e-Infrastructures & RDA
for data intensive science
Pre-RDA plenary workshops
22 September 2015, CNAM Paris, France

AGENDA

08:00	**Registration**				
09:00 - 10:00	**Opening Plenary Session**				*Amphi Paul Painleve*
09:00 - 09:05	Welcome & Workshop Introduction, **Hilary Hanahoe**, RDA Europe & Trust-IT Services Ltd				
09:05 - 09:35	**European Digital Single Market Strategy: changing the game with a service oriented e-infrastructure, Augusto Burgueño-Arjona,** Head of e-Infrastructures Unit, DG Communications Networks, Content and Technology - European Commission				
09:35 - 10:00	**Open Data is not enough, Mark Parsons,** RDA Secretary General				
10:00 - 17:30	**Parallel tracks**				
	Infrastructure for Understanding the Human Brain	**Data and computing infrastructures for open scholarship**	**Service orientation to data and high-performance computing infrastructures**	**Research Data infrastructures for Environmental related Societal Challenges**	**Data & Computing infrastructures for Global Linguistic Resources**
18:00 - 19:00	**Outcomes from each stream & wrap up Session** *Amphi Paul Painleve* *Moderated by **Carlos Morais-Pires**, Scientific Data Infrastructures, e-Infrastructures Unit, DG Communications Networks, Content and Technology - European Commission* *Intervention by **William L. Miller**, Advanced Cyberinfrastructure (ACI) Division, Directorate for Computer & Information Science & Engineering, National Science Foundation (NSF), US* *Interventions from each of the 5 streams rapporterus on main outcomes and recommendations*				
19:00 - 21:00	Networking Dinner & Entertainment				

Infrastructure for Understanding the Human Brain

10:00 - 11:30	**Session 1: Active Data Repositories and Federations** A core challenge in neuroscience is providing reliable, performant and scalable solutions with authentication, metadata support, persistent identifiers and data-type specific services. Here we will review some of the leading options for neuroscience data sharing repositories and options from other domains. Discussion will be around common requirements and potential solutions for deploying neuroscience data repositories across Europe and the world. A related Bird of a Feather session is planned during the RDA Plenary (https://rd-alliance.org/bird-feather-volumetric-imaging-services-vis-p6-...). • EGI Data Federations - Lukasz Dutka, Tiziana Ferrari, Yin Chen, Fabrizio Furano, EGI • Federation of HPC Resources - Fabien Delalondre, HBP/BBP • DVID: Distributed, Versioned, Image-Oriented Datastore - William Katz, Janelia Farms, HHMI
11:30 - 12:00	Coffee break
12:00 - 13:30	**Session 2: Multiscale Federated Data Mining** With the production of large amounts of imaging data comes the challenge of measuring and extracting features and properties of the brain from this data. Only by quantifying specific properties of the nervous system from this large-scale imaging data will it be possible to analyze, cluster, model and predict nervous system properties including cells, synapses and connectivity. This session will focus on the challenges of analyzing large datasets without moving them, machine learning, machine vision and other methods to analyze brain imaging data across scales. • Large-scale EM Connectome Reconstruction in the Adult Fly Brain - Stephen Plaza, Janelia Farms, HHMI • Images to Graphs for Inference - Will Gray, OpenConnectome

- Scaling Up Collaborative Circuit Reconstruction with CATMAID - Andrew Champion, Janelia Farms, HHMI
- BigNeuron: Large-Scale 3D Neuron Reconstruction from Optical Microscopy Images - Hanchuan Peng, Allen Institute
- neuGRID: an online environment for neuroimaging studies - Paolo Bosco, IRCCS Fatebenefratelli

13:30 - 14:30	Lunch
14:30 - 16:00	**Session 3: Knowledge Management and Search**

As large datasets are mined, analyzed and clustered, the key properties, relationships, categories and structure in the data is transformed into knowledge. As this knowledge is generated it is essential to have rich online databases and search engines to disseminate this knowledge, make it accessible and searchable. Data-driven ontologies can then be contrasted with expert domain knowledge in order to drive the knowledge discovery process forward. This session will explore a variety of tools developed for online knowledge management and search including defining, curating and maintaining ontologies, navigating structured information and search strategies for complex data. A related Bird of a Feather session is planned during the RDA Plenary (https://rd-alliance.org/bird-feather-controlled-vocabulary-and-ontology-...).

- Neuroscience Information Framework - Maryann Martone, NIF, INCF Node US
- The Development of a Neuroimaging Data Model (NIDM) from the INCF Data Sharing Task Force - JB Poline, UC, Berkeley, INCF Node France
- The Experimental Factor Ontology Utilities in Knowledge Discovery and Sharing - Sirarat Sarntivijai, EMBL

16:00 - 16:30	Coffee break
16:30 - 18:00	**Session 4: Online Collaboration Environments**

Collaboration is becoming a requisite part of modern science in many domains. New web technologies are enabling collaboration through data sharing, provenance tracking, reproducible workflows, communication, hyperlinks and online training. Scientific social networks are also emerging as an important tool for collaboration. This session will explore the factors enabling successful online collaboration and the pitfalls to avoid.

- The Human Brain Project Collaboratory - Jeff Muller, HBP/BBP
- The Synapse Collaborative Platform - Arno Klein, Sage Bionetworks
- What can be learned from the UK CARMEN project? - Leslie Smith, CARMEN, INCF Node UK
- A quick review of the Jupyter notebook - JB Poline, UC, Berkeley, INCF Node France
- The Loop Scientific Social Network - Kamila Markram/Fred Fenter, Frontiers/Loop Social Network

Data and computing infrastructures for open scholarship
<div align="right">**Amphi JB Say**</div>

10:00 - 11:30	**Session 1: Policies and principles of Open Scholarship**

- Introduction / EC perspective - Jarkko Siren, Project Officer, DG Communications Networks, Content and Technology, European Commission
- Policies to promote open science: evidence from OECD countries - Giulia Ajmonemarsan, Directorate for Science, Technology and Innovation, OECD
- Digital Infrastructure for Reproducibility in Science - William Gunn, Director of Scholarly Communications, Mendeley
- Data and computing infrastructures for open scholarship – a legal perspective - Thomas Margoni, Intellectual Property Law, University of Stirling

11:30 - 12:00	Coffee break
12:00 - 13:30	**Session 2: Linking up science: Interoperability aspects and services**

- Reminiscing about fifteen years of interoperability efforts - Herbert Van de Sompel, Digital Library Research & Prototyping Los Alamos National Laboratory, Research Library
- THOR – Towards seamless identifier interoperability across people, data, and research artefacts - Sunje Dallmeier-Tiessen, CERN, Scientific Information Services
- Moving content across the OpenAIRE infrastructure boundaries - Paolo Manghi, CNR - OpenAIRE technical manager
- Research Data Switchboard: find connections to your data - Amir Aryani, Australian National Data Service (ANDS)

13:30 - 14:30	Lunch
14:30 - 16:00	**Session 3: Tools, services, ideas as enablers of Open Science**

- The Open Science Framework and SHARE - Andew Sallans, Center of Open Science

- Data Packages: Decoupling Publishers and Consumers - Rufus Pollock, President and Founder, Open Knowledge
- Fostering the next generation of sustainable software and reproducible research practices in the scientific community - Karthik Ram, Quantitative ecologist, Berkeley Institute for Data Science, University of Berkeley, CA
- Zenodo : Making open data and open software an achievable reality for all scholars - Tim Smith, CERN Library Services

16:00 - 16:30	Coffee break
16:30 - 17:30	**Session 4: Training and education facilitating Open Science**

- Training for Open Science: the FOSTER approach - Eloy Rodriguez, Director of the University of Minho Documentation Services
- The LERU Roadmap and the potential for Open Data - Paul Ayris, Director of UCL Library Services
- EDISON – Coordination and cooperation to establish new profession of Data Scientist for European Research and Industry - Yuri Demchenko, Senior Researcher, University of Amsterdam

17:30 - 18:00	**Round the table discussion: Rethinking infrastructures for open scholarship**

- Jean-Pierre Bourguignon - ERC President
- Marie Farge – CNRS, Director of Research
- Gabi Lombardo – Science Europe, Senior Scientific Officer for the Social Sciences Scientific Committee
- Kathleen Shearer – Executive director of COAR (Confederation of Open Access repositories)

Service orientation to data and high-performance computing infrastructures   **Amphi Robert Faure**

10:15 - 10:30	**Welcome, Augusto Burgueño Arjona , Head of e-Infrastructures Unit, DG Communications Networks, Content and Technology - European Commission**
10:30 - 11:30	**A service orientation to High Performance & Distributed Computing (Part I) / Session I** *HPC and high throughput computing as a service* - Short keynote by Ed Seidel, Director National Center for Supercomputing Applications (US) Short presentations from leading European computing initiatives (5 minutes)

- PRACE Partnership for Advanced Computing in Europe – Alison Kennedy, PRACE Board of Directors, AISBL, EPCC (UK)
- ETP4HPC The European Technology Platform for High Performance Computing –Jean-Philippe Nominé, ETP4HPC Office, CEA (FR)
- Helix Nebula The Science Cloud - Bob Jones, Helix Nebula coordinator, CERN
- EGI European Grid Infrastructure – Tiziana Ferrari, Technical Director, EGI.eu (NL)
- INDIGO-DataCloud – Davide Salomoni, IndigoDataCloud project coordinator, INFN (IT)

Q&A

11:30 - 12:00	Coffee Break
12:00 - 13:30	**A service orientation to High Performance & Distributed Computing (Part II) / Session II** Short presentations form leading computing initiatives outside Europe (5 minutes)

- XSEDE Extreme Science and Engeneering Dicovery Environment – John Towns, XSEDE Project Director, NCSA (US)
- ComputeCanada – Mark Dietrich, ComputeCanada CEO (CA)
- RNP / Brazil – Leandro Ciuffo, RNP (BR)
- Data intensive science and data management challenges Panel *moderated by Karim Chine, Cloud ERA LTD*

13:30 - 14:30	Lunch
14:30 - 16:00	**Virtualizing infrastructures for Data / Session III** Data landscape, challenges and trends - Short keynote: Peter Wittenburg, Senior Advisor Data Systems, Max Planck Computing and Data Facility Short presentations from leading data initiatives

- EUDAT Collaborative Data Infrastructure - Rob Baxter, EPCC (UK)
- OpenAIRE – Paolo Manghi, OpenAIRE Technical Manager, CNR (IT)
- Phenomenal, an e-Infrastructure for Clinical Metabolomics Data – Kenneth Haug, PheNoMenal project coordinator, EBI (UK)
- Australian National Data Service, Enabling Discoverability and Reusability of Research Data – Amir Aryani, Project Manager, ANDS (AUS)
- DataONE Ensuring Access, Use and Reuse of Science Data – Rebecca Koskela, University of New Mexico, DataONE Executive Director (US)
- The National Data Service – John Towns, NCSA (US)

	• Research Organisation of Information and Systems - Hideaki Takeda, ROIS (JP)
	Short Q/A (10mn)
	Connecting sites, people and content / Session IV
	• GEANT Networking Technologies and Services – John Chevers, GEANT Chief Business Development Officer, DANTE (UK)
	• AARC Building the European AAI Ecosystem – Licia Florio, AARC project coordinator, GEANT (NL)
	• THOR Technical and Human Infrastructure for Open Research – Adam Farquhar, THOR Project Coordinator, the British Library, (UK)
16:00 - 16:30	Coffee Break
16:30 - 17:15	**Data Commons Panel** *moderated by Jean-Claude Guédon, University of Montreal*
17:15 - 18:00	**Workshop wrap-up** – Alan Blatecky

Research Data infrastructures for Environmental related Societal Challenges Amphi Jean Prouvé

10:00-10:15	**Welcome and Objectives:** Donatella Castelli, CNR ISTI; Odile Hologne, INRIA; Michel Schouppe, Policy Officer Climate action and Earth observation, DG Research & Innovation - European Commission
10:15-11:30	**Session 1 - Data infrastructure to address Climate change, marine and agricultural sustainability**
	• Data and computing infrastructures for climate modelling - Sylvie Joussaume, CNRS, Institut Pierre Simon Laplace
	• Closing the data/information gaps for marine policy: from local to global perspectives - Julian Barbiere, UNESCO-IOC
	• Data requirements for addressing environmental challenges faced by agriculture - Michèle Tixier Boichard, INRA
11:30-12:00	Coffee Break
12:00-13:30	**Session 2 - Operational solutions, best practice and trends in infrastructures across the globe (Part I)**
	• The Copernicus Climate Change Service of the EU - Baudouin Raoult, ECMWF
	• The European Marine Observation and Data Network (EMODnet) - Philip P.E. Weaver, Seascape Consultants Ltd
	• Building on iMarine for fostering Innovation, Decision making, Governance and Education - Donatella Castelli, CNR ISTI
	• DataONE: Facilitating Discovery, Access and Use of Scientific Data - William Michener, University of New Mexico
	• GEOSS Common Infrastructure and the Discovery and Access Broker - Stefano Nativi, CNR
13:30-14:30	Lunch
14:30-16:00	**Session 3 - Operational solutions, best practice and trends in infrastructures across the globe (continued)**
	• AgInfra – Information Infrastructure for agriculture: lessons learned in the agINFRA project - Johannes Keizer, FAO
	• INSPIRE and Environment Data in the EU - Andrea Perego, EC JRC
	• Australia's Integrated Marine Observing System: Data use and uptake. Can we close the loop? - Sebastien Mancini, IMOS – University of Tasmania, Australia
	• Environmental Related Data Intensive R&D Activities and Infrastructures in South Africa - Anwar Vahed, CSIR, South Africa
	• DIAS, Data Integration and Analysis System and agricultural decision support - Seishi Ninomya, Institute of Sustainable Agro-ecosystem Services, University of Tokyo, Japan
16:00-16:30	Coffee Break
16:30-17:20	**Session 4 – Panel and group discussion on Data and computing infrastructures for more social-aware science**
	• Hervé Caumont ,Terradue
	• Jean-Pierre Vilotte, Institut de Physique du Globe - CNRS & Belmont Forum
	• Gelsomina Pappalardo, CNR-IMAA & ESFRI ENV SWG chair
	• Michel Schouppe, DG Research & Innovation - European Commission
	• Augusto Burgueño-Arjona, Head of e-Infrastructures Unit, DG Communications Networks, Content and Technology - European Commission
17:20-17:30	**Summary and Conclusions**, Donatella Castelli, CNR ISTI; Odile Hologne, INRIA; Michel Schouppe , Policy Officer Climate action and Earth observation, DG Research & Innovation - European Commission

Data & Computing infrastructures for Global Linguistic Resources   Room 17 Access 21.2.32

10:00 - 11:30	**Session 1: Scientist Views on Rome and Paris Papers** • Dieter Van **Uytvanck Welcome & Introduction** • Bridget Almas - Perseus, Open Philology • Erhard Hinrichs **–** CLARIN • John McNaught - NaCTeM • Tatjana Gornostaja - Tilde
11:30 - 12:00	*Coffee Break*
12:00 - 14:30	**Session 2: Data Center and Infrastructure Views on Rome and Paris Papers** • Khalid Choukri – ELRA • Chris Cieri – LDC • Alastair Dunning - Europeana • Jane Hunter - eResearch Lab, University of Queensland • Panel Discussion
13:30 - 14:30	Networking Lunch
14:30	**Panel Discussion** (continued)
14:40 - 16:00	**Session 3: Humanities on Rome and Paris Papers** • Sandra Collins- National Library of Ireland • Beth Plale - Hathi Trust Research Center • Stelios Piperidis – ILSP • Panel Discussion
16:00 - 16:30	Coffee break
16:30 - 18:00	**Panel Discussion** • Presentation of Summary Messages • Panellist Intervention • Audience Interventions • Final Summary Messages