View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Columbia University Academic Commons

CHAPTER 12*

Curation of Scientific Data at Risk of Loss

Data Rescue and Dissemination

Robert R. Downs and Robert S. Chen

Data rescue offers an opportunity for digital repositories, including institutional repositories, data archives, and scientific data centers, to provide access to potentially valuable scientific data that is at risk of being lost. Rescue may be valuable not only to restore access to data of past scientific interest, such as environmental observations or social surveys, but also to recover historic information about the state of knowledge and science at the time the data was collected or assembled. Scientific data may need to be rescued at any stage along the data life cycle, and the extent of data curation that was completed prior to a data rescue effort may vary, depending on the circumstances that led to the need for data rescue. The level of effort required to complete a data rescue depends largely on the condition of the data being rescued, the availability and quality of data documentation and provenance information, and the accessibility of the data producers. In extreme cases, data organization and documentation are poor, and those knowledgeable about how the data was collected or developed are no longer available. In some cases, collections of data sets may need to be rescued from an existing archive that is no longer sustainable. In short, scientific data may be at risk of loss for a variety of reasons, and a data rescue effort can present new challenges for data curation and dissemination operations.

* Copyright The Trustees of Columbia University in the City of New York, 2016. Creative Commons Attribution 4.0 License, CC BY (https://creativecommons.org/licenses/by/4.0/).

We report here on a recent effort by the NASA Socioeconomic Data and Applications Center (SEDAC) to rescue the Millennium Ecosystem Assessment (MA) collection of scientific data as a case study on the issues raised by a data rescue effort from an existing archive that had not fully curated the original data. The MA was an international survey of the world's ecosystems conducted by the scientific community in 2001–2005 involving more than 1,300 experts from around the world. As part of the MA, a diverse set of environmental and socioeconomic data was assembled and integrated in order to enable scientific analysis and assessment in support of policy and decision making. This data was held by the US Geological Survey (USGS) National Biological Information Infrastructure (NBII), which was terminated by the US government in early 2012.[1] This case study describes what happened to the data after the MA was completed, why data rescue was subsequently needed, the process used to decide on the data rescue effort, and the subsequent issues and challenges addressed in rescuing the MA data. The core preservation need for the MA collection is described along with the tradeoffs involved in conducting the data rescue. Based on the case study, we summarize lessons learned from the data rescue effort, including lessons for projects that create or collect data, for repositories that acquire data from such projects, and for those engaged in rescuing data. Of course, whether there will be significant scientific or historical benefit resulting from this rescue effort remains to be seen.

Benefits of Data Rescue

Data repositories that work closely with the scientific community are likely to encounter opportunities to conduct data rescue activities that could contribute to science by facilitating the use of legacy data for new studies. The term *data rescue* refers to efforts that enable the sustained use of data that otherwise might go unused. The World Meteorological Organization has defined data rescue as "the ongoing process of 1. preserving all data at risk of being lost due to deterioration of the medium and; 2. digitizing current and past data into computer compatible form for easy access."[2]

Data rescue needs to occur before the data in question becomes completely inaccessible or unusable, and ideally should occur while those scientists or others familiar with the data are still available to provide important information about the data, its origin, collection, and management, and its quality. Data rescue can enable studies that would not otherwise be possible without the rescued data.[3] For example, legacy data can fill gaps about events and anomalies that might not be part of a longitudinal study. In summarizing several data rescue efforts, Griffin noted that "legacy data may be the best, sometimes the only, sources of information about those critical departures from the norm."[4] As another example, data rescued from various publications of 1855 and 1856 and from weather station

records of the era, along with other sources, has revealed extreme precipitation events that occurred during that period in the Iberian Peninsula.[5]

Scientific data rescue efforts also offer opportunities for repositories to improve their collections and contribute to the infrastructure, advancement, and application of science. Climate records for countries in the Mediterranean region from the past few centuries are currently being inventoried and rescued to facilitate longitudinal climate assessments and predictions.[6] Many important long-term climate data series have been developed from historical records, such as those available from the Climatic Research Unit at the University of East Anglia, the Climate Data Library of the International Research Institute for Climate and Society (IRI), and the US National Climatic Data Center (NCDC).[7] This data has been critical not only to the advancement of science, but also to international assessments conducted by the Intergovernmental Panel on Climate Change (IPCC).[8]

Rescue also may be used to recover historic information about the state of knowledge and science at the time the data was collected or assembled. For example, historians or political scientists may be interested in understanding the level of scientific awareness and understanding at important points in decision making that requires significant scientific input.[9] Another possible benefit results when the cost of the data rescue represents a fraction of the cost of any new data collection.[10] In such cases, data rescue could offer an efficient alternative to new data acquisition, saving time and money.

Challenges of Data Rescue for Repositories

A data rescue effort offers unusual challenges for repositories, such as scientific data centers and archives, which routinely work with data producers and user communities to curate data and improve its potential for use by the communities that they serve. A data rescue could be required as a result of various circumstances, such as media decay and obsolescence, laboratory closure, absence of documentation and data quality information, non-digital data capture, and missed opportunities to capture data within a data management system.[11] Data rescue efforts can be quite diverse, reflecting the different kinds of data that have been collected, the effects of time and technological change, and the availability of resources for obtaining the data and enabling its sustained use by an identified community. Complex data rescue efforts can involve developing automatic correction and conversion methods for recovering data, for example from multiple satellite instruments or creating metadata from forty-year-old tapes to study sea ice during the 1960s.[12] Furthermore, data rescue could require collection, digitization, and quality control of historical data from various sources that are

no longer publicly available, including historical records from obsolete analog instruments and handwritten observations obtained from historical documents, such as ship logbooks and signal stations that create a comprehensive time series climate record.[13] Most of these situations mean that normal processes for properly managing the life cycle of scientific data cannot be carried out in a routine manner due to inadequate data management during parts of the data life cycle.

Often, knowledge about the context of the data being rescued is not readily available. Ideally, such knowledge can be gathered from publications or technical documents describing the data, or else obtained from members of the original study team or others intimately familiar with the data. For example, handwritten materials that have faded or are illegible pose challenges that can be mitigated if members of the original data collection team can help interpret the materials or fill in the information gaps.[14] Furthermore, Knapp, Bates, and Barkstrom warned "that without the active participation from the complete chain of data provider, archive, and users, data sets will atrophy and become unusable."[15] However, when a decision has been made to rescue a particular data set or collection, the rescuing repository may not know about relevant sources of information and may not be aware of who was involved in creating and managing the data or how to reach them—assuming they are still available to be reached!

In the absence of complete information about the data in need of rescue or assistance from those who possess knowledge of the data and its provenance, a data rescue effort may require divergence from rigorous data curation and quality assurance practices, such as those that are usually completed within a scientific data center. In cases where information about scientific data and its quality is limited, tradeoffs may be necessary to balance the desire for scientific rigor or completeness, the requirements of potential uses and users, and the available resources at hand. The adoption and use of specialized hardware and software may be needed, and the required capabilities for conducting a data rescue could be different for each data set in need of rescue. Furthermore, data rescue efforts in developing countries, even though they could be of significant value, are prone to conditions that pose risks for data preservation (even for current data management efforts), and developing countries typically do not have the resources to conduct data rescue efforts.[16]

Repository Considerations for Data Rescue

Scientific data may need to be rescued at any stage along the data life cycle, and the extent of data curation that was completed prior to the data rescue effort may vary. Whereas some data rescue initiatives involve digitization of data from analog form, rescue of data from the last half-century can involve remastering to

convert digital data from older databases, formats, and media.[17] The condition of the data and associated documentation that are in need of rescue will likely affect the level of effort required to make the rescued data usable. For example, significant effort may be necessary when data values have not been properly collected into a data set or curated, and the data producers are no longer available. On the other hand, a properly curated and usable collection of data rescued from an archive that is no longer sustainable may take only a small amount of effort to ingest and assimilate into a new repository.

Although it might be ideal to bring older or orphaned data sets up to current standards of data management, doing so could consume resources that are needed to manage current data that could have many more users, uses, and scientific or societal benefits. In this case, consider a basic data rescue strategy that includes digital preservation of the data files, identification and preservation of *critical* documentation, and preparation of appropriate preservation and discovery metadata. While development of *complete* documentation would be ideal, a high priority for documentation should be the identification of data ownership information and, if possible, securing of dissemination rights from the owners if the owners can be identified and reached. This strategy ensures that data is not lost forever; on the other hand it leaves some onus on future users to invest time and effort to obtain any additional information about the data needed to interpret and use the data appropriately to meet their own objectives.

Observations from a data rescue effort by a scientific data center can help inform future data rescue efforts in their decision-making process. This case study of a data rescue effort, which was completed in 2015 by the NASA Socioeconomic Data and Applications Center (SEDAC),[18] provides insight into the issues, challenges, and choices that future data rescue efforts might encounter. SEDAC routinely acquires, manages, preserves, and prepares data about human interactions in the environment for dissemination to scientific communities, decision makers, and the public. The case study describes how the collection of data was identified, assessed, and selected for the data rescue effort. The workflow of the data rescue, including planning, preparation, organization, review, and dissemination of the collection, is also described. Successful aspects of the described data rescue are discussed to inform future data rescue efforts and to suggest opportunities for repositories to plan for and complete their own data rescue efforts.

Rescue of the Millennium Ecosystem Assessment (MA) Data

The Millennium Ecosystem Assessment (MA) data was developed as part of a worldwide appraisal of ecosystems and conducted under the auspices of the Unit-

ed Nations by more than 1,300 scientists between 2001 and 2005. The data was gathered from multiple sources and assembled for analysis, forming the basis for a series of influential reports on the state of the world's ecosystems issued in 2005.[19] The data included version 3 of SEDAC's Gridded Population of the World (GPW) data set as well as "alpha" versions of several other SEDAC data sets that were made available to the MA in advance of formal release. All of the data was originally held by the National Biological Information Infrastructure (NBII) program of the United States Geological Survey (USGS). However, the US Congress cut the budget for NBII beginning in the 2012 federal fiscal year, leading to closure of the NBII's main website and associated nodes in January 2012.[20]

At that time, SEDAC recognized that there was scientific and historical value in the MA collection of data, and that this data was at high risk of being permanently lost due to the NBII's termination. Several SEDAC staff members had been involved in the MA and the NBII, and were therefore knowledgeable about the origins of the data and who had been involved. An initial assessment was conducted to determine the relevance of the data to specific SEDAC mission objectives and to meeting future user needs. It was determined that the socioeconomic scenarios developed for the MA would be of high interest to SEDAC users and that other MA data could be of interest to user communities concerned with climate impacts, adaptation, and vulnerability; environmental sustainability; agricultural and forest productivity; and land use and land cover change.

SEDAC acquired copies of the MA data in 2012 from an individual who had worked with the NBII for a preliminary review. The initial inventory of the collection identified 43 possible data products in 92 data files, with a total volume of approximately 1.75 gigabytes. The files were not well-documented and did not include any data set–level metadata or permissions documentation, reflecting the limited attention given to formal data management during the MA. Additional documentation, provenance information, and methodological details for the data were sought by e-mail from members of the data creation teams, with limited success. Many MA scientists were not available or had limited recollection of specific information about the collection contents. SEDAC determined that it would take substantial staff time (over multiple years) to archive and document all of the 43 data products individually with appropriate provenance and context information, and that in some cases important information might not be recoverable. In most cases, data had been superseded by more recent versions, so the primary interest in the data would be historical.

In light of these factors, and considering its other data development, management, and dissemination priorities, SEDAC decided to propose a basic data rescue effort that could enable future discovery and use of the MA collection. In May 2013, the SEDAC User Working Group (UWG), an advisory group of scientists, representative users, and other experts that meets annually,[21] approved

SEDAC's plan to archive and disseminate the MA collection with limited additional value-added efforts.

To streamline the data rescue effort, SEDAC organized the MA files thematically into six data sets for online dissemination: MA Biodiversity, MA Climate and Land Cover, MA Ecosystems, MA Population, MA Rapid Land Cover Change, and MA Scenarios. These six MA data sets contain the original MA files in their original formats with supplementary information obtained from various sources. SEDAC staff members worked intensively to clarify authorship and dissemination rights, working with the relevant report or chapter authors. However, SEDAC decided to refer users to the published MA assessment reports for detailed information on the scientific background of the data and its use in the MA analysis. The data and the MA assessment reports were analyzed to create a collection description and a summary and metadata record for each of the MA data sets.

Prior to dissemination, each data set in the MA collection was accessed and analyzed to ensure that the data quality was not compromised and that the data could be accessed by interested users. Each data set also received an internal "alpha" review by SEDAC scientists and staff, followed by a "beta" scientific and technical review by selected external users including members of the SEDAC UWG. The SEDAC Configuration Management Board (CMB) reviewed all comments received and ensured that corrections to collection and data set descriptions and to metadata were completed prior to public release. Each data set in the MA collection was archived to ensure preservation prior to dissemination.

Dissemination of the Millennium Ecosystem Assessment (MA) Data

Within the structure of the SEDAC website, a data collection was established to provide access to the MA collection (http://sedac.ciesin.columbia.edu/data/collection/ma). The collection description on the MA collection webpage explains that as the result of "a data rescue effort, minimal documentation and support is provided,"[22] to notify potential users that the data sets in the MA collection might not meet their expectations. As for other SEDAC collections, the MA collection webpage then links to the landing page for each data set (see figure 12.1), which contains a data set description and a recommended data citation, including an assigned Digital Object Identifier (DOI). Webpages for data download, documentation, and metadata are linked from each data set landing page. The data download page links to a zip file containing the data files for that data set in their original formats. The documentation webpage displays the titles to all five of the 2005 MA reports, with links to each of those reports. Each data set has

a full metadata record compliant with the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) schema, which can be displayed in various formats. The MA collection is available for free to all users from the SEDAC website, but users are required to log in using NASA's Earthdata login service in order to download data.

FIGURE 12.1
Landing page of the data set, Millennium Ecosystem Assessment: MA Population. Source: Millennium Ecosystem Assessment, Millennium Ecosystem Assessment: MA Population. (Palisades, NY: NASA Socioeconomic Data and Applications Center [SEDAC], 2005). doi:10.7927/H4CF9N1K.

Lessons Learned

The MA collection data rescue experience is instructive for data producers who create or collect data, for repositories that acquire data from such projects, and for repositories that will be rescuing data. In addition to informing other data rescue efforts, the lessons of the MA collection data rescue also offer insight into potential risks of data management that can be mitigated by repositories through better coordination with data producers and anticipation of users' needs.

Clearly, data rescue would be much easier if data producers conducted due diligence during data creation and collection projects to ensure that all data produced has been properly prepared for preservation to enable its continuing use by others who are not part of the data study team. Such preparation should include the creation of complete documentation and provenance information. But in the absence of full documentation, even basic information on files, data sources, and the names and contact information of those involved would facilitate future preservation. Much time and effort can be wasted when such basic information is missing. Similarly, clearly identifying authorship and intellectual property rights is much more straightforward to do at the time when data is produced rather than years or decades later. Data repositories can provide guidance and tools for data producers to enhance their data documentation and provide users with more comprehensive information about the data, its collection, and its potential for use. Earlier involvement of data managers in national and international scientific research and assessment programs could also improve the development of appropriate data management policies, procedures, and incentives and increase the likelihood that resources would be allocated for their implementation.[23]

In many cases, research groups or assessment teams assemble data from multiple sources and integrate this data with their own, producing value-added data sets, models, or other research outputs. Again, clear documentation of these steps and careful attention to version control of both inputs and outputs are important in order to improve transparency and traceability of results. Such efforts are often neglected due to the assumption that input data is already sufficiently documented or due to time and resource limitations and competing priorities. More extensive use of workflow management tools and self-documenting data transformation and analysis packages may help address this problem in the future, as would publisher and funder requirements to deposit data in an approved archive in order to make data openly available.

Data repositories that acquire data from data producers and accept responsibility for the management of such resources need to discuss the opportunities for broad public dissemination with the data producers and come to an agreement regarding the expectations and responsibilities of both parties. As part of such negotiations with the data producers, the data repository should request and receive nonexclusive intellectual property rights that will allow anyone to archive, use,

integrate, and disseminate the data without restrictions, as long as attribution is provided for the source of the data. SEDAC tries to negotiate such unrestrictive rights for the data that it acquires so that the same rights can be offered to its users. These rights are described in each data set's online documentation and metadata.

It is also of course critical that data repositories take long-term data stewardship seriously, even if their primary focus is support for current data needs. They should attempt to develop appropriate preservation metadata in addition to discovery metadata for their holdings so that key information needed to understand and use data are not lost. Potential time-based dependencies should be identified to avoid losses due to media deterioration, technological obsolescence, or destruction schedules.[24] Information about the quality of the data and the results of data quality assessments should be accessioned with the data. Likewise, any rights agreement or other licenses obtained for the data should be archived. Repositories should manage their data holdings in accordance with the Open Archival Information Systems (OAIS) framework.[25] Data repositories need to conduct ongoing assessments of their data holdings to ensure that their data holdings have been properly prepared and effectively managed to enable usability by the communities served, even if the data is not planned for transfer to another facility. Plans for the sustainability or transition of the data infrastructure and holdings should be established by the repository so that access to the data can continue in the event of the termination of funding or operational authority of the repository. In the long run, it would be ideal for all data repositories to meet one or more standards for data stewardship, such as the Data Seal of Approval, the Trustworthy Repositories Audit & Certification (TRAC), or ISO 16363:2012, Space data and information transfer systems—Audit and certification of trustworthy digital repositories.[26] SEDAC has worked to meet the TRAC and ISO 16363:2012 standards, including a collaboration with the Columbia Libraries to ensure a long-term institutional home for all of SEDAC's data holdings.

Like the repositories that acquire data from producers, data repositories that engage in data rescue efforts need an established selection-and-appraisal process to select the data for curation and determine the appropriate level of service for continuing use of the data. A complete assessment of the candidate data rescue should be conducted to identify the effort and resources needed to meet basic preservation goals versus additional investments to meet current preservation and usability standards and expectations. When considering competing priorities for limited budgets, the potential value of scientific data to future scientific, historical, and policy research and applications should be considered both for data rescue and for current data management. Alternatively, it may be worth exploring whether members of the scientific community or another repository or entity might be able to contribute to or support the data rescue.

Discussion and Conclusion

Unlike typical data curation efforts that are conducted at scientific data centers, data rescue may well require divergence from regular data curation procedures as tradeoffs may be necessary. The extent of such divergence may depend on the state of the data when it is acquired as well as on the availability of the data producers and data documentation. With the passage of time, the difficulty of any particular data rescue will inevitably increase, as data, documentation, and sources of information become more difficult if not impossible to access.

It is therefore important to move quickly when the need for a data rescue has been identified. In the case described here, SEDAC benefited from the relatively quick recognition of the need for a data rescue effort, that is, within one to two years of the NBII closure. However, the effort was also hampered by the poor state of the data more than seven years after the completion of the MA. Early identification of candidates for data rescue and the initiation of immediate action should increase the success of data rescue efforts. Similarly, the MA data rescue effort benefited from the familiarity that some SEDAC staff members had with the data being rescued. Such familiarity helped facilitate access to key scientists and critical information needed to document the data and determine access rights. Repositories, data centers, and archives that have worked with data that is at risk or with the associated scientific communities may be better positioned to take on data rescue activities in these areas.

Acknowledgments

The work reported in this chapter was conducted with support from the National Aeronautics and Space Administration (NASA) under contract NNG13HQ04C for the Socioeconomic Data and Applications Distributed Active Archive Center (DAAC). We acknowledge the extensive efforts of many members of SEDAC's staff in carrying out the MA data rescue, especially that of information scientist Xiaoshi Xing.

Notes

1. US Geological Survey, "NBII to Be Taken Offline Permanently in January," *USGS Access Newsletter* 14, no. 3 (Fall 2011), http://www.usgs.gov/core_science_systems/Access/p1111-1.html.
2. L. S. Tan, S. Burton, R. Crouthamel, A. van Engelen, R. Hutchinson, L. Nicodemus, T. C. Peterson, F. Rahimzadeh, *Guidelines on Climate Data Rescue*, WMO/TD No. 1210, ed. Paul Llansó and Hama Kontongomde (Geneva, Switzerland: World Meteorological Organization, 2004), http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-55.pdf.

3. Manola Brunet and Phil Jones, "Data Rescue Initiatives: Bringing Historical Climate Data into the 21st Century," *Climate Research* 47, no.1 (2011): 29, doi:10.3354/cr00960.
4. R. Elizabeth Griffin, "When Are Old Data New Data?" *GeoResJ* 6 (June 2015): 93, doi:10.1016/j.grj.2015.02.004.
5. F. Domínguez-Castro, Alexandre M. Ramos, Ricardo García-Herrera, and Ricardo M. Trigo, "Iberian Extreme Precipitation 1855/1856: An Analysis from Early Instrumental Observations and Documentary Sources," *International Journal of Climatology* 35, no. 1 (2015): 142–53, doi:10.1002/joc.3973.
6. M. Brunet, P. D. Jones, S. Jourdain, D. Efthymiadis, M. Kerrouche, and C. Boroneant, "Data Sources for Rescuing the Rich Heritage of Mediterranean Historical Surface Climate Data," *Geoscience Data Journal* 1, no. 1 (2014): 61–73, doi:10.1002/gdj3.4.
7. University of East Anglia (UEA) Climatic Research Unit, "Data," accessed November 23, 2015, http://www.cru.uea.ac.uk/data; International Research Institute for Climate and Society (IRI), "Climate Data Library," accessed November 23, 2015, http://iri.columbia.edu/resources/data-library/; US National Climatic Data Center, accessed November 23, 2015, http://www.ncdc.noaa.gov.
8. Intergovernmental Panel on Climate Change (IPCC), "Data Distribution Centre," accessed November 23, 2015, http://www.ipcc-data.org/observ/clim/ar4_global.html.
9. Paul N. Edwards, *A Vast Machine* (Cambridge, MA: MIT Press, 2010).
10. S. J. Hawkins, L. B. Firth, M. McHugh, E. S. Poloczanska, R. J. H. Herbert, M. T. Burrows, M. A. Kendall et al., "Data Rescue and Re-use: Recycling Old Information to Address New Policy Concerns," *Marine Policy* 42 (November 2013): 91–98, doi:10.1016/j.marpol.2013.02.001.
11. S. Levitus, "The UNESCO-IOC-IODE 'Global Oceanographic Data Archeology and Rescue' (GODAR) Project and 'World Ocean Database' Project," *Data Science Journal* 11 (2012): 46–71, doi:10.2481/dsj.012-014; Hawkins et al., "Data Rescue and Re-use"; Kenneth R. Knapp, John J. Bates, and Bruce Barkstrom, "Scientific Data Stewardship: Lessons Learned from a Satellite-Data Rescue Effort," *Bulletin of the American Meteorological Society* 88, no. 9 (2007): 1359–61, doi:10.1175/BAMS-88-9-1359; Patrick C. Caldwell, "Tide Gauge Data Rescue," *Proceedings of The Memory of the World in the Digital Age: Digitization and Preservation* (Vancouver, British Columbia, Canada: UNESCO, 2013), 134–149. http://ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf; Jamus Collier, Stefanie Schumacher, Cornelia Behrens, Amelie Driemel, Michael Diepenbroek, Hannes Grobe, Taewoon Kim, Uwe Schindler, Rainer Sieger, and Hans-Joachim Wallrabe-Adams, "Rescued from the Deep: Publishing Scientific Ocean Drilling Long Tail Data," *GeoResJ* 6 (June 2015): 17–20, doi:10.1016/j.grj.2015.01.003.
12. David G. Gallaher, Garrett Campbell, Walter Meier, John Moses, and Dennis Wingo, "The Process of Bringing Dark Data to Light: The Rescue of the Early Nimbus Satellite Data," *GeoResJ* 6 (June 2015): 124–34, doi:10.1016/j.grj.2015.02.013.
13. Frank Kaspar, Birger Tinz, Hermann Mächel, and Lydia Gates, "Data Rescue of National and International Meteorological Observations at Deutscher Wetterdienst," *Advances in Science and Research* 12, no. 1 (2015): 57–61, doi:10.5194/asr-12-57-2015.
14. Karen M. Fallas, Robert B. MacNaughton, and Matthew J. Sommers, "Maximizing the Value of Historical Bedrock Field Observations: An Example from Northwest Canada," *GeoResJ* 6 (June 2015): 30–43, doi:10.1016/j.grj.2015.01.004.

15. Knapp, Bates, and Barkstrom, "Scientific Data Stewardship," 1359.
16. Richard Munang, Johnson N. Nkem, and Zhen Han, "Using Data Digitalization to Inform Climate Change Adaptation Policy: Informing the Future Using the Present," *Weather and Climate Extremes* 1 (September 2013): 17–18, doi:10.1016/j.wace.2013.07.001.
17. Lesley Wyborn, Leslie Hsu, Kerstin Lehnert, and Mark A. Parsons, "Guest Editorial: Special Issue: Rescuing Legacy Data for Future Science," *GeoResJ* 6 (June 2015): 106–7, doi:10.1016/j.grj.2015.02.017.
18. NASA Socioeconomic Data and Applications Center, "Millennium Ecosystem Assessment (MA)," 2005, accessed September 26, 2015, http://sedac.ciesin.columbia.edu/data/collection/ma.
19. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Synthesis* (Washington DC: Island Press, 2005), http://www.unep.org/maweb/documents/document.356.aspx.pdf.
20. US Geological Survey, "NBII to Be Taken Offline."
21. NASA Socioeconomic Data and Applications Center (SEDAC), "User Working Group," accessed November 23, 2015, http://sedac.ciesin.columbia.edu/user-working-group.
22. NASA Socioeconomic Data and Applications Center (SEDAC), "Millennium Ecosystem Assessment (MA)," 2005, http://sedac.ciesin.columbia.edu/data/collection/ma.
23. Mustapha Mokrane and Mark A. Parsons "Learning from the International Polar Year to Build the Future of Polar Data Management," *Data Science Journal* 13 (2014): PDA88–PDA93, doi:10.2481/dsj.IFPDA-15.
24. Robert R. Downs, "Data Rescue at a Scientific Data Center," (presentation at the Best Practices Exchange [BPE] 2015, Harrisburg, PA, October 19–21, 2015), Columbia University Academic Commons, doi:10.7916/D8B857Q3.
25. Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, Recommended Practice CCSDS 650.0-M-2, Magenta Book, Issue 2 (Washington, DC: CCSDS Secretariat, June 2012), http://public.ccsds.org/publications/archive/650x0m2.pdf.
26. Data Seal of Approval homepage, accessed November 23, 2015, http://datasealofapproval.org/en/; Center for Research Libraries, "TRAC Metrics," accessed November 23, 2015, http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac; International Organization for Standardization, "ISO 16363:2012: Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories," February 15, 2012, accessed November 23, 2015, http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.

Bibliography

Brunet, M., P. D. Jones, S. Jourdain, D. Efthymiadis, M. Kerrouche, and C. Boroneant. "Data Sources for Rescuing the Rich Heritage of Mediterranean Historical Surface Climate Data." *Geoscience Data Journal* 1, no. 1 (2014): 61–73. doi:10.1002/gdj3.4.

Brunet, Manola, and Phil Jones. "Data Rescue Initiatives: Bringing Historical Climate Data into the 21st Century." *Climate Research*, 47, no.1 (2011): 29–40. doi:10.3354/cr00960.

Caldwell, Patrick C. "Tide Gauge Data Rescue." *Proceedings of The Memory of the World in the Digital Age: Digitization and Preservation,* 134–149. Vancouver, British Columbia, Canada: UNESCO 2013, 2012. http://ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf.

Center for Research Libraries, "TRAC Metrics," accessed November 23, 2015, http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac.

Collier, Jamus, Stefanie Schumacher, Cornelia Behrens, Amelie Driemel, Michael Diepenbroek, Hannes Grobe, Taewoon Kim, Uwe Schindler, Rainer Sieger, and Hans-Joachim Wallrabe-Adams. "Rescued from the Deep: Publishing Scientific Ocean Drilling Long Tail Data." *GeoResJ* 6 (June 2015): 17–20. doi:10.1016/j.grj.2015.01.003.

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*. Recommended Practice, CCSDS 650.0-M-2, Magenta Book, Issue 2. Washington, DC: CCSDS Secretariat, June 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf.

Data Seal of Approval homepage, accessed November 23, 2015, http://datasealofapproval.org/en/.

Domínguez-Castro, F., Alexandre M. Ramos, Ricardo García-Herrera, and Ricardo M. Trigo. "Iberian Extreme Precipitation 1855/1856: An Analysis from Early Instrumental Observations and Documentary Sources." *International Journal of Climatology* 35, no. 1 (2015): 142–53. doi:10.1002/joc.3973.

Downs, Robert R. "Data Rescue at a Scientific Data Center." Presentation, Best Practices Exchange (BPE) 2015, Harrisburg, PA, October 19–21, 2015. Columbia University Academic Commons, doi:10.7916/D8B857Q3.

Edwards, Paul N. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press, 2010.

Fallas, Karen M., Robert B. MacNaughton, and Matthew J. Sommers. "Maximizing the Value of Historical Bedrock Field Observations: An Example from Northwest Canada." *GeoResJ* 6 (June 2015): 30–43. doi:10.1016/j.grj.2015.01.004.

Gallaher, David, G. Garrett Campbell, Walter Meier, John Moses, and Dennis Wingo. "The Process of Bringing Dark Data to Light: The Rescue of the Early Nimbus Satellite Data." *GeoResJ* 6 (June 2015): 124–34. doi:10.1016/j.grj.2015.02.013.

Griffin, R. Elizabeth. "When Are Old Data New Data?" *GeoResJ* 6 (June 2015): 92-97. doi:10.1016/j.grj.2015.02.004.

Hawkins, S. J., L. B. Firth, M. McHugh, E. S. Poloczanska, R. J. H. Herbert, M. T. Burrows, M. A. Kendall et al. "Data Rescue and Re-use: Recycling Old Information to Address New Policy Concerns." *Marine Policy* 42 (November 2013): 91–98. doi:10.1016/j.marpol.2013.02.001.

Intergovernmental Panel on Climate Change (IPCC), "Data Distribution Centre," accessed November 23, 2015, http://www.ipcc-data.org/observ/clim/ar4_global.html.

International Organization for Standardization, "ISO 16363:2012: Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories," February 15, 2012, accessed November 23, 2015, http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.

International Research Institute for Climate and Society (IRI), "Climate Data Library," accessed November 23, 2015, http://iri.columbia.edu/resources/data-library/.

Kaspar, Frank, Birger Tinz, Hermann Mächel, and Lydia Gates. "Data Rescue of National and International Meteorological Observations at Deutscher Wetterdienst." *Advances in Science and Research* 12, no. 1 (2015): 57–61. doi:10.5194/asr-12-57-2015.

Knapp, Kenneth R., John J. Bates, and Bruce Barkstrom. "Scientific Data Stewardship: Lessons Learned from a Satellite-Data Rescue Effort." *Bulletin of the American Meteorological Society* 88, no. 9 (2007): 1359–61. doi:10.1175/BAMS-88-9-1359.

Levitus, S. "The UNESCO-IOC-IODE 'Global Oceanographic Data Archeology and Rescue' (GODAR) Project and 'World Ocean Database' Project." *Data Science Journal* 11 (2012): 46–71. doi:10.2481/dsj.012-014.

Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: Synthesis*. Washington DC: Island Press, 2005. http://www.unep.org/maweb/documents/document.356.aspx.pdf.

Mokrane, Mustapha, and Mark A. Parsons. "Learning from the International Polar Year to Build the Future of Polar Data Management." *Data Science Journal* 13 (2014): PDA88–PDA93. doi:10.2481/dsj.IFPDA-15.

Munang, Richard, Johnson N. Nkem, and Zhen Han. "Using Data Digitalization to Inform Climate Change Adaptation Policy: Informing the Future Using the Present." *Weather and Climate Extremes* 1 (September 2013): 17–18. doi:10.1016/j.wace.2013.07.001.

NASA Socioeconomic Data and Applications Center (SEDAC). "Millennium Ecosystem Assessment (MA)." 2005. http://sedac.ciesin.columbia.edu/data/collection/ma.

———. "Millennium Ecosystem Assessment: MA Population." 2005. doi:10.7927/H4CF9N1K.

———. "User Working Group," accessed November 23, 2015, http://sedac.ciesin.columbia.edu/user-working-group.

Tan, L. S., S. Burton, R. Crouthamel, A. van Engelen, R. Hutchinson, L. Nicodemus, T. C. Peterson, F. Rahimzadeh. *Guidelines on Climate Data Rescue*. WMO/TD No. 1210. Edited by Paul Llansó and Hama Kontongomde. Geneva, Switzerland: World Meteorological Organization, 2004. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-55.pdf.

US National Climatic Data Center, accessed November 23, 2015, http://www.ncdc.noaa.gov.

University of East Anglia (UEA) Climatic Research Unit, "Data," accessed November 23, 2015, http://www.cru.uea.ac.uk/data.

US Geological Survey. "NBII to Be Taken Offline Permanently in January." *USGS Access Newsletter* 14, no. 3 (Fall 2011). http://www.usgs.gov/core_science_systems/Access/p1111-1.html.

Wyborn, Lesley, Leslie Hsu, Kerstin Lehnert, and Mark A. Parsons. "Guest Editorial: Special Issue: Rescuing Legacy Data for Future Science." *GeoResJ* 6 (June 2015): 106–7. doi:10.1016/j.grj.2015.02.017.