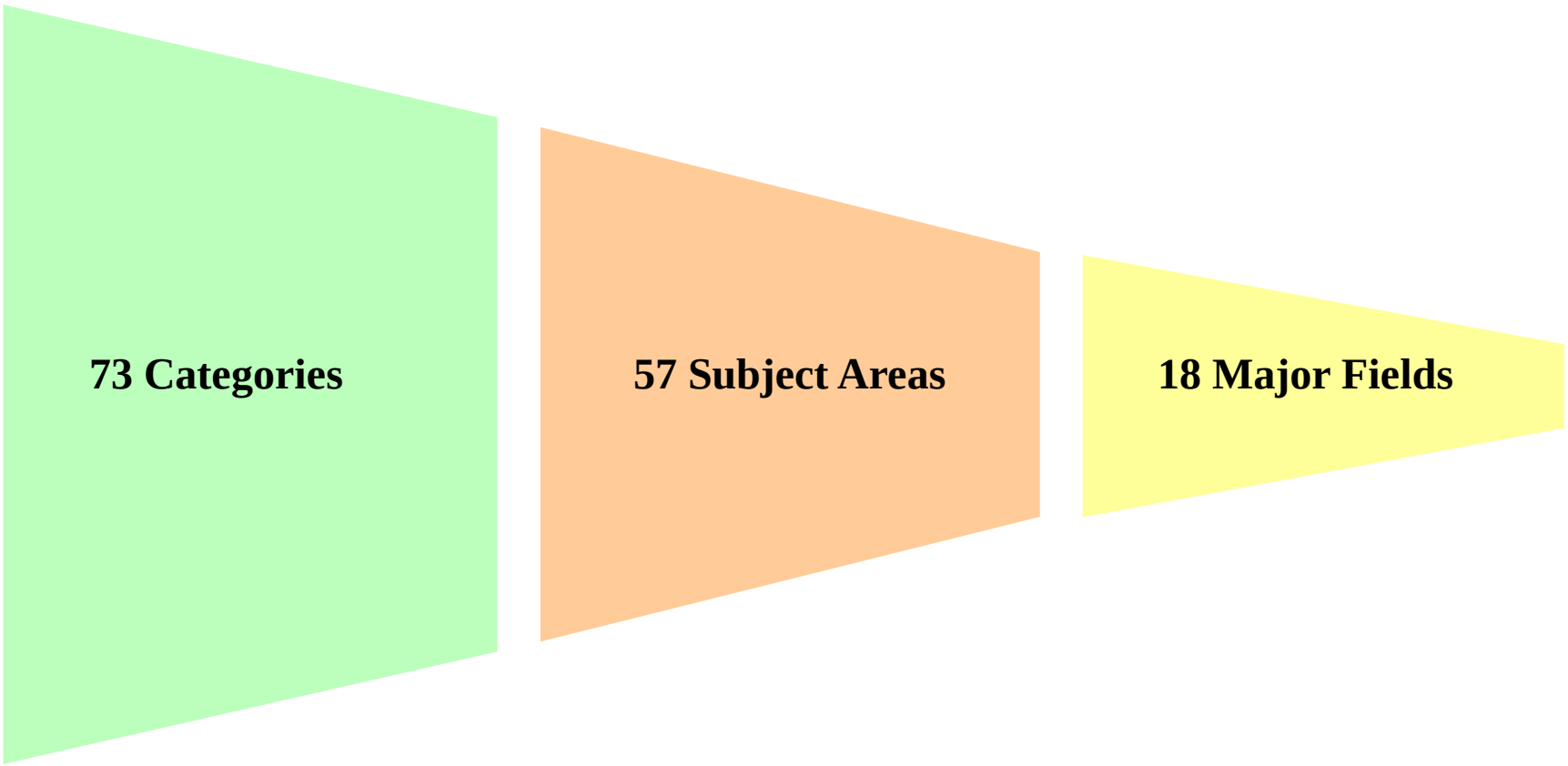
- **Collected citations of SEDAC data cited with remote sensing data**
  - Received alerts of journal articles citing SEDAC data published during 2011 – 2016
  - Searched each article for citations or mentions of remote sensing data using standard terminology list (conducted routinely by the same person on a weekly basis) and verified use
- **Obtained classifications to identify disciplines of each citing journal**
  - Obtained Web of Science® (WoS) Category assignments
  - Obtained Web of Knowledge® Subject Classification of WoS Categories (WoK5.3) and Equivalent General Categories and Subject Areas
  - Obtained ScienceWatch® Field Definitions of Major Fields
- **Identified multidisciplinary use of Co-cited data**
  - Identified WoS Categories assigned to journals citing SEDAC data, and used Scopus®, journal titles, and publisher sites when Categories were not assigned
  - Paired assigned WoS Categories to Equivalent General Categories and Subjects
  - Identified Major Fields corresponding to assigned WoS Categories and Subjects
  - Normalized journals with WoS Categories, (WoK5.3) and Equivalent General Categories and Subject areas, and Field Definitions of Major Fields
  - Identified Categories, Subjects, and Major Fields of journals citing SEDAC data

Methodology adapted from Downs and Chen. 2015

Consistent Co-Citation of SEDAC Data with Remote Sensing Data

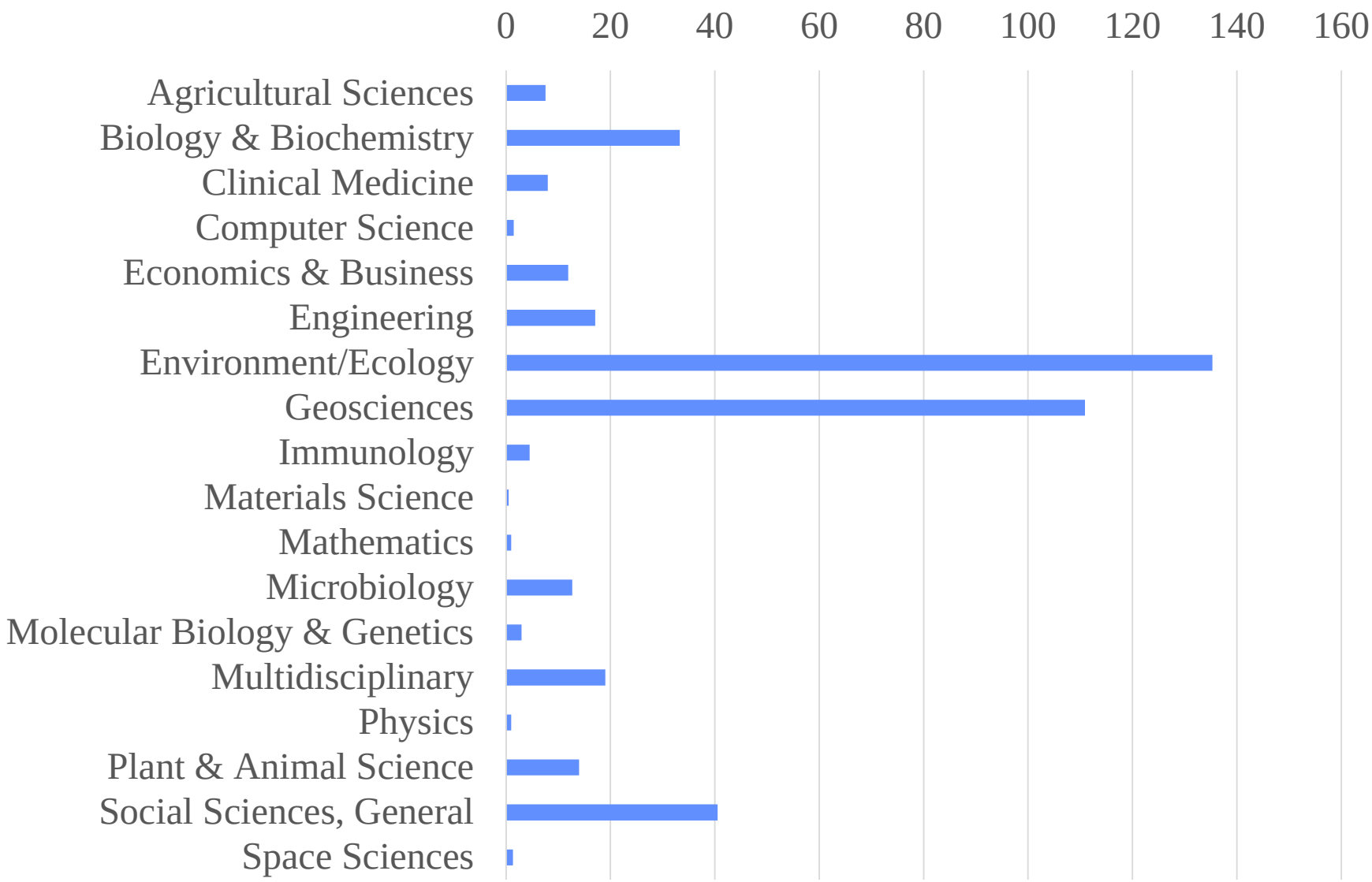


423 Articles Co-Citing SEDAC Data with Remote Sensing Data in 199 Journals, 2011-2016

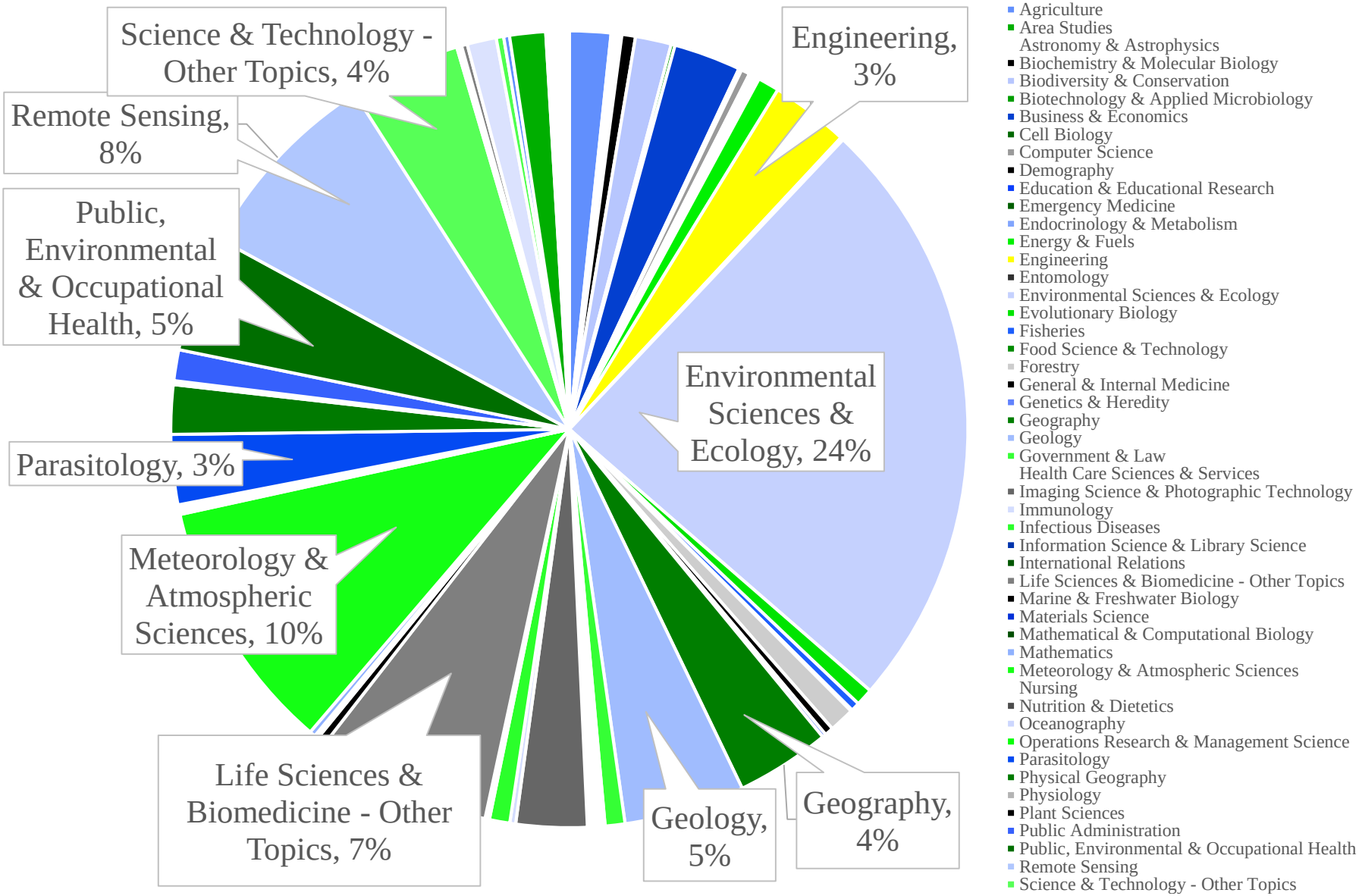


For context, all citations of SEDAC data from 2004-2013 were published in journals representing 216 categories, 92 subject areas, and 22 major fields.

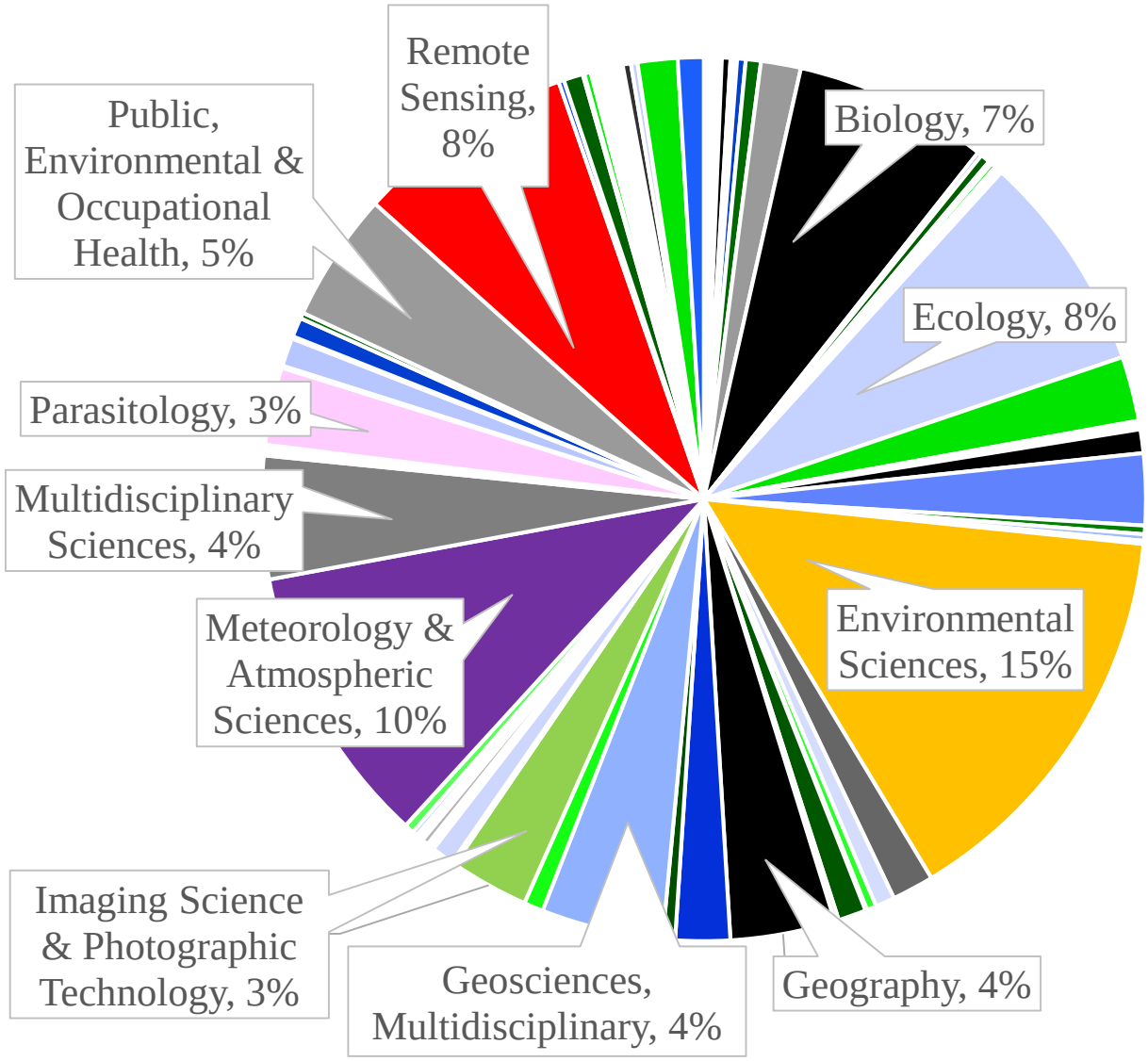
18 Major Fields of Journals with Articles Co-Citing SEDAC Data with Remote Sensing Data, 2011-2016



57 Subject Areas of Journals with Articles Co-Citing SEDAC Data with Remote Sensing Data, 2011-2016



73 Categories of Journals with Articles Co-Citing SEDAC Data with Remote Sensing Data, 2011-2016



- Agricultural Economics & Policy
- Agricultural Engineering
- Agriculture, Multidisciplinary
- Agronomy
- Area Studies
- Asthma
- Astronomy & Astrophysics
- Biochemistry & Molecular Biology
- Biodiversity Conservation
- Biology
- Biotechnology & Applied Microbiology
- Business, Finance
- Cell Biology
- Computer Science, Information Systems
- Computer Science, Interdisciplinary Applications
- Demography
- Ecology
- Economics
- Education, Scientific Disciplines
- Emergency Medicine
- Endocrinology & Metabolism
- Energy & Fuels
- Engineering, Environmental
- Engineering, Geological
- Engineering, Multidisciplinary
- Entomology
- Environmental Sciences
- Environmental Studies
- Evolutionary Biology
- Fisheries
- Food Science & Technology
- Forestry
- Genetics & Heredity
- Geography
- Geography, Physical
- Geology
- Geosciences, Multidisciplinary
- Health Care Sciences & Services
- Imaging Science & Photographic Technology
- Immunology
- Infectious Diseases
- Information Science & Library Science
- International Relations
- Law
- Limnology
- Marine & Freshwater Biology
- Materials Science, Multidisciplinary

Journals containing 6 or more Articles Co-Citing SEDAC Data and Remote Sensing Data, 2011-2016

Journal Title	Co-Citing Articles
PLoS ONE	24
Remote Sensing of Environment	17
Remote Sensing	14
Atmospheric Chemistry and Physics	10
Environmental Research Letters	10
International Journal of Remote Sensing	10
Environmental Science & Technology	9
Journal of Geophysical Research	9
Applied Geography	7
Geospatial Health	6
Global Ecology and Biogeography	6
International Journal of Applied Earth Observation and Geoinformation	6
Parasites & Vectors	6
Proceedings of the National Academy of Sciences	6

Observations

- Number of joint citation papers has not increased as quickly as the total number of SEDAC data citations from 2011-16
- Higher joint usage in natural science journals compared with social science journals (not correcting for total number of articles published)
- Multidisciplinary journals (PLoS ONE, PNAS) also publish papers that cite both types of data
- Some health journals also have numerous papers, mainly in area of mapping disease prevalence and vectors

Further Work

- What type and degree of integration is reflected by joint citations?
- Are there patterns in joint citations and integration over time?
- Are citation methods such as DOIs becoming more common?
- If so, do they have an impact on results?

Taxonomy of Data Citation
Cited but not used
Used as background or context
-- Used in figure only
Used in study design
-- Hypothesis/theory development
-- Sample selection
-- Other
Used in trend or spatial analysis
Used in statistical model
-- Statistically significant?
-- Total number of variables
Used in simulation model
-- Key component or variable
-- Minor variable or parameter
-- Baseline or boundary condition
Used for validation purposes
Used in research translation
-- Making results relevant to policy
-- Enabling use in applications
-- Cited in conclusion/discussion

Chen, Downs, and Schumacher, 2013

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