Ecological Informatics 29 (2015) 33–44

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Ecological data sharing

William K. Michener *

College of University Libraries & Learning Sciences, MSC04 2815, University of New Mexico, NM 87131-0001, United States

ARTICLE INFO

Article history:
Received 4 March 2015
Received in revised form 18 June 2015
Accepted 24 June 2015
Available online 2 July 2015

Keywords:
Data publication
Data sharing
Information technology
Metadata
Open access
Policy

ABSTRACT

Data sharing is the practice of making data available for use by others. Ecologists are increasingly generating and sharing an immense volume of data. Such data may serve to augment existing data collections and can be used for synthesis efforts such as meta-analysis, for parameterizing models, and for verifying research results (i.e., study reproducibility). Large volumes of ecological data may be readily available through institutions or data repositories that are the most comprehensive available and can serve as the core of ecological analysis. Ecological data are also employed outside the research context and are used for decision-making, natural resource management, education, and other purposes. Data sharing has a long history in many domains such as oceanography and the biodiversity sciences (e.g., taxonomic data and museum specimens), but has emerged relatively recently in the ecological sciences.

A review of several of the large international and national ecological research programs that have emerged since the mid-1900s highlights the initial failures and more recent successes as well as the underlying causes—from a near absence of effective policies to the emergence of community and data sharing policies coupled with the development and adoption of data and metadata standards and enabling tools. Sociocultural change and the move towards more open science have evolved more rapidly over the past two decades in response to new requirements set forth by governmental organizations, publishers and professional societies. As the scientific culture has changed so has the cyberinfrastructure landscape. The introduction of community-based data repositories, data and metadata standards, software tools, persistent identifiers, and federated search and discovery have all helped promulgate data sharing. Nevertheless, there are many challenges and opportunities especially as we move towards more open science. Cyberinfrastructure challenges include a paucity of easy-to-use metadata management systems, significant difficulties in assessing data quality and provenance, and an absence of analytical and visualization approaches that facilitate data integration and harmonization. Challenges and opportunities abound in the sociocultural arena where funders, researchers, and publishers all have a stake in clarifying policies, roles and responsibilities, as well as in incentivizing data sharing. A set of best practices and examples of software tools are presented that can enable research transparency, reproducibility and new knowledge by facilitating idea generation, research planning, data management and the dissemination of data and results.

Contents

* Tel.: +1 505 220 3123; fax: +1 505 277 2541.
E-mail address: william.michener@gmail.com.

http://dx.doi.org/10.1016/j.ecoinf.2015.06.010

1. Introduction

Data sharing is the practice of making data available for use by others. Ecologists are increasingly generating and sharing immense amounts of data as part of the research enterprise. The data are derived from direct human observations in the field and recorded in notebooks and other media, laboratory observations, remote and in situ sensors, and instruments that are employed to measure particular attributes of biota (e.g., presence, temperature) and the physical environment (e.g., air, soil, water) such as rainfall, solar radiation, soil moisture, and pH. Ecologists often use shared data that originate from other scientists for comparative purposes or to augment their data collections, for synthesis efforts such as meta-analysis, for parameterizing models, and for verification of results (i.e., study reproducibility). In some cases, shared data may be the only data or the best data that are readily available. Data are also used outside the research context. Many non-researchers use available data for decision-making, natural resource management, education, and other purposes.

Some science domains such as oceanography and taxonomy have a relatively long tradition of data sharing. For example, the International Oceanographic Data and Information Exchange of the Intergovernmental Oceanographic Commission (IOC) of UNESCO was established in 1961 to facilitate the international exchange of oceanographic data and information exchange (http://www.iode.org). The IOC has enabled the creation of more than 80 oceanographic data centers in IOC countries.

Data sharing in ecology, on the other hand, has evolved slowly and is only now becoming common practice. In this paper, I first describe the history of data sharing in ecology, primarily focusing on several of the large international and national (primarily USA) ecological research programs that have emerged since the mid-1900s. Second, I examine the sociological aspects of data sharing, especially the perceived impediments and benefits, and review the role of societies, funders, and journals in changing the culture of data sharing. Third, I review the role of cyberinfrastructure in supporting data sharing including data repositories, software tools, persistent identifiers, and federated search and discovery. Last, I discuss the future of data sharing and conclude with a set of best practices for sharing ecological data.

2. Ecological data and a brief history of data sharing

In a review of historic ecological data, Bowser (1994) categorized ecological data into three types: (1) planned—i.e., well-planned and well-documented long-term data such as the long-term records of atmospheric CO_2 from Mauna Loa, Hawaii (Keeling et al., 1976) and the Hubbard Brook watershed studies in New Hampshire (Likens et al., 1977) that were relatively rare at the time; (2) opportunistic—i.e., data that are collected to achieve short-term goals over a discrete funding period and are commonly encountered in the literature; and (3) serendipitous—i.e., data that are not for testing a scientific hypothesis such as weather data collected by private citizens, fish and wildlife harvest data, and other types of data. Bowser (1994) recounted efforts that began in 1979 at the North Temperate Lakes Long-Term Ecological Research site to retrieve and use data previously collected in Wisconsin lakes including the data sets generated in the pioneering limnology studies by Birge and Juday that led to more than 400 publications over a period of seven decades (see Juday and Hasler, 1946). Bowser (1994) summarized the state of the historic data as:

> "*The scope, degree of documentation, quality, and availability of different data sets varies widely. Both published and unpublished data sets have strengths and weaknesses. Data discontinuity, whether from single or multiple sources, makes data calibration difficult. Quality control is uneven, at best, and is often undocumented. Instrumentation changes have been rapid and intercalibration with new techniques is not practiced as commonly as would be hoped.*"

Such data challenges are not unexpected in an emerging, but relatively young scientific discipline. Prior to and during the first half of the 20th century, individuals or a small number of researchers performed most ecology studies over a short time period and with limited funding. Other than the data published as tables in a manuscript, data sharing was not the norm. Few, if any, data collection and data management standards existed or were followed for documenting (i.e., ascribing metadata), quality assuring (i.e., quality assurance/quality control; QA/QC), and organizing (i.e., database management) data. This situation began to change in the 1960s in response to the emergence of "big ecology" (sensu Coleman, 2010) programs that followed in the footsteps of the International Geophysical Year of 1957–58, an international earth sciences research effort that included a focus on meteorology and oceanography.

Coleman (2010) provides a detailed history of many of the large ecological and environmental research programs from the 1950s through the present including the International Biological Program (IBP), the Long-Term Ecological Research Program (LTER) and International LTER Program (I-LTER), and the National Center for Ecological Analysis and Synthesis (NCEAS). The timeline and characteristics of these and other programs that extend to the present day (i.e., Global Biodiversity Information Facility (GBIF), National Ecological Observatory Network (NEON), and Ocean Observatories Initiative (OOI)) are presented in Table 1. The included programs are similar in that the U.S. National Science Foundation partially or wholly funded them and they reflect the transition from short-term (i.e., 1–3 years), low-cost, minimally-staffed projects to long-term (i.e., decade or longer), high-cost, multi-institutional and multi-national projects that serve a large group of

Table 1
Timeline and characteristics of large ecological and related environmental programs since the 1960s (organized chronologically).

Program name	Timeline	Location	Characteristics of program	Website (if available) and references
IBP	1964–1974	International	Large-scale ecosystem ecology studies in multiple biomes funded through several sources	Hagen (1992); Coleman (2010)
LTER	1980–ongoing	USA, Antarctica	Ongoing, multi-decadal ecology studies in >24 ecosystems	http://www.lternet.edu/; Coleman (2010); Michener and Waide (2009)
I-LTER	1993–ongoing	International; 40 member countries or regions	Long-term, site-based research and monitoring in different ecosystems	http://www.ilternet.edu/; Coleman (2010); Michener and Waide (2009)
NCEAS	1995–ongoing	University of California center that hosts USA and international scientists	Analysis and synthesis of existing ecological data	https://www.nceas.ucsb.edu/; Hackett et al. (2009); Hampton and Parker (2011)
GBIF	2001–ongoing	Denmark	Center that provides free and open access to biodiversity data worldwide	http://www.gbif.org
NEON	2006–ongoing	USA	Networked sites in terrestrial and aquatic ecosystems across 20 USA ecoclimatic domains	http://www.neoninc.org/; Schimel et al. (2011)
OOI	2009–ongoing	Coastal, regional and global	Networked infrastructure of sensor systems that measure the physical, chemical, geological and biological variables in the ocean and seafloor	http://oceanobservatories.org/

stakeholders. The programs vary from large site-based ecosystem research efforts (i.e., IBP, LTER, I-LTER) to centers that support ecological synthesis (i.e., NCEAS) and provide access to global biodiversity data (i.e., GBIF) to networked sensor systems in different regions of the ocean (i.e., OOI) and landmass (i.e., NEON). Specific examples are provided below that demonstrate how data sharing practices evolved over the period encompassing these large research programs.

2.1. International Biological Program

The International Biological Program (IBP) represented one of the early (1964–1974), large multinational efforts to understand ecosystem patterns and processes and was exceptional in that it was multidisciplinary in scope, covered a broad range of biomes, and included an integral modeling effort (Coleman, 2010; Hagen, 1992; McKee, 1970). The IBP ran for a shorter period from 1967–1974 in the USA and included grassland, coniferous and deciduous forest, Arctic and alpine, and desert sites (Coleman, 2010). IBP proved to be quite innovative for the time and resulted in many significant achievements including several successful inter-biome synthesis efforts, adoption of a holistic approach to ecosystem research, the incorporation of whole system experiments, and the formation of new theories such as the stream continuum concept (Coleman, 2010). Despite the documented successes with respect to synthesis, attempts to formulate uniform IBP data policies "met with near complete failure from the outset, to the extent that data policies and protocols were never elaborated nor even agreed to in principle" (Porter and Callahan, 1994). Consequently, most IBP data are difficult or impossible to discover and acquire today.

2.2. LTER and ILTER

The U.S. Long-Term Ecological Research (LTER) Network was created in 1980 by the U.S. National Science Foundation and has now grown to include more than two-dozen sites located in Alaska, the continental USA, Puerto Rico, French Polynesia, and Antarctica (Michener and Waide, 2009). The U.S. LTER model served as the basis for the International LTER Program that was founded in 1993 and has since grown to 40 member networks (see https://ilternet.edu). Early in the history of LTER, the National Science Foundation required each LTER site to develop a data management program, although policies and implementation were left to the discretion of the individual sites (Porter, 2010; Porter and Callahan, 1994).

During the first decade of LTER (1980–89), many sites hired data managers and established site-specific programs, but data were typically neither discoverable nor shared outside the site (Michener et al., 2011). This lack of data sharing was, in part, due to the commonly held view that data use should solely be at the discretion of the data collectors and their collaborators (Michener et al., 2011). Two innovations in 1990 began to change this perspective. First, a data catalog describing core data sets available at every LTER site was published making it possible, for the first time, to discover what data were available, where the data were collected, and who collected the data (Michener et al., 1990). Second, the first formal guidelines for LTER site data management policies were issued in 1990 and included ten provisions that should be included in each site's data management policy (Porter, 2010; Porter and Callahan, 1994). The guidelines covered roles and responsibilities of data contributors and data users. Moreover, the guidelines emphasized the importance of creating comprehensive metadata, adhering to QA/QC standards, preserving data for the long-term, and making data available in a timely fashion. However, specific details such as time limits for making data available and other details were to be determined by each individual site; consequently, individual LTER sites created policies that were highly variable with respect to data access and the responsibilities of data users (Porter and Callahan, 1994, see Table 13.3, page 199).

In 1997, the LTER Network adopted a network-wide policy that was based on the commonalities in data policies across the sites. The LTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement, which enacted network-wide data policies, was approved by the LTER Coordinating Committee April 6, 2005 (Michener and Waide, 2009; Porter, 2010). This policy strengthened the 1997 policy by defining the responsibilities of the data collector and generally limiting data embargo periods to no more two years after the data were collected. The formal adoption of data sharing policies plus the establishment of Ecological Metadata Language (EML) as the LTER metadata content standard (Andelman et al., 2004) facilitated the LTER Network in providing easy access to approximately 20,000 data packages (i.e., data plus metadata), almost a quarter of which (4538) were contributed by the LTER sites (https://portal.lternet.edu/nis; accessed 3 Feb 2015).

The LTER Program has been instrumental in bringing scientists together to develop standard field and laboratory methods such as for soils (Robertson et al., 1999) and primary productivity measurements (Fahey and Knapp, 2007) that promote data integration and synthesis. In addition, many LTER and ILTER sites have completed volumes that synthesize the ecological research at individual sites as well as across multiple similar sites (e.g., Knapp et al., 1998; Shachak et al., 2004).

2.3. NCEAS

The National Center for Ecological Analysis and Synthesis (NCEAS) was created in 1995 to advance ecological knowledge through collaboration, synthesis and data sharing (Baskin, 1997; Hackett et al., 2009). NCEAS was ground-breaking in the sense that researchers brought

existing data to the Center where small groups of scientists (usually 8–15) collaborated on synthesizing data and information during multiple several-day-long working group meetings scattered over a two to three year period (Hampton and Parker, 2011). NCEAS developed an informatics staff that assisted the working groups in manipulating, documenting, analyzing and preserving the data brought to the Center. In addition, NCEAS staff played a key role in developing *Morpho*—metadata management software that is now widely used to create metadata for ecological data (Andelman et al., 2004) and KNB, the Knowledge Network for Biocomplexity data repository that is used by NCEAS working group members and others to archive ecological and related data (https://knb.ecoinformatics.org). More than 2200 peer reviewed synthesis publications have resulted to date and the NCEAS model has now been emulated at numerous synthesis centers worldwide (see https://nceas.ucsb.edu and http://www.synthesis-consortium.org).

2.4. Data sharing since 2000

An increasing number of Long-Term Research Networks (LTRNs), Ecological Observatory Networks (EONs), and Coordinated Distributed Experiments and Observations Networks (CDEOs) have emerged internationally, mostly over the past two decades, to collect and synthesize biodiversity and ecological data at regional, continental, and global scales (Peters et al., 2014). Some of the notable developments since 2000 include:

- In 2001, the Global Biodiversity Information Facility was established to facilitate the sharing of biodiversity data and information across national borders and, in 2007, the global data portal was launched (http://www.gbif.org).
- In 2002, the Ocean Biogeographic Information System was established providing access to marine biodiversity data and information worldwide (Zhang and Grassle, 2003). Since 2004, an international network of regional OBIS nodes has developed providing specialized services within the different regions (http://www.obis.org).
- The 2005 LTER Policy provided the framework for the International LTER Network (ILTER) Data Policy (2008) that focused on data release, access and use of ILTER data by the international community (http://www.ilternet.edu).
- Between 2015 and 2017, two large environmental observatories funded by the U.S. National Science Foundation are projected to be fully operational. The National Ecological Observatory Network (http://neoninc.org) and the Ocean Observatories Initiative (http://www.oceanobservatories.org) will provide free access to data products from terrestrial and freshwater sites and from ocean and coastal sites, respectively.

Despite this existing and developing infrastructure, Peters et al. (2014) noted that ecologists are unevenly prepared to address regional- to continental-scale questions due to the lack of a data sharing culture, non-standard data collection methods and data and metadata formats, and inattention to documenting the provenance (i.e., where the data came from and how they were derived) of derived data products. Section 3 examines the sociocultural issues surrounding data sharing.

3. Sociology of data sharing

Data sharing has evolved slowly and unevenly due to a mix of incentives, disincentives and the emergence of enabling technologies. Below, I examine some of the perceived impediments to data sharing, highlight benefits that can be derived through increased data sharing, and discuss many of the key drivers of sociological change.

3.1. Perceived impediments to data sharing

Researchers perceive many potential impediments to data sharing. First and foremost, they jealously value their time and have real concerns about the requisite time, labor and expertise to share data (Campbell et al., 2002; Tenopir et al., 2011). Researchers are also concerned about the potential for misinterpretation and misuse of data (Campbell et al., 2002; Davis et al., 2001; Hilgartner, 1997; Hilgartner and Brandt-Rauf, 1994; Kervin et al., 2014). Nevertheless, recent surveys indicate that most environmental and ecological scientists are willing to share their data, but they are challenged by a lack of experience with data management and insufficient training, a paucity of effective and easy-to-use metadata management tools, lack of awareness of standards, and absence of institutional support and resources for data management (Kervin et al., 2014; Tenopir et al., 2011). Furthermore, numerous real and perceived legal constraints to sharing data exist such as different governmental and international approaches to copyright, the complexity of intellectual property rights and confidentiality issues, and uncertainty about the law (NSB, 2012; Reichman and Uhlir, 2003; Uhlir and Schröder, 2007).

3.2. Benefits of data sharing

In his review of the history of big ecology, Coleman (2010) highlighted the value of the LTER program and observed that "the collection of comprehensive field data and careful archiving, with suitable metadata (what the data are about, and their provenance) pays big dividends for the entire body of scientific researchers, and the wider human community as well." Others have noted the benefits that are derived for the public good. First, data sharing accelerates the pace of science by enabling researchers to discover and re-use relevant data, combine data from multiple sources, and ask new questions (Butler, 2006; Hampton et al., 2013; Whitlock, 2011). Opportunities for novel collaborations are created and time and money are saved since data are not necessarily re-collected multiple times. Second, public trust increases as science is made more transparent and findings can be reproduced and verified (Beardsley, 2010; South and Duke, 2010; Whitlock et al., 2010). Third, it has further been argued that access to research data represents one of our human rights (Duke and Porter, 2013; Duke et al., 2011). Uhlir and Schröder (2007) reiterated many of these points by noting that closed data systems that inhibit data sharing have many hidden costs including: contributing to higher research costs and lost opportunity costs; adding barriers to innovation; reducing the effectiveness of cooperation, education and training; suboptimal data quality; and widening the gap between developed and developing countries.

Researchers also benefit from the credit attributed to them when their archived data are cited and used by others (Parsons et al., 2010). Recent studies demonstrate that citation rates of publication increase when the research data are shared (Piwowar and Chapman, 2010; Piwowar et al., 2007).

3.3. Drivers of sociological change

Funders, journals and professional societies can each drive sociological change with respect to data sharing. Establishing and enforcing mandates for data archiving greatly increase the likelihood that data will be available for the long-term (Vines et al., 2013). Moreover, funders, publishers and professional societies can all contribute to incentivizing and reducing barriers to data sharing by providing credit, supporting education, establishing community standards for data and data sharing, and streamlining approaches to data submission.

3.3.1. The role of funders

Funders play a central role in driving the culture of the science enterprise. For example, US government research sponsors must now ensure that all research output resulting from funded projects be made publicly

available (OSTP, 2013). Prior to this policy, individual agencies developed independent policies. The policy enacted April 1, 2001 at the U.S. National Science Foundation, for example, was "NSF … expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work (NSF, 2001, page 17)." However, practical interpretation and enforcement of the policy varied widely within the agency. Attitudes towards sharing research data began changing with the America Competes Act that was signed August 9, 2007 by President Bush and required civilian federal agencies to provide guidelines, policy and procedures, to facilitate and optimize the open exchange of data and research between agencies, the public and policymakers. President Obama reauthorized the Act in 2011.

Data sharing has been central to many areas of research and sponsors are increasingly recognizing the costs associated with collecting certain types of data as well as the need to increase the scientific return on investment. For instance, the Australian Antarctic Program has had a comprehensive data policy since 1999. The most recent policy (2014) states "that each supported expeditioner is required to acknowledge that data and physical samples collected from the Antarctic, subantarctic and Southern Ocean are the property of the Commonwealth of Australia …" (https://www1.data.antarctica.gov.au/aadc/about/data_policy.cfm). The policy is exemplary in that it explicitly defines data, roles and responsibilities, embargo periods, and, even, how field and laboratory notebooks and samples are to be managed.

3.3.2. The role of publishers

The creation of the Joint Data Archiving Policy (JDAP; Box 1) was a milestone that has led to significant changes in the practice of data sharing. JDAP was developed in 2010 by several leading journals in the fields of evolution and ecology and provides the basis for a policy that requires that data supporting publications be made publicly available. JDAP has been adopted by numerous ecology journals including *American Naturalist* (Whitlock et al., 2010), *Molecular Ecology* (Rieseberg et al., 2010), *Biotropica* (Bruna, 2010), *Ecological Monographs* (Ellison and Baldwin, 2011), and *Functional Ecology* (Fox et al., 2014). Other journals and publishers such as *Science* (Hanson et al., 2011), *Nature* (Anonymous, 2014), and the Public Library of Science (Bloom et al., 2014) have adopted similar policies that require authors to share the data that support the findings reported in published articles in their journals. Costello et al. (2013) recommended that all data be published and proposed a multi-step peer-review workflow whereby data quality assurance would continually increase. Lin and Strasser (2014) recommended that publishers continue to expand their role in increasing access to data by streamlining and incentivizing data sharing as well as by creating and enforcing mandatory data availability policies.

Box 1
The joint data archiving policy from the dryad digital repository (4 Feb 2015; http://datadryad.org/pages/jdap).
"[Journal] requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as [list of approved archives here]. Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period up to a year after publication. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species."

Several data journals have emerged to provide a dedicated venue for authors to publish data and associated metadata. In 2005, the Ecological Society of America began publishing peer-reviewed *Data Papers* that consisted of ecological data and detailed metadata that were accompanied by an abstract in the journal *Ecology* (Kervin et al., 2013). BioMed Central established *GigaScience* in 2012 to support the publication of biomedical and life sciences data, including ecological data (http://www.gigasciencejournal.com/). In 2014, *Scientific Data* (from Nature Publishing Group; see Anonymous, 2013) and the *Geoscience Data Journal* (John Wiley & Sons Ltd. and the Royal Meteorological Society; see Allan, 2012) began publishing the detailed descriptors (i.e., metadata) of valuable scientific datasets that are archived elsewhere in community-recognized data repositories or general-science repositories such as the Dryad Digital Repository.

3.3.3. The role of scientists and professional societies

The U.S. National Research Council has published data sharing principles including the roles of researchers, publishers and professional societies (NRC, 2003, 2009). Scientists and professional societies such as the Earth Science Information Partners (ESIP) and Force11 have been leaders in recommending guidelines for data citation (ESIP: http://wiki.esipfed.org/index.php/index.php/Interagency_Data_Stewardship/Citations; Force11: https://www.force11.org/datacitation). The American Geophysical Union has continued to strengthen its data sharing policy over time and in 2013 the policy was revised to include the expectation that data be available as soon as the article is available online (Hanson and van der Hilst, 2014). Many professional societies encourage data sharing through their associated journals (see 3.3.2), as well as facilitating data sharing through training and professional development opportunities offered in association with society meetings. University libraries may also promote data sharing by providing training and access to institutional data repositories and data management guides (Adamick et al., 2012; King, 2007; Treloar et al., 2012).

4. The role of cyberinfrastructure

Prior to the 1980s, most data were shared with other researchers through in-person exchanges of data by physically mailing hard-copy data, punched cards, or data tapes via the postal service. Various types of software, hardware and networking infrastructure, especially the Internet and World Wide Web, have facilitated data sharing. In this section, I particularly focus on: (1) metadata standards and software tools, (2) persistent unique identifiers, and (3) data repositories.

4.1. Metadata standards and software tools

In a large study of the data management practices employed by ecologists and environmental scientists, Tenopir et al. (2011) noted that most researchers either did not use existing metadata standards or they created their own idiosyncratic approach. Some of the consequences of not using metadata standards include uneven documentation that does not support data use or data reproducibility as well as difficulty or inability to discover metadata and associated data. Several metadata standards have evolved along with tools that support metadata creation and management. In the mid- to late 1990s, a U.S. interagency committee developed the Content Standards for Digital Geospatial Metadata ([FGDC] Federal Geographic Data Committee Biological Data Working Group and USGS Biological Resources Division, 1999; FGDC, 1994, 1998) that were subsequently refined in 2003 by the Technical Committee of the International Organization for Standardization (ISO) as ISO 19115 (see http://www.iso.org/) and latterly through the addition of associated profiles such as the Biological Data Profile that could be more easily applied to particular types of data. Also, during the mid-1990s, significant effort was devoted to identifying metadata content descriptors that were more relevant to the ecological sciences (Michener et al., 1995, 1997). Such efforts were a precursor to the

development of Ecological Metadata Language (EML) that was a comprehensive suite of modules that supported data discovery, data use and interpretation, and automated processing (Fegraus et al., 2005; Jones et al., 2001).

Standards are useful constructs, but accompanying software tools are also necessary to facilitate metadata creation and management. NetCDF (Network Common Data Form), for example, was developed by the University Corporation for Atmospheric Research in the late 1980s and 1990s primarily for the earth science community. It includes software libraries and a platform-independent, self-documenting data format that enables the creation, sharing and use of array-based data via a diverse array of application software (http://www.unidata.ucar.edu/software/netcdf/). Metavist was created in 2004 to support geospatial metadata creation following the FGDC Content Standards for Digital Geospatial Metadata (https://metavist2.codeplex.com/). During the same period, Morpho was created to support manual and semi-automated metadata creation for biological, ecological and environmental data using the EML standard (Higgins et al., 2002). Other developments such as controlled vocabularies and thesauri helped researchers standardize keywords and optimize discovery of their documented data products (Michener, 2006).

4.2. Persistent unique identifiers and altmetrics

Digital object identifiers (DOIs) are standardized character strings that are used to uniquely identify digital objects such as citations, data sets and metadata documents. DOIs contain metadata about the object including location information, such as uniform resource locators (URLs). The naming convention has evolved and become more generic over time making it to possible to easily, permanently and unambiguously identify objects (e.g., journal articles, reports, books and data) associated with specific DOIs (see http://www.doi.org/). DOIs can be acquired at nominal charge through different DOI registration agencies that exist such as CrossRef for citations and DataCite for data packages (Brase, 2009). Life Science Identifiers (LSIDs) represent another unique identifier similar to DOIs that are used in the biodiversity and biomedical communities for resolving biological entities such as taxonomic names and concepts (see http://wiki.tdwg.org/twiki/bin/view/GUID/LSID).

The Open Researcher and Contributor ID (ORCID) is a nonproprietary alphanumeric code to permanently and unambiguously identify humans such as authors of journal articles and creators of data sets. ORCIDs are managed by the ORCID organization. ORCIDs enable scientists to receive appropriate attribution for their scholarly creations. This is important since human names are neither unique nor permanent.

The existence of both DOIs and ORCIDs makes it possible for specific individuals to be permanently and uniquely associated with products resulting from their creative work such as publications (e.g., books, journal articles) as well as software code, data products, web pages, and presentations. Altmetrics represent non-traditional metrics that have emerged to track an individual's scholarly impact by tracking their cumulative creative output as well as how often and by whom such products are referred to via social and news media, downloads and views, and traditional citations. Two services that calculate and track altmetrics include Altmetric.com and ImpactStory (Piwowar, 2013).

The availability of DOIs, ORCIDs and altmetrics means that researchers can now receive attribution and credit for all of their scholarly works. Such an advance creates the opportunity for data and metadata authorship and subsequent use of the data as documented by citations, downloads and views, tweets and other mentions in social media to be counted towards overall scholarly impact, including tenure and promotion decisions. Thus, data products and other approaches to disseminating results can now be thought of as first class citizens in the scientific enterprise.

4.3. Data repositories

A data repository or data archive has been defined as "a permanent collection of data sets with accompanying metadata such that a variety of users can readily acquire, understand, and use the data" (Olson and McCord, 2000). Hundreds of data repositories have emerged across all science domains and disciplines. Table 2 highlights the breadth of data repositories that hold data that are especially relevant for the ecological sciences. These repositories cover a broad range of material including specific areas such as climate and terrestrial and marine biodiversity data. Many are associated with national and international data collection and research programs such as Antarctic and Long-Term Ecological Research. Others such as Dryad, figshare and the Knowledge Network for Biocomplexity are more generic and allow deposition of data associated with a broad array of peer-reviewed journal articles and research programs.

One of the challenges associated with a burgeoning number of data repositories is knowing where to most effectively deposit data for long-term preservation as well as where to find relevant archived data. There are two approaches to resolving this problem. First, the Registry of Research Data Repositories (https://re3data.org) provides a searchable database of more than 1000 data repositories that cover all science domains (Pampel et al., 2013). Second, federated data systems like DataONE provide a uniform interface that enables users to easily search for data that are stored in a large number of ecologically-relevant data repositories (Michener and Jones, 2012; Michener et al., 2012).

5. Future of data sharing

Ecological data sharing has evolved slowly since the 1950s and is increasingly a pre-requisite for funding by research sponsors (Section 2). Data sharing has increased in response to sociocultural changes (Section 3) and the availability of supporting information technologies (Section 4). Despite these improvements, challenges remain. For example, a review of peer-reviews of data papers (i.e. data and metadata) published in the Ecological Society of America's *Data Papers* from 2004–2012 indicated that most authors did not provide metadata that was sufficient to support interpretation and re-use of the data (Kervin et al., 2013). Similarly, a recent survey of managers of environmental and ecological data repositories demonstrated that data contributors frequently made errors with respect to how data were organized (83% of the time) and documented (79% of the time) (Kervin et al., 2014).

Ecological data can be expected to grow non-linearly in volume and importance. In this Section, I envision many of the changes in information technology, sociocultural attitudes towards data and specific tools that can improve research efficiencies, promulgate data sharing and advance the pace of ecology as a science. I recommend several best practices that can advance the creation, sharing, discovery and re-use of valuable ecological data.

5.1. A vision for the future

In thinking about how ecologists can more effectively address continental scale questions, Peters et al. (2014) envisioned "an über network to allow users to seamlessly identify and select, analyze, and interpret data from sites regardless of network affiliation, funding agency, or political affinity, to cover the spatial variability and extent of regional- to continental-scale questions." Such a vision requires that data not only be shared and discoverable, but that they also are extremely well documented. In particular, data sharing would be much easier if metadata and data were both standardized and tightly coupled. Ideally, potential users would also be able to easily assess the data provenance and fitness-for-use (including data quality, scale, etc.). In addition, seamless analysis requires more effective semantic mediation tools that facilitate the harmonization of data that are represented at different scales or in different units.

Many advances have occurred recently that can bring this vision to reality. DataONE, for example, is a federated data system that functions as an über network allowing researchers to more easily discover and access data that are held at numerous data repositories that are associated with different research networks, institutions, and governments (Michener et al., 2012). Efforts are underway in DataONE

Table 2
Examples of data repositories that hold ecological and related environmental data.

Data repository	Focus	Web URL	Sponsoring institution/country
Atlas of Living Australia (Atlas)	The Atlas provides a wide variety of data and information on all the known species in Australia aggregated from museums, herbaria, community groups, government departments, individuals and universities.	http://www.ala.org.au/	Administered by Commonwealth Scientific and Industrial Research Organisation (CSIRO) on behalf of all the partners and the funding organization/Australia
Australian Antarctic Data Center (AADC)	AADC manages and provides access to the diverse environmental and biological data collected in Antarctica.	https://www1.data.antarctica.gov.au/	Australian Antarctic Division, Australian Government/Australia
Botanic Gardens Conservation International (BGCI)	BGCI provides a global database of living plants, seed and tissue collections aggregated from a network of more than 600 botanic gardens in more than 120 countries.	https://www.bgci.org	BGCI — a membership-based organization with supported by corporate and conservation partners, supporters, patrons and members
DataONE	DataONE supports discovery and access to a federation of data repositories worldwide that host biological, environmental and earth science data.	https://www.dataone.org/	Largely funded by the US National Science Foundation/USA
Dryad digital repository	Dryad enables researchers to publish data that underpin findings reported in scientific journals with a current focus on evolution, ecology, and the life and biomedical sciences.	https://datadryad.org	Dryad Inc./USA
Environmental Information Data Centre (EIDC)	The EIDC provides access to data and tools related to integrated research in terrestrial and freshwater ecosystems and their interaction with the atmosphere.	http://www.ceh.ac.uk/	Natural Environment Research Council/UK
Earth Observation System Data and Information System (EOSDIS)	EOSDIS supports discovery and processing of earth science data from satellite, aircraft and field campaigns.	https://earthdata.nasa.gov/	National Aeronautics and Space Administration/USA
figshare	figshare enables researchers to preserve and share their research outputs, including figures, datasets, images, and videos.	http://figshare.com/	Digital Science, a Macmillan Publishers company/UK
Global Biodiversity Information Facility (GBIF)	GBIF provides a single point of access to data related to more than 1.5 million species, collected worldwide over three centuries.	http://www.gbif.org/	GBIF Secretariat, Denmark/multiple nations
Knowledge Network for Biocomplexity (KNB)	KNB is an international data repository that supports data deposition (with DOIs), data discovery, and metadata creation via an online Data Registry or Morpho (a downloadable desktop tool).	https://knb.ecoinformatics.org/	Largely funded by the US National Science Foundation/USA
Long Term Ecological Network (LTER) Network Data Portal (LTER-NDP)	LTER-NDP provides access to approximately 20,000 data sets (with DOIs) along with tools for uploading and evaluating data packages and viewing the provenance of LTER data.	https://portal.lternet.edu/nis/	LTER Network Office/US National Science Foundation/USA
National Center for Atmospheric Research (NCAR) Community Data Portal (CDP)	The CDP provides access to 8000 + data collections from climate studies and field campaigns as well as software and tools.	http://cdp.ucar.edu/	U.S. National Science Foundation/USA
National Geophysical Data Center (NGDC)	NGDC provides access to and services for geophysical data from the marine terrestrial environment, as well as earth observations from space.	http://www.ngdc.noaa.gov/	National Oceanic and Atmospheric Administration/USA
National Geoscience Data Centre (NGDC)	NGDC provides access to licensed subsurface data (e.g., geology, groundwater, marine) and 3D models and free, open data via the OpenGeoscience portal.	http://www.bgs.ac.uk/services/ngdc/	British Geological Survey, Natural Environment Research Council/UK
National Oceanographic Data Center (NODC)	NODC was established in 1961 and holds in situ and remotely sensed physical, chemical, and biological data from coastal and ocean waters.	http://www.nodc.noaa.gov/	National Oceanic and Atmospheric Administration/USA
National Snow and Ice Data Center (NSIDC)	NSIDC archives and distributes data from satellite and field programs that focus on the cryosphere.	https://www.nsidc.org/	Cooperative Institute for Research in Environmental Sciences at the University of Colorado Boulder; funded by NASA, NSF, NOAA/USA
Ocean Biogeographic Information System (OBIS)	OBIS supports discovery and access of data about the diversity, distribution and abundance of life in oceans.	http://www.iobis.org/	UNESCO/multiple countries and institutions
Ocean Data Portal (ODP)	ODP provides access to oceanographic data held by the International Oceanographic Data and Information Exchange (IODE) global network of National Oceanographic Data Centres.	https://www.oceandataportal.org/	IODE program, UNESCO/multiple countries and institutions

(continued on next page)

Table 2 (*continued*)

Data repository	Focus	Web URL	Sponsoring institution/country
Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) for Biogeochemical Dynamics	The ORNL DAAC provides data and information relevant to biogeochemical dynamics, ecological data, and environmental processes as well as tools for working with data.	http://daac.ornl.gov/	Operated by ORNL Environmental Sciences Division and funded by NASA/USA
Pangaea	Pangaea archives, publishes and distributes earth system data.	http://www.pangaea.de/	Hosted by Alfred Wegener Institute at the Helmholtz Center for Polar and Marine Research and the Center for Marine Environmental Sciences at the University of Bremen; Supported by the European Commission, Federal Ministry of Education and Research, Deutsche Forschungsgemeinschaft, and the International Ocean Discovery Program/Germany
Core Science Analytics and Synthesis (CSAS) Geoportal	CSAS Geoportal provides access to a wide array of geological, environmental and biological data.	http://www1.usgs.gov/csas/geoportal/	U.S. Geological Survey/USA
Terrestrial Ecosystem Research Network (TERN)	The TERN portal provides access to Australia's terrestrial ecosystem data.	http://portal.tern.org.au/	Supported by the Australian Government through the NationalCollaborative Research Infrastructure Strategy
VertNet	VertNet helps researchers discover, capture, and publish biodiversity data	http://www.vertnet.org/	Operated by University (U.) California, U. Colorado, U. Kansas, Tulane U. and supported by NSF/USA

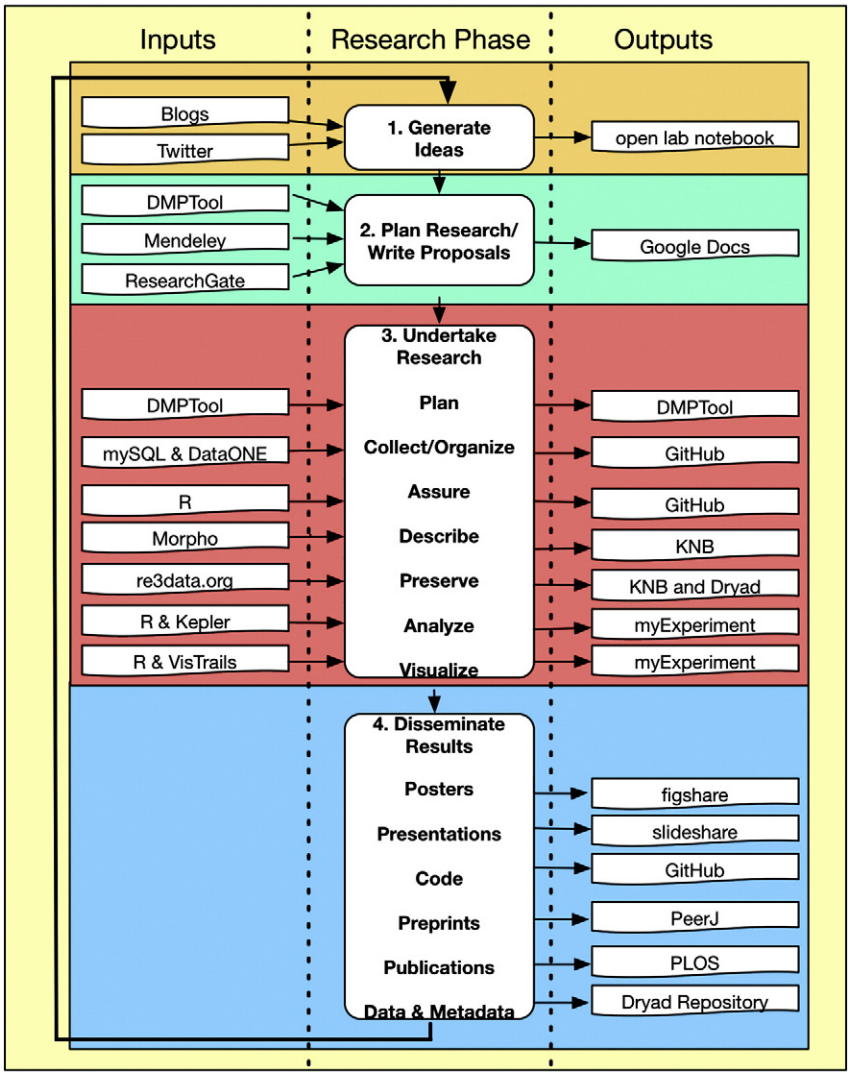

Fig. 1. Four phases of the research life cycle including process steps (middle column): (1) idea generation; (2) research planning and writing proposals; (3) undertaking research and managing data throughout its life cycle from data management planning through analysis and visualization; and (4) disseminating results through multiple media. *Left column* includes examples of tools and information sources that may inform the research enterprise (i.e. inputs). *Right column* includes examples of repositories, web sites, publication outlets and tools where research products may be shared with others (i.e. outputs).

to enable provenance tracking that will allow researchers to see not only the precursors to a data product but also how others subsequently used those data to generate new data products. Likewise, there is increased attention aimed at developing new ontologies and semantic mediation tools that can support more precise discovery and recall of data and better enable automated or semi-automated data harmonization and integration (Madin et al., 2008).

Technology can only move us so far in realizing this new vision for data sharing. Ecological data and other data remain difficult to discover, access, and use due to licensing and intellectual property right concerns, insufficient documentation, and lack of comprehensive provenance information. In clarifying and responding to these issues, the Open Knowledge and the Open Definition Advisory Council (7 October 2014) defined an Open Work as "a set of three key principles:

- Open License: The work must be available under an open license (as defined in the following section but this also includes freedom to use, build on, modify and share).
- Access: The work shall be available as a whole at no more than a reasonable one-time reproduction cost, preferably downloadable via the Internet without charge.
- Open Format: The work must be provided in a convenient and modifiable form such that there are no unnecessary technological obstacles to the performance of the licensed rights. Specifically, data should be machine-readable, available in bulk, and provided in an open format or, at the very least, can be processed with at least one free/libre/open-source software tool."

Many tools now exist that can support the creation of open works as defined above (Hampton et al., in press). Fig. 1 illustrates four key elements of the research life cycle from idea generation, through project planning, to data generation and interpretation and, lastly, publication and dissemination of results as well as examples of various open science tools and repositories that can facilitate idea generation and data sharing, interpretation and use by others. In addition to their familiarity with a subset of the literature, scientists may first generate initial ideas based on their reading of science blogs and twitter and the primary peer-reviewed literature and other sources that are highlighted in those blogs and tweets (Darling et al., 2013); as the ideas mature they may then share them with others and seek feedback via an open lab notebook. Second, the ideas may undergo refinement and be incorporated into research proposals based on their reading of additional manuscripts discovered via Mendeley and ResearchGate; the proposal text, including a draft data management plan, may then be shared with colleagues who contribute their ideas via Google Docs. Third, an array of tools (e.g., DMPTool, mySQL, DataONE, R, Morpho, re3data.org, Kepler and VisTrails) may be used as ecologists undertake their research and manage the data; subsequently, data, new software and algorithms, and workflows may be shared with colleagues and others via numerous outlets (e.g., DMPTool, GitHub, KNB, Dryad, and myExperiment). Lastly, results and analytical procedures may be disseminated in multiple ways including posters (via figshare), presentations (via slideshare), code repositories (e.g., GitHub), preprint services (e.g., PeerJ), open publications (e.g., PLoS), and digital repositories (e.g., Dryad).

Scratchpads exemplifies a state-of-the-art solution for publishing and disseminating data and related products for the biodiversity sciences—one that will ideally be emulated in the ecological sciences. It provides tools and an innovative, online virtual research environment for biodiversity science that enables researchers to create a unique website; publish, link and share structured data; build a research network; and collaborate with peers in building databases, creating reference collections, and publishing papers (Smith et al., 2011). Automated linking and sharing of ecological data would be greatly facilitated through the further development and adoption of internationally agreed upon domain ontologies.

5.2. Best practices for data sharing

Data sharing will continue to permeate the scientific culture in response to the establishment and enforcement of sponsor and publisher mandates, encouragement and provision of training by professional societies, promotion via incentives such as attribution and incorporation into the tenure and promotion process, and the increased availability of enabling information technologies. Based on the lessons learned by examining peer-reviews of data publications (Kervin et al., 2013) and surveying repository managers (Kervin et al., 2014) and environmental scientists and ecologists (Tenopir et al., 2011), it is clear that data sharing and re-use can best be promulgated if several simple best practices are followed.

5.2.1. Create and follow a data management plan

Before a project gets underway, researchers should have a plan for how the data will be managed. Plans should cover: (1) data collection and processing methods, organization in tables or databases, and relevant access and use policies; (2) quality assurance and quality control procedures; (3) metadata creation and management; (4) data preservation; (5) integration, analysis, synthesis and dissemination; (6) relevant policies including data sharing plans; and (7) a budget that explicitly details costs (i.e., time and money) for preparing, documenting, and archiving data. Data may include a diverse array of raw and processed data records as well as physical samples, biotic specimens, publications, models and software. Although the plan can initially be tailored to research sponsor requirements and page limitations, it should be recognized that a comprehensive and usable plan would almost certainly benefit from additional documentation as well as frequent review and revision (Strasser et al., 2011). The DMPTool and DMPonline are tools that make it easy for researchers to create an initial data management plan that meets the requirements for a particular research sponsor in the USA and UK, respectively (see https://dmptool.org/ and https://dmponline.dcc.ac.uk/). The DMPTool also allows one to share a plan with the project team and publish it openly for broader viewing and attribution.

5.2.2. Establish data sharing and attribution policies

Data originators and data users benefit when everyone has a clear understanding of their rights and responsibilities. This can be done informally by stating relevant policies on a project website or more formally by adopting specific licenses. For example, Creative Commons copyright licenses offer several standardized alternatives for controlling and communicating with the public about how creative works are shared and used (https://creativecommons.org/licenses/). Licenses range from those that maximize content dissemination such as "CC0" (i.e., work is in the public domain and all rights are waived) and "CC By" (i.e., requires that credit go to the creator) to others that are more restrictive such as "CC BY-NC-ND" that allows others to download and share a work as long as credit is given to the creator and if the work is neither changed nor used commercially. The Dryad digital repository, for example, has adopted the CC0 license to facilitate the discovery, reuse, and citation of archived data, and also provides users with recommendations for how data products should be cited in the literature. Regardless of what licenses or policies are adopted, all project participants should participate in the discussion and decision-making to maximize input, reduce confusion, and achieve buy-in.

5.2.3. Fully document the data

Data products cannot be re-used unless the context, structure, collection and processing methods, and quality of the data are sufficiently documented. Ideally, all aspects of the data are documented throughout the entire project from planning and hypothesis formulation through

QA/QC and metadata creation through analysis and dissemination. Creating comprehensive metadata is most effective when researchers are routinely documenting data collection, processing and analysis activities. New tools such as open lab notebooks allow research notes and data to be published online as they are created. Metadata management is facilitated when standards such as ISO 19115 and EML are adopted and comprehensive, user-friendly tools like Morpho are employed to create, manage, and disseminate the project's metadata.

5.2.4. Preserve the data, software, and workflows

Data, software, and analytical workflows (i.e., procedures followed during the data acquisition, integration and analysis phases) must necessarily be preserved, discoverable and accessible before others can use them. Table 2 listed many commonly used data repositories, most of which are free (or have low costs for data deposits) and open to the research community such as Dryad, figshare, and KNB. Software and models can be deposited in community archives such as GitHub and the Community Surface Dynamics Modeling System (CSDMS). GitHub is a web-based repository hosting service that supports version control, source code management, access control, and social networking type features (see https://github.com/). CSDMS is one example of a repository that supports the deposition and dissemination of models pertaining to earth surface patterns and processes (see http://csdms.colorado.edu/). Increasingly, scientists are developing and managing their analytical workflows (i.e. the steps involved in acquiring, integrating, processing, and analyzing and visualizing data) in workflow environments such as Kepler (https://kepler-project.org/), Taverna (www.taverna.org), and VisTrails (http://www.vistrails.org/). Such workflows can then be preserved and shared via myExperiment (http://www.myexperiment.org/), a repository and social website that enables scientists to contribute to a pool of workflows that can be reused and repurposed by other scientists. Although workflows are unlikely to replace the methods sections of journal articles in the future, citable workflows (e.g., associated with a DOI and deposited in a community repository) can be expected to enhance methods sections and promote transparency and reproducibility.

5.2.5. Publish and disseminate the data and related products

During the history of printed publications, professional societies and journal publishers increasingly restricted the length of journal articles as well as the types of article content that could be published. The advent of the Internet, web services and archives enabled data appendices and supplements (e.g., lengthy tables, algorithms and code, pictures and maps) to be published electronically without greatly adding to the publication costs and page charges for authors. Presently, it is possible and, in some cases, a requirement by publishers and research sponsors that authors "publish" the data, data management plans, software, models and workflows in various community repositories (Table 2, Sections 5.2.1 and 5.2.4, Fig. 1), research proposals and data publications.

6. Conclusion

Despite the fact that data are generally viewed as being valuable products of the science enterprise, they have not always been treated as such. Data publication and sharing have only recently emerged as community norms and data products typically receive only cursory mention during tenure and promotion deliberations. In this paper, I examined the history of data sharing as well as the barriers and solutions. Information technologies have greatly advanced our ability to preserve, manage and disseminate data, code and models. The availability and peer-review of such materials can greatly enhance the quality of data and research results and better support science transparency and reproducibility.

Challenges remain. In the technical realm, there is a critical need for tools that manage and depict data quality and provenance information about data products. Seamless data integration across studies remains problematic and new statistical and visualization approaches are needed that allow one to discern possible mismatches in scale (spatial and temporal) and units of measurement as well as identify duplicate records that are generated via multiple pathways to the aggregator (i.e., de-duplication); innovative semantic tools and provenance tracking systems are needed to address the challenge (see Section 5.1). Similarly, the absence of user-friendly tools that can automatically or semi-automatically generate metadata continue to hinder researchers in creating the comprehensive metadata that are sufficient to enable data interpretation and repurposing. Community-driven organizations such as the Federation of Earth Science Information Partners (http://www.esipfed.org/), the Research Data Alliance (https://www.rd-alliance.org/) and Taxonomic Databases Working Group (http://www.tdwg.org/) are expected to influence data sharing by identifying, developing and promoting practical technology solutions, standards and guidelines, and good practices. Likewise, concerted governmental attention to the challenges is critical. The Australian National Data Service, for example, offers a holistic, multidisciplinary approach to research data sharing that embraces technology, standards, open access, and education (http://www.ands.org.au/).

In the sociocultural realm, significant attention needs to be paid to increasing scientific data and information literacy and acknowledging researchers that are employing good practices. Moreover, deliberations among researchers, academic and governmental institutions, and publishers and research sponsors are necessary to delineate roles and responsibilities related to further promulgating data sharing, implementing effective training, and building and sustaining the requisite cyberinfrastructure. New solutions to scientifically and societally relevant challenges require that we bring all relevant data from the past, present and future to the table.

Acknowledgments

This work was supported by the NSF IIA-1301346, IIA-1329470, and ACI-1430508. Special thanks to Professor Friedrich Recknagel for his encouragement and input, as well as B. Kimbell and two anonymous reviewers for their insightful suggestions.

References

[FGDC] Federal Geographic Data Committee, 1994. Content Standards for Digital Geospatial Metadata (June 8). Federal Geographic Data Committee, Washington, D.C.
[FGDC] Federal Geographic Data Committee, 1998. FGDC-STD-001-1998. Content Standard for Digital Geospatial Metadata (revised June 1998). Federal Geographic Data Committee, Washington, D.C.
[FGDC] Federal Geographic Data Committee Biological Data Working Group, USGS Biological Resources Division, 1999. Content Standard for Digital Geospatial Metadata — Biological Data Profile, FGDC-STD-001.1-1999. Federal Geographic Data Committee, Washington, D.C.
[NRC] National Research Council, 2003. Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. National Academies Press, Washington, DC.
[NRC] National Research Council, 2009. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. National Academies Press, Washington, DC.
[NSB] National Science Board, 2012. Digital Research Data Sharing and Management (NSB-11-79). National Science Foundation, Arlington, Virginia.
[NSF] National Science Foundation, 2001. Grant General Conditions (GC-1). National Science Foundation, Arlington, Virginia.
[OSTP] Office of Science and Technology Policy, 2013. http://www.whitehouse.gov/administration/eop/ostp (accessed 15 Feb 2015).
Adamick, J., Reznik-Zellen, R.C., Sheridan, M., 2012. Data management training for graduate students at a large research university. J. eScience Librariansh. 1 (3). http://dx.doi.org/10.7191/jeslib.2012.1022 (Article 8).
Allan, R., 2012. Geoscience data. Geosci. Data J. http://dx.doi.org/10.1002/gdj.3.
Andelman, S.J., Bowles, C.M., Willig, M.R., Waide, R.B., 2004. Understanding environmental complexity through a distributed knowledge network. BioScience 54, 243–249. http://dx.doi.org/10.1641/0006-3568(2004)054[0240:UECTAD]2.0.CO;2.
Anonymous, 2013. Announcement: launch of an online data journal. Nature 502, 142. http://dx.doi.org/10.1038/502142a.
Anonymous, 2014. Data-access practices strengthened. Nature 515, 312. http://dx.doi.org/10.1038/515312a.
Baskin, Y., 1997. Center seeks synthesis to make ecology more useful. Science 275, 310–311.
Beardsley, T., 2010. The biologist's burden. BioScience 60, 483.

Bloom, T., Ganley, E., Winker, M., 2014. Data access for the open access literature: PLOS's data policy. PLoS Biol. 12, e1001797. http://dx.doi.org/10.1371/journal.pbio.1001797.

Bowser, C.J., 1994. Historic data sets: lessons from the past, lessons for the future. In: Michener, W.K., Brunt, J.W., Stafford, S.G. (Eds.), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor & Francis, London, pp. 155–179.

Brase, J., 2009. DataCite: a global registration agency for research data. Proceedings of the Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology (COINFO 2009). Institute for Electrical and Electronics Engineers, pp. 257–261.

Bruna, E.M., 2010. Scientific journals can advance tropical biology and conservation by requiring data archiving. Biotropica 42, 399–401. http://dx.doi.org/10.1111/j.1744-7429.2010.00652.x.

Butler, D., 2006. Mashups mix data into global service: is this the future for scientific analysis? Nature 439, 6–7.

Campbell, E.G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., Blumemthal, D., 2002. Data withholding in academic genetics: evidence from a national survey. J. Am. Med. Assoc. 287, 473–480.

Coleman, D.C., 2010. Big Ecology: The Emergence of Ecosystem Science. University of California Press, Berkeley and Los Angeles, CA.

Costello, M.J., Michener, W.K., Gahegan, M., Zhang, Z.-Q., Bourne, P., 2013. Biodiversity data should be published, cited and peer-reviewed. Trends Ecol. Evol. 28, 454–461.

Darling, E.S., Shiffman, D., Côté, I.M., Drew, J.A., 2013. The role of Twitter in the life cycle of a scientific publication. Ideas Ecol. Evol. 6 (32–43), 2013. http://dx.doi.org/10.4033/iee.2013.6.6.f.

Davis, M.A., Tilman, D., Hobbie, S.E., Lehman, C.L., Reich, P.B., Knops, J.M.H., Naeem, S., Ritchie, M.E., Wedin, D.A., 2001. Public access and use of electronically archived data: ethical considerations. Bull. Ecol. Soc. Am. 82, 90–91.

Duke, C.S., Porter, J.H., 2013. The ethics of data sharing and reuse in biology. Bioscience 63, 483–489.

Duke, C.S., Middendorf, G., Wyndham, J., 2011. Science as a human right: ESA and the AAAS science and human rights coalition. Bull. Ecol. Soc. Am. 92, 61–63.

Ellison, A.M., Baldwin, J.D., 2011. Editorial. Ecol. Monogr. 81, 1–2.

Fahey, T.J., Knapp, A.K., 2007. Principles and Standards for Measuring Primary Production. Oxford University Press, New York.

Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86, 158–168.

Fox, C.W., Irschick, D.J., Knapp, A.K., Thompson, K., Baker, L., Meyer, J., 2014. Functional ecology: moving forward into a new era of publishing. Funct. Ecol. 28, 291–292. http://dx.doi.org/10.1111/1365-2435.12254.

Hackett, E.J., Parker, J.N., Conz, D., Rhoten, D., Parker, A., 2009. Ecology transformed: the national center for ecological analysis and synthesis and the changing patterns of ecological research. In: Olson, G.M., Zimmerman, A., Bos, N. (Eds.), Scientific Collaboration on the Internet. MIT Press, Boston, pp. 277–296.

Hagen, J.B., 1992. An Entangled Bank: The Origins of Ecosystem Ecology. Rutgers University Press, New Brunswick.

Hampton, S.E., Parker, J.N., 2011. Collaboration and productivity in scientific synthesis. BioScience 61, 900–910.

Hampton, S.E., Strasser, S.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162. http://dx.doi.org/10.1890/120103.

Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K., Mudge, J., Pourmokhtarian, A., Schildhauer, M., Woo, K.H., Zimmerman, N., 2015. The Tao of Open Science for Ecology. Ecosphere (in press).

Hanson, B., van der Hilst, R., 2014. AGU's data policy: history and context. Eos 95, 337. http://dx.doi.org/10.1002/2014EO370008.

Hanson, B., Sugden, A., Alberts, B., 2011. Making data maximally available. Science 331, 649.

Higgins, D., Berkley, C., Jones, M., 2002. Managing heterogeneous ecological data using Morpho. Proceedings of the 14th International Conference on Scientific and Statistical Database Management, July 24–26.

Hilgartner, S., 1997. Access to data and intellectual property: scientific exchange in genome research. Intellectual Property Rights and the Dissemination of Research Tools in Molecular Biology. National Academy Press, Washington, D.C., pp. 28–39.

Hilgartner, S., Brandt-Rauf, S.I., 1994. Data access, ownership and control: toward empirical studies of access practices. Knowledge 15, 355–372.

Jones, M.B., Berkley, C., Bojilova, J., Schildhauer, M., 2001. Managing scientific metadata. IEEE Internet Comput. 5, 59–68.

Juday, C., Hasler, A.D., 1946. List of publications dealing with Wisconsin limnology 1871–1945. Trans. Wis. Acad. Sci. Arts Lett. 36, 469–490.

Keeling, C.D., Bacastow, R.B., Bainbridge, A.E., Ekdahl Jr., C.A., Gunther, P.R., Waterman, L.S., Chin, J.F.S., 1976. Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. Tellus 28, 538–551.

Kervin, K.E., Michener, W.K., Cook, R.B., 2013. Common errors in ecological data sharing. J. eScience Librariansh. 2 (2). http://dx.doi.org/10.7191/jeslib.2013.1024 (Article 1).

Kervin, K., Cook, R.B., Michener, W.K., 2014. The backstage work of data sharing. ACM International Conference on Supporting Groupwork (GROUP) 2014, Sanibel Island, FL, USA http://dx.doi.org/10.1145/2660398.2660406.

King, G., 2007. An introduction to the Dataverse Network as an infrastructure for data sharing. Sociol. Methods Res. 36, 173–199.

Knapp, A.K., Briggs, J.M., Hartnett, D.C., Collins, S.L., 1998. Grassland Dynamics: Long-Term Ecological Research in Tallgrass Prairie. Oxford University Press, New York.

Likens, G.E., Borman, F.H., Pierce, R.S., Eaton, J.S., Johnson, N.M., 1977. Biogeochemistry of a Forested Ecosystem. Springer-Verlag, New York.

Lin, J., Strasser, C., 2014. Recommendations for the role of publishers in access to data. PLoS Biol. 12, e1001975. http://dx.doi.org/10.1371/journal.pbio.1001975.

Madin, J.S., Bowers, S., Schildhauer, M.P., Jones, M.B., 2008. Advancing ecological research with ontologies. Trends Ecol. Evol. 23, 159–168. http://dx.doi.org/10.1016/j.tree.2007.11.007.

McKee, J., 1970. International biological program. Science 170, 471–472. http://dx.doi.org/10.1126/science.170.3956.471.

Michener, W., 2006. Meta-information concepts for ecological data management. Ecol. Inform. 1, 3–7.

Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27, 85–93.

Michener, W.K., Waide, R.B., 2009. The evolution of collaboration in ecology: lessons from the United States Long Term Ecological Research Program. In: Olson, G.M., Zimmerman, A., Bos, N. (Eds.), Scientific Collaboration on the Internet. MIT Press, Boston, pp. 297–310.

Michener, W.K., Miller, A.B., Nottrott, R., 1990. Long Term Ecological Research Core Data Set Catalog. Belle W. Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC.

Michener, W.K., Brunt, J.W., Helly, J., Kirchner, T.B., Stafford, S., 1995. Demystifying metadata. In: Gross, K.L., Pake, C.E., Allen, E., Bledsoe, C., Colwell, R., Dayton, P., Dethier, M., Helly, J., Holt, R., Morin, N., Michener, W., Pickett, S.T.A., Stafford, S. (Eds.), Final Report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED): Volume I. Text of the Report. Ecological Society of America, Washington, DC, pp. 40–62.

Michener, W.K., Brunt, J.W., Helly, J., Kirchner, T.B., Stafford, S.G., 1997. Non-geospatial metadata for the ecological sciences. Ecol. Appl. 7, 330–342.

Michener, W.K., Porter, J., Servilla, M., Vanderbilt, K., 2011. Long term ecological research and information management. Ecol. Inform. 6, 13–24.

Michener, W.K., Allard, S., Budden, A., Cook, R.B., Douglass, K., Frame, M., Kelling, S., Koskela, R., Tenopir, C., Vieglais, D., 2012. Participatory design of DataONE—enabling cyberinfrastructure for the biological and environmental sciences. Ecol. Inform. 11, 5–15.

Olson, R.J., McCord, R.A., 2000. Archiving ecological data. In: Michener, W.K., Brunt, J.W. (Eds.), Ecological Data: Design, Management, and Processing. Blackwell Science Ltd., London, pp. 117–141.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., et al., 2013. Making research data repositories visible: the re3data.org registry. PLoS One 8, e78080. http://dx.doi.org/10.1371/journal.pone.0078080.

Parsons, M.A., Duerr, R., Minster, J.-B., 2010. Data citation and peer-review. Eos 91, 297–298.

Peters, D.P.C., Loescher, H.W., SanClements, M.D., Havstad, K.M., 2014. Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5, 29. http://dx.doi.org/10.1890/ES13-00295.1.

Piwowar, H., 2013. Altmetrics: value all research products. Nature 493, 159. http://dx.doi.org/10.1038/493159a.

Piwowar, H.A., Chapman, W.W., 2010. Public sharing of research datasets: a pilot study of associations. J. Inform. 4, 148–156.

Piwowar, H.A., Day, R.S., Fridsma, D.B., 2007. Sharing detailed research data is associated with increased citation rate. PLoS One 2, e308. http://dx.doi.org/10.1371/journal.pone.0000308.

Porter, J.H., 2010. A brief history of data sharing in the U.S. Long Term Ecological Research Network. Bull. Ecol. Soc. Am. 91, 14–20.

Porter, J.H., Callahan, J.T., 1994. Circumventing a dilemma: historical approaches to data sharing in ecological research. In: Michener, W.K., Brunt, J.W., Stafford, S.G. (Eds.), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor & Francis, London, pp. 193–202.

Reichman, J.H., Uhlir, P.F., 2003. Contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. Law Contemp. Probl. 66, 315–462.

Rieseberg, L., Vines, T., Kane, N., 2010. Editorial and retrospective 2010. Mol. Ecol. 19, 1–22. http://dx.doi.org/10.1111/j.1365-294X.2009.04450.x.

Robertson, G.P., Bledsoe, C.S., Coleman, D.C., Sollins, P., 1999. Standard Soil Methods for Long-term Ecological Research. Oxford University Press, New York.

Schimel, D., Keller, M., Berukoff, S., Kao, R., Loescher, W.W., Powell, H., Kampe, T., Moore, D., Gram, W., 2011. NEON Science Strategy: Enabling Continental-scale Ecological Forecasting. NEON, Inc., Boulder, Colorado.

Shachak, M., Gosz, J.R., Pickett, S.T.A., Perevolotsky, A., 2004. Biodiversity in Drylands: Toward a Unified Framework. Oxford University Press, New York.

Smith, V.S., Rycroft, S., Brake, I., Scott, B., et al., 2011. Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. Zookeys 150. http://dx.doi.org/10.3897/zookeys.150.2193 (Special Issue: e-Infrastructures for data publishing in biodiversity science, pp. 53–70).

South, D.B., Duke, C.S., 2010. Will a data registry increase professional integrity? J. For. 108, 370–371.

Strasser, C., Cook, R., Michener, W., Budden, A., Koskela, R., 2011. Promoting data stewardship through best practices. In: Jones, M.B., Gries, C. (Eds.), Proceedings of the Environmental information Management Conference 2011 (EIM 2011). University of California, Santa Barbara, pp. 126–131.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., et al., 2011. Data sharing by scientists: practices and perceptions. PLoS One 6, e21101. http://dx.doi.org/10.1371/journal.pone.0021101.

Treloar, A., Choudhury, G.S., Michener, W., 2012. Contrasting national research data strategies: Australia and the United States. In: Pryor, G. (Ed.), Managing Research Data. Facet Publishing, London, pp. 173–203.

Uhlir, P.F., Schröder, P., 2007. Open data for global science. Data Sci. J. 6, OD36–OD53.
Vines, T., Andrew, R., Bock, D., Franklin, M., Gilbert, K., et al., 2013. Mandated data archiving greatly improves access to research data. FASEB J. 274, 1304–1308. http://dx.doi.org/10.1096/fj.12-218164.
Whitlock, M.C., 2011. Data archiving in ecology and evolution: best practices. Trends Ecol. Evol. 26, 61–65.
Whitlock, M.C., McPeek, M.A., Rausher, M.D., Rieseberg, L., Moore, A.J., 2010. Data archiving. Am. Nat. 175, 145–146. http://dx.doi.org/10.1086/650340.
Zhang, Y., Grassle, J.F., 2003. A portal for the Ocean Biogeographic Information System. Oceanol. Acta 25, 193–197.