

Building Global Partnerships

Research Data Alliance 2nd **Plenary Meeting**

September 16-18, 2013 National Academy of Sciences Washington DC, US







The Research Data Alliance (RDA) Second Plenary Meeting is sponsored by RDA/US through grants from the United States of America National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST).



Sponsors for RDA also include RDA/EU (sponsored by the European Commission) and the Australian National Data Service.

Microsoft Research

RDA thanks Microsoft Research for generously sponsoring our Monday evening reception and poster session.



Current Second Plenary program is online at http://www.rd-alliance.org. Follow the meeting on Twitter (@resdatall, #RDAPlenary).

TABLE OF CONTENTS

Plenary Logistics	3
Program Overview	6
Program Details	11
Plenary I	
Plenary II	
Plenary III	
Plenary IV	
Breakout Sessions I	
Plenary V	
Plenary VI	
Breakout Sessions II	
RDA Plenary 2 Poster Session	
RDA Technical Advisory Board Candidates	



Plenary Logistics

Dates and Times:

Mon., Sept. 16, 2013	09:00 - 17:30	Plenary at the National Academies of Science
Mon., Sept.16, 2013	19:00 - 20:30	Reception and Poster Session at the Marriott
Tues., Sept.17, 2013	09:00 - 18:15	Breakout Sessions at the Washington Marriott
Tues., Sept.17, 2013	19:30 - 22:30	Conference Dinner on the Odyssey Cruise Ship
Wed., Sept.18, 2013	09:00 - 12:30	Half-day Plenary at the NAS
Wed., Sept.18, 2013	14:00 - 17:00	Breakout Sessions at the Washington Marriott

Updates to the agenda will be made available online at: <u>https://rd-alliance.org/programme.html</u>

National Academy of Sciences (NAS)

2101 Constitution Avenue NW, Washington DC 20418 (Auditorium)

Plenary sessions are being held at the NAS. Please bring a government-issued photo identification or passport with you to get through NAS Security. *Please plan to check in on Monday morning at the NAS to receive your conference badge.*

NAS is Metro-accessible via the Foggy Bottom Station (Blue Line; 0.7 miles). Very limited parking is available at the NAS. The parking lot entrance is on 21st Street between C Street NW and Constitution Avenue. Parking is strongly discouraged at the NAS. For information on Metro rates and schedules, please visit <u>www.wmata.com</u>

Washington Marriott

1221 22nd Street NW, Washington DC 20037 (Second Floor Ballrooms)

The reception and poster session on Monday evening, as well as breakout sessions on Tuesday and Wednesday, are being held at the Washington Marriott. The Washington Marriott is approximately one mile from the NAS. If you need assistance with a hotel reservation at the Washington Marriott, please contact Trinh Lieu at trinh@cmpinc.net.

The Marriott is Metro-accessible via Dupont Circle Station (Red Line; 0.3 miles) and Foggy Bottom Station (Blue Line; 0.4 miles). Self-parking in the hotel garage with full in and out privileges is \$7.00 USD hourly, \$36.00 USD daily.

Internet Access

At NAS, complimentary wireless internet is available on an open network, no password required. At the Washington Marriott, complimentary internet access is available in the lobby and guest rooms; however internet in the breakout sessions will require a passcode. Please see the conference registration desk for more information.



RDA Secretariat & Web Help Desk:

If you have questions about RDA, the website, and our online tools, or if you would like information about working with RDA, please visit the RDA Help Desk. We are here to assist you:

- September 16, NAS: 13:15 13:45 (lunch) and 15:30 16:00 (coffee break)
- September 17, Marriott: 11:00 13:00 & 15:00 17:00
- September 18, NAS: 10:30 11:00 (coffee break)

Conference Meals

Breakfast will be provided each day of the conference, prior to the agenda start time. Please note that at the NAS, food and beverages are not permitted inside the Auditorium so please plan to arrive early to enjoy breakfast. Lunch and snacks will be provided Monday and Tuesday, in addition to the Monday evening reception and Tuesday conference dinner. Please note that lunch on Wednesday is on your own. More details about the conference dinner, including transportation to the cruise ship, can be found in this packet.

RDA Poster Session Set-Up

The RDA Plenary 2 poster session takes place on Monday September 16th at the Evening Reception sponsored by Microsoft Research which takes place at 19:00 in the West End Ballroom on the second floor of the Washington Marriott. Set up of posters takes place immediately after the end of the plenary sessions and all posters should be mounted by 7pm. Please note your poster number in the Poster Session Listing later in this packet; the number corresponds to the poster board assigned to you.

Live Webcasting and Archival

The Plenary will be live webcasted on both Monday and Wednesday at the following URL: <u>http://www.tvworldwide.com/events/rda/130916</u>. Please encourage those unable to attend in person to participate and ask questions online. The webcast will be archived and available by the end of the week and will remain available for one year.

QR Code and Online Access

For easy access to the RDA website and up-to-date Plenary information and agenda, scan this code with your smartphone or tablet or go to <u>http://www.rd-alliance.org/</u>







RDA Second Plenary Conference Dinner Tuesday, September 17, 2013 Odyssey Dinner Cruise

Complimentary transportation is available! Buses depart from the Washington Marriott at 18:45

Boarding Location: 600 Water Street SW, Washington DC **Boarding Time:** 18:30 – 19:30 **Cruise Time:** 19:30 – 22:00 **Dress Code:** Conference attire is acceptable on the cruise

Not taking the bus?

The Odyssey departs from the Gangplank Marina at the corner of 6th and Water Streets. For parking, turn into the Gangplank Lot off of Water Street. Parking is limited and on a first come first serve basis. Before boarding, you will need to pick up a boarding pass at the cruise ticket booth, located next to the ship.

Please visit www.OdysseyCruises.com for more information.





Program Overview

Monday September 16th | National Academy of Sciences Building

08:00 - 09:00	Registration and Continental Breakfast
09:00 - 09:40	Plenary I a - Welcome and Opening Keynote
	Chair: Farnam Jahanian NSF Assistant Director for the Computer and
	Information Science and Engineering (CISE) Directorate
09:00 - 09:05	Welcome - Fran Berman, Co-chair RDA, Chair RDA/US
09:05 - 09:15	Introduction and Opening of Plenary - Farnam Jahanian, NSF Assistant
	Director for the Computer and Information Science and Engineering (CISE)
	Directorate
09:15 - 09:35	Tom Kalil, Deputy Director for Technology and Innovation, White House
	Office of Science and Technology Policy
09:40 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing
09:40 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management,
09:40 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins University
09:40 - 10:30 09:40 - 10:05	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins University Open and Machine Readable: Living In The New Default World
09:40 - 10:30 09:40 - 10:05	 Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins University Open and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage Bionetworks
09:40 - 10:30 09:40 - 10:05 10:05 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins UniversityOpen and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage BionetworksFuelling and Transforming Evidential Cultures of Research
09:40 - 10:30 09:40 - 10:05 10:05 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins UniversityOpen and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage BionetworksFuelling and Transforming Evidential Cultures of Research Carole Palmer, Director, Center for Informatics Research in Science and
09:40 - 10:30 09:40 - 10:05 10:05 - 10:30	 Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins University Open and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage Bionetworks Fuelling and Transforming Evidential Cultures of Research Carole Palmer, Director, Center for Informatics Research in Science and Scholarship (CIRSS) and Professor Graduate School of Library and Information
09:40 - 10:30 09:40 - 10:05 10:05 - 10:30	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins University Open and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage Bionetworks Fuelling and Transforming Evidential Cultures of Research Carole Palmer, Director, Center for Informatics Research in Science and Scholarship (CIRSS) and Professor Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign
09:40 - 10:30 09:40 - 10:05 10:05 - 10:30 10:30 - 11:00	Plenary I b - Benefits and Possibilities of Open Data Sharing Chair: Sayeed Choudhury Associate Dean for Research Data Management, The Johns Hopkins UniversityOpen and Machine Readable: Living In The New Default World John Wilbanks, Chief Commons Officer, Sage BionetworksFuelling and Transforming Evidential Cultures of Research Carole Palmer, Director, Center for Informatics Research in Science and Scholarship (CIRSS) and Professor Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-ChampaignCoffee Break

11:00 -12:45	Plenary II - RDA Business including news from Council and the firstOrganisational AssemblyThis is the general business meeting for RDA. We will introduce new Councilmembers and Working Groups. We will update the membership on keyWorking Group activities, evolving organizational structures, and other RDAbusiness. This session will also include the first Organizational Assembly,allowing the membership to better understand the role of privateorganizations as members in RDA.Chair: Herman Stehouwer, RDA Secretariat
11:00 -11:30	Council and Secretariat updates: introduce new members, bids for Plenary 4, other business
11:30 - 12:00	Updates from currently active and recognized WGs
12:00 -12:45	Organizational Assembly - Member organizations of RDA
12:45 - 13:45	Networking Lunch

13:45 -15:30Plenary III - Building Global Partnerships. The Perspective of Funders and
Affiliate Organisations
This session seeks to get perspectives from governments and organizations



	related to RDA on the overarching theme of building global partnerships. A panel of Government agency representatives from the EC, Australia, and the US will discuss their perspectives on how agencies can collaborate and promote the development of a data sharing infrastructure. An organizational panel consisting of a few data-related organizations, both broad and disciplinary in scope, will be discussing how they can most effectively interact with RDA highlighting actionable ideas or means for connection. Active participation from the audience will be encouraged. Chair: Chris Greer, NIST Acting Director, Cyber Physical Systems and Smart Grid Program Office, Engineering Laboratory
14:00 -14:45	RDA Colloquium Panel Discussion Mark Suskin, Deputy Director, Division of Advanced Cyberinfrastructure (CISE/ACI), US National Science Foundation Kostas Glinos, European Commission, Head e-Infrastructure Clare McLaughlin, Counsellor - Education and Science, Australian Embassy to Belgium and Luxembourg and Mission to the European Union and NATO in Brussels
14:45 -15:30	Affiliate organizations Panel Discussion Sara Graves, CODATA Curt Tilmes, ESIP Phil Archer, W3C Jan Brase, DataCite
15:30 -16:00	Coffee Break

this session plays in h the interi TAB and H Chair: Lar (CNBI)	on, you will learn what the TAB is all about and the critical role it elping RDA Working Groups conduct their activities. You'll meet m members of the hear from the candidates running for election to the Board. Try Lannom, VP, Corporation for National Research Initiatives
collabora	tion
The Tech	hical Advisory Board provides the technical roadmap for RDA. In



Tuesday September 17th | Washington Marriott

08:00 - 09:00	Continental Brea	Continental Breakfast						
09:00 -11:00	Breakout Sessio	on I						
Room	Salon A	Salon B	Salon C	Salon D	Salon E	Salon F	Salon G	
09:00 - 11:00	Data Foundations & Terminology	RDA/WDS Certification	Engagement	Agricultural Data	RDA for Newcomers	Legal Interoperability	RDA Colloquium last half with Council (by invitation)	
11:00 - 11:15	Coffee break							

11:15 - 13:15		Breakout Session	II						
Room	Salon A	Salon B	Salon C	Salon D	Salon E	Salon F	Salon G	Salon H	Logan
11:15 - 12:15	Data Foundations & Terminology	Global Registry of Trusted Data Repositories and Services	Engagement	Agricultural Data	Data Citation	Legal Interoperability	Economic Models & Infrastructures for Federated Materials Data Management	Cost data proposal on publishing dataS. Callaghan	Council (by invitation)
12:15 - 13:15	Practical Policy	Global Registry of Trusted Data Repositories and Services	Marine Data	Agricultural Data	Data Citation	Toxicogenomics Interoperability	Research Data Provenance		Council (by invitation)
13:15 - 14:00	٨	Networking Lunch							

RDA 2nd Plenary Documentation



14:00 - 16:00	Breakout Se	ession III						
Room	Salon A	Salon B	Salon C	Salon D	Salon E	Salon F	Salon G	Salon H
14:00 - 15:00	Practical Policy	RDA/WDS Publishing Data	PID Types	Metadata	Brokering	Big Data Analytics	Organizational Members business meeting	
15:00 - 16:00	Practical Policy	RDA/WDS Publishing Data	PID Types	Metadata	Brokering	Big Data Analytics	Long Tail of Research Data IG	
16:00 - 16:15	Coffee break	(

16:15 - 18:15	Breakout Sess	sion IV						
Room	Salon A	Salon B	Salon C	Salon D	Salon E	Salon F	Salon G	Salon H
16:15 - 17:15	Community Capability Model	RDA/WDS Publishing Data Subgroup: Services	Data Type Registries	PID Types	Preservation e- Infrastructure	BoF: Chemical Safety	Long Tail of Research Data IG	
17:15 - 18:15	Community Capability Model	RDA/WDS Publishing Data Subgroup: Workflow	Data Type Registries	PID Types	Preservation e- Infrastructure	BoF: Education and Training in Data Intensive Science		
19:00 - 21:30	Conference Dir	nner						



Wednesday September 18th | National Academy of Sciences Building

08:00 - 09:00	Continental Breakfast
09:00 - 10:30	Plenary V - Initial Meeting Results and Working Group Reports Chair: Carlos Morais-Pires, European Commission
10:30 - 11:00	Coffee Break

11:00 - 12:00	Plenary VI: Community discussion Facilitator: Mark Parsons, RDA Secretariat
12:00 - 12:30	Closing and Announcement of Plenary 3 Andrew Treloar, Director of Technology, Australian National Data Service
12:30 - 14:00	Lunch on your own and travel to Hotel

14:00 - 17:00	Working G	roups, Inter	est Groups	,Task Forces, ar	nd BoFs at the	Washin	gton Marriott
Room	Salon A	Salon B	Salon C	Salon D	Salon E	Salon F	NAS
14:00 - 17:00	Libraries and Research Data	Technical Advisory Board	Working Group Chairs meeting	Digital Practices in History and Ethnography	Cloud Computing and Data Analysis Training for the Developing World		Data Citation Harmonization (@NAS)
17:00	Closing						

PLENARY LOCATION:

National Academy of Sciences Building 2101 Constitution Avenue, NW Washington, DC 20418 Meeting Room: Auditorium (1st floor) Directions: <u>http://www.nationalacademies.org/about/contact/na 069684.html</u> Metro Station: Foggy Bottom (Blue Line) - approximately 5 blocks

Note: You will be required to present a government-issued PHOTO ID (or passport) to security. Please use the Constitution Avenue entrance which is at 2101 Constitution Avenue.

BREAKOUT SESSION LOCATION :

Washington Marriott 1221 22nd Street NW Washington, DC 20037 Directions: <u>http://www.marriott.com/hotels/maps/travel/waswe-washington-marriott/</u> Metro Stations: Dupont Circle (Red Line) - 0.3 Miles, Foggy Bottom (Blue Line) - 0.4 Miles



Program Details Plenary I Monday September 16th | National Academy of Sciences Building

09:00-09:35 Plenary Ia: Welcome and Opening Keynote

Chair: Farnam Jahanian, NSF Assistant Director for the Computer and Information Science and Engineering (CISE) Directorate

09:00-09:05 Welcome - Fran Berman, Co-Chair RDA, Chair RDA/US



Francine Berman is the Hamilton Distinguished Professor in Computer Science at Rensselaer Polytechnic Institute. She is a Fellow of the ACM and the 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." Prior to joining Rensselaer, Berman was High Performance Computing Endowed Chair at UC San Diego. From 2001 to 2009, Berman served as

Director of the San Diego Supercomputer Center. From 2007-2010, she served as co-Chair of the US-UK Blue Ribbon Task Force for Sustainable Digital Preservation and Access. From 2009 to 2012, Berman served as Vice President for Research at Rensselaer, stepping down in 2012 to lead U.S. participation in the Research Data Alliance. Berman has been recognized by the Library of Congress as a "Digital Preservation Pioneer", as one of the top women in technology by BusinessWeek and Newsweek, and as one of the top technologists by IEEE Spectrum.

09:05-09:15 Introduction and Opening of Plenary - Farnam Jahanian, NSF Assistant Director for the Computer and Information Science and Engineering (CISE) Directorate



Farnam Jahanian leads the National Science Foundation Directorate for Computer and Information Science and Engineering (CISE). He guides CISE, with a budget of over \$850 million, in its mission to uphold the nation's leadership in scientific discovery and engineering innovation through its support of fundamental research in computer and information science and engineering and of transformative advances in cyberinfrastructure. Dr. Jahanian is on leave from the University of Michigan, where he holds the

Edward S. Davidson Collegiate Professorship and served as Chair for Computer Science and Engineering from 2007 – 2011 and as Director of the Software Systems Laboratory from 1997 – 2000. His research on Internet infrastructure security formed the basis for the Internet security company Arbor Networks, co-founded in 2001, where he served as Chairman until its acquisition in 2010. Dr. Jahanian holds a master's degree and a Ph.D. in Computer Science from the University of Texas at Austin. He is a Fellow of the Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), and the American Association for the Advancement of Science (AAAS).



09:15-09:35 Tom Kalil, Deputy Director for Technology and Innovation for the White House Office of Science and Technology Policy and Senior Advisor for Science, Technology and Innovation for the National Economic Council.



Tom Kalil is the Deputy Director for Technology and Innovation for the White House Office of Science and Technology Policy and Senior Advisor for Science, Technology and Innovation for the National Economic Council. In this role, Tom serves as a senior White House staffer charged with coordinated the government's technology and innovation agenda. Prior to serving in the Obama Administration, Tom was Special Assistant to the Chancellor for Science and Technology at the University of California,

Berkeley. In 2007 and 2008, Tom was Chair of the Global Health Working Group for the Clinton Global Initiative. Previously, Tom served for 8 years in the Clinton White House, ultimately as the Deputy Assistant to the President for Technology and Economic Policy, and the Deputy Director of the National Economic Council. Prior to joining the White House, Tom was a trade specialist at Dewey Ballantine. He also served as the principal staffer to the Chair of the SIA Technology Committee, Gordon Moore. Tom received a B.A. in Political Science and International Economics from the University of Wisconsin at Madison, and completed graduate work at Tufts University's Fletcher School of Law and Diplomacy.

09:40-10:30 Plenary Ib: Benefits and Possibilities of Open Data Sharing

Chair: Sayeed Choudhury, Associate Dean for Research Data Management, The Johns Hopkins University



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 also the Director of Operations for the Institute of Data Intensive Engineering and Science (IDIES) based at Johns Hopkins. He is a Senior Presidential Fellow with the Council on Library and Information Resources and a member of the ICPSR Council, DuraSpace Board, the Digital Library Federation advisory committee and Library of Congress' National Digital

Stewardship Alliance Coordinating Committee. He has been a Lecturer in the Department of Computer Science at Johns Hopkins and a Research Fellow at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign.

09:40-10:05 Open and Machine Readable: Living In The New Default World -John Wilbanks, Chief Commons Officer, Sage Bionetworks



John Wilbanks is the Chief Commons Officer at Sage Bionetworks. He has worked at Harvard's Berkman Center for Internet & Society, the World Wide Web Consortium, the US House of Representatives, and Creative Commons. John is a past affiliate of MIT's Project on Mathematics and Computation and also started a bioinformatics company called Incellico, which is now part of Selventa. He sits on the Advisory Boards for Boundless Learning, Genomic Arts, Curious, GenoSpace, and Genomera. He also serves

as a senior advisor to the National Coordination Office, and as a Senior Fellow at both the



Ewing Marion Kauffman Foundation and FasterCures. John holds a degree in Philosophy from Tulane and studied modern letters at the Sorbonne.

10:05-10:30 Fuelling and Transforming Evidential Cultures of Research - Carole Palmer, Professor, Graduate School of Library and Information Science, University of Illinois



Carole Palmer is Director of the Center for Informatics Research in Science and Scholarship (CIRSS) and Professor in the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. Her research investigates the information processes and cultures of scientific and scholarly communities to inform development of cross-disciplinary

digital research collections and the curation of research data for long-term use. She currently leads teams studying technical and policy requirements for the curation of open access research data in the sciences, and strategies for national and international aggregation and interoperability of digital cultural heritage collections. Current projects include Site-Based Data Curation for Geobiology at Yellowstone National Park and the Data Curation Education in Research Centers initiative, in partnership with the National Center for Atmospheric Research. She was co-PI on the NSF Data Conservancy effort from 2009-2012 and PI on the IMLS Digital Collections and Content project from 2007-2013. She has been leading education and outreach initiatives in data curation since 2005, including a master's specialization, a series of summer institutes for practicing information professionals, and a number of sciences study committee on Future Career Opportunities and Educational Requirements for Digital Curation and previously served on the study committee on Building Cyberinfrastructure for Combustion Research.

Plenary II

Monday September 16th | National Academy of Sciences Building

<u>RDA Business including news from Council and the first Organisational</u> <u>Assembly</u>

This is the general business meeting for RDA. We will introduce new Council members and Working Groups. We will update the membership on key Working Group activities, evolving organizational structures, and other RDA business. This session will also include the first Organizational Assembly, allowing the membership to better understand the role or private organizations as members in RDA.

Chair: Herman Stehouwer, RDA Secretariat & Max Planck Institute for Psycholinguistics, Coordinator for RDA/Europe & RDA Secretariat

Dr.ir. Herman Stehouwer has a background in computer science and computational linguistics.



At the Max Planck Institute for Psycholinguistics he has been responsible for the technical aspects of the CLARIN search infrastructure for CLARIN-D as well as having the responsibility for a number of TLA software projects (such as the VLO). In the last year he has been involved with setting up the RDA. Now that the RDA is moving to a steady state he is the Europe



representative in the RDA Secretariat next to his regular RDA and RDA/Europe work.

11:00 - 11:30	Council and Secretariat updates: introduce new members, bids for Plenary 4, other business
11:30 - 12:00	Updates from currently active and recognized WGs
12:00 - 12:45	Organizational Assembly - Member organizations of RDA

Plenary III

Monday September 16th | National Academy of Sciences Building

13:45 - 15:30 Plenary III - Building Global Partnerships. The Perspective of **Funders and Affiliate Organizations**

Chair: Chris Greer, NIST Associate Director for Program Implementation

13:45 - 14:00 - Welcome and Introductions - Chris Greer, NIST Associate Director for Program Implementation



Chris Greer (Director, Smart Grid and Cyber-Physical Systems Program Office) brings with him extensive leadership experience from academia, the NSF, the White House, and NIST. Chris previously served as Associate Director for Programs in the NIST Information Technology Laboratory (ITL) and Acting Senior Advisor for Cloud Computing. In these positions, he was responsible for strategic planning for information technology initiatives across ITL, including its data and cloud computing efforts. Prior

to joining NIST, Chris served as Assistant Director for Information Technology R&D in the White House Office of Science and Technology Policy (OSTP) and Cybersecurity Liaison to the National Security Staff. His responsibilities there included networking and information technology research and development, cybersecurity, and digital scientific data access. He has also served as Director of the National Coordination Office for the Federal Networking and Information Technology Research and Development (NITRD) Program. This program coordinates IT R&D investments across the Federal government, including the cyber-physical systems research portfolio. Prior to undertaking government service, Chris was a member of the tenured faculty at the University of California, Irvine, where his research focused on gene expression. He holds a Ph.D. degree in biochemistry from the University of California, Berkeley and a BA degree from the University of California, San Diego.

14:00 - 14:45 RDA COLLOQUIUM PANEL DISCUSSION

RDA Colloquium Panel Discussion - Mark A. Suskin, Deputy Director, Office of **Cyberinfrastructure, U.S. National Science Foundation**



Mark A. Suskin is currently Deputy Director of NSF's Division of Advanced Cyberinfrastructure which supports research, development, acquisition, and operation of advanced cyberinfrastructure to enable advances in 21st century science and engineering research and education. Dr. Suskin received a Ph.D. in theoretical atomic physics from the Johns Hopkins University in 1987 and has worked as a researcher and manager of the Atomic Transition Probability Data Center at the



National Institute of Standards and Technology; as a subcontractor to major aerospace corporations in the optimization of low- and high-thrust trajectories for exploratory interplanetary spacecraft as part of NASA's Space Exploration Initiative; and as an analyst in the International Security and Space Program of the Congressional Office of Technology Assessment (OTA).

RDA Colloquium Panel Discussion - Kostas Glinos, European Commission, Head e-Infrastructure



Kostas Glinos has been with the European Commission since 1992. He leads the e-Infrastructure unit of the Directorate General for Communications Networks, Content and Technology since 1 January 2009. From 2003 to 2008 he was Head of the Embedded Systems and Control unit and interim Executive Director of the ARTEMIS Joint Undertaking. Previously he was deputy head of Future and Emerging Technologies. Before joining the Commission Kostas worked with multinational

companies and research institutes in the U.S., Greece and Belgium. He holds a diploma in Chemical Engineering from the University of Thessaloniki, a PhD from the University of Massachusetts and a MBA in investment management from Drexel University.

RDA Colloquium Panel Discussion - Clare McLaughlin, Counsellor - Education and Science, Australian Embassy to Belgium and Luxembourg and Mission to the European Union and NATO in Brussels



Clare's role is to develop and implement strategies to strengthen Australia's engagement with the education, science, research and innovation systems in Europe. Immediately prior, Clare was General Manager, Research Funding and Infrastructure Branch in the Science and Research Division of the Australian Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education. The Branch administered research and research training block grant funding; and

managed research infrastructure projects funded from the National Collaborative Research Infrastructure Strategy (NCRIS) and the Education Investment Fund (EIF), including the Super Science Initiative The Branch also supported Australian involvement in the OECD Global Science Forum and international collaboration on research infrastructure, particularly with the European Union. Clare managed the Science Policy team in the development of the 2011 Research Infrastructure Roadmap and the Strategic Framework for Research Infrastructure Investment. She was Manager of the eResearch team from 2007-10, implementing the eResearch elements of NCRIS and the Super Science Initiative. In a long career in the Australian Public Service, she has worked on programs, projects and policy related to copyright, data, access to publicly funded research, and advanced ICT infrastructure including high performance computing, network infrastructure and collaboration tools



14:45 - 15:30 AFFILIATE ORGANIZATIONS PANEL DISCUSSION



Affiliate organizations Panel Discussion – Sara Graves, CODATA

Dr. Sara James Graves is the Director of the Information Technology and Systems Center, the University of Alabama System Board of Trustees University Professor and Professor of Computer Science at the University of Alabama in Huntsville.

Her current service includes membership on the National Academy of Sciences Board on Research Data and Information, the Executive Committee of CODATA, the International Council for Science: Committee on Data for Science and Technology; the Science Advisory Board for the Oak Ridge Climate Change Science Institute: the Board of Trustees for the Southeastern Universities Research Association (SURA), and is a Founding Member, National Oceanic and Atmospheric Administration Science Advisory Board Data Archiving and Access Requirements Working Group.

Dr. Graves directs research and development in large-scale distributed information systems, data mining and knowledge discovery, high performance computing and networking, semantic and collaborative technologies, analysis and visualization of biological and geosciences data. She has been the Principal Investigator on many research projects with DOD, DOE, NASA, NOAA, NSF, and other entities. Her degrees are in Computer Science and Mathematical Sciences, and she has served as chair or member of over 100 Ph.D. and M.S. committees.

Affiliate organizations Panel Discussion - Curt Tilmes, ESIP



Curt Tilmes is the Technical Lead for the Global Change Information System, detailed from NASA's Goddard Space Flight Center to the U.S. Global Change Research Program. At GSFC he worked on science data processing systems for the Earth Observing System Data and Information System, including that of the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Ozone Monitoring Instrument (OMI).Curt is involved with NASA's Earth Science Data Systems Working

Groups and the Federation of Earth Science Information Partners where he chairs the Data Stewardship Committee and serves on the Executive Committee. His research has been focused on provenance and reproducibility of scientific data and he currently serves as NASA's representative to the W3C Provenance Working Group.Dr. Tilmes received B.S. degrees in Computer Engineering and Computer Science from Virginia Tech, M.S. degrees in Computer Science and System Engineering from Johns Hopkins University, and his Ph.D. in Computer Science from the University of Maryland, Baltimore County.

Affiliate organizations Panel Discussion - Phil Archer, W3C



Phil Archer joined the W3C staff after many years representing one of its member organisations. Chair of the first Incubator Group, a lightweight mechanism for developing new ideas for the Web, he went on to chair the POWDER Working Group whose standards form part of the Semantic Web technology suite. Alongside this work he was an original member of the Mobile Web Best Practices Working Group and was editor or



acknowledged contributor to 6 of that group's publications. Although he continues to be involved with the mobile Web, Phil's primary focus since 2010 has been on eGovernment and open data. In particular he works on improving interoperability between data sets whilst advocating the principles and return on investment available through the open data movement. As well as work at the W3C, his career has encompassed broadcasting, teaching, linked data publishing, copy writing, and, perhaps incongruously, countryside conservation. The common thread throughout has been a knack for communication, particularly communicating complex technical ideas to a more general audience.

Plenary IV

Monday September 16th | National Academy of Sciences Building

<u>Plenary IV - Technical Advisory Board and general technical collaboration</u>

The Technical Advisory Board provides the technical roadmap for RDA. In this session, you will learn what the TAB is all about and the critical role it plays in helping RDA Working Groups conduct their activities. You'll meet the interim members of the

TAB and hear from the candidates running for election to the Board.



Chair: Larry Lannom, VP, Corporation for National Research Initiatives (CNRI)

Larry Lannom is Director of Information Services and Vice President at the Corporation for National Research Initiatives (CNRI), where he works with organizations in both the public and private sectors to develop experimental and pilot applications of advanced networking and

information management technologies. His current work is focused on CNRI's Digital Object Architecture, which is based on the concept of the digital object, a uniform approach to representing digital information across computing and application environments, both now and into the future. He is responsible for the development and ongoing evolution of a series of infrastructure components needed to implement the architecture. This includes a high performance resolution system, known as the Handle System that maps identifiers into current state information about the digital objects being identified, repositories for storing digital objects and from which they may be accessed, and metadata registries for managing collections of digital objects across one or more repositories. These efforts, primarily funded through research grants, have borne considerable fruit in the real world of operational information management, from the use of the Handle System in the electronic publishing and digital library world to the use of digital object registries in the entertainment industry.



Breakout Sessions I

Tuesday September 17th | Washington Marriott

• Agricultural Data

• Big Data Analytics

This is the second BoF and a follow on from the first BoF that has been held at the 1st RDA Plenary in Gothenburg, Sweden.

The BoF will review the developments since the first BoF both in attendees' organisations and in international bodies dealing with the standards and policy development in the area of Big Data and Data Intensive Science and technologies. The BoF will provide an overview of the Big Data Architecture Framework that incorporates the main components of the Big Data Ecosystem and discuss how it can be used for defining the stakeholder model, skills profile and curriculum scope.

The BoF will discuss the prospective Interest Group charter possible cooperation framework for inter-organisational cooperation in developing educational and training programs on Data Intensive Science and e-Infrastructure.

• BoF: Chemical Safety

The aim of establishing a working group for chemical safety is to enhance data sharing, access, and analysis capabilities with a focus on existing toxicogenomics data. Alternatives to animal testing and adoption of rapid assay methods have expanded in the last decade driven by the need to identify hazards and prioritize testing for an increasing number of environmental chemicals as described in the Toxicity Testing in the 21st Century report by NRC in 2007 and the requirements of the EU REACH regulations. It is this context that genomic technologies are being applied to study the impact that chemical exposures may have on public health, referred to here as toxicogenomics. Rapid deployment of these technologies has created an influx of toxicogenomic research data and related databases. Meaningful application of these data to understand the potential chemical effects on human toxicity pathways is a particular challenge facing the public health community. Extrapolation of the existing data from in vitro results to an in vivo system, to untested chemical space, and expanding the knowledge of toxicity pathways is being undertaken by a broad based research community consisting of statisticians, data scientists, toxicologists, genomics experts, and others. The chemical safety working group proposes to contribute to the larger chemical safety effort by beginning to address the obstacles related to sharing of toxicogenomic data held within existing databases. In the Toxicogenomics Interoperability interest group, we conducted an e-questionnaire on data uses and limitations within the community. The goal of this e-questionnaire was to better understand how the data are currently being used in the field, so that a roadmap towards database interoperability would be grounded in the current work and needs of ongoing toxicogenomic and chemical safety research. In this session, we will discuss current and future steps to integrate the results of the e-questionnaire into a case statement for an RDA chemical safety working group.

• Education and Training in Data Intensive Science

• Brokering

• Community Capability Model

It is often difficult for researchers, organisations and funders to assess their capability for pursuing data-intensive research now, and in the future. To address this, UKOLN Informatics and Microsoft Research Connections are working in partnership to develop a Community Capability Model Framework for Data-Intensive Research, building upon the principles



described in The Fourth Paradigm. The work has direct relevance to the aims of the Research Data Alliance. The CCM Interest Group will lead on progressing the application of the Model across a range of disciplines to make it usable and provide a picture of how different communities are tackling data-intensive research. This first meeting of the IG has three main aims: 1) to update on progress with the Community Capability Model Framework (CCMF) and application of the tools, 2) to review the prototype CCMF Profile Template and 3) to plan the first disciplinary workshop from the nominated domains (Astronomy, Environment, Social Sciences). The ultimate aim is to gather evidence of the value of the CCMF across a wide range of disciplines and domains to inform the production of the CCMF Value & Impact Case Study for the RDA. Project website: http://communitymodel.sharepoint.com/

Meeting Agenda

16.15	Introduction and RDA Context	Liz Lyon, UKOLN Informatics, University of Bath
16.20	CCMF – Summary of progress Introducing the Community Capability	Manjula Patel UKOLN Informatics
	Model Framework & Tools	
16.35	Applying the CCMF – some lessons learnt	Kenji Takeda, Microsoft Research Connections
16.50	Group work – Reviewing the Profile Template Tool and Feedback (Look and feel? Questions? Usability?)	Scott Brandt, Purdue Univ
17.30	Taking the CCM out to domains ("lite") Which ones? Who will lead? How to feedback to the IG? Discussion session	Liz Lyon, UKOLN Informatics
18.00	Planning the 1 st "deep dive" disciplinary workshop: Embracing the Environment?	Bill Michener, University of New Mexico
18.15	Closing remarks and IG DONM	Liz Lyon, Kenji Takeda

• Data Citation

This WG aims at providing solutions for enabling the citation of data in settings where huge volumes of data are involved, where data is changing, i.e. being amended, updateed/corrected, and where thus traditional approaches of citing an entire static volume of data is not sufficient.

We will build upon the results of the first WG meeting in Gothenburg where we discussed different types of data and dynamics. During the break-out meeting we aim at identifying scenarios and resulting requirements for machine-actionable data citation, discuss the advantages and disadvantages of approaches identified so far, and see whether any concrete pilots emerge. We will also discuss the cross-links to other WGs to identify issues related to PID systems, metadata to be associated with such a data citation, etc. These requirements will also provide input to the Data Citation Harmonization meeting on Wed afternoon as part f an activity co-ordinated by FORCE 11, bringing together key projects and groups working on different aspects of data citation.



• Data Foundations & Terminology

The Data Foundation and Terminology Working Group describes a basic abstract data organization model which can be used to derive a reference data terminology that can be used across communities and stakeholders to better synchronize conceptualization, enable better understanding within and between communities and stimulate tool building.

The following is the working agenda for the 2 DFT Breakout session being held Tuesday 17th September in the Washington Marriott (1221 22nd Street NW)

Washington, DC) - Salon A. There will be a teleconferencing link provided for remote attendees, from our RDA WG page for those registered as part of the WG.

DFT Session 1 (9-11) Co-Chairs Peter Wittenburg and Gary Berg-Cross

Overview of the DFT Breakout Session- Goals and Plan	Gary/Peter
Formal State of the WG and Case Statement	Gary
Brief overview of Draft Docs and their schedule	Peter
Review Doc1 -Model Overview	Peter
Review Doc2 – Data Model Analysis	Peter
Review Doc3 - Analysis of Workflow	Gary
Review Doc4 - Synthesis	Peter
General Discussion (including remote participants)	All
Any changes to 2nd Session	All
Session - Part 2 (11.15-12.15)	

DFT Session - Part 2 (11:15-12:15)

Draft Method on Vocabulary Development Process (Comments on Vocabulary Development by Joe Hourcle (NASA)	Gary
Discussion of Proposed Process	
Proposal on how to start Doc5 seeded with 3-4 terms	Peter
Persistent Identifier	
Digital Object - Data Object	
Collection - Data Set - Aggregation	
Repository (and related Policies?)	
Discussion of the above and other Key Vocabulary Ideas	
Soliciting ideas for candidate vocabulary items	
Liaison with other WGs and their topics including follow on work	k.
Formalizing action items & getting commitments for work, -e.g. n	reviews of the draft
documents and vocabulary and also discuss WG	
Plan for follow up virtual meeting.	

• Data Type Registries

The two-hour Data Type Registries (DTR) breakout session will begin with a brief introduction for newcomers to the group. We will then move then to the consideration of use cases. A initial set have been published in the DTR section of the RDA web site. Additional use cases will be solicited. A proposed data model will be presented for discussion and an initial registry implementation will be described. Finally, potential relationship with other RDA WGs and other efforts will be discussed.

• Economic Models & Infrastructures for Federated Materials Data Management

The proposed Materials Data Interest Group will hold a meeting Tuesday, September 17 11:15-12:15 to discuss its proposed <u>charter</u>, focusing on: (1) (A) federated data management infrastructures and (B) economic models of sustainability currently practiced in



the materials community; and 2) ideas for a potential collaborative undertaking with CODATA TG on Materials Data that would be of value to the materials community.

• Engagement

In its breakout session, the RDA-Engage interest group will build upon the successful session and a very insightful discussion during the first RDA plenary in March 2013. One or two guest speakers from other organizations, such as the ESIP Federation, will share their experiences with community engagement, followed by the brainstorming session on the next steps for RDA-Engage in creating advantages and capacities for data exchange. ---

• Global registry of trusted data repositories and services

See RDA/WDS Certification.

Interoperability	
Welcoming remarks and introductions	Paul Uhlir NAS, US
Review of white paper on legal interoperability of data Paul U and annotated bibliography - Discussion	Jhlir
Interest Group website - Discussion	Natalia Manola U. Athens, GR
Workshop plans - Discussion	Paul Uhlir
Selection and assessment criteria for case studies - Discussion	Enrique Alonso Gov. of Spain, Bob Chen Columbia U., US
Break	,
Instructional materials - Discussion	Enrique Alonso
Other issues for consideration by the group Summary of actions Closing remarks <i>Meeting adjourns</i>	Paul Uhlir Anita Eisenstadt Paul Uhlir
	Interoperability Welcoming remarks and introductions Review of white paper on legal interoperability of data Paul U and annotated bibliography - Discussion Interest Group website - Discussion Workshop plans - Discussion Selection and assessment criteria for case studies - Discussion Break Instructional materials - Discussion Other issues for consideration by the group Summary of actions Closing remarks Meeting adjourns

• Long tail of research data IG

Universities and research institutions are becoming increasingly interested in collecting and providing access to datasets produced at their institution that do not fall within the scope of other discipline-based, or government repositories. The aim of this Interest Group is to develop a set of good practices for managing research data archived in the university context. The scope of the topic will be limited to the data generated in universities and research institutions and the role of institutional repositories and libraries as agents of the institutional data management.

Attendees are asked in advance to identify a dataset produced at an institutions, for which the researcher has archived the data locally or is looking for a place to archive the dataset. Please describe the dataset according to the following elements: domain (research area), format, size, doi (yes or no), any access restrictions (i.e. privacy OR readability). These examples will contribute to the development of a number of dataset profiles that will help us better understand the nature of research data that constitutes the so called "long tail". **Agenda**



1. Welcome: How did the interest group (IG) come about, what are the broad aims of the group.

2. What is the long tail? Presentation by Wolfram Horstmann and general discussion

3. Introductions and each person will describe their dataset (profiles will be developed based on these examples)

4. Review and discuss draft objectives for the IG:

- Define the scope of datasets that will be addressed in this project.
- Develop a number of use cases, based on a range of disciplinary practices and other approaches.
- Map the current repository landscape: to categorize the types of repositories that do exist along a number of axes such as domain, open/closed, data formats, etc.; to have a understand the capacity of existing repositories to collect small datasets
- Identify gaps
- Review data federation approaches that provide mechanisms for supporting discovery across the myriad of existing repositories (distinguish domain specific attributes from generalizable practices)
- Identify and publish good practices.
- Identify skills and competencies for those managing research data in university repositories.

5. Next steps for the IG

• Marine Data

The break-out session for the RDA Marine Data Interest Group will take place on Tuesday, September 17 at 12:15 to 13:25 in Salon C. The meeting will discuss the draft charter and in particular the activities planned for the group with regards to assessing the current state of the art for standards, formats, best practice etc. in use for marine data management including the documentation, assessment and dissemination of these standards etc. to the wider user community.

The BoF session will also discuss the strategy for the expansion of the IG membership and also how the RDA MarineData group will engage with other relevant RDA working and interest groups.

• Metadata

14:00-14:10 - Welcome

14:10-14:20 - Brief overview/history, MASDIG progress

Useful links:

Disciplinary Metadata, see: DCC <u>http://www.dcc.ac.uk/resources/metadata-standards</u> CERIF metadata model: <u>http://www.eurocris.org/Index.php?page=CERIFreleases&t=1</u> 14:20-14:30 - Vision/ideas for MIG

14:30-15:00 - CAMP activities + outcomes; MASDIR – next steps

15:00-15:15 - Setting priorities, action items MASDIR, wrap-up of MASDIR discussion

15:15-15:45 - MIG overview and discussion

15:45-16:00 - Summarize action items; closing remarks

Organizational Members business meeting

• PID types

The PID Information Types WG session will focus on extracting an initial set of types from the use cases that have been gathered from various communities so far. The use case documents need to be examined in more detail and recurring themes must be identified that can be used to define an initial set of types during the session. Furthermore, in the session we will work on



the high-level architecture of the API and sort out the roles of various stakeholders that are important for practical adoption. Activities during the session will be creative, technically detailed and focused on solving practical problems.

• Practical Policy

The Practical Policy sessions on Tuesday 12:15-13:15 and 14:00-16:00 will cover an introduction to policy-based data management, participation in the group, and discussions of mapping English language rules to policies, organization of testbeds, review of replication policies, categorization of policies, and next steps.

https://www.rd-alliance.org/working-groups/practical-policy-wg.html

• Preservation e-Infrastructure

"The Preservation e-Infrastructures session will review the current status of relevant ongoing work together with the documents and plans which have been collected. Short presentations will be made by participants, followed by discussions. In the final period of the session a number of core commonalities in concepts and approaches should be identified, on which to develop the subsequent work of the group.

A webcon should allow those not able to attend in person to contribute to the discussions."

Research Data Provenance

• RDA for Newcomers

This session provides an overview of the RDA and how members can get involved. Representatives from the Secretariat, Council, and Technical Advisory Board will summarize the RDA organizational structure, governance, and Working and Interest groups. They explain how to join and start RDA groups and how to use RDA tools. There will be ample time for questions and discussion.

• RDA/WDS Certification (with Global registry of trusted data repositories and services)

The Interest Group on Certification of Digital Repositories organises a combined session in which the planned activities of the WDS-DSA Working Group will be presented and the set-up of a global registry of trusted data repositories and services will be discussed.

The Working Group aims to explore and develop a DSA-WDS partnership with the objectives of realizing efficiencies, stimulating more certifications, and having a greater impact on the community. Next to that, the work performed in common by DSA and WDS in this WG will be a starting point for next steps including other organisations, which deal with repository certification.

The BOF session on a global registry of trusted data repositories and services will explore and discuss the needs and requirements for such a registry as well as identify possible stakeholders and key players in this field. The intention is conceive a first draft for a WDS/RDA WG.

The session will close with a general, open discussion on the broader IG goals, in which new ideas, other initiatives, or (inter-)national developments can be presented.

• RDA/WDS Publishing Data

14:00-14:30 RDA/WDS Publishing Data Interest Group

Co-chairs: Michael Diepenbroek, Jonathan Tedds

The Publishing Data Interest Group l brings together all stakeholders involved in publishing research data including researchers, discipline specific and institutional data repositories, academic publishers, funders and service providers. Every effort will be made to get a good



representation from related international programmes, their working groups and other private or institutional activities involved in this area. We will build on existing resources, reports and other shared experiences from the different stakeholders and will nurture more specific and targeted working groups addressing practical aspects in publishing research data.

As such, the Publishing Data Interest Group can be regarded as a broad and inclusive forum for interested individuals to contribute to and test, validate and promote the findings of the Working Groups. In particular we plan to address the implementation of workflows for publishing data and therefore help establish appropriate supporting infrastructure including bibliometrics, costings and other relevant services.

This session will update participants on the work to date through the joint RDA-WDS Working Groups proposed and invite further expressions of interest in these or other areas of common interest.

14:30-15:00 RDA/WDS Publishing Data Working Group Costs and Cost recovery

Co-Chairs: Ingrid Dillo

This proposed RDA-WDS Joint Working Group has its focus on cost recovery models. How do you fund a long-term infrastructure when research funders are unwilling to take on recurring cost? This means we need cost recovery models for digital repositories: what has been done, and what has been shown to work? We also need to look at data publication and the relationship between publishers (traditional/non-traditional), data centres, libraries, and computer centres.

This approach of course requires a good understanding of the current work on cost models, but the focus of this group will be on understanding the various cost recovery models / business models that are available. We need to develop an understanding of the various business models that help provide this infrastructure. This means we need to answer questions like: What are the costs? What models exist for long-term sustainability? Under what conditions are different business models most appropriate/necessary?

The activities of the group will focus on the following components:

- Surveying current cost models (building on work of APARSEN and 4C);
- Filling in current knowledge about actual costs (taking a range of WDS and DSA members);
- Survey policies and implications for funding (what are the key policy developments and their implications);
- Researching and analysing various approaches to cost recovery/business models (survey data centres and data publication initiatives; how are costs recovered (phenomenon of DBs charging for access; and how are roles, responsibilities and costs divided within the data infrastructure, e.g. front office/back office model));
- Presenting conclusions about the potential appropriateness of different cost recovery models to different situations (what specific aspects of the data, the institution, or a combination of both work for which recovery models?) and about the potential of data publication initiatives fitting into a cost recovery strategy.



In this session the draft case statement of the group will be presented and the participants will be asked to give their comments and to provide clear examples of use cases and relevant cost recovery models.

15:00-15:30 RDA/WDS Publishing Data Working Group Bibliometrics

Co-Chairs: Kerstin Lehnert

Bibliometrics are an essential ingredient for the broad adoption of data publication as part of scholarly communication. Bibliometrics provide incentives for investigators to publish their data, and will thus catalyze the necessary culture change toward comprehensive data sharing. The proposed RDA/WDS Publishing Data WG Bibliometrics aims to define the concept of data metrics as a way to quantitatively track the citation of published data in research and assess their use and impact. The WG will explore feasible options for indicators and their implementation across the broad range of stakeholders in digital data infrastructures, and develop a plan for adoption. Activities of the WG will be directed toward the following goals:

- Summarize current and emerging data citation practices of all stakeholders (journals, data centers, funders, societies), taking advantage of recent studies related to this topic;
- Understand and articulate necessary organizational and cultural changes in the scholarly publishing system needed to foster proper attribution of data sets;
- Evaluate and report on possible models how data citations can be successfully tracked and measured, based on existing and emerging approaches;
- Identify and report on possible barriers for the implementation and adoption of data citation and data metrics solutions.

In this session we will present the draft case statement of the group, gather feedback as well as information on existing practices, use cases, and ongoing or emerging studies, and solicit ideas for further engagement.

15:30-16:00 RDA/WDS Workflows for Publishing Data Working Group

Chair: Jonathan Tedds and Merce Crosas

The objectives of this proposed RDA-WDS Joint Working Group are to investigate the range of possible workflows for publishing data, in particular reporting on:

- Current workflows for archiving and publishing data
- The role of Quality Assurance / Quality Control and peer-review in the publication process
- The role of science publishers and journals in the data publication process
- Barriers to implementation

In particular we will build on the work carried out to date in various international settings, including the UK Jisc funded PREPARDE project which provided reports on journal workflows from author submission of datasets and papers, through review to publication. This project also investigated data repository workflows from ingestion of data, through data centre



technical review, to DOI assignment to dataset. While the focus was on mature examples in the Earth Sciences the reports also considered alternative paradigms which will be extended through the work of this group.

The group aims to identify workflow models for data publication across an appropriate range of use cases in different disciplines. We will provide the international research community with clear examples that may be adapted for use, in each case identifying the varying stakeholders and their different roles and responsibilities as well as the likely associated resource and cost implications, where possible.

This session will update participants on the work to date through the proposed joint RDA-WDS Workflows Working Group and invite further expressions of interest, e.g. in providing clear examples of use cases and relevant workflows.

16:00-16:15 Coffee break

16:15-16:45 RDA/WDS Publishing Services Working Group

Chair: Hylke Koers

The objective of this proposed RDA-WDS Joint Working Group is to address processes, workflows, and solutions that currently exist between individual parties within the data publication landscape, and investigate how these can be lifted to one-for-all services that increase interoperability, decrease systemic inefficiencies, and power new tools and functionalities to the benefit of researchers. Such processes may lie at different moments in the data publication workflow, including the submission, editorial, review, and publication process.

As a point of focus, the proposed working group takes particular interest in the problem of limited interoperability between data repositories, scholarly journal publication platforms, and tools for bibliometric analysis - in particular the absence of a common framework for cross-referencing data sets and published articles on different levels of granularity. This poses barriers and induces inefficiencies for the interlinking and contextualization of journal articles and data sets.

Key deliverables include:

- Inventory of interlinking, cross-referencing, and other tools and processes relevant to data publication currently in place. An analysis of pro's and con's, with an emphasis on scalability and do-ability.
- Gap analysis, including an analysis of needs & use cases for key stakeholders (data repositories, journal publishers, providers of bibliographic services, funding bodies, research institutions, the researcher).
- Recommendations for a one-to-all cross-resolving service that benefits the stakeholders in data publishing. These recommendations will include technical aspects, organizational, governance, and costs.
- An operational service (beta-release) for cross-referencing data sets and articles.

16:45-17:15 Reports from other groups



17:15-18:15 General Discussion and Wrap up

Chair: Ingrid Dillo and Mustapha Mokrane

• RDA Colloquium

This session will be conducted in two segments with participation in both segments by invitation only. We invite government representatives of current and prospective RDA Colloquium government agencies to meet in the first session from 09:00 to 09:45 to discuss shared interests in data infrastructure and global cooperation. We invite the members of the RDA Council to joint these government representatives for the second session to be held from 09:55 to 11:00 to discuss interactions between the Council and the Colloquium.

• Toxicogenomics Interoperability

The objective of the Toxicogenomics Interoperability Interest Group is to define opportunities toward enhancing interoperability of existing toxicogenomics databases. Toxicogenomics is the application of genomic technology to study the impact that chemical exposures may have on public health. Historically chemical toxicity assessment has relied on time-consuming and costly animal exposure experiments. Toxicogenomics-based cellular models have been proposed and adopted as an alternative to animal testing, resulting in a rapid increase in these data, and associated databases. Meaningfully applying this data to understand the potential chemical effects on human toxicity pathways is a particular challenge facing the public health community. Application of the data requires that the data be accessible and interoperable for scientists to interrogate. Several large databases have been identified during the joint US – European Union "Workshop on Identifying Opportunities for Global Integration of Toxicogenomics Databases" held at the National Institute for Environmental Health Sciences in Durham, North Carolina, USA on June 26 and 27, 2013. During this event there was general agreement among database owners to engage in a collaborative process toward interoperability, which led to the creation of an RDA interest group. To guide the direction of any future database interoperability efforts, user stories have been solicited from existing stakeholder groups via an e-questionnaire focused on soliciting user stories and data limitations from the scientific community. In this session, the importance of toxicogenomic data within the context of public health, as well as the analysis of the collected user stories will be used to produce a number of use cases and other requirements of a roadmap toward interoperability will be presented and discussed.

Plenary V

Wednesday September 18th | National Academy of Sciences Building

Plenary V - Initial Meeting Results and Working Group Reports

Chair: Carlos Morais-Pires, European Commission



Carlos Morais Pires is scientific Officer, Excellence in Science DG/CONNECT and coordinator, Scientific Data e-Infrastructures at DGconnect European Commission. He is with the unit GÉANT & e-Infrastructure which manages the e-Infrastructures part of the Capacities programme of the 7th European framework programme for research and development.

He has a PhD degree on Telecom Engineering taken in Torino in 1996. His professional experience on Information and Communication Technologies includes teaching



Computer Networks and Multimedia at the Lisbon University and, for the past 10 years in the European Commission, managing research and development programmes on Information and Communication Technologies and ICT infrastructures for research.

Plenary VI

Wednesday September 18th | National Academy of Sciences Building

Plenary VI: Community Discussion

Facilitator: Mark Parsons, RDA Secretariat & Managing Director of the US Component of the Research Data Alliance and the Rensselaer Center for the Digital Society



Mark Parsons is the Managing Director of the US Component of the Research Data Alliance and the Rensselaer Center for the Digital Society. He focusses on stewarding research data and making them more accessible and useful across different ways of knowing. He has been leading major data stewardship efforts for more than 20 years, and received the American Geophysical Union Charles S. Falkenberg Award as an advocate of robust data stewardship as a vital component of Earth

system science and as an important profession in its own right. Prior to joining Rensselaer, Parsons was a Senior Associate Scientist and the Lead Project Manager at the National Snow and Ice Data Center (NSIDC). While at NSIDC, he defined and implemented their overall data management process and led the data management effort for the ICSU/WMO International Polar Year 2007-2008. He is currently active in several international committees while helping lead the Research Data Alliance in its goal of accelerating innovation through data exchange. His research interests include the role of scientific social interaction in the success, development, and extension of data sharing networks.

Closing and Announcement of Plenary 3

Andrew Treloar, Director of Technology, Australian National Data Service



Dr. Andrew Treloar is the Director of Technology for the Australian National Data Service (ANDS) (http://ands.org.au/), with particular responsibility for demonstrating the value of bringing together data from different disciplines to answer new questions, and international engagement. In 2008 he led the project to establish ANDS. Prior to that he was associated with a number of e-research projects as Director or Technical Architect: ARCHER (http://archer.edu.au/ - an e-Research support environment), DART (http://dart.edu.au - data acquisition and

analysis), and ARROW (http://arrow.edu.au/ - institutional repository software), as well as the development of an Information Management Strategy for Monash University. His research interests include data management, institutional repositories and scholarly communication. He never seems to be able to make enough time for practising his 'cello, or reading, but does



try to prioritise talking to his chickens and working in his vegetable garden and orchard. Further details at http://andrew.treloar.net/ or follow him on Twitter as @atreloar.

Andrew holds a Bachelor of Arts with first-class honours, majoring in Germanic Languages and Linguistics, a Graduate Diploma in computer science, a Master of Arts in English Literature and a Ph. D. with the thesis topic Hypermedia Online Publishing - The Transformation of the Scholarly Journal.

Breakout Sessions II

Wednesday September 18th | Washington Marriott

• Agricultural Data

• Cloud Computing and Data Analysis Training for the Developing World

Title: Proposal to form a working group on cloud computing provision for developing world science.

At this point cloud computing presents an opportunity for developing world science to sidestep their infrastructural difficulties. Through access to large amounts of computing and public data sets, developing world science could make cutting-edge contributions to Science in general and to the domains that are of importance to them. The key difficulties are the financial resources to accessing cloud computing platforms and the relevant teaching materials to get them started. The working group we propose will have three goals.

In the first instance financial models for cloud computing use will be developed so operational expenditure can be fully understood. Teaching materials for data science and cloud computing will be put together and finally a proposal for a fully funded sustainable cloud computing platform for developing world science will be developed.

• Data Citation Harmonization (@ NAS)

https://www.rd-alliance.org/filedepot/folder/150?fid=184

Wednesday, 18 September, Members' Room, National Academy of Sciences

13:15	Welcoming remarks and introductions	[?]	
13:30	Review of synthesis group process since]	uly Maryann Marton	e
13:45 Uhlir	Synthesis Principles and examples - Discussion	Dan Cohen and Paul	
15:15	Break		
15:30	Synthesis Principles and examples	Merce Crosas and Maryann Martone	
	- Discussion		
17:15	Adjourn		



Thursday, 19 September, Board Room, National Academy of Sciences

8:45	Review of Synthesis Principles and examples	Micah Altman and Joan Starr
	- Discussion	
10:00	Detailed dissemination plan	Sarah Callaghan
10:45	Break	
11:00	Detailed dissemination plan (continued)	Bonnie Carroll?
- Cover letter	S	
- Assignment	S	
	- Milestones	
11:30	Other issues and closing remarks	Bonnie Carroll?
11:50	Meeting ends	

• Digital Practices in History and Ethnography IG

RDA's Digital Practices in History and Ethnography Interest Group was formed in summer 2013, extending from conversation over many years responsive to an array of developments – in the American Anthropological Association around open access issues, for example, within the museum community, and around projects such as Open Folklore, Digital Himalaya and Dissertation Reviews. We are just beginning to enroll members, identify issues for discussion and lay out a work process. This session will be used to develop a work process and plan for the coming year, giving special attention consideration to the ways this group can interrelated with other RDA groups. Members of other groups are welcome. We especially welcome participation by those interested in the special data challenges of the humanities and interpretive social sciences. Data in this area are heterogeneous, often qualitative, found and difficult to categorize, share and integrate. Flexible ontologies, workflows and modes of representation are needed, recognizing the special importance of evolving methodologies and research designs as interpretive research progresses. Deep interdisciplinary collaboration will be required to advance data practices, standards, analytic tools and infrastructure in history, ethnology and allied fields - with transformative implications for both the humanities and data science.

- Libraries and Research Data
- Technical Advisory Board Business Meeting
- Working Group Chairs meeting



RDA Plenary 2 Poster Session

at Microsoft Research sponsored Evening Reception – Monday September 16, 2013 at 19:00 – Washington Marriott Hotel

ID	Poster Title	Contact Person	Link to Abstract
1	Australian National Data Service: Creating a national research data asset	Andrew Treloar, Australian National Data Service	https://www.rd- alliance.org/filedepot_downloa d/694/206
2	Building e-infrastructure and services for the agricultural research community	Johannes Keizer; Vassilis Protonatarios- FAO;Agro- Know	https://www.rd- alliance.org/filedepot_downloa d/694/225
3	Chemistry and Materials Data	Richard Kidd, Royal Society of Chemistry	https://www.rd- alliance.org/filedepot_downloa d/694/214
4	Data Curation Education in Research Centers (DCERC)	Matthew Mayernik, University Corporation for Atmospheric Research (UCAR)	https://www.rd- alliance.org/filedepot_downloa d/694/223
5	Data Foundations and Terminology (DFT) RDA Work Group (WG)	Peter Wittenburg & Gary Bergh-Cross, DFT RDA Work Group	https://www.rd- alliance.org/filedepot_downloa d/694/190
6	Data Life Cycle Labs in LSDMA	Rainer Stotzka, Karlsruhe Institute of Technology	https://www.rd- alliance.org/filedepot_downloa d/694/205



ID	Poster Title	Contact Person	Link to Abstract
7	Data sharing as a "broader impact": Results from the Site-Based Data Curation Project at Yellowstone National Park	Carole L. Palmer, G. Sayeed Choudhury, Andrea K. Thomer - University of Illinois; Digital Research and Curation Center at the Sheridan Libraries, Johns Hopkins University	https://www.rd- alliance.org/filedepot_downloa d/694/218
8	Data Type Registries	Larry Lannom, CNRI	https://www.rd- alliance.org/filedepot_downloa d/694/210
9	DataONE	William Michener, University of New Mexico	https://www.rd- alliance.org/filedepot_downloa d/694/197
10	Defining Big Data Architecture Framework	Yuri Demchenko, University of Amsterdam, System and Network Group	https://www.rd- alliance.org/filedepot_downloa d/694/229
11	Digital Infrastructure for the Arts and Humanities	Daniel Kurzawe, Göttingen State and University Library	https://www.rd- alliance.org/filedepot_downloa d/694/212
12	DMPTool: Expert Resources and Support for Data Management Planning	Patricia Cruse & Carly Strasser , California Digital Library	https://www.rd- alliance.org/filedepot_downloa d/694/199
13	EarthServer - Big Earth Data Analytics	Peter Baumann, Jacobs University	https://www.rd- alliance.org/filedepot?cid=149 &fid=154



ID	Poster Title	Contact Person	Link to Abstract
14	ENGAGE – An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens	Evangelos Argyzoudis, INTRASOFT International S.A.	https://www.rd- alliance.org/filedepot_downloa d/694/207
15	EUDAT - Towards a pan-European Collaborative Data Infrastructure	Nagham Salman, BSC	https://www.rd- alliance.org/filedepot_downloa d/694/155
16	Exchange for Local Observations and Knowledge of the Arctic: An Infrastructure for the Collection, Preservation, and Sharing of Local and Traditional Knowledge	Julia Collins, National Snow and Ice Data Center	https://www.rd- alliance.org/filedepot_downloa d/694/215
17	From BIOTIFF to JSON: Mobilizing Raw Scientific And Clinical Data For The Web	Peter Pennefather, West Suhanic, Laboratory For Collaborative Diagnostics, University Of Toronto	https://www.rd- alliance.org/filedepot_downloa d/694/170
18	From Data to Discovery to A Better World: The National Consortium for Data Science (NCDS)	Stan Ahalt, RENCI (UNC-Chapel Hill)	https://www.rd- alliance.org/filedepot_downloa d/694/159
19	GigaGalaxy: A GigaSolution for reproducible and sustainable genomic data publication and analysis	Scott Edmunds, Laurie Goodman, Susanna-Assunta Sansone - GigaScience/BGI	https://www.rd- alliance.org/filedepot_downloa d/694/211
20	Global Dimensions: Access to Foreign Government/IGO Data for Research	Bernard F. Reilly, Center for Research Libraries	https://www.rd- alliance.org/filedepot_downloa d/694/171



ID	Poster Title	Contact Person	Link to Abstract
21	iMarine, Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources	Donatella Castelli, CNR-ISTI	https://www.rd- alliance.org/filedepot_downloa d/694/234
22	Independent Evaluation of a Scientific Data Center for Compliance with the ISO 16363 Requirements for Audit and Certification of Trustworthy Digital Repositories	Robert R. Downs and Robert S. Chen, Center for International Earth Science Information Network (CIESIN), Columbia University	https://www.rd- alliance.org/filedepot_downloa d/694/209
23	Informatics for Phase-Based Materials Data	James Warren, NIST	https://www.rd- alliance.org/filedepot_downloa d/694/187
24	Infrastructure and Methodologies to Facilitate Semantic Graph-search on Technological Information in a Social-media Style	Talapady N Bhat, NIST	https://www.rd- alliance.org/filedepot_downloa d/694/160
25	ISNI and linked data	Laura Dawson, Jeff Baer, PROQUEST/BOWKER & ISNI International Authority (ISNI-IA)	https://www.rd- alliance.org/filedepot_downloa d/694/173
26	Making Data Citable	Brigitte Hausstein, GESIS Leibniz Institute for the Social Sciences	https://www.rd- alliance.org/filedepot/folder/1 49?fid=162
27	Metadata Standards Directory Working Group	Rebecca Koskela, DataONE, University of New Mexico	https://www.rd- alliance.org/filedepot_downloa d/694/224



ID	Poster Title	Contact Person	Link to Abstract
28	Neuroscience Information Framework	Jeffrey Grethe, Neuroscience Information Framework	https://www.rd- alliance.org/filedepot_downloa d/694/235
29	OmniMea: Using Personal Repositories to Support Research Data Dissemination for Individual Researchers	James C. French, Allison L. Powell , Corporation for National Research Initiatives	https://www.rd- alliance.org/filedepot_downloa d/694/204
30	OPENAIREPLUS: 2nd Generation of Open Access Infrastructure for Research in Europe	Najla Rettberg, University of Göttingen	https://www.rd- alliance.org/filedepot_downloa d/694/186
31	Opening IFPRI Research Data	Nilam Prasai & Luz Marina Alvare, International Food Policy Researsch Institute	https://www.rd- alliance.org/filedepot_downloa d/694/221
32	PID Information Types	Timothy DiLauro Data Conservancy / John Hopkins University, Tobias Weigel German Climate Computing Center (DKRZ) / University of Hamburg	https://www.rd- alliance.org/filedepot_downloa d/694/195
33	Privacy Tools for Sharing Research Data	Salil Vadhan, Michael Wojcik , Micah Altman- Harvard University , MIT	https://www.rd- alliance.org/filedepot_downloa d/694/227
34	RDA Community Capability Model Interest Group	Manjula Patel, UKOLN Informatics, University of Bath	https://www.rd- alliance.org/filedepot_downloa d/694/228



ID	Poster Title	Contact Person	Link to Abstract
35	RDA Europe - The European Plug-in into RDA	Hilary Hanahoe, Trust-It Services	https://www.rd- alliance.org/filedepot_downloa d/694/233
36	RDA-Engage Interest Group: Creating Advantages and Capacities for Data Exchange	Inna Kouper, Indiana University	https://www.rd- alliance.org/filedepot_downloa d/694/217
37	Representing Humanities Research Data using Complementary Provenance Models	Bridget Almas, Tufts University	https://www.rd- alliance.org/filedepot_downloa d/694/158
38	Research shared: ZENODO	Tim Smith, CERN	https://www.rd- alliance.org/filedepot_downloa d/694/201
39	Rolling Deck to Repository (R2R): Supporting Global Data Access Through the Ocean Data Interoperability Platform (ODIP)	Robert Arko, Columbia University	https://www.rd- alliance.org/filedepot_downloa d/694/220
40	Scientific Data – helping you publish, discover and reuse research data	Ruth Wilson, Andrew Hufton, Susanna-Assunta Sansone- Nature Publishing Group & University of Oxford e-Research Centre	https://www.rd- alliance.org/filedepot_downloa d/694/208
41	Sharing Irish Placenames as Linked Open Data	Sandra Collins, Digital Repository of Ireland	https://www.rd- alliance.org/filedepot_downloa d/694/188

R	D	A	€
RESEARCH	DATA	ALLI	ANCE

ID	Poster Title	Contact Person	Link to Abstract
42	SOCoP NSF INTEROP Project on Geospatial Semantics	Gary Berg-Cross, SOCoP	https://www.rd- alliance.org/filedepot_downloa d/694/161
43	Terra Populus: Integrated Data on Population and Environment	Tracy Kugler, Minnesota Population Center, University of Minnesota	https://www.rd- alliance.org/filedepot_downloa d/694/203
44	The ENVRI Reference Model	Alun Preece, Cardiff University	https://www.rd- alliance.org/filedepot_downloa d/694/185
45	The ESIP Federation: Enabling Cross-Community Innovation	Carol B. Meyer, Erin Robinson, Foundation for Earth Science	https://www.rd- alliance.org/filedepot_downloa d/694/191
46	The ICSU World Data System: Trusted Data Services For Global Science	Mustapha Mokrane, ICSU World Data System	https://www.rd- alliance.org/filedepot_downloa d/694/172
47	The ORCID iD: Optimizing Discoverability of Research and Scholarship	Rebecca Bryant, ORCID	https://www.rd- alliance.org/filedepot_downloa d/694/196
48	The Research Data Alliance as a Double-Loop Organization	Bruce Caron, New Media Research Institute	https://www.rd- alliance.org/filedepot_downloa d/694/174



ID	Poster Title	Contact Person	Link to Abstract
49	The UC Curation Center (UC3): Developing Tools & Services for Managing Research	Patricia Cruse & John Kunze, California Digital Library	https://www.rd- alliance.org/filedepot_downloa d/694/198
50	The Worldwide DPHEP Collaboration for Long- Term Data Preservation	Jamie SHIERS, CERN	https://www.rd- alliance.org/filedepot?cid=149 &fid=153
51	Toward enhancing interoperability of toxicogenomics databases	Diana Hendrickx, Maastricht University	https://www.rd- alliance.org/filedepot_downloa d/694/192
52	Wagging the long tail of research data: improving the visibility and interoperability of disparate datasets	Kathleen Shearer, Najla Rettberg - Confederation of Open Access Repositories, OpenAIRE	https://www.rd- alliance.org/filedepot_downloa d/694/226
53	XTXS: A Comprehensive Database of Country-Level, Education-Related Data for Comparative and International Research	David C. Miller, American Institutes for Research (AIR)	https://www.rd- alliance.org/filedepot_downloa d/694/213



RDA Technical Advisory Board Candidates

Bridget Almas - Tufts University, United States Peter Baumann - Jacobs University & Rasdaman GmbH, Germany Simon J D Cox - CSIRO, Australia Robert R. Downs –Earth Institute, Columbia University, United States Giuseppe Fiameni - CINECA, Italy Peter Fox - Rensselaer Polytechnic Institute, United States David W. Fulker - OPeNDAP, United States John Kunze - California Digital Library, United States Michael Lautenschlager - German Climate Computing Centre (DKRZ), Germany Stefano Nativi - National Research Council, Italy Susanna-Assunta Sansone - University of Oxford e-Research Centre, United Kingdom Jennifer Schopf - Indiana University, United States Jamie Shiers – CERN, Switzerland Rainer Stotzka - Karlsruhe Institute of Technology (KIT), Germany



Bridget Almas - Tufts University, United States



Bridget Almas, Senior Software Developer and Architect, Perseus Digital Library, Tufts University, United States

Bridget has worked in software development since 1994, in roles which have covered the full spectrum of the software development life cycle, focusing since 2007 in the fields of language study and digital humanities. She has designed and developed multiple applications and services deployed in various SOA environments, and her skills include the programming languages Java, Javascript, XQuery, XSLT, Ruby and Perl, as well as broad experience with databases and XML technologies. She has served as the technical lead on various projects and has extensive experience in project management.

In her current role at Tufts University, Bridget is the lead software developer and architect for the Perseus Digital Library. In this role, she has represented and contributed on behalf of the Perseus Digital Library on several multi-institution projects, including the Bamboo Project funded by the Mellon Foundation, the Libraries and the Transformation of the Humanities funded by the Institute of Museum and Library Services, and NEH/DFG Bilateral Digital Humanities-funded Hellespont Project. Starting September 1, 2013, Bridget will be the lead developer on The Perseids Project, a two-year project funded by The Mellon Foundation to expand the Perseids collaborative editing platform with support for use cases including classroom collaboration on digital editions, scholary curation of texts, and development of dynamic syllabi.

Bridget also has a background in the study of foreign languages, including French and Mandarin Chinese and is one of the primary programmers and founding board members of the open source Alpheios Project, developing tools and resources for the study of classical languages.

- Contribution to RDA -

In putting forth my candidacy for the Technical Advisory Board of the Research Data Alliance, I intend to bring a new technical perspective from the humanities, with a particular focus on the supporting the data creation, preservation and access needs of undergraduate researchers in the humanities. My role as lead developer for the internationally supported Perseus Digital Library enables me to contribute to the board a unique exposure to a wide variety of humanities use cases across diverse fields of study including modern and classical languages, linguistics, political science and history. I also bring practical experience from outside of the academic community, with an extensive background in commercial software development that includes ten years of experience at Ovid Technologies, serving the needs of the medical and scientific research community.



Peter Baumann - Jacobs University & Rasdaman GmbH, Germany



Peter Baumann - Professor of Computer Science and Leader of the Large-Scale Scientific Information Systems Research Group, Jacobs University; founder/CEO of Rasdaman GmbH, Germany

Peter Baumann is Professor of Computer Science at Jacobs University where he leads the Large-Scale Scientific Information Systems Research Group and founder/CEO of rasdaman GmbH, a hi-tech research spinoff.

Since 1992, he is researching on Big Science Data – specifically, on services for massive multidimensional gridded data. He is Principal Architect of the rasdaman Big Array Analytics engine, a multi-parallel scalable raster server extending SQL with n-D raster operators. This work effectively has pioneered today's database research domain of Array Databases, which is generally seen as a core contributor to Big Science Data solutions. He has ample experience ranging from theory over scalable architectures to science data applications in Earth sciences (remote sensing, ocean, atmosphere, geology, ...), Life science (human brain, genetics, ...) to Space science (cosmology simulations, virtual observatories, planetary geology, ...), and longterm data preservation.

In standardization, Peter is fostering conceptual models and interfaces for Big Data in science and engineering. In the Open Geospatial Consortium (OGC), he is editor of OGC's unified spatio-temporal Coverage model and Web Coverage Service (WCS) suite standards.

A main quest for Peter is harmonization between different standards, based on technically sound and practice-proven principles. He has connected the WCS, WCPS, WPS, WFS, SOS standards through a common underlying spatio-temporal model. He has integrated temporal coordinate systems with spatial coordinates, and OGC operates its Coordinate System resolver running a server developed by his group.

In ISO, he is member of the SQL Working Group where is has initiated work on Array SQL. He is Council Member of the Commission for Geo Information in IUGS and member in the Federation of Earth Science Information Partners (ESIP). Further, he is Charter Member in OSGeo, a not-for-profit organization supp-orting development of open source geospatial software, and promoting its widespread use.

In projects, he is working with NASA, ESA, and many vendors, agencies, and data centers worldwide. Over two decades of such work has given him a thorough understanding of the needs and viewpoints of these players with respect to scientific and engineering data gathering, offering, and use. Consequently, he is running projects in upstream research (eg, ESA funded and EU ICT) and downstream operational services, such as the EU FP7 e-Infrastructure project EarthServer which has evolved into a transatlantic collaboration for Big Earth Data.

In outreach, Peter is not limiting himself to scientific publications. He is regularly publishing in newspapers etc. to educate the general public about data issues and their progress. Currently, he is working on a mini-series documentary about Big Earth Data in TV, a project



which two main German TV stations, ZDF and Arte, have embraced enthusiastically and with substantial financial investment.

- Contribution to RDA -

The Research Data Alliance mission of making Data Work across barriers very much characterizes Peter's work and contributions since 1990. Unleashing the power of data, in particular when they are voluminous, coming at fast rates, with high inherent complexity, and with quality issues, was and is at the heart of his holistic perspective. His group's recognized lead in Array Database technology and application is a consequence of his deep, sustained devotion to lowering barriers in (big and small) data access.

However, he is not stopping with technical considerations, but actively working across community boundaries, communicating with techies, data providers and practitioners, as well as policy and standards makers.

Being deeply rooted in both academia and industry, Peter can bring in a balanced, unbiased view on all stakeholder interests. Notably is his strong involvement in the open-source community, specifically OSGeo.

Therefore, as a Technical Advisory Board member Peter will be able to contribute to technical harmonization and advance of interoperability, standards, Big Data, and related areas where he is already member of the corresponding interest RDA groups. Further, he will establish a point of liaison of RDA with related stakeholders like OGC, ISO SQL, and OSGeo.



Simon J D Cox - CSIRO, Australia



Simon J D Cox, Senior Principal Research Scientist, CSIRO Australia.

Simon Cox trained in geophysics, with a PhD from Columbia following degrees from Cambridge and Imperial College London. His engagement with informatics began in the Australian Geodynamics CRC, where Simon moved its reporting onto the emerging World Wide Web, including deployment of a web-mapping system for Australian geology and geophysics in 1995. The challenge of content management for the AGCRC led to an interest in metadata-based systems, and Simon's engagement with standards when he joined the Dublin Core Advisory Council. Work on XML for mineral exploration data led to leadership in the GeoSciML project in collaboration with a number of geological surveys. An interest in tying these into broader interoperability systems led to engagement with the Open Geospatial Consortium, where he co-edited the Geography Markup Language (GML) standard, and developed Observations and Measurements (0&M) as a common language for in situ, ex situ and remote sensing. 0&M became an ISO standard and now forms the basis for operational systems in diverse fields including air-traffic, climate and weather, water data transfer and environmental monitoring applications. In 2009-10 he spent a year at the EC Joint Research Centre in Italy working on integration of GEOSS and INSPIRE. He has served on the council of the IUGS Commission for Geoscience Information and the International Association for Mathematical Geosciences, and amongst other roles currently serves on the OGC Architecture Board, the ISO/TC 211 Ontology Management Group, the steering committee of the Ocean Data Interoperability Platform (ODIP), the International GeoSample Number (IGSN) Implementing Organization, and the Australian Government Linked Data Working Group. Simon is a Senior Principal Research Scientist in CSIRO Land and Water, based in Melbourne, working on a variety of projects across environmental informatics and spatial data systems.

- Contribution to RDA -

As part of the RDA TAB Simon would provide strong links into related initiatives being undertaken under a variety of auspices:

•the geospatial standard community, through leadership roles in the Open Geospatial Consortium and ISO Technical Committee 211, in particular through his chairing the group responsible for persistent identifiers in both organizations

•the earth sciences through longstanding involvement in geoscience data exchange, with International Union Geological Sciences through GeoSciML, International Association Mathematical Geosciences, Ocean Data Interoperability Platform and International Geosample Number, as well as the Earth and Space Science Informatics sections of American Geophysical Union and European Geosciences Union



•the statutory data sector, through involvement with Australian initiatives in geosciences and water resources data, the Australian Government Linked Data Working Group, and the Australian-New Zealand Foundation Spatial Data Framework

•high-level international collaborations, through strong links with European Commission – Joint Research Centre, and participation in US EarthCube projects

•Institutional data publication, through CSIRO.

He has had a long involvement with some of the core technologies related to web data distribution, including XML, RDF, URIs, and most recently Linked Data.



Robert R. Downs -Earth Institute, Columbia University, United States



Robert R. Downs - Senior Digital Archivist and Senior Staff Associate Officer of Research, Center for International Earth Science Information Network (CIESIN), The Earth Institute, Columbia University

Dr. Robert R. Downs serves as a senior staff associate officer of research at Columbia University and as 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. He holds the PhD in Information Management from the Stevens Institute of Technology and develops, manages, and conducts research on systems that support science, scholarship, and learning by focusing on scientific data management and stewardship, data policy, digital preservation, software reuse, and system design and evaluation. He has taught management and computer science courses, has written for various publications, and has presented on many topics at conferences and workshops. He has been serving as a member of the Editorial Board of the CODATA Data Science Journal (DSJ) since 2009, as Vice-Chair of the Columbia University Institutional Review Board (IRB) since 2006, and as a Scientific Member of the IRB since 2002. He has served as principal investigator, co-principal investigator, or manager for various projects and in leadership roles for many conferences and associations, including the Federation of Earth Science Information Partners (ESIP) and the NASA Earth Science Data Systems (ESDS) Working Groups, for which he currently serves as Vice-Chair for the Open Source Working Group. He contributes to the development of international standards for trustworthy digital repositories with the Consultative Committee for Space Data Systems (CCSDS) and leads the EarthCube Education and Workforce Development Special Interest Group (SIG) for the National Science Foundation. Dr. Downs is a Senior Member of the Association for Computing Machinery (ACM) and a member of the American Geophysical Union (AGU), the American Society for Information Science and Technology (ASIS&T), and the International Association for Social Science Information Services and Technology (IASSIST).

- Contribution to RDA -

My experiences serving in various roles, often using limited resources to address evolving needs of diverse institutions and those they serve, can contribute to the practices, research, and technologies that the Research Data Alliance (RDA) Technical Advisory Board (TAB) faces. Experiences in managing natural and social science data, from data policy and strategy to dissemination, preservation, and impact, enable me to bring depth and breadth to data stewardship issues. My research and experience on software adoption, development, management, and reuse provides systems perspectives. My IRB experience provides a unique view on data issues related to a wide range of both social and natural science research involving human subjects. Collaborating on standards development and committees gives me an international perspective. I bring the diverse skills and interests that are needed to serve a



community of members who have many responsibilities for the data and documents that represent great value for the future of science, scholarship, and learning. By electing me to the RDA TAB, all RDA members will have an experienced, collaborative, and well-rounded advisor who is honored to provide leadership and support for their efforts to improve practices for the stewardship, dissemination, and preservation of scientific data and research-related resources.



Giuseppe Fiameni - CINECA, Italy



Giuseppe Fiameni - SuperComputing Applications and Innovation Department, CINECA, Italy

Giuseppe Fiameni holds a degree in Computer Science from the University of Bologna and joint the SuperComputing Department of CINECA (largest Italian supercomputing centre) as technology consultant on 2004. Over the years he has been contributing to many European (PRACE, DEISA, EMI, etc.) and National projects maturing a strong experience in High Performance Computing infrastructure and large data management and analysis systems. He has extensive experience in the field of computational sciences, parallel architectures, parallel programming models, scaling applications and system performance evaluation.

G.F. is currently leading the "Middleware for HPC Services" group of the SuperComputing Applications and Innovation department which is responsible for the implementation of data and computational services and contribute to the evolution of the CINECA computational and data infrastructure. He is actively contributing to the EUDAT and the Human Brain project, member of two RDA Working groups (Practical Policy and Big Data Analytics) and member of the ICT board of the EPOS project (www.epos-eu.org).

- Contribution to RDA -

I personally think that experience in managing and producing vast amount of scientific data, varying in nature and form, might be of great value in the settlement and definition of RDA directions. Furthermore, the perspective of an HPC centre, operating production services for the public and the private sector, could be fundamental for RDA Working Groups to produce implementable and durable results.



Peter Fox - Rensselaer Polytechnic Institute, United States



Peter Fox - Tetherless World Constellation Chair, Professor of Earth and Environmental Science and Computer Science, Director of the Information Technology and Web Science Program at Rensselaer Polytechnic Institute

Peter Fox is Tetherless World Constellation Chair and Professor of Earth and Environmental Science and Computer Science and Director of the Information Technology and Web Science Program at Rensselaer Polytechnic Institute. He holds B.Sc. (Hons) and Ph.D. in Applied Mathematics from Monash University. He spent 17 years at the High Altitude Observatory of the National Center for Atmospheric Research as Chief Computational Scientist. Fox's research specializes in the fields of solar and solar-terrestrial physics, computational and computer science, information technology, and distributed semantic data frameworks. This research utilizes state-of-the-art modelling techniques, internet-based technologies, including the semantic web, and applies them to large-scale distributed scientific repositories addressing the full life-cycle of data and information within specific science and engineering disciplines as well as among disciplines. Fox is currently PI for the Integrated Ecosystem Assessment, the Deep Carbon Observatory Data Science, Global Change Information System, and the GeoData projects and co-PI on 5 others. Fox has spent over 28 years bridging science and distributed data and information systems to support community activities utilizing use case driven design. Fox leads working groups for the semantic web cluster of the Earth Science Information Partners, is chair of the International Union of Geodesv and Geophysics Union Commission on Data and Information and past chair of the AGU Special Focus Group on Earth and Space Science Informatics, is an associate editor for the Earth Science Informatics journal, is a member of the editorial boards for Computers in Geosciences, Geoscience Data Journal and Nature's Scientific Data Journal. Fox served on the International Council for Science's Strategic Coordinating Committee for Information and Data. Fox was awarded the 2012 Martha Maiden Lifetime Achievement Award for service to the Earth Science Information community, and the 2012 European Geoscience Union, Ian McHarg/ Earth and Space Science Informatics Medal. http://tw.rpi.edu/web/Person/PeterFox

- Contribution to RDA -

In addition to my bio, I have extensive experience in, and knowledge of, socio-technical systems/ virtual organizations. I have served in capacities as both technical and strategic advisor for many "data" activities other than my own, over my career. As examples, these range from: technical advisory board member for OPeNDAP in its early years (1996-2003); the first ICSU Strategic Committee on Information and Data (pre-SCCID) that determined the fate of the World Data Centers and Services, and gave guidance to CODATA; as a User Working Group member for the NASA DAAC Socio-Economic Data Applications Center; chair of the Virtual Observatory Working Group for the electronic Geophysical Year; long-term member of NASA's Earth Science Data System Working Group; new member of NOAA's Data Archiving and Access Requirements Working Group (under NOAA Science Advisory Board); consultant



to USGS on Data Release; convener of the NSF Geodata inter-agency workshop activity for Earthcube; all the way to advisory board member for the Digital Antiquity project, CyberSHare - an NSF CREST-funded Center of Excellence, and the Open Geospatial Consortium Interoperability Institute (now in stasis). I would bring a steady, well-considered, and integrated global view of research data, balancing technical advice with socio-cultural and organizational realities.



David W. Fulker - OPeNDAP, United States



David W. Fulker - President, Open-source Project for a Network Data Access Protocol OPeNDAP, Inc.

President of the nonprofit Open Source Project for a Network Data Access Protocol (OPeNDAP), Dave Fulker has focused his career on serving scientists and science educators via software advances. His teams have combined leading-edge technologies with end-user service, underpinned by expertise in both technical and social aspects of the information age. Dave directed the Unidata Program (at the University Corporation for Atmospheric Research) from inception in 1984 until 2002, overseeing development of the Network Common Data Form (netCDF) and other data management, analysis and visualization software that now is considered fundamental geoscience infrastructure, also benefitting other domains. Unidata is seen as an exemplar of community participation and data sharing, especially in higher education, and the program's reach (even beyond its technologies) is increasingly international. Before leading Unidata and becoming (founding) Executive Director of the National Science Digital Library, Dave spent 18 years in software-development and team leadership at the National Center for Atmospheric Research, where he was corecipient of NCAR's first Technology Advancement Award. His role at OPeNDAP is a natural follow-on to these earlier experiences and includes winning awards (as principal investigator) from NSF and NOAA under which OPeNDAP advances the state of the art in Web-based data access services.

Dave is a Fellow of the American Meteorological Society (AMS) and recipient of the AMS Cleveland Abbe Award and the Educom Medal for Technology in Education. Dave holds Master and Bachelor of Arts degrees in Mathematics from the University of Colorado, is a professional trumpeter (jazz and classical), and served 8 years as President and Board Chair for the Boulder Philharmonic Orchestra, where he earlier played principal trumpet for 34 seasons.

- Contribution to RDA -

The Research Data Alliance purpose—implementing the technology, practice, and connections that make data work across barriers—characterizes much of Dave Fulker's career, which has lowered barriers in these contexts:

• During 18 years at NCAR, Dave addressed limitations of early computers and lack of software experience among scientists of the era. His abstractions and libraries matched how scientists frame problems and reach understanding.

• His next 18 years at Unidata enabled meteorology departments' transitions into the digital realm. Unidata participants—among academic peers—often were the most advanced users of computers and Internet, emerging as a well-connected and pace-setting community of data users.

• At the National Science Digital Library Dave addressed distinctions between physical and digital libraries, with the latter needing to characterize a newly rich set of connections (similarities and contrasts) among digital artifacts.



Presiding over OPeNDAP extends Dave's commitment to technologies that lower barriers to data use and integration. He is now PI on an NSF/EarthCube Building Blocks project to develop a protocol that enables clients to invoke data-acquisition services rich with capabilities for pre-retrieval operations.



John Kunze - California Digital Library, United States



John Kunze - Associate Director, UC Curation Center, California Digital Library, United States

John Kunze is an Associate Director at the University of California Curation Center in the California Digital Library. With a background in computer science and mathematics, his current work focuses on data set curation, "data papers", creating long-term durable information object references using ARK identifiers and other identifiers using the N2T resolver, publishing tabular datasets, archiving websites, and promoting lightweight metadata standards (Dublin Core "Kernel" metadata and crowd-source vocabularies). His work has been supported by National Science Foundation, the Library of Congress, the Gordon and Betty Moore Foundation, and Microsoft Research.

Previously, he contributed heavily to the standardization (RFC's, NISO specifications) of URLs, Dublin Core metadata, and the Z39.50 search and retrieval protocol. In an earlier life he created UC Berkeley's first campus-wide information system, which was an early rival and client of the World Wide Web. Before that he was a BSD Unix hacker whose work survives in today's Linux and Apple systems.

- Contribution to RDA -

I would bring a wealth of experience, practical insights, and creative thinking to the Research Data Alliance's Technical Advisory Board and the charge to evolve the RDA Technical Roadmap and assist the Working Groups.

Being around online data systems for all of my professional life, I have deep experience of all sides of the data curation problem: from designing and building, to supporting and troubleshooting, to using and teaching, to managing, promoting, and standardizing.

My work at the California Digital Library (CDL) has been an extension of the library world's historic and centuries old mission to archive, curate, and preserve all forms of scholarly output, most recently to datasets, methods, software, etc. In current curation world, persistent identifiers and metadata -- two of my specializations -- are critical components.

I also bring deep experience with the Internet Engineering Task Force (IETF) culture and processes, dating from the early 1990s and including publication of six RFCs and numerous Internet Drafts.

The library work has also forced becoming skilled at developing and identifying solutions that are simple, practical, open-source, and low-cost. Working within these constraints has produced granular solutions ("micro-services") that can be combined in various ways (eg, Bagit and Pairtree, both specified as Internet Drafts).



Michael Lautenschlager - German Climate Computing Centre (DKRZ), Germany



Dr. Michael Lautenschlager - Head of German Climate Computing Centre DKRZ , Department Data Management & Director World Data Center, Germany

Education:

Masters Degree in Physics Ph.D. in Meteorology (Turbulence Parameterisation in Helical Flows)

Employment:

- 1986-1991: Max-Planck Institute for Meteorology (MPI-M), Hamburg (Climate model development and application with emphasis on paleo-climate)
- 1991-2000: German Climate Computing Centre (DKRZ), Hamburg (Climate modelling and scientific data management)
- 2000-2009: Model and Data Group (M&D) at MPI-M Establishment of ICSU World Data Center Climate (WDCC, URL: <u>www.wdc-climate.de</u>)
- Begin of 2010: M&D moved to DKRZ as additional department "Data Management" Continuation of life cycle management, long-term archiving, and inter-disciplinary utilisation of Earth system research data with emphasis on climate modelling (DKRZ Bundesstrasse 45a, D-20146 Hamburg, Germany, URL: <u>www.dkrz.de</u>)

Relation to IPCC Data Management:

- IPCC activity started with the formation of the IPCC Data Distribution Centre (DDC) together with Tim Carter and Mike Hulme as WG II and III support in the Second Assessment Report (SAR).
- DDC manager and member of TGICA since start of the DDC.
- The WDCC at DKRZ hosts the GCM data part of the DDC and acts as IPCC reference data archive for the SAR, TAR and AR4. Recently the reference archive has been completed by data from the First Assessment Report (FAR) and the AR5 archive completion is under way.

- Contribution to RDA -

The overall topic for more than 20 years is bridging the gap in scientific data management between technical possibilities and scientific requirements which is connected with improvements of data infrastructures on one hand and development in scientific work flows on the other hand. DKRZ is running the WDCC as long-term data archive for climate with more than 1 PetaByte and 4500 registered users. In the average the WDCC handles 75,000 data downloads per months which are related to a data volume of 20 TB per month (<u>http://www.dkrz.de/daten-en/wdcc/statistics</u>). The WDCC serves inter-disciplinary data requirements which are naturally related to climate research with all its contributions from



various climate disciplines ranging from Physics, Chemistry, Meteorology and Oceanography over the more general Geosciences to Biology, Ecology, Hydrology, Forestry and Agriculture. Over the last years special emphasis is on

- Data life cycle management
- Data transition from the scientific domain over long-term archiving to data integration into scientific literature
- Usage of Persistent Identifiers (PID) with respect to EPIC PIDs and DataCite DOIs and the underlying Handle Server infrastructure
- Big Data handling while big means archive size in the range of PetaByte with millions of individual data entities
- Cooperation in the Implementation and management of a globally distributed climate data federation (Earth System Grid Federation) and development of the underlying data infrastructure



Stefano Nativi - National Research Council, Italy



Stefano Nativi - Head of the Earth & Space Informatics laboratory (ESSI-lab), National Research Council of Italy - Institute of Atmospheric Pollution Research -(CNR-IIA), Head of the CNR-IIA Division at the Area di Ricerca di Firenze, Italy

PRESENT POSITION: IIA-CNR (Institute of Atmospheric Research Pollution of the National Research Council of Italy), head of the Division at the Area di Ricerca di Firenze and head of the Earth & Space Informatics laboratory (ESSI-lab).

RESEARCH INTERESTS:

Multidisciplinary interoperability; research cyber(e)-infrastructures; Earth and Space Science Informatics; Brokering services, architecture, and technologies; Open Data Access; Big Data.

PROFESSIONAL PREPARATION:		
University of Florence (Italy)	Master degree (Laurea) in Electronic Engineering	1990
University of Padua	Professional Engineer	1990
University of Florence (Italy)	Ph.D. "Methods and Technologies for the Environmental Monitoring"	1993
Bristol University (UK)	Post-Doctoral Research Fellowship: "Remotely sensed data fusion and	1994-1995
	passive microwave techniques for atmospheric water cycle study"	
UNIDATA of UCAR/NCAR (USA)	Distinguish Visitor as "Expert of Geo-Information Standard and Data	2001-2003
	Model"	

OTHER RELEVANT POSITIONS HELD:

2012-	Member of the Infrastructure Implementation Board (IIB) of GEO (Group on Earth Observation)
2009–2012	Co-chair of the Science & Technology Committee of GEO
2013-	Member of the Open Data Access WG of Science Europe, representing the CNR
2008–2013	Co-founder and President of the Earth & Space Science Informatics division of EGU (European Geosciences Union)
2005–	Member of the Metadata Core Drafting Team of the INSPIRE European Directive (Infrastructure for Spatial
	Information in Europe)
2005–	PI and co-PI of research and development projects funded by: the European Commission, European Space
	Agency, and NSF.
2006–	External Lecturer of Geoinformation Infrastructure in Earth Observation and Earth System Science, master in
	Geomatics, Friedrich Schiller University in Jena – Germany
2004-2012	Professor of Systems for land management, Master in Informatics, University of Padua -Italy
2001-2011	Professor of Web services management, Information Engineering degree, University of Florence -Italy
1998–2001	Professor of Software engineering, degree in Informatics, University of Padua -Italy

- Contribution to RDA -

[Barbara Ryan, GEO] Dr. Nativi has been the driving force behind a major change in information management and technology for the Group on Earth Observations (GEO) and the Global Earth Observation System of Systems (GEOSS). This change has singularly resulted in more than 10,000,000 Earth observation resources becoming discoverable and accessible



through our information infrastructure. Without his leadership, this would not have occurred. In addition to his astute technical experience, Dr. Nativi has a wealth of knowledge regarding the governance aspects of linking, integrating and accessing global information assets.

[Arnoldas Milukas, EC RTD] Dr. Nativi is very well connected with the IT community in Europe. He has notably been involved as leader of technical activities in major Community-funded collaborative research projects such as EUROGEOSS, GEOWOW and GEOVIQUA. In these projects, he has demonstrated his high competence in the field of multidisciplinary interoperability for System of Systems and, more widely, regarding technologies to manage, harmonize, and share Earth and Space Science resources. For these reasons, Dr. Nativi could provide very useful links to what has already been developed within GEOSS.

[Max Craglia, EC JRC] We believe his knowledge and experience will be invaluable assets for the success of the RDA.

Dr. Stefano Nativi will contribute to RDA-TAB with his leadership on brokering solutions for multidisciplinary interoperability reflecting the needs of the various disciplines.



Susanna-Assunta Sansone - University of Oxford e-Research Centre, United Kingdom



Susanna-Assunta Sansone - Associate Director and Principal Investigator, University of Oxford e-Research Centre, Consultant for the Nature Publishing Group (NPG), United Kingdom

I am an Associate Director and Principal Investigator at the University of Oxford e-Research Centre and a Consultant at Nature Publishing Group (NPG). For the last thirteen years, my research interests and activities have centered on the capture, management and curation of biological, biomedical and health sciences data; ensuring value is not lost and facilitating the interpretation and the corroboration of results, making for more efficient practice.

I hold a PhD in Molecular Biology from Imperial College, London, UK. After spending few years in privately-held vaccine company as a senior scientist, I moved to the European Bioinformatics Institute (EMBL-EBI), in Cambridge, UK, where for nine years I led a team of curators and developers, coordinating several international data management projects. With my team in Oxford, I continue to research and develop methods and software for curating and managing data with and for a variety of academic and industrial communities.

I am also a consultant at NPG, providing expertise and advice open access data publication products. I am the Honorary Academic Editor of Scientific Data (http://www.nature.com/scientificdata), the new NPG open access data publication platform; my role is to foster relationships with the research and technical communities, help defining standards for data representation, publication and reuse.

I also work with funding agencies and pre-competitive informatics initiatives to develop and promote uptake of community-developed standards, such as minimum reporting requirements, ontology/terminologies and exchange formats (e.g. doi: 10.1038/ng.1054 and doi: 10.1016/j.drudis.2011.09.013). I am a central player in the grass-roots data standardization movement; I am co- founder of the MIBBI and the BioSharing initiatives (http://biosharing.org) and co-chair of the upcoming "Community-Based Standards Efforts" event, part of the NIH Big Data to Knowledge initiative (http://bd2k.nih.gov/workshops.html). I have extensive experience in service on non-profit Boards of several standardization and advocacy initiatives, including the Genomic Standards Consortium (http://gensc.org), iCommons (http://icommons.org) and Data Dryad (http://datadryad.org).

- Contribution to RDA -

Being at the heart of RDA activities, discussing issues with stakeholders and moving to implement solutions, greatly interests me. I believe that shared, annotated research data and methods prevent unnecessary repetition of work; reproducible research offers new discovery opportunities by underpinning science of the future. But even when data are publicly available, published results are often not reproducible or reusable by independent investigators due to incomplete annotation. For these reasons, I am a strong advocate of the open 'data commoning' culture and supporter of community-driven standardization efforts.



RDA's long-term success depends on the degree to which all stakeholders (researchers in the academic and industrial arenas, service providers, publishers, funder agencies etc.) adopt the recommendations of its Working and Interest Groups. To further availability, discoverability and interpretability of research data we must address both technical issues as well as the 'social engineering' challenge.

If elected to the RDA TAB, I will bring my direct expertise in the capture, management and curation of data in the biological, biomedical and health sciences; these are also areas where a wealth of (non interoperable) standards for reporting datasets exists (see http://www.biosharing.org/standards). I will also help promoting the growth of technologies and reward mechanisms for improving data curation, sharing and publication, (re)use of existing community standards, where possible, and their harmonization and integrations.



Jennifer Schopf - Indiana University, United States



Jennifer Schopf - Director of International Networking, Indiana University, United States In 2013, Dr. Schopf joined the Indiana University networking team, a leader in national and international connectivity. She oversees the NSF-funded America Connects to Europe (ACE) and Trans Pacific (TransPac3) awards, which provide network infrastructure for international collaborations. She will oversee the next generation of these projects to support the collaborative needs for science and large data transfers. She is uniquely qualified in this role as a prior NSF program officer, who was part of the refactoring of the International Networking Research Collaboration program, and helped to develop domestic pragmatic networking solicitations, all which focused on enabling science and data sharing through sustained infrastructure. She was also part of the team that drafted the NSF guidance requiring Data Management Plans to be included in all submitted proposals.

In addition to her experience at NSF, Schopf was a member of the Woods Hole Oceanographic Institution's Ocean Informatics team, where she assisted with CI and data planning across the institute, a senior member of the Globus distributed software team for over 7 years, and the Director for the DOE SciDAC Center for Enabling Distributed Petascale Science. She also spent 3½ years working at the UK National eScience Center in Edinburgh, Scotland. In between NSF and IU, she spent a year as a program manager and analyst for the IEEE Computer Society. Her research focuses on performance, anomaly detection, and user requirements gathering, primarily in distributed systems with relation to large data use. She has co-edited a book, coauthored over 50 refereed papers, and given over 100 invited talks. She received her BA in Mathematics and Computer Science (with general and departmental honors) from Vassar College and her MS and PhD in Computer Science and Engineering from UC San Diego, where she was a student of Dr. Francine Berman.

- Contribution to RDA -

Dr. Schopf has significant experience both in the research needs for large-scale data use and in working with volunteer organizations. Her technical expertise includes work in understanding user requirements in data and collaboration as part of her role with the Globus team, her time as a program officer at NSF, and in her current role at IU. She has been involved with several efforts that developed and evolved roadmaps for technical and governmental agencies (including NSF in the US and JISC in the UK). She will bring a unique background to the current board members and will be able to advocate from the perspective of networking for data sharing, especially in international environments. In addition, she was part of the core team that started the Grid Forum, now known as the Open Grid Forum, and spent 5 years on the steering committee of that organization. In that role, along with the chair of the organization, she defined the policies and procedures for Working Group startup, review, and retirement, and oversaw the initial use and adaptation of those procedures. She has a depth of experience evaluating the impact and effectiveness of projects, from both technical and political points of view.



Jamie Shiers - CERN, Switzerland



Jamie Shiers - Information Technology Department, CERN & Manager of the Data Preservation for Long-Term Analysis in High Energy Physics (DPHEP)

High Energy Physics (DPHEP) project that involves all of the major (HEP) laboratories and experiments worldwide and collaborates actively with other disciplines. DPHEP is focusing on solutions to long-term exa-scale data preservation: data-related standards are clearly of key importance.

He has been involved in Large-Scale Data Management, Object and Relational Database systems and services, distributed application design, implementation and support, operations and services for many years. Recently, he led the effort to harden the worldwide LHC Computing Grid (WLCG) services that helped scientists to turn data into discoveries in record time, using petascale distributed resources.

He has experience in a variety of standardization activities, including those directly related to data (the Object Data Management Group and the IEEE Storage Systems working group) as well as computer languages (ISO/IEC JTC1/SC22/WG5 – Fortran).

He has served on the Programme Committees for numerous conferences, including the upcoming PV 2013 (ensuring long-term Preservation and adding Value to scientific and technical data), numerous IEEE Mass Storage Systems and Massive Storage Systems and Technologies conferences and is Vice-President for European Activities of the IEEE Computer Societies Executive Committee on Mass Storage Systems.

He has participated in numerous EU-funded projects, including the Enabling Grids for EsciencE (EGEE) series, the European Grid Initiative Design Study, and led the Work Package focusing on "Heavy User Communities" in the EGI-InSPIRE project (May 2010 to April 2013). The latter included, but was not limited to, Life Science, Earth Science, Astronomy and Astrophysics communities.

HEP: High Energy Physics DPHEP: Data Preservation (for Long-Term Analysis) in HEP WLCG: Worldwide LHC Computing Grid LHC: Large Hadron Collider ISO: International Organization for Standardization EGI: European Grid Infrastructure EGI-InSPIRE: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

- Contribution to RDA -

Extensive experience in diverse data-related projects, small and large, involving numerous disciplines has convinced me of the value of cooperation, collaboration, and the role of



standards. My current responsibilities, which include ensuring solutions for long-term data and knowledge preservation up to the exa-scale, and in fostering peer to peer interdisciplinary collaboration, fit well with the goals of the RDA. As such, I am strongly motivated to be considered as a candidate for the RDA Technical Advisory Board. The "added value" that I would bring is firstly this experience – which ranges from hands-on requirements gathering, application design and implementation to project leadership, community building and service delivery. My proven ability to listen, build consensus, evangelize and deliver production systems according to strict deadlines and with limited resources should help further the RDA's goals and be instrumental in the take-up of its output across numerous communities. My background has included being first a Data User, then later Technologist, Policy Maker and Manager. With respect to my host organization – CERN – it is perhaps today best described as Data Generator / Manager, representing primarily, but not exclusively, Natural Sciences (High Energy Physics).



Rainer Stotzka - Karlsruhe Institute of Technology (KIT), Germany



Rainer Stotzka - Leader of "Software Methods" department and software development group "Big Data", Institute for Data Processing and Electronics, Karlsruhe Institute of Technology (KIT), Germany

Dr. Rainer Stotzka studied technical cybernetics in Stuttgart and received his Ph.D. degree in computer science from the University of Mannheim, Germany, in 1995. At Karlsruhe Institute of Technology (KIT), Institute for Data Processing and Electronics, he leads the department "Software Methods" and a software development group "Big Data". Within the Helmholtz activity "Large Scale Data Management and Analysis" he is responsible for the "Data Life Cycle Lab Key Technologies" inducing novel research data management methods for many scientific communities. He is involved in various national and international research projects and initiatives, e.g. DARIAH, eCodicology, EUDAT, RDA and the German Alliance for Research Data Management. Within the last five years he and his research groups published more than 40 papers in peer-reviewed conferences and journals.

- Contribution to RDA -

RDA has the potential to grow into the leading institution in research data management. It will influence most of the scientific communities. I would like to foster actively the common idea of Open Research Data. After analysis and publication of scientific results, research data should be publicly available to enable reproducibility of the results and reuse in new scientific contexts. For this purpose data has to be reconditioned causing additional efforts and costs. On one hand these costs have to be recognized and supported by the funding agencies, on the other hand common and approved recipes for effective data organization and powerful tools have to be developed. Within RDA we have the potential to work on both sides and to clear the way for better and more reliable and long-term research data management.

With some ten years of experience as a data scientist and the head of the "Data Life Cycle Lab Key Technologies" I am closely connected to many scientific communities, e.g. humanities, physics and biology. Receiving and handling their special needs in data management and analysis I recognize common requirements and technologies for RDA. Furthermore I will disseminate the results and recommendations of RDA to the scientific communities and foster their realization.