

Measuring the Impact of Digital Repositories

February 28-March 1, 2017

Participant Introductions

Bruce Ambacher's career spans four decades at NARA, George Mason University, and the University of Maryland. His responsibilities included service as acting chief of NARA's digital preservation unit and as court-appointed preservation officer for the PROFS, Iran-Contra, and Clinton email collections. He represented NARA on the Federal Geographic Data Committee and helped develop federal and international geospatial standards. He was NARA's representative for the OAIS Reference Model. He co-chaired the development of TRAC. He helped develop the trustworthy digital repositories standards ISO 16363 and ISO 16919. He began teaching courses in archives and digital preservation at George Mason University in 1984, became an adjunct professor at the University of Maryland in 2000 and a Visiting Professor between 2007 and 2013. He has consulted on digital preservation for industry and cultural humanities institutions. He is a Research Affiliate of the Digital Curation Innovation Center at the University of Maryland

Mary Barlow is head of the Strategic Project Management Office at the European Bioinformatics Institute (EMBL-EBI), and has oversight of impact reporting and the development of infrastructure business cases. Prior to taking on this role, Mary served as a Programme Manager for the multi-million pound investment in EMBL-EBI by the UK government's Large Facilities Capital Fund. This programme included the construction of new office space and the on-going public procurement of ICT infrastructure to support EMBL-EBI's growing public databases. Mary's work prior to EMBL-EBI focused on ICT integration and intelligent buildings. She began her career working on broadcast projects in BBC Technology, after obtaining her degree in Maths and Engineering at Nottingham University.

Chaitan Baru is Senior Advisor for Data Science, CISE Directorate, National Science Foundation. He co-chairs the NSF working group on Harnessing the Data Revolution "Big Idea"; is advisor to the NSF Big Data Regional Innovation Hubs and Spokes program (BD Hubs/Spokes); manages the cross-Foundation NSF BIGDATA program; and, is member of the NSF Transdisciplinary Research in Principles of Data Science (TRIPDS) program. He co-chairs the Big Data Inter-agency Working Group of the Networking and IT R&D program, White House Office of Science and Technology Policy, and is a primary co-author of the Federal Big Data R&D Strategic Plan, released May 2016. He is also a member of the NITRD Data Science Interagency Working Group and a member of the Federal Data Cabinet, representing NSF. He is General Chair for ICDE 2017—33rd IEEE International Conference on Data Engineering (ICDE 2017), to be held April 19-22, 2017, in San Diego, CA.

Avi Bender is the Director of the National Technical Information Service (NTIS), a Bureau within the Department of Commerce. The Bureau is a fee-for-service organization that assists Federal agencies

with their mission critical data priorities through an innovative joint venture partnership with the private sector. Avi has extensive experience in digital media in both the private sector and the Federal government. Most recently, he served as the Chief Technology Officer of the US Census Bureau.

Francine Berman is the Edward P. Hamilton Distinguished Professor in Computer Science at Rensselaer Polytechnic Institute (RPI). She is a Fellow of the ACM, AAAS, and IEEE. In 2009, Berman was the inaugural recipient of the ACM/IEEE-CS Ken Kennedy Award for "influential leadership in the design, development, and deployment of national-scale cyberinfrastructure." In 2015, she was nominated by President Obama and confirmed by the U.S. Senate to become a member of the National Council on the Humanities. Berman is former Director of the San Diego Supercomputer Center, former HPC Endowed Chair at UC San Diego, former Vice President for Research at RPI, and former co-Chair of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. She currently serves as U.S. lead of the Research Data Alliance (RDA), a community-driven international organization created to accelerate research data sharing world-wide.

Helen M. Berman is a Board of Governors Professor Emerita of Chemistry and Chemical Biology at Rutgers, The State University of New Jersey. From 1998-2014, she was the Director of the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). RCSB PDB is a member of the Worldwide Protein Data Bank (wwPDB) that manages the PDB archive of information about the structures of proteins, nucleic acids, and complex assemblies. She also serves in leadership roles for the EMDatabank, the Structural Biology Knowledgebase, and the Nucleic Acid Database. She received her Ph.D. in Chemistry in 1967 from the University of Pittsburgh under the direction of George Alan Jeffrey. Prior to joining the Rutgers faculty in 1989, she was a faculty member at the Fox Chase Cancer Center. In addition to her work on structural databases and ontologies, she has had an active research career in structural biology with a particular focus on nucleic acid-containing systems and collagen. She is the recipient of several awards, including the Benjamin Franklin Award for Open Access in the Life Sciences, the DeLano Award for Computational Biosciences, and the Carl Brändén Award.

Wo Chang is the Digital Data Advisor for the NIST Information Technology Laboratory (ITL). His responsibilities include data interoperability; promoting a vital and growing Big Data community at NIST and with external stakeholders in the commercial, academic, and government sectors. Mr. Chang currently the Convener of the ISO/IEC JTC 1/WG9 on Big Data, co-chairs the NIST Big Data Public Working Group. In the past, Mr. Chang was manager of the Digital Media Group in ITL with duties included oversees key projects in digital data archival and preservation, management of electronic health records, motion image quality, and multimedia standards. Chang was the Deputy Chair for the US INCITS L3.1, chaired other key projects for MPEG, participated HL7 and ISO/IEC TC215 for health informatics, IETF protocols development, and original member of the W3C's SMIL and developed one of the SMIL reference software. Mr. Chang's research interests include many CPUs/Cores/GPUs high performance analytics and computing, digital data mashup, metadata description, multimedia synchronization, and Internet protocols.

G. Sayeed Choudhury is the Associate Dean for Research Data Management and Hodson Director of the Digital Research and Curation Center at the Sheridan Libraries of Johns Hopkins University. He is a President Obama appointee to the National Museum and Library Services Board. Choudhury is a member of the Executive Committee for the Institute of Data Intensive Engineering and Science (IDIES) based at Johns Hopkins. He is also a member of the Board of the National Information Standards Organization (NISO) and a member of the Advisory Board for OpenAIRE2020. He has been a member of the National Academies Board on Research Data and Information, the ICPSR Council, the DuraSpace

Board, Digital Library Federation advisory committee, Library of Congress' National Digital Stewardship Alliance Coordinating Committee, Federation of Earth Scientists Information Partnership (ESIP) Executive Committee and the Project MUSE Advisory Board.

Mark Conrad is an archives specialist in the Systems Engineering Branch of the National Archives and Records Administration's (NARA) Office of Information Services. He works with computer scientists and engineers from around the world and NARA staff to ensure that NARA takes advantage of the latest relevant technological developments in carrying out its mission. He has 26 years experience working with electronic records and data. He is a member of the working group that produced ISO 14721 - Open archival information system (OAIS) -- Reference model, ISO 16363 - Audit and certification of trustworthy digital repositories, and ISO 16919 - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories. Mark was a Visiting Fulbright Scholar in the Archives Department of University College Dublin, Ireland, where he taught courses on electronic records issues. He also taught electronic records management at the University of Dundee, Scotland.

Mercè Crosas is the Chief Data Science and Technology Officer at the Institute for Quantitative Social Science (IQSS) at Harvard University. She has more than 10 years of experience leading the Dataverse project and more than 15 years of experience building data management and analysis systems in industry and academia. She is part of numerous committees and working groups focus on research data management, data citation and standards. Crosas is currently co-PI of the Dataverse Project, with IQSS faculty director Gary King, and supervises the Zelig project for statistical analysis, Consilience for text analysis, the Data Science Services and Data Curation team at IQSS. She currently collaborates with the Harvard Privacy Tools project led by Salil Vadhan (Harvard), the Provenance project with Margo Seltzer (Harvard), the Structural Biology Grid Data project with Piotrek Sliz (Harvard Medical School), and the Massachusetts Open Cloud (Orran Krieger and Piyana Saowarattitada, Boston University), among others. Crosas holds a Ph.D. in Astrophysics and a B.S. in Physics. More at mercecrosas.com and [@mercecrosas](https://twitter.com/mercecrosas).

Patricia (Trisha) Cruse is the Executive Director of DataCite, a leading global non-profit organization that provides persistent identifiers (DOIs) for research data. Trisha comes to her position with a strong commitment and passion for data sharing. As Executive Director, her role is to advance DataCite's mission, build strategic partnerships and work with a diverse community of stakeholders. Prior to joining DataCite, Trisha was the Director of the University of California Curation Center (UC3) at the California Digital Library (CDL) and launched products such as the DMPTool. In addition, Trisha is a co-PI on the DataONE initiative and co-leads sustainability and governance activities. Trisha holds an MLIS from the University of California, Berkeley.

Robert R. Downs is senior digital archivist and acting head of cyberinfrastructure and informatics research and development at CIESIN, the Center for International Earth Science Information Network, a research and data center of the Earth Institute of Columbia University. Holding the PhD in Information Management from the Stevens Institute of Technology, he is Vice-Chair of the Columbia University Morningside Institutional Review Board, a Senior Member of the Association for Computing Machinery (ACM), co-leader of the GEOSS Evolve Data Management Principles team, co-chair of the RDA Interest Group on Repository Platforms for Research Data, a member of the Data Archive Ingest group of the Consultative Committee for Space Data Systems, and on the Editorial Board of the CODATA Data Science Journal. Other memberships include the American Geophysical Union (AGU), the Association for Information Science and Technology (ASIS&T), and the International Association for Social Sciences Information Services and Technology (IASSIST).

Anita de Waard has a degree in low-temperature physics from Leiden University, and worked in Moscow before joining Elsevier as a physics publisher in 1988. Since 1997, she has worked on bridging the gap between science publishing and computational and information technologies, collaborating with academic groups in Europe and the US. From January 2006 onwards, de Waard has been conducting research through the University of Utrecht, on a discourse analysis of scientific narrative, with an emphasis on finding key epistemic components in biological text. Her past accomplishments include working on a semantic model for research papers, and co-founding the interdisciplinary member organization 'FORCE11: The Future of Research Communications and E-Science' (Force11.org). For her current remit as VP of Research Data Collaborations, Anita is developing cross-disciplinary frameworks for sharing data and tools to store, share and search experimental outputs, in collaboration with academic and government groups, in the US and Europe. She cochaired the RDA Working Group on Cost Recovery Models for Data Centers and is currently chairing the RDA Data Discovery Interest Group, is vice-chair of the Steering Committee of the National Data Service.

Mark D. Flood did his undergraduate work at Indiana University in Bloomington, where he majored in finance (B.S., 1982), and German and economics (B.A., 1983). In 1990, he earned his Ph.D. in finance from the Graduate School of Business at the University of North Carolina at Chapel Hill. He has taught finance and business at universities in the U.S. and Canada, and worked as a financial economist on issues of regulatory policy and risk management at the Federal Reserve Bank of St. Louis, the Office of Thrift Supervision, the Federal Housing Finance Board, and the Federal Housing Finance Agency. He was a founding member of the Committee to Establish a National Institute of Finance. He is currently a Research Principal in the Treasury's Office of Financial Research.

Susan K. Gregurick is the Division Director for Biomedical Technology, Bioinformatics and Computational Biology (BBCB) in NIH's National Institute of General Medical Sciences (NIGMS). Her mission in BBCB is to advance research in computational biology, behavioral and data sciences, mathematical and biostatistical methods, and biomedical technologies in support of the NIGMS mission to increase understanding of life processes. Prior to joining the NIH, Susan was a program manager for the Department of Energy where she oversaw the development and implementation of the DOE Systems Biology Knowledgebase, which is a framework to integrate data, models and simulations together for a better understanding of energy and environmental processes. During Susan's academic career she was a Professor of Computational Biology at the University of Maryland, Baltimore County and her research interests include dynamics of large biological macromolecules. Susan holds a Ph.D. in Computational Chemistry and her areas of expertise are computational biology, high performance computing, neutron scattering and bioinformatics.

Tony Hey did not submit a short introductory bio but you can read about him at: https://en.wikipedia.org/wiki/Tony_Hey .

Eva Huala was a founding member of TAIR (the Arabidopsis Information Resource) and has served as its Director since 2004. In 2013 she co-founded and became Executive Director of the nonprofit organization Phoenix Bioinformatics, dedicated to exploring new funding approaches to sustainably support research infrastructure. Earlier work included graduate work in nitrogen fixation and postdoctoral research on the regulation of flowering and in phototropism. Dr. Huala has a bachelor's degree in Biology from the University of California Santa Cruz and a PhD in Genetics from Harvard University.

Yusuke Kuwayama is a Fellow at Resources for the Future (RFF) and Director of Socioeconomic Studies for the Consortium for the Valuation of Applications Benefits Linked with Earth Science (VALUABLES), a cooperative agreement between RFF and the National Aeronautics and Space Administration (NASA). Kuwayama's research focuses on the economics of water resource management and the value of Earth science information. His work has been published in outlets such as the Journal of Environmental Economics and Management, Land Economics, the American Journal of Agricultural Economics, and Hydrogeology Journal. He received his Ph.D. in Agricultural and Applied Economics and M.S. in Economics from the University of Illinois as well as an A.B. in Economics from Amherst College.

Margaret Levenstein is Director of ICPSR, the Interuniversity Consortium for Political and Social Research, Research Professor at the Survey Research Center in the Institute for Social Research and at the School of Information, and Adjunct Professor of Business Economics and Public Policy at the Stephen M. Ross School of Business. She has taught economics at the University of Michigan since 1990. She also serves as Executive Director of the Michigan Federal Statistical Research Data Center and co-PI of the Michigan Center for the Demography of Aging, where she has responsibility for access to restricted data resources. She is the Associate Chair of the American Economic Association's Committee on the Status of Women in the Economics Profession and past President of the Business History Conference. She is PI of CenHRS, a Sloan Foundation-funded project building an enhancement to the Health and Retirement Study based on linkages to administrative and survey data on HRS employers and co-workers. She received a Ph.D. in Economics from Yale University and a B.A. from Barnard College, Columbia University. She is the author of numerous studies on competition and collusion, the development of information systems, and using "organic" data to improve social and economic measurement. Her project using Tweets to predict unemployment is updated weekly at <http://econprediction.eecs.umich.edu/study>. You can see her discuss her research on the impact of the 1930s Great Depression on innovative firms in the Midwest at <http://www.youtube.com/watch?v=g8Ms7s-tPM4>.

Jennifer Lin has fifteen years' experience in product development, project management, community outreach, and change management within scholarly communications, education, and the public sector. She is currently Director of Product Management at Crossref, a scholarly infrastructure provider, developing metadata services that make scholarly content easy to find, cite, link, and assess. She previously worked for PLOS where she oversaw product strategy and development for their data program, article-level metrics initiative, and open assessment activities. Jennifer earned her PhD at Johns Hopkins University. [ORCID profile](#).

Peter Lyster is a program manager in computational biology at the National Institute of Health (NIH)/National Institute of General Medicine Sciences (NIGMS), Division of Biomedical Technology, Bioinformatics, and Computational Biology. For the past year and a half he was Deputy Director of NITRD NCO. At the NIH, Dr. Lyster managed a broad program in biomedical informatics and computational biology. He also served as the lead of the National Institute of General Medical Sciences (NIGMS) National Centers for Systems Biology (NCSB) program and was co-lead of the Big Data (BD2K) Centers of Excellence in Big Data Science. He was formerly a Program Director at the National Institute of Biomedical Imaging and Bioengineering. Prior to joining NIH, Dr. Lyster worked at the Jet Propulsion Laboratory in Pasadena, California, where he laid the groundwork for some of the scientific applications of the Federal High Performance Computing and Communications (HPCC) initiative. He then became Principal Investigator in the HPCC program at NASA Goddard Space Flight Center at the University of Maryland. As such, he led the effort to use massive computing power to improve data collection for weather and climate models. Dr. Lyster received his Ph.D. in Physics and Computational Science from

Cornell University. His thesis topic was computational plasma (nuclear) physics. His expertise is in computational science and technology in a range of domain areas that have societal relevance: nuclear energy research, climate and weather research, high-performance computing, networking and communications, and biomedicine and life sciences.

Jo McEntyre is Team Leader for Literature Services at the European Bioinformatics Institute (EMBL-EBI), where she is responsible for developing public resources around full-text life science research articles such as Europe PMC. Her primary focus is on the integration of the open access literature with open life sciences data, to develop new ways for researchers to access and discover useful information. As Executive Co-Chair of the ELIXIR Data Platform, with Christine Durinx (Swiss Institute of Bioinformatics) Jo is leading on the processes and impact indicators to be used for the identification of ELIXIR Core Data Resources. Before joining the EMBL-EBI, Dr McEntyre was a scientist at the NCBI, NIH, where she worked on various literature-related resources and website usability. Before that, she was the Editor of the review journal Trends in Biochemical Sciences (TiBS).

Mustapha Mokrane is the Executive Director of the World Data System (WDS) since 2012. In this capacity, he cooperates with and contributes to global activities in the field of research data. Mustapha serves on several international committees such as the Belmont Forum e-Infrastructure and Data Management Steering Committee, the Organizational Advisory Board of the Research Data Alliance (RDA) and the Data Sharing Principles Task Force of the Group on Earth Observations (GEO). Previously, he worked at the International Council for Science (ICSU) as Science and Information Technology Officer in charge of the coordination of ICSU's Scientific Data and Information activities and the liaison with its partners. He trained as molecular biologist in Marseille, France, after moving from Algeria his home country. His scientific background covers genetics and bioinformatics and he developed a strong interest for scientific data and information challenges. He holds a Ph.D. in Molecular Biology from the Aix-Marseille University.

Mary Moulton joined the National Transportation Library (NTL) in 2011 following a career in the private sector. NTL's digital repository serves as central clearinghouse and archive for publicly accessible transportation information, including federally funded research results, statistical data products, and other information needed for transportation decision-making at the Federal, State, and local levels. As Digital Librarian, Mary leads the development and enhancement of digital repository services, information organization, digital curation, and discovery and use of NTL resources. In addition to an MLIS degree, she has a BS in Plant Science and MS in Entomology.

Lucy Nowell is a computer scientist and program manager in the Office of Advanced Scientific Computing Research (ASCR) within the Department of Energy's Office of Science. She manages a broad spectrum of ASCR-funded computer science research, with emphasis on scientific data management, analysis and visualization. Previously she served as a Program Director in the Office of Cyberinfrastructure at the National Science Foundation and as a research Program Manager for the Department of Defense, managing projects related to information analysis and visualization. Formerly a theatrical designer, her own research lies at the intersection of computer science, psychology, and visual art. Her MS and Doctorate in Computer Science are from Virginia Tech. She has a BA and MA in Theatre from the University of Alabama – Tuscaloosa and the MFA from the University of New Orleans.

Kim D. Pruitt is a Senior Staff Scientist at the National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health. She received her B.S. degree from Syracuse University and her Ph.D. degree in Genetics and Development from Cornell University in 1990. She lead

the NCBI Reference Sequences (RefSeq) project since its inception in 1999 and oversaw the projects growth from annotating the initial draft human genome assembly to providing an internationally recognized genome annotation resource for all super-kingdoms. She is currently the Director of NCBI's Data Services Division and is involved with strategic planning and oversight of literature, sequence, medical genetics, and chemistry products and services.

Dean N. Williams: For almost three decades, I have been the Chair or Principal Investigator leading several DOE projects related to "Big Data" initiatives, including the Earth System Grid Federation (ESGF)[1], the Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT)[2], and the International Climate Network Working Group (ICNWG)[3]. These multi-agency international collaborations provide interactive views of future climate changes based on projected natural and human factors. These projects were also asked to support the Intergovernmental Program and Climate Change, work which was honored with a 2007 Nobel Peace Prize, awarded jointly to former Vice President Al Gore and the IPCC team. The goal of my work has been to enhance interoperability between common components and analysis tools, and to enable end-to-end simulations and analysis workflows for science communities. This includes large-scale project management for diverse interagency scientific research and next generation solutions.