

Chapter 2

Data Foundation for Actionable Science



Ziheng Sun

Contents

1	Introduction.....	31
2	Data Categories and Availability for Actionable Science.....	33
2.1	Satellite Data.....	33
2.2	In Situ Data.....	36
2.3	Model Simulation Data.....	37
2.4	Citizen Science Data.....	41
2.5	Social Media Data.....	43
3	Data Discovery and Retrieval.....	45
4	Data Preprocessing and Cleaning.....	45
5	Data Integration and Management in Environmental Sciences.....	47
6	Continuous Operation and Maintenance of Data Stream.....	47
7	Challenges in Data Community to Support Actionable Science.....	49
8	Conclusion.....	50
9	Lessons Learnt.....	50
10	Open Questions and Brainstorming Solutions for Future.....	51
	References.....	51

1 Introduction

Good quality data is essential in climate and environment research as the foundation for making informed decisions and taking accurate action (Wiens et al. 2009). Our climate and environmental systems are complex and highly interconnected, and changes in one area can have cascading effects on others (Lawrence et al. 2020).

Z. Sun (✉)

Center for Spatial Information Science and Systems, Department of Geography and
Geoinformation Science, George Mason University, Fairfax, VA, USA

e-mail: zsun@gmu.edu

Accurate and comprehensive data is required for understanding these systems and predicting their behavior under different scenarios. Without reliable data, it is difficult to make climate and environmental policies, and forecast the potential consequences of those decisions. Climate and environmental research often involves large amounts of data from multiple sources, including remote sensing, in situ measurements, and modeling. The integration and analysis of these data require advanced computational tools and techniques, which can only be applied effectively if the data is high quality and properly documented. Poor quality or incomplete data can lead to inaccurate or biased results, which can have serious consequences if utilized in real-world decision-making and policy development (Ruiz-Benito et al. 2020). For example, consider the case of sea level rise, a concerning issue for coastal communities, and accurate predictions are essential for developing effective adaptation strategies. However, sea level is a complex variable that is influenced by many different factors, including thermal expansion, melting ice sheets and glaciers, and changes in ocean currents (Golledge 2020). To accurately model sea level rise, researchers must integrate data from multiple sources, including satellite measurements of ocean height, in situ measurements of ocean temperature and salinity, and models of ice sheet dynamics (Meyssignac et al. 2019; Cook et al. 2023). If any of this data is inaccurate or incomplete, it can lead to incorrect predictions of sea level rise and potentially disastrous consequences for coastal communities and public distrust of the coastal sciences.

Without accurate and reliable data, it is difficult to identify the root causes of environmental problems and develop effective solutions (Sun et al. 2021). Climate data can be sophisticated, large in scale, and come from various sources, such as satellites, weather stations, or citizen science projects. Managing, cleaning, and integrating the data is a daunting task, requiring significant expertise and resources. There is often a lack of standardized protocols and tools for collecting and processing data, and can lead to inconsistencies, making it difficult to compare and combine datasets from different sources (Zimmerman 2008). Issues often arise around data access and sharing, particularly when data is collected by government agencies or private companies. Access to data may be restricted due to privacy concerns, national security issues, or commercial interests, making it difficult for researchers to obtain the data they need to conduct their studies. Moreover, funding for climate and environmental research may be limited, leading to a lack of investment in necessary data infrastructure with incomplete or outdated datasets, which may not accurately reflect current conditions and trends.

Collecting and analyzing high-quality data requires specialized equipment, trained personnel, and infrastructure support. For example, a single oceanographic research vessel (Brett et al. 2020) can cost tens of millions of dollars, and operating costs can run into the millions per year. The cost of satellite missions can also be significant, with some missions costing several hundred million dollars. These costs can be a significant barrier to entry for researchers and institutions, particularly those in developing countries or with limited funding.

The chapter introduces the importance of collecting, storing, and processing data in an accessible, transparent, and replicable way. It overviews best practices in

environmental data management and interdisciplinary collaboration and open data sharing. For example, assume there is a research project aiming to assess the impacts of deforestation on biodiversity in a tropical rainforest. In order to effectively conduct this research, the scientists need access to high-quality data on the forest ecosystem, including information on the tree species, soil composition, water cycles, and animal populations (Giam 2017). This data must be collected and stored in a way that allows for easy access and analysis. To ensure the reliability of the data, the researchers must use standardized methods for collecting and analyzing the data, and they must carefully document their procedures to ensure transparency and replicability. They must also consider ethical considerations, such as obtaining informed consent from any individuals or communities affected by the research. Once the data is collected, it must be stored in a secure and accessible database that allows for easy sharing and collaboration among researchers. This promotes interdisciplinary collaboration and enables scientists from different disciplines to work together to address complex environmental challenges. In summary, the chapter emphasizes the importance of good quality data in actionable science for climate and environment, and provides guidance on best practices for data collection, storage, and processing. This is critical for ensuring that our efforts to address climate change and environmental degradation are based on sound scientific data and can effectively inform decision-making and policy development.

2 Data Categories and Availability for Actionable Science

Based on the sources and the collection methods, scientific datasets can be divided into the following categories: satellite data, in situ data, model simulation data, citizen science data, and social media data. This section will overview each category and analyze their current availability for actionable science.

2.1 *Satellite Data*

The sensors onboard satellites can provide a lot of information on land use, vegetation, ocean temperature, atmospheric composition, and many Earth's surface, ocean, and atmosphere processes. Some widely used satellite datasets include Moderate Resolution Imaging Spectroradiometer (MODIS) (Justice et al. 2002), Landsat series (Tucker et al. 2004), Sentinel series (Spoto et al. 2012), Suomi National Polar-orbiting Partnership (Suomi NPP) (Weng et al. 2012), ICESat (Schutz et al. 2005), Global Precipitation Measurement (GPM), SWOT (Biancamaria et al. 2016), etc. Table 2.1 lists some popular available satellite datasets that can be publicly retrieved and used for climate and environment research.

Many reasons draw scientists to usually first turn to satellite image datasets to look for those that can fit their research purposes. Satellites can cover vast areas of

Table 2.1 Available satellites for actionable science

Satellite name	Operator	Specs	Application
GRACE and GRACE-FO	NASA and DLR	Revisit: 30 days	Earth gravity field, hydrology, and ice
Landsat	USGS	Revisit: 16 days Resolution: 30 m	Land use and land cover, forest, and environmental studies
Sentinel-2	ESA	Revisit: 5 days Resolution: 10–20 m	Vegetation, crop, and land use change detection
MODIS (Aqua and Terra)	NASA	Revisit: 1–2 days Resolution: 250–1000 m	Atmospheric composition, land cover, sea surface temperature, and vegetation indices
GOES	NOAA	Resolution: 0.5 km Revisit: 5–15 min	Weather forecasting, storm, hurricane, and atmospheric studies
Suomi NPP	NASA	Resolution: 750 m Revisit: 16 days	Weather, climate, sea surface temperature, vegetation indices, and aerosol concentrations

the Earth's surface and provide data on a global scale (Wulder et al. 2008). It can monitor changes in the environment over time, identify patterns and trends, and make predictions about future conditions. While circling Earth by themselves (some satellites stay unmoving relative to Earth if they are in Earth's synchronous orbit) (Boain 2004), satellites can provide continuous monitoring of the environment over the years, which is especially important for identifying areas of high risk for natural disasters like floods, hurricanes, and wildfires. Another unbeatable advantage is that satellite data can cover remote areas difficult to access on foot or by vehicle and very important for monitoring changes in biodiversity, forest cover, and other critical ecosystem services. Also, for projects with a limited budget for data collection, publicly available satellite datasets could be a lifesaver. While satellite data can be expensive to acquire and process, it is often more cost-effective than ground-based monitoring. The data providers like NASA, NOAA, and USGS often preprocessed their datasets into different levels, and if scientists know about them and when to use them and how to connect the data to their data analytics, that would avoid a lot of duplicated work and is a financially wise choice (Ebert-Uphoff et al. 2017). For example, the Landsat satellites operated by NASA and the US Geological Survey have been providing images of the Earth's surface since 1972 (Wulder et al. 2016), and used for mapping land use, monitoring deforestation, and tracking changes in the cryosphere. The derived information has been used by governments, NGOs, and individuals to make informed decisions about land use, natural resource management, and climate change mitigation and adaptation strategies. Another example is the European Space Agency's Sentinel-1 satellite, which provides radar imagery to monitor changes in land cover, detect oil spills, and track changes in sea ice, and

disaster response efforts and help governments and NGOs respond to environmental emergencies.

The availability of satellite data varies depending on the mission and data type. Many government space agencies provide free satellite datasets for scientific and environmental research purposes. For example, NASA, the US Geological Survey (USGS), and the European Space Agency (ESA) all have missions that provide freely accessible data to the public. Generally, the low-to-medium resolution data are less sensitive and more easily available than the high-to-very-high resolution data. The Landsat program, a joint mission between NASA and USGS provides free and open access to over 40 years of satellite imagery data for land-use monitoring, natural resource management, and environmental monitoring and generated billions of dollars of benefits to the public and greatly advanced the Earth science developments. Other NASA missions that provide free data include the Moderate Resolution Imaging Spectroradiometer (MODIS), the Atmospheric Infrared Sounder (AIRS), and the Ozone Monitoring Instrument (OMI). The ESA also provides a range of freely available satellite data through their Sentinel missions, which are part of the Copernicus program (Thépaut et al. 2018). The Sentinel-1, Sentinel-2, and Sentinel-3 missions provide data on land cover, vegetation health, sea level, and sea ice. The Japan Aerospace Exploration Agency (JAXA) provides access to their Advanced Land Observing Satellite (ALOS) and the Global Precipitation Measurement (GPM) mission (Shimada et al. 2009). Compared to government-funded missions offering free satellite data, there are commercial satellite operators whose data will require a fee such as DigitalGlobe, Planet Labs, and DarkSky. These companies provide very-high-resolution (<1 m) satellite imagery for commercial and government purposes, such as disaster responses. In some cases, satellite data is completely restricted and only accessible to authorized users or purchasers due to national security concerns. For example, data from spy satellites operated by military and intelligence agencies will not (or never) be publicly available. China has a large and growing satellite program, including civilian satellites such as the Gaofen (Li et al. 2017) and ZY Earth observation satellites partially available for purchase, while others, such as the military reconnaissance satellites, are completely restricted and openly sharing is a very serious offense. Russia has a similar program like the Resurs-P Earth observation satellites. The India Remote Sensing program has launched several EO satellites like Resourcesat-2 (Rao et al. 2006), Cartosat-3, and Oceansat-3, and manages them via the Indian Space Research Organization. Several download portals are provided like the National Remote Sensing Centre (NRSC) and the Indian Earth Observation Data Gateway (IEODG). The Canadian Space Agency (CSA) Earth Observation program includes the RADARSAT Constellation satellites (Thompson 2015) and provides spatial data infrastructure (SDI). For scientific purposes, most datasets are good to use. However, please note that access to RADARSAT-2 and RCM data may be subject to licensing requirements and restrictions, and users may need to obtain permission or pay for commercial use.

2.2 *In Situ Data*

It refers to the data collected on the ground by in situ devices, such as weather station data, stream gauges, soil moisture probes, water quality sensors, carbon dioxide sensors, seismometers, spectrometers, GPS receivers, fluorimeters, acoustic Doppler current profilers, air quality sensors, and thermometers. These sensors are critical to measure temperature, humidity, air pressure, wind speed and direction, precipitation, water content in soil, etc. Table 2.2 presents a nonexhaustive list of the common sensor types and the variables they are designed to monitor.

You can see that all these data are highly valuable by providing detailed information about specific locations and thus are often used for validation and calibration of satellite data or model simulation as “ground truth.” A single sensor has too many uncertainties and could be inaccurate, and the data is unreliable without

Table 2.2 In situ sensor types and their target monitoring variables

Sensor type	Variables
Weather station	Temperature, humidity, air pressure, wind speed and direction, and precipitation
Soil moisture probe	Water content in soil
Stream gauge	Flow of water in rivers and streams
Water quality sensor	Water pH, dissolved oxygen, and turbidity
Carbon dioxide sensor	Atmospheric CO ₂ concentrations
Seismometers	Ground motion caused by earthquakes and other seismic events
Spectrometers	Spectrum of light reflected or emitted by the Earth’s surface
GPS receivers	Location, velocity, and direction of motion of objects
Fluorimeters	Fluorescence of chlorophyll in plants
Acoustic Doppler current profiler	Water velocity and direction in the ocean
Air quality sensor	Concentration of pollutants such as ozone, nitrogen oxides, and particulate matter in the atmosphere
Thermometers	Temperature
LiDAR	The height and structure of vegetation, terrain, and buildings
Flux towers	Exchanges of carbon, water, and energy between the Earth’s surface and the atmosphere
Buoys	Floating devices to monitor weather conditions, water temperature, salinity, and ocean currents
Ice core drills	Extract ice cores from glaciers and ice sheets
Tree cores	Extract cores from trees to study tree growth
Ocean gliders	Autonomous underwater vehicles to collect temperature, salinity, and currents
Animal tagging devices	Track the movements and behavior of animals in their natural habitats
Ground-penetrating radar	Subsurface geology and soil structure
Photometers	Amount of light in a specific location
Soil respiration chambers	Amount of carbon dioxide released by soil

professional calibration. To make solid and reliable observations, one common practice is for scientists to usually create ground observation networks with similar sensors and calibrate and clean the data into standard products. Ground observation networks can provide spatially extensive and long-term coverage of environmental parameters over large areas, and often use standardized protocols and equipment to ensure consistency and reliability of the data collected. Examples of in situ datasets include the Global Historical Climatology Network (GHCN) (Menne et al. 2012) and the National Water Quality Assessment (NAWQA) program (Gilliom et al. 1995). Table 2.3 contains the well-known ground observation networks we have set up so far.

These are just some dedicated well-known networks. Many other ground observation networks are also set up for many other purposes, such as Landsat has its own ground stations for satellite calibration. NASA operates the Near-Earth Network (NEN) (Schairer et al. 2016), a global network of ground stations that provides support for the agency's Earth-observing missions to transmit data from satellites to Earth. It is worthwhile to search for any available networks and reach out to the operator to check for availability.

2.3 *Model Simulation Data*

Besides the real observations collected remotely and in the field, there is a great amount of data generated by climate and environmental models, such as global circulation models (GCMs) (Dosio et al. 2015), regional climate models (RCMs) (Rummukainen 2010), and ecosystem models. These models can simulate the behavior of complex systems and provide projections of future climate and environmental conditions. Examples of model output datasets include the Coupled Model Intercomparison Project (CMIP) (Eyring et al. 2016) and the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns 2009). We list some commonly seen models in Table 2.4. Model-generated data can help make science more actionable by providing projections of future climate and environmental conditions based on different initial conditions of greenhouse gas emissions and other human activities. These projections can inform decision-making in various sectors, such as agriculture, water management, and infrastructure planning, by providing insights into potential future risks and opportunities. Compared to satellite and in situ data, model-generated data can provide detailed information on processes that cannot be directly measured and also give consistent, continuous, and long-term records of environmental conditions. Besides, model output does not have the hard restrictions on looking into the future and can make projections of future environmental conditions, allowing for better planning and decision-making. Another benefit is that models can fill in the gaps in spatial and temporal coverage where there are no satellite or in situ observations, and is a great resource to test hypotheses and validate observations. Satellite data provides global coverage and can monitor changes over time, but may have limited resolution and accuracy. In

Table 2.3 Ground observation network list

Name	Operator	Sensors	Variables	Availability	Coverage
Global Historical Climatology Network (GHCN)	NOAA	Weather stations	Temperature, precipitation, pressure, and humidity	GHCN-Monthly and GHCN-Daily are on the NCEI website	100,000 stations in 180 countries 1–175 years
National Water Quality Assessment (NAWQA)	USGS	Gauges	Nutrients, pesticides, sediment, water temperature, pH, and more	USGS National Water Information System (NWIS) website https://fluxnet.org/	Specific watersheds and aquifers in the United States 1970 to present
FLUXNET	Various	Eddy covariance towers	Carbon dioxide, water vapor, and energy fluxes between the land surface and atmosphere		Global coverage 1990 to present
OceansITES	Various	Moored buoys, profiling floats, gliders, etc.	Temperature, salinity, currents, nutrients, carbon dioxide, and more	OceansITES website	Global coverage 1990s to present
NEON	NSF	Various	Climate, geology, hydrology, and vegetation	NEON data portal	United States 2013 to present
ICOS	European research infrastructure consortium (ERIC)	Various sensors	Carbon dioxide, methane, and nitrous oxide	ICOS carbon portal	Europe 2015 – present
Ocean observatories initiative (OOI)	NSF	Moored, autonomous, and cabled sensors	Temperature, salinity, pressure, dissolved oxygen, pH, chlorophyll, turbidity, currents, and seismic activity	https://ooinet.oceanobservatories.org/	2014 – present Pacific and Atlantic oceans
International Soil Moisture Network (ISMN)	Various	Soil moisture sensors	Soil moisture at various depths	https://ismn.geo.tuwien.ac.at/	Global 2007 – present

Name	Operator	Sensors	Variables	Availability	Coverage
Snow Telemetry (SNOTEL)	USDA	Automatic monitoring station with many sensors	Snow water content, snow depth, and barometric pressure, soil moisture, solar radiation, etc.	https://www.nrcs.usda.gov/wps/portal/wcc/home/	Western United States, Alaska, and Puerto Rico 1978 – present
Atmospheric Radiation Measurement (ARM)	U.S. DOE	Various sensors	Atmospheric radiation and cloud properties	https://www.arm.gov/	The United States and internationally 1989 – present
Global Atmosphere Watch	World Meteorological Organization (WMO)	Various sensors	Greenhouse gas and air pollutants	https://public.wmo.int/en/programmes/global-atmosphere-watch-programme	>65 countries 1960s – present
AirNow	EPA	Various	Particulate matter, ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide	aimow.gov	The United States Real time
CHARM (Canadian High Arctic Research Station Monitoring)	Polar Knowledge Canada (POLAR)	Weather stations	Air quality, water quality, permafrost temperature, and biodiversity	https://arcticportal.org/maps/download	Canada's high arctic Real time

Table 2.4 A list of some example climate and environment models

Name	Provider	Description	Coverage
Coupled Model Intercomparison Project (CMIP)	WCRP	A framework for coordinated climate model experiments	Globe
North American Regional Climate Change Assessment Program (NARCCAP)	NARCCAP	A collaboration of climate modeling groups to produce high-resolution regional climate model simulations over North America	North America
Community Earth System Model (CESM)	NCAR	A global climate model that simulates the Earth's system by integrating atmosphere, ocean, land, and sea ice components	Globe
Global Land Data Assimilation System (GLDAS)	NASA	A land surface modeling system that assimilates observational data to produce a suite of land surface variables	Globe
Community Earth System Model Large Ensemble	NCAR	A large ensemble of simulations from the CESM climate model that provides a range of possible future climate scenarios	Globe

situ data provides high-resolution and accurate measurements of environmental parameters, but may have limited spatial coverage. Model data can fix both issues and produce both high spatial and temporal resolution datasets with the best continuity and completeness.

Despite the advantages, the existing models are still improving and cannot say they are already good enough (in many cases they are not). One common issue is that model results are calculated based on a simplified simulation of the real-world complex systems and may not fully capture all relevant processes or interactions. The results are sensitive to input parameters and theory assumptions, and contain biases or errors at local and regional scale, which can impact their effectiveness in decision-making. Also, the model outputs are sometimes confusing and hard to visualize and interpret, and not as easily accessible or understandable as satellite or in situ observations for nonexpert users. The uncertainty in numerical models mainly comes from three places: the model inputs, parameterization, and model structure. Also, the spatial and temporal resolution of numerical models is usually coarse because running complex numerical models can require significant computational resources, which may limit the ability to generate results at high resolution or with large ensembles of simulations. These uncertainties and drawbacks can affect the use of model output in real-world decision-making and planning. In particular, decision-makers may be hesitant to rely on model output when there are significant uncertainties in the results. Additionally, the limited spatial and temporal resolution of models may make it difficult to apply model results at the local scale, where decisions are often made.

The availability of model outputs is usually good as long as the model operators have a place to store the data in a publicly accessible server for people to download. It depends on the specific model and the producer. Some models have well-established processes for making their output available to the public, while others may have limited accessibility or require special permissions to access. Many numerical models used in climate and environmental science have established data distribution systems, such as the Earth System Grid Federation (ESGF) (Cinquini et al. 2014), which provides access to a wide range of model output data from the CMIP and other modeling efforts. However, there can still be challenges in accessing and using the data, particularly for nonexperts or those with limited computational resources. Additionally, the sheer volume of data generated by some models can make it difficult to store and distribute the data efficiently. The data publishers usually use a moving-window approach, meaning the data can be stored for a limited amount of time, such as 2 weeks or a month, before being automatically deleted to free up storage space. This approach ensures that the most recent and relevant data are available to users while also managing the storage and maintenance of the data repository. However, it is important to note that this approach may not be suitable for all applications, and the specific needs of each user group should be carefully considered before implementing any data management strategy.

2.4 *Citizen Science Data*

This includes data collected by members of the public, often through crowdsourcing and citizen science projects (Table 2.5). Citizen science data is valuable to make science actionable as it enables a large number of people to contribute to scientific research and monitoring. These projects not only increase the volume of data but also provide a way to engage and educate the public on environmental issues. Citizen science projects can collect data on a wide range of variables such as temperature, precipitation, air quality, water quality, and dust storms. Famous projects like the eBird project (Sullivan et al. 2014) encourage citizen scientists to collect data on bird sightings, and the iNaturalist project (Nugent 2018) invites people to collect data on species to study biodiversity. This data can be used to address some of the grand challenges, such as climate change, biodiversity loss, and habitat degradation. For example, USA-NPN (Betancourt et al. 2005) data can help to identify patterns and changes in phenology, the timing of plant and animal life cycle events, or help identify the areas where pollution levels are high, where invasive species are spreading. They can foster a sense of ownership and stewardship of the environment for the public, which is critical for ensuring long-term sustainability efforts and public engagement.

Similar to the other data sources, some disadvantages are commonly attached to citizen science datasets. Although most citizen science projects have well planned for all the details and keep protocols up for participants to follow to ensure the data quality, there are still many things that would go wrong in the fields. It is common

Table 2.5 A list of sample citizen science projects and their data topics

Name	Description	Data topic
eBird	A project by the Cornell Lab of Ornithology and the National Audubon Society that invites birdwatchers to contribute their observations to a global database of bird sightings	Bird species
Globe Observer	A project by NASA that invites citizen scientists to contribute data on cloud cover, land cover, and mosquito habitats	Cloud cover, land cover, and mosquito habitats
iNaturalist	A project by the California Academy of Sciences that invites people to document and share observations of plants, animals, and other living things	Species identification
AirVisual	A project by IQAir that invites people to contribute data on air quality by sharing information on air pollution in their area	Air quality
Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS)	A project that invites citizen scientists to measure and report precipitation data in their area	Precipitation
Old Weather	A project by the National Oceanic and Atmospheric Administration (NOAA) that invites people to transcribe historical weather data from ship logs dating back to the nineteenth century	Historical weather data
Citizen Weather Observer Program (CWOP)	A project that invites citizen scientists to contribute weather data from their personal weather stations	Weather data
Floating forests	A project by the University of California, Santa Barbara, that invites people to identify kelp forests from satellite imagery	Kelp forest identification
PhytoMap	A project by NASA's Goddard Space Flight Center that asks citizen scientists to help identify phytoplankton in ocean-color images	Phytoplankton distribution
Project BudBurst	A project by the Chicago Botanic Garden that asks citizen scientists to observe and record the timing of leafing, flowering, and fruiting of plants in their area	Vegetation growth
Snow Tweets	A project by the National Weather Service that asks citizens to report snowfall amounts in their area via Twitter	Snowfall
Dust Bowl Diary	A project by the University of Nebraska-Lincoln that asks citizens to transcribe diaries kept by people who lived through the Dust Bowl of the 1930s	Dust Bowl incidents
Hurricane Watch Net	A project by amateur radio operators that provides communication support during hurricanes	Meteorological data

(continued)

Table 2.5 (continued)

Name	Description	Data topic
PenguinWatch	A project by the University of Oxford that asks citizens to help identify penguins in camera trap images from the Antarctic	Penguin populations
IceWatch USA	A project by the University of Maine that asks citizens to observe and report on ice conditions in their area	Ice data
Earthwatch	A nonprofit organization that provides opportunities for citizens to participate in environmental research projects around the world	Climate change, wildlife conservation, and marine ecology
USA National Phenology Network (NPN)	A project engages citizen scientists in collecting data on phenology, or the timing of life cycle events such as flowering, leaf emergence, and migration	Phenology data including flowering, fruiting, leafing, migration, and hibernation

that citizen science data may not meet the same rigorous standards as data collected by professional scientists. Using citizen science could introduce uncertainty and errors in the analysis, make the models divert from the correct path, and can be difficult to draw valid conclusions. Another major issue is that the participants for one project might not be completely random (sampling bias) and distribution is biased toward certain areas or certain types of observations (Kosmala et al. 2016). For example, many citizen science projects have more data in urban areas, and lack of data in rural or remote areas. Compared to the ground observation network, citizen science projects mostly do not cover all regions or all variables of interest. Also, as citizen scientists have not gone through strict academic training, the participants do not have the same level of expertise as professional scientists and might lead to misunderstanding or wrong labels.

Despite these drawbacks, citizen science data can still be a valuable resource for actionable science in climate and environment. It is important to carefully consider the strengths and limitations of the data, and to use appropriate methods to analyze and interpret the results before considering fully adopting the data in serious actionable science projects. Almost all the citizen science datasets are freely available on their project website. The datasets are made available to the public with the goal of promoting scientific research, education, and environmental conservation. Some projects even make their data available in real time through web portals or mobile apps. However, some projects may have restrictions on data access, particularly if the data involves sensitive information or endangered species.

2.5 Social Media Data

Social media platforms are such powerful tools that have played a significant role in our society and touch all the aspects of our lives and the environment we are living in (Lewandowsky et al. 2019). It can be a significant tool to make science more

Table 2.6 A list of popular social media datasets and their application

Name	Launch	Active user	Media format	Open data policy
Facebook	2004	2.85 billion (2021)	Text, images, and videos	Restricted
Twitter (X)	2006	330 million (2019)	Text, images, and videos	Developer account for academia
Instagram	2010	1 billion (2021)	Images and videos	Developer API
YouTube	2005	2 billion (2021)	Videos	Developer API for research
Reddit	2005	52 million (2021)	Text, images, and videos	Developer API
TikTok	2016	1.1 billion (2021)	Short-form videos	Developer API for research
LinkedIn	2002	740 million (2021)	Text, images, and videos	Developer API for research

actionable in tackling climate and environmental challenges. It can monitor and track environmental events in real time, such as wildfires, floods, and storms, and help provide early warning and response to natural disasters, and mitigate their impact on communities. In addition, it can engage the public in discussions about climate and environmental issues, raise awareness and educate people about the impacts of climate change, and encourage them to take action. If a filter and algorithm is applied, social media could be another source to collect data on environmental conditions, such as air quality or hurricane wind speed. Similar to citizen science projects, scientists can use social media to find volunteers and allow them to engage in scientific research. The data is definitely the most important asset of the social media companies, from which a lot of useful information can be concluded on public opinions and behaviors related to climate and environmental issues. Table 2.6 contains the major social media platforms that are most popular at present and their open data policy for academic research.

Similar to citizen science datasets, social media data has issues with data quality, biases in sampling, and also some new problems like data privacy concerns, accessibility to the full-size datasets, and other legal and ethical considerations. For example, people may use different hashtags or keywords to describe the same phenomenon, or may post misleading information intentionally or unintentionally. During the 2018 California wildfires, social media users shared images and videos that were not related to the fires (Du et al. 2019), which led to confusion and hampered response efforts. Furthermore, social media may not capture environmental issues in areas with limited Internet access or low social media use. It tends to be more prevalent among younger and more tech-savvy individuals. The user demographics can vary across different regions and countries, which could lead to an incomplete picture of environmental or climate issues in those areas. The availability of social media datasets varies across platforms, but most of them have developer API opened for academic research purposes. Also, there are legal and ethical considerations associated with using social media data, such as copyright laws and data ownership, and the researchers need to consider all these factors before they decide to involve the data to do actionable science. The social media companies'

attitude toward the data openness also plays an important role, for example, Twitter recently changed its API policies, which has limited the availability of free historical tweet data for researchers (Bruns 2019).

3 Data Discovery and Retrieval

It is important to have a clear understanding of the scientific question or problem being addressed and the relevant data sources. To identify the data needs, we first need to scope the project's variables, the spatial and temporal extents, to make the data discovery process easy. For example, if the scientific question is related to understanding the impact of climate change on a particular species of plant, the relevant variables might include temperature, precipitation, and CO₂ concentrations. The spatial extent might be defined by the range of the plant species, while the temporal extent might be defined by the historical record of climate data. Once the data needs are defined, scientists need to search for the data to fill in them. There are various data portals, databases, and repositories that can provide access to climate and environmental data. Some example data portals are the National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information (NCEI), NASA Earth Observing System Data and Information System (EOSDIS) (Ramapriyan et al. 2010), and the European Centre for Medium-Range Weather Forecasts (ECMWF) Climate Data Store (Palmer et al. 1990).

After relevant data sources are identified, scientists need to assess their quality to ensure they meet the project's needs. Data quality assessment involves examining various factors like accuracy, completeness, consistency, and reliability. After the data quality assessment is complete, the next step is to access and retrieve the data. The process of data retrieval may vary depending on the data source, and data may be available in various formats like text, image, or binary. The data may be downloaded directly from the portal or may require registration and authentication. Once the data is retrieved, it needs to be stored and managed properly. This involves organizing the data, ensuring proper metadata and documentation, and implementing appropriate data backup and security measures.

4 Data Preprocessing and Cleaning

Data preprocessing and cleaning are essential in preparing data for actionable science to ensure that the data is accurate, complete, and in the right format for analysis. Otherwise, the data may contain errors, missing values, or inconsistencies that could affect the results of the analysis and ultimately the actions taken. Data preprocessing involves a variety of techniques to transform the data into a format suitable for analysis. This may include techniques such as data normalization, scaling, and feature extraction. Data cleaning involves identifying and correcting errors or

inconsistencies in the data, such as missing values, outliers, and duplicates. In the context of climate and environmental data, preprocessing and cleaning may involve correcting for measurement errors, identifying and filling in missing values, and removing outliers or erroneous data points. For example, in satellite data, preprocessing might involve correcting for atmospheric interference or adjusting for changes in instrument calibration over time. In citizen science data, cleaning might involve identifying and removing data points that are clearly incorrect or identifying patterns of data that suggest errors or biases. Here are some common steps:

- Check for missing values, incomplete data or gaps in the time-series data.
- Ensure that the data is of high quality and free from errors, outliers, or any inconsistencies that may affect the analysis.
- When using different sources of data, they may have different scales and units. Normalization and scaling techniques can be used to convert the data to a standard scale, facilitating comparisons between variables.
- The data should be formatted consistently to make it easier to read and understand. For example, consistent date formats and time zones should be used to make sure the data is easily comparable and to avoid errors.
- Transformations like log, power, and square root can be used to reduce the influence of outliers and make the data more normally distributed, which can make statistical analysis more reliable.
- Sometimes datasets may be too large to handle or may contain unnecessary data. In such cases, data reduction techniques like principal component analysis and singular value decomposition or clustering can be used to reduce the dimensionality of the data and remove noise.
- Document all the preprocessing and cleaning steps taken, including any decisions made, code used, and any changes made to the original data. The documentation is essential for reproducibility and transparency.
- Ensure that all the data layers have the same spatial reference system (SRS) and projection. Mismatched SRS can cause alignment and accuracy issues.
- The spatial resolution of the data should be appropriate for the analysis being performed. For example, if analyzing changes in land cover, a coarse resolution may not be suitable.
- The data format should be compatible with the software being used for analysis. Common geospatial data formats include shapefiles, GeoTIFFs, and netCDF.
- Missing or null values should be identified and handled appropriately. They can be replaced with interpolated values or removed depending on the context of the analysis.
- Outliers should be identified and evaluated for their impact on the analysis. They can be removed or adjusted if necessary.
- Different data layers should be integrated appropriately, accounting for differences in resolution, scale, and data type.
- Quality control checks should be performed to ensure data accuracy and consistency, like comparing data with ground truth observations or other reference data sources.

- Metadata should be reviewed and updated as necessary to ensure that data is properly documented and can be easily understood and reused by others.
- Version control should be used to track changes to the data over time, especially if the data is being updated or modified regularly.

The expected output of data preprocessing and cleaning is a clean and consistent dataset that can be directly used for analysis and modeling. The cleaned dataset should be free of missing values, outliers, duplicates, and errors, and have consistent formatting and units. Once the data has been preprocessed and cleaned, it can be directly used for analysis and modeling. This saves time and resources that would otherwise be spent manually addressing data quality issues during analysis. Additionally, a clean and consistent dataset enables scientists to compare and combine data from multiple sources more easily, which can lead to more comprehensive and insightful analysis.

5 Data Integration and Management in Environmental Sciences

Data integration involves combining data from multiple sources to produce a comprehensive dataset. Data management usually refers to the storage, retrieval, and manipulation of large volumes of data. Data integration and management is usually comprised of several steps: data preparation, data transformation, and data fusion. Data preparation involves cleaning and preprocessing data to ensure that it is accurate, complete, and consistent. Data transformation is responsible for converting data into a format that is compatible with the analysis tools and methods that will be used. Data fusion means combining data from different sources to produce a comprehensive dataset that can be used for analysis.

Data management is critical for ensuring that data is accessible and usable over the long term. Common practices involve the use of data repositories and archives that can store large volumes of data and provide access to it over time. It also uses standards and protocols for data formatting, storage, and sharing to ensure that data is interoperable and can be used by others. Metadata is widely used to provide information about the data, such as the location, time, and method of collection, and helps to ensure that the data is accurately interpreted and used. Metadata also facilitates data discovery and sharing as it allows other researchers to understand the context and quality of the data.

6 Continuous Operation and Maintenance of Data Stream

Continuous operation and maintenance of data streams are essential for ensuring the quality and reliability of data used in actionable science for climate and the environment (Becker et al. 2015). Data streams are typically collected over long periods of

time and from multiple sources, and it is important to ensure that the data is continuously monitored, validated, and maintained to ensure its accuracy and consistency.

To achieve continuous operation and maintenance, several steps can be taken:

- Data streams need to be monitored on a continuous basis to ensure that they are operating correctly and that the data being collected is valid and reliable. This can be done using various monitoring tools, such as dashboards, alerts, and automated checks.
- Data quality control involves reviewing the data for errors, inconsistencies, and outliers, and taking corrective action if necessary. This can involve data cleaning, data validation, and data verification to ensure that the data is accurate and reliable.
- Regular maintenance and upkeep of data streams are critical to ensure that the data is continuously available and up to date. This includes ensuring that the data collection system is functioning correctly, replacing any faulty equipment, and ensuring that the data is backed up and securely stored.
- Continuous improvement includes identifying and addressing any gaps or issues in the data collection process and making improvements to ensure that the data is of the highest quality possible. This can involve improving data collection methods, upgrading equipment, and implementing new data analysis techniques.

For example, the Global Forest Watch provides near-real-time information on forest loss and gain around the world (Harris et al. 2016) using satellite imagery and data from local sources. The information is updated monthly and can be used to inform policies and interventions to prevent deforestation and promote reforestation. Another project like Smartfin involves equipping surfboards with sensors to collect data on ocean conditions such as temperature, salinity, and pH. The data is uploaded to a cloud-based platform in near real time and can be used by researchers and policymakers to better understand the impacts of climate change on the ocean. Also, many cities around the world have installed air quality sensors to continuously monitor the levels of pollutants such as particulate matter, ozone, and nitrogen dioxide. Policymakers can use the data to identify areas of high pollution and implement interventions to improve air quality.

However, the challenges for maintaining continuous data streams include

- Regularly updating data sources requires regular calibration, validation, and verification of the data, as well as monitoring and detection of any anomalies or errors that may occur.
- Hard to manage the large volume of data generated by continuous data streams. The data needs to be stored, processed, and analyzed in a timely and efficient manner, and this requires a robust data management infrastructure.
- The technical infrastructure required for continuous data streams can be complex and expensive to maintain. The equipment and sensors used to collect the data need to be regularly maintained and calibrated to ensure accurate and reliable data. Additionally, data transmission and storage systems need to be reliable and secure to prevent data loss or corruption.

- It requires a significant investment in funding and resources. This includes the cost of equipment, infrastructure, and personnel. Funding may be limited, and there may be a shortage of skilled personnel available to manage and operate the data streams.
- Continuous data streams are often collected from multiple sources, and integrating this data into a single usable format can be challenging. Data may be collected at different resolutions, time intervals, and spatial scales, and may need to be normalized or transformed before integration.
- Making data accessible and sharing it with others can also be a challenge. Data owners need to ensure that appropriate data sharing agreements are in place, and that the data is properly documented and annotated with metadata to facilitate its use by others.

7 Challenges in Data Community to Support Actionable Science

From the data provider perspective, challenges in supporting actionable science in climate and environment include:

- Data providers need to make sure that the data is reliable, accurate, and accessible to researchers and stakeholders via investing in data infrastructure and maintenance, data documentation, and data standardization efforts.
- Consider the different data formats, standards, and metadata used by various data sources and ensure that their data can be integrated with other datasets.
- Check that the to-be-released datasets are compliant with privacy regulations and that the data they provide is secure.
- Facilitate data sharing and collaboration among researchers and stakeholders. This can involve developing policies and incentives to encourage data sharing and collaboration, as well as developing platforms and tools that enable data sharing and collaboration.
- Secure long-term funding to support data infrastructure, maintenance, and management. This involves developing funding models that ensure sustainable data provision and management, as well as engaging stakeholders to secure support for ongoing data management efforts.
- Keep pace with advances in technology and new data sources. This can involve investing in new data collection methods and technologies, as well as updating data storage and management systems to accommodate new data types and formats.

8 Conclusion

In this chapter, we discussed the importance of data in actionable science for climate and environment and covered various aspects, including data sources, data discovery, retrieval, preprocessing, cleaning, integration, and continuous operation and maintenance. We also discussed the challenges faced by the climate data community in supporting actionable science. We highlighted the crucial role of data in actionable science, enabling researchers and decision-makers to make informed decisions about climate and environmental challenges. Different data sources available like citizen science data and social media data are introduced. The details of various data portals, databases, and repositories that can provide access to climate and environmental data are also provided.

The importance of metadata in data integration and management, providing essential information about the data and ensures accurate interpretation and use, is also mentioned. The challenges in continuous operation and maintenance of data streams and the importance of addressing them for the success of actionable science are discussed. In addition, the challenges faced by the data community in supporting actionable science are outlined too, including the need for interdisciplinary collaboration, data quality and consistency, data sharing, and data security and privacy. In the future, it is foreseeable mandatory path for entire climate science community to collectively address these challenges to ensure that actionable science can continue to provide innovative and practical solutions to climate and environmental challenges.

9 Lessons Learnt

- Data collection and preprocessing could be tedious and challenging for actionable science that relies on accurate and reliable data to achieve successful implementation of climate and environmental actions.
- Citizen science data can provide valuable insights but may come with limitations and challenges that need to be addressed before use in actionable science.
- Metadata is essential for data management and sharing. Properly documented metadata helps ensure data accuracy, reproducibility, and facilitates data discovery and sharing (Sun et al. 2013).
- Continuous operation and maintenance of data streams are important to ensure long-term availability of data and prevent data degradation.
- Collaboration between data providers and end users is critical for actionable science. End users should also be involved in the process to ensure that the data meet their specific requirements.
- Open data policies and data sharing platforms facilitate the access and sharing of data.

- Data integration and management enable the combination of data from different sources, providing a comprehensive and holistic view of climate and environmental issues.
- The challenges in data management and sharing require interdisciplinary collaboration and coordination between scientists, policymakers, and data providers to work closely on.

10 Open Questions and Brainstorming Solutions for Future

As we move toward a more data-driven approach to tackling climate and environmental challenges, there are still several open questions and areas for improvement in the data foundation for actionable science. Here are some potential solutions and ideas in the future:

- While there are many open data portals available, not all datasets are easily accessible in a user-friendly manner. We need to work toward improving open data access and promoting standardization in data formats and utilization.
- To make the most out of the data available, we need to focus on integrating and fusing multi-source data. This can be challenging due to the differences in data formats and metadata, and efforts are much needed to improve data interoperability.
- The field of data science is constantly evolving, and emerging technologies such as machine learning and artificial intelligence can be utilized to quickly extract insights from large datasets. It is important for data providers to stay up to date with emerging technologies and incorporate them into actionable science projects.
- Citizen science can be a valuable source of data, especially in areas where official data collection is difficult or expensive. To engage more citizen scientists, we need to improve data literacy and promote open data access and friendly collection toolkits.
- As more data becomes available, it is mandatory to strengthen data management and sharing practices. This includes developing common metadata standards and promoting open data sharing policies.

References

- Becker, David, Trish Dunn King, and Bill McMullen. 2015. Big data, big data quality problem. In *2015 IEEE international conference on big data (big data)*, 2644–2653. IEEE.
- Betancourt, Julio L., Mark D. Schwartz, David D. Breshears, Daniel R. Cayan, Michael D. Dettinger, David W. Inouye, Eric Post, and Bradley C. Reed. 2005. Implementing a US national phenology network. *Eos Trans. AGU* 86(51), 539–539, <https://doi.org/10.1029/2005EO510005>

- Biancamaria, Sylvain, Dennis P. Lettenmaier, and Tamlin M. Pavelsky. 2016. The SWOT mission and its capabilities for land hydrology. *Remote Sensing and Water Resources*: 117–147.
- Boain, Ronald J. 2004. AB-Cs of sun-synchronous orbit mission design. 14 AAS/AIAA Space Flight Mechanics Conference, Maui, Hawaii, February 8–12, <https://citeseeerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=43ffd02cbb61b0381f6b4ae3e18d0c9a536d446a>
- Brett, Annie, Jim Leape, Mark Abbott, Hide Sakaguchi, Ling Cao, Kevin Chand, Yimnang Golbuu, Tara J. Martin, Juan Mayorga, and Mari S. Myksovoll. 2020. Ocean data need a sea change to help navigate the warming world. *Nature* 582 (7811): 181–183.
- Bruns, Axel. 2019. After the ‘APIcalypse’: Social media platforms and their fight against critical scholarly research. *Information, Communication & Society* 22 (11): 1544–1566.
- Cinquini, Luca, Daniel Crichton, Chris Mattmann, John Harney, Galen Shipman, Feiyi Wang, Rachana Ananthakrishnan, et al. 2014. The earth system grid federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems* 36: 400–417.
- Cook, Sue, Keith W. Nicholls, Irena Vaňková, Sarah S. Thompson, and Benjamin K. Galton-Fenzi. 2023. Data initiatives for ocean-driven melt of Antarctic ice shelves. *Annals of Glaciology*: 1–6.
- Dosio, Alessandro, Hans-Jürgen Panitz, Martina Schubert-Frisius, and Daniel Lüthi. 2015. Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: Evaluation over the present climate and analysis of the added value. *Climate Dynamics* 44: 2637–2661.
- Du, Hanxiang, Long Nguyen, Zhou Yang, Hashim Abu-Gellban, Xingyu Zhou, Wanli Xing, Guofeng Cao, and Fang Jin. 2019. Twitter vs news: Concern analysis of the 2018 California wildfire event. In *In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, 207–212. IEEE.
- Ebert-Uphoff, Imme, David R. Thompson, Ibrahim Demir, Yulia R. Gel, Anuj Karpatne, Mariana Guereque, Vipin Kumar, Enrique Cabral-Cano, and Padhraic Smyth. 2017. A vision for the development of benchmarks to bridge geoscience and data science. In *17th international workshop on climate informatics*.
- Eyring, Veronika, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* 9 (5): 1937–1958.
- Giam, Xingli. 2017. Global biodiversity loss from tropical deforestation. *Proceedings of the National Academy of Sciences* 114 (23): 5775–5777.
- Gilliom, Robert J., William M. Alley, and Martin E. Gurtz. 1995. *Design of the National Water-Quality Assessment Program: Occurrence and distribution of water-quality conditions*. Vol. 1112. US Geological Survey.
- Golledge, Nicholas R. 2020. Long-term projections of sea-level rise from ice sheets. *Wiley Interdisciplinary Reviews: Climate Change* 11 (2): e634.
- Harris, Nancy, Rachael Petersen, Crystal Davis, and Octavia Payne. 2016. “INSIDER: Global Forest Watch and the Forest Resources Assessment, Explained in 5 Graphics.” World Resources Institute. Accessed on Sep 8, 2023. <https://www.wri.org/technical-perspectives/insider-global-forestwatch-and-forest-resources-assessment-explained-5-graphics>
- Justice, C.O., J.R.G. Townshend, E.F. Vermote, E. Masuoka, R.E. Wolfe, Nazmi Saleous, D.P. Roy, and J.T. Morisette. 2002. An overview of MODIS land data processing and product status. *Remote Sensing of Environment* 83 (1–2): 3–15.
- Kosmala, Margaret, Andrea Wiggins, Alexandra Swanson, and Brooke Simmons. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14 (10): 551–560.
- Lawrence, Judy, Paula Blackett, and Nicholas A. Cradock-Henry. 2020. Cascading climate change impacts and implications. *Climate Risk Management* 29: 100234.
- Lewandowsky, Stephan, John Cook, Nicolas Fay, and Gilles E. Gignac. 2019. Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition* 47: 1445–1456.

- Li, Feng, Lei Xin, Yi Guo, Dongsheng Gao, Xianghao Kong, and Xiuping Jia. 2017. Super-resolution for GaoFen-4 remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 15 (1): 28–32.
- Mearns, Linda O. 2009. North American Regional Climate Change Assessment Program (NARCCAP). In *Presentation at workshop on water infrastructure sustainability and adaptation to climate change, USEPA*. Slide presentation available at: http://www.epa.gov/nrmrl/wswrd/wqm/wrap/pdf/workshop/A1_Mearns.pdf.
- Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. 2012. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* 29 (7): 897–910.
- Meyssignac, Benoit, Tim Boyer, Zhongxiang Zhao, Maria Z. Hakuba, Felix W. Landerer, Detlef Stammer, Armin Köhl, et al. 2019. Measuring global ocean heat content to estimate the earth energy imbalance. *Frontiers in Marine Science* 6: 432.
- Nugent, Jill. 2018. iNaturalist. *Science Scope* 41 (7): 12–13.
- Palmer, T.N., C. Brankovic, F. Molteni, S. Tibaldi, L. Ferranti, A. Hollingsworth, U. Cubasch, and E. Klinker. 1990. The European Centre for Medium-range Weather Forecasts (ECMWF) program on extended-range prediction. *Bulletin of the American Meteorological Society* 71 (9): 1317–1330.
- Ramapriyan, Hampapuram K., Jeanne Behnke, Edwin Sofinowski, Dawn Lowe, and Mary Ann Esfandiari. 2010. Evolution of the earth observing system (EOS) data and information system (EOSDIS). *Standard-Based Data and Information Systems for Earth Observation*: 63–92.
- Rao, M., J.P. Venkata, Ram Rattan Gupta, and K. Thyagarajan. 2006. RESOURCESAT-2: A mission for earth resources management. In *GEOSS and next-generation sensors and missions*, vol. 6407, 119–126. SPIE.
- Ruiz-Benito, Paloma, Giorgio Vacchiano, Emily R. Lines, Christopher P.O. Reyer, Sophia Ratcliffe, Xavier Morin, Florian Hartig, et al. 2020. Available and missing data to model impact of climate change on European forests. *Ecological Modelling* 416: 108870.
- Rummukainen, Markku. 2010. State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change* 1 (1): 82–96.
- Schaire, Scott, Bill Horne, James Schier, Harry Shaw, George Bussey, Peter Celeste, Yen F. Wong, et al. 2016. NASA near earth network (NEN) and space network (SN) CubeSat communications. In *14th international conference on space operations*, 2598.
- Schutz, Bob E., H. Jay Zwally, Christopher A. Shuman, D. Hancock, and John P. DiMarzio. 2005. Overview of the ICESat mission. *Geophysical Research Letters* 32 (21).
- Shimada, Masanobu, Takeo Tadono, and Ake Rosenqvist. 2009. Advanced Land Observing Satellite (ALOS) and monitoring global environmental change. *Proceedings of the IEEE* 98 (5): 780–799.
- Spoto, Francois, Omar Sy, Paolo Laberinti, Philippe Martimort, Valerie Fernandez, Olivier Colin, Bianca Hoersch, and Aime Meygret. 2012. Overview of sentinel-2. In *2012 IEEE international geoscience and remote sensing symposium*, 1707–1710. IEEE.
- Sullivan, Brian L., Jocelyn L. Aycrigg, Jessie H. Barry, Rick E. Bonney, Nicholas Bruns, Caren B. Cooper, Theo Damoulas, et al. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation* 169: 31–40.
- Sun, Ziheng, Peng Yue, Hu Lei, Jianya Gong, Liangpei Zhang, and Lu Xianchang. 2013. GeoPWProv: Interleaving map and faceted metadata for provenance visualization and navigation. *IEEE Transactions on Geoscience and Remote Sensing* 51 (11): 5131–5136.
- Sun, Ziheng, Liping Di, Hui Fang, Liying Guo, Xicheng Tan, Lili Jiang, and Zhongxin Chen. 2021. Agro-geoinformatics data sources and sourcing. *Agro-Geoinformatics: Theory and Practice*: 41–66.
- Thépaut, Jean-Noël, Dick Dee, Richard Engelen, and Bernard Pinty. 2018. The Copernicus programme and its climate change service. In *IGARSS 2018–2018 IEEE international geoscience and remote sensing symposium*, 1591–1593. IEEE.

- Thompson, Alan A. 2015. Overview of the RADARSAT constellation mission. *Canadian Journal of Remote Sensing* 41 (5): 401–407.
- Tucker, Compton J., Denelle M. Grant, and Jon D. Dykstra. 2004. NASA's global orthorectified Landsat data set. *Photogrammetric Engineering & Remote Sensing* 70 (3): 313–322.
- Weng, F., X. Zou, X. Wang, S. Yang, and M.D. Goldberg. 2012. Introduction to Suomi national polar-orbiting partnership advanced technology microwave sounder for numerical weather prediction and tropical cyclone applications. *Journal of Geophysical Research: Atmospheres* 117 (D19).
- Wiens, John A., Diana Stralberg, Dennis Jongsomjit, Christine A. Howell, and Mark A. Snyder. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences* 106 (Supplement_2): 19729–19736.
- Wulder, Michael A., Joanne C. White, Samuel N. Goward, Jeffrey G. Masek, James R. Irons, Martin Herold, Warren B. Cohen, Thomas R. Loveland, and Curtis E. Woodcock. 2008. Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment* 112 (3): 955–969.
- Wulder, Michael A., Joanne C. White, Thomas R. Loveland, Curtis E. Woodcock, Alan S. Belward, Warren B. Cohen, Eugene A. Fosnight, Jerad Shaw, Jeffrey G. Masek, and David P. Roy. 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment* 185: 271–283.
- Zimmerman, Ann S. 2008. New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values* 33 (5): 631–652.