An accessible infrastructure

Professors Carol Tenopir and **Suzie Allard**, and **Mike Frame** elucidate how their studies into perceptions of data sharing among scientists are enhancing research collaboration and progress

Can you begin by outlining the objectives of Data Observation Network for Earth (DataONE)?

CT, SA & MF: DataONE is a data infrastructure that provides open, persistent, robust and secure access to Earth observational data. Supported by the US National Science Foundation as one of the initial DataNets, DataONE ensures the preservation, access and use of multi-scale, multidiscipline and multinational science data through three primary cyberinfrastucture elements, and a broad education and outreach programme.

DataONE Coordinating Nodes are located at the University of New Mexico, University of California, Santa Barbara, and University of Tennessee, in collaboration with Oak Ridge National Laboratory. Our mission is to enable the creation of new science and knowledge through universal access to data about life on Earth and the environment that sustains life.

What is the role of the Usability and Assessment (U&A) Working Group?

CT, SA & MF: The U&A Working Group conducts assessments of data sharing practices, perceptions and barriers. These include formal surveys with DataONE stakeholders and user experience tests of the DataONE system and toolkit.

How important is data sharing in the modern research environment and why is there a necessity for understanding current scientific practice regarding data sharing in research?

CT, SA & MF: Research data access, sharing, reuse and preservation are driven by the growing interdisciplinary nature of scientific problems, mandates from funders and publishers for data deposition, and the increasing availability of subject-based or institutional-based data repositories. Scientists need to understand the systems and tools available to them throughout the research data lifecycle, while DataONE seeks to understand their attitudes and needs regarding data access, sharing and reuse. Understanding current practices, needs, incentives and barriers to data sharing and reuse by scientists is vital for those who are building a culture of data sharing.

Could your DataONE survey techniques be applied to assessing other areas of innovation such as technology transfer and industrial data sharing practices?

CT, SA & MF: The habits, beliefs and practices of data sharing do have some similarity across different domains. However, some domains have very different practices; for instance, the physics community has supported sharing and open access data for years. International copyright, technology transfer and other related laws and regulations often complicate the generalisation of sharing practices across different areas of R&D. However, the techniques used by DataONE – which include identifying key stakeholder organisations, developing formal survey instruments, using interviews or key focus groups and developing comprehensive analysis – are relevant to all domains and could be readily applied.

How have collaborations with DataONE members and academic groups contributed to your research?

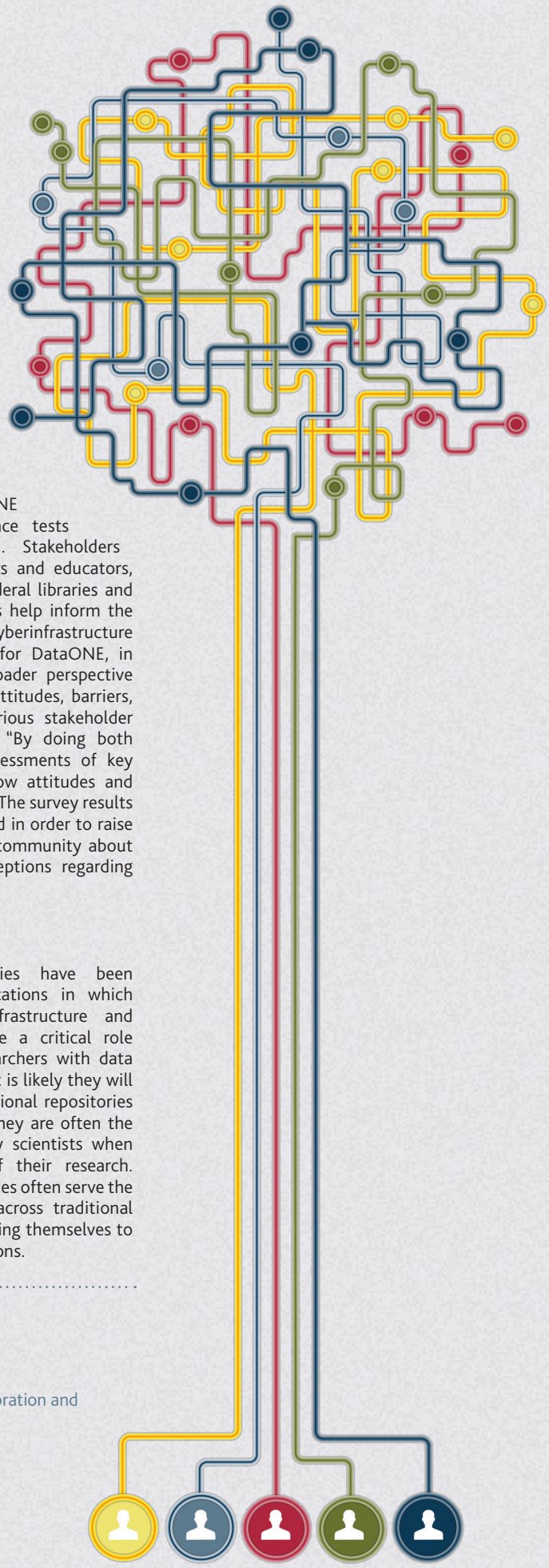
CT, SA & MF: Collaboration is essential for conducting and analysing assessments and, indeed, for the success of DataONE. We have tested pilot surveys with our collaborators and ask for assistance from the broad DataONE community to distribute the scientist surveys within their institutions and organisations. Through our biannual DataONE working group meetings, we have also sought input into survey development and analysis. This has provided a wide range of perspectives and insights into the analysis of our survey results, and helped us identify new tools and training programmes that may be required to meet these identified results. We have often seen our DataONE collaborators apply or incorporate some of the findings within their various undergraduate- or graduate-level courses.

Professor Tenopir, you have published over 200 peer-reviewed journal articles and you also participate in numerous international conferences. In which upcoming events will the results from your DataONE surveys be disseminated?

CT: Some results from our surveys with academic libraries, academic librarians and scientists will be presented in early November at the annual meeting of the Association of Information Science & Technology (ASIST) in Seattle. Suzie Allard and I will be participating in two sessions at ASIST, entitled 'Responding to Emerging Data Workforce Demand: Harnessing Data Center Expertise' and 'Transforming the Data Landscape: Connecting Data, Policies and Communities'. At the end of November, I will also be presenting results from our surveys with academic libraries and librarians at a conference in the Library & Information Studies Centre at the University of Cape Town, South Africa. Later this year, we will submit papers from these surveys to peer-reviewed journals and deposit the datasets to appropriate data repositories, such as Dryad, ONEShare and/or the University of Tennessee's institutional repository. In addition, we intend to publish results of the US Geological Survey (USGS) in an Open File report in early 2015.

The data drive

Mandates from funders and publishers are generating a growing demand for science data deposition and access to repositories. The **Usability and Assessment Working Group** for the Data Observation Network for Earth project is investigating the practices and perceptions of data sharing by scientists to improve future data management



THE PAST FEW decades have seen significant advances in Earth sciences, generated by vast numbers of environmental studies that have amassed volumes of datasets. However, this presents challenges for storing, identifying and analysing relevant information. Future scientific progress in environmental sciences depends on the implementation of effective data management systems that allow data to be stored, accessed and shared. Indeed, data sharing is vital for validating existing results, encouraging collaborations and developing innovative new research projects. There is therefore an urgent need to scrutinise current data sharing practices and launch informed strategies to improve them.

In this context, three researchers – Professors Carol Tenopir and Suzie Allard, and Mike Frame – are collaborating to evaluate the perceptions and practices that underpin the sharing of scientific data, as well as the barriers that may prevent it from happening. Working under the auspices of the Data Observation Network for Earth (DataONE) project – a collaborative network that aims to provide open, persistent, robust and secure access to Earth observation data – Tenopir and Frame are leading the Usability and Assessment (U&A) Working Group. Tenopir and Allard, both based at the University of Tennessee, USA, and Frame, Chief of Computational Science in the US Geological Survey (USGS), are all members of the DataONE leadership team.

DATAONE ASSESSMENT

The surveys conducted by the U&A Working Group comprise formal, periodic assessments of key stakeholder groups in the DataONE project and user experience tests of the DataONE system. Stakeholders include scientific researchers and educators, as well as academic and federal libraries and librarians: "Our assessments help inform the development of both the cyberinfrastructure and educational materials for DataONE, in addition to providing a broader perspective on data sharing practices, attitudes, barriers, incentives and roles of various stakeholder groups," Tenopir discloses. "By doing both baseline and follow-up assessments of key groups, we hope to see how attitudes and practices change over time." The survey results are published and distributed in order to raise awareness in the scientific community about current practices and perceptions regarding the use and reuse of data.

A KEY ROLE

Academic research libraries have been pinpointed as primary locations in which to base robust data infrastructure and services. Indeed, they have a critical role to play in supporting researchers with data management; for instance, it is likely they will have partnered with institutional repositories for academic outputs and they are often the first point of call for many scientists when in the planning stages of their research. Additionally, academic libraries often serve the whole institution, cutting across traditional subject boundaries and lending themselves to multidisciplinary collaborations.

Advantages of data sharing

- Reanalysis of data enables the validation of results
- The use of different approaches to analyse the same data encourages collaboration and advances scientific understanding
- Widespread availability of data prevents fabrication of results
- Data availability is a time-saving resource for scientists

A pioneering project

Data Observation Network for Earth (DataONE) is a collaborative effort across multiple institutions that aims to provide access to and preserve datasets related to the environmental sciences. As a distributed network of data centres, scientific networks and organisations, it enables members to expose their data within the DataONE community, as well as providing computing resources and services such as data replication. Community engagement and educational outreach are also central components of DataONE. Ultimately, the initiative aims to forge a stronger understanding of the environment and conserve life on Earth.

In 2011 and 2014, Tenopir and her team conducted empirical investigations into the practices of research data services (RDS) of librarians in US and Canadian academic institutions. The surveys focused on uncovering the attitudes of librarians towards providing RDS and their views on the importance of the services in their respective libraries. Tenopir's team also aimed to identify the factors that enhance and prevent librarian engagement in data storage and sharing. While the results highlighted that few academic institutions offer RDS at present, many have plans to expand this area in the future. Although it is still early days, some libraries have already begun to develop their RDS by helping various university research organisations implement data management plans and find appropriate data repositories. Some of these libraries are even beginning to provide initial training on available data management tools and are aiming to serve as long-term archives for research data. One good example is the US Institute of Museum and Library Services (IMLS) which has supported the development of data management education in several schools of information sciences, including the University of Tennessee.

FACING THE CHALLENGES

The results of the surveys highlighted that lack of time and money were perceived as the biggest barriers to data sharing. However, Tenopir and her team believe that these barriers could be overcome by equipping managers and librarians with the tools to assist scientists with data description and deposition through the different stages of the science data life cycle.

After time and money, a third barrier appears to be the desire for data restriction. While most scientists reported their willingness to share at least some of their data with few or no restrictions, in some cases restrictions may be desired. This barrier could be overcome by building strict embargoes into data repositories and developing measures that ensure each researcher receives credit for their own dataset when cited by others: "Accessibility and preservation of data for the future is important, even when immediate reuse is not possible and citing data in a standard way needs to become part of research practice, just as citing published papers is now," Tenopir states. Furthermore: "When data repositories are linked to increased use and when a combination of data managers and data repositories make it more convenient to deposit, locate and reuse data, then data sharing is also likely to increase".

FUTURE AIMS

The DataONE surveys have been made available to many researchers in both academic and governmental institutions, flagging up useful insights into current data sharing practices and informing those who are building systems and services that enable data management. For example, USGS – an organisation within the US Department of the Interior, primarily consisting of researchers specialising in Earth science issues in the US – has reshaped its data management policies due to a survey conducted with its data managers and researchers based on the DataONE survey questions.

Looking ahead, the hope is that USGS's approach has set a precedent for the future of data storage, access and sharing, both in the US and internationally. In collaboration with DataONE partners, the University of Tennessee is educating a new generation of research data specialists. And as DataONE continues to amass huge quantities of Earth science data, it is likely that more researchers will join the network in order to contribute their own results and access available expertise: "As the assessments show, many scientists are willing to share their data, but they may not know how or where to make this happen," Tenopir points out. "Crucially, DataONE helps to build a culture of data sharing in all stages of the data lifecycle."

INTELLIGENCE

DATA SHARING BY SCIENTISTS: PRACTICES AND PERCEPTIONS

OBJECTIVES

- To conduct periodic surveys of key stakeholder groups, notably scientists and librarians, to assess data sharing practices, perceptions and barriers
- To help the broader research community understand the prevailing practices and attitudes surrounding research data management, use and reuse

KEY COLLABORATORS

For a full list of former and current Usability and Assessment (U&A) Working Group members, please visit: **www.dataone.org/working_groups/usability-and-assessment**

FUNDING

This work was supported by Data Observation Network for Earth (DataONE), National Science Foundation (NSF) award #0830944 under a Cooperative Agreement.

CONTACT

Carol Tenopir
Chancellor's Professor, School of Information Sciences
U&A Working Group co-Leader

College of Communication and Information
University of Tennessee
1345 Circle Park Drive
451 Communications Building
Knoxville, TN 37996-0341, USA

T +1 865 974 7911
E ctenopir@utk.edu

http://bit.ly/1ymh9Ea

CAROL TENOPIR is Chancellor's Professor and Board of Visitors Professor at the School of Information Sciences, University of Tennessee. She is a member of the DataONE Leadership Team and co-lead of the DataONE U&A Working Group, and PI on many grants that study how researchers access and use scholarly information and data.

SUZIE ALLARD is Associate Dean of Research for the College of Communication and Information, Director of the Center for Information and Communication Studies and Professor in the School of Information Sciences at the University of Tennessee. Allard is a member of the DataONE Leadership Team, co-PI of DataONE and PI on many educational grants for science data management.

MIKE FRAME is Chief of Computational Science, US Geological Survey. He is a member of the DataONE Leadership team and co-lead of the DataONE U&A Working Group. He has served as PI on several NSF projects, international councils and initiatives.

THE UNIVERSITY of TENNESSEE UT KNOXVILLE