eSSENCE at the 11th IEEE international conference on e-Science, Aug 31-Sep 4 in Munich, Germany

by Carmen Medina, Uppsala University

Ludwig-Maximilian University in Munich

The annual IEEE eScience conference was held this year in Munich, Germany in the venues of the Ludwig-Maximilian University. A curious fact one can't stop noticing is the spelling of the term e-Science in the conference name, which, along the years, has been written with a hyphen (2010, 2011, 2013, 2014) or without a hyphen (2012, 2015), depending on the organizers. In this report I shall follow this year's "de-hyphened" version for consistency.

According to the organizers, the objective of this conference is to promote and encourage all aspects of eScience and its associated technologies, applications, and tools. Here eScience is considered in its broadest meaning as any technical, organizational or cultural aspect of the *digital revolution of research* that promotes innovation in collaborative, computationally or data-intensive research across all disciplines, throughout the research lifecycle.

Some of the application fields in the conference, where eScience has fundamentally transformed the way of doing research and has created specialized eScience disciplines, are:

- Arts, humanities (including Digital Humanities) and e-Social science
- Bioinformatics and e-Health
- Physical Sciences and Engineering
- Climate, Environmental & Earth Sciences

eScience itself has become a topic of specialized research, engineering and outreach activities, such as:

- Data science, data management and digital repositories
- Cyberinfrastructure, including novel hardware, software and services
- eScience in the cloud
- Education and eScience practice

Programme at a glance

The conference itself consisted of three intensive days with two parallel sessions morning and afternoon, and one or more keynote lectures during the day. There was also an evening poster session and a visit to the Leibnitz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities.

Two more days ware dedicated to specialized workshops: the *1st International Workshop on Interoperable infrastructures for interdisciplinary big data sciences* (IT4RIs 15); and the *1st Workshop on e-Science Research leading to negative results*.

During the last day there was the third plenary meeting of PLAN-E, the Platform of eScience Centers in Europe. The documentation of this meeting can be downloaded from their site: http://plan-europe.eu/.

There were 45 accepted papers to the conference (see APPENDIX 1) from a wide variety of disciplines and subjects. However, some research areas were heavily represented at the conference: eScience in the cloud or Cloud Computing, which includes many examples of eScience Services, generic and specialized; GPU computing, Workflows, and the expansive area of Citizen Science. The most common areas of application include the Humanities, where text analysis is a strong field; The Natural Sciences, with Life Sciences leading the application fields, followed by Medicine and Biology, Astronomy and Seismology. A special emphasis in this conference was made on applications for Climate and Environmental research, to which a whole day of sessions was dedicated.

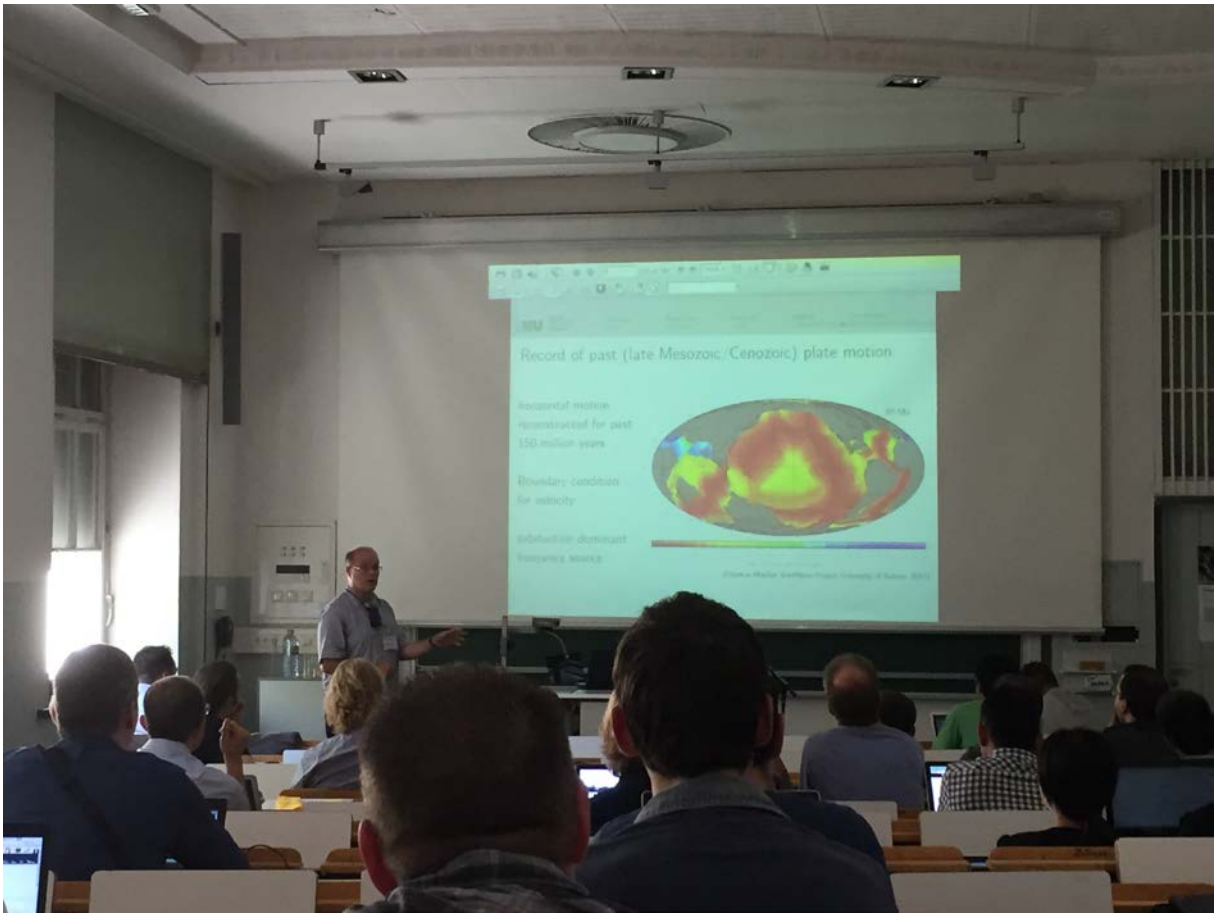
eSSENCE participation

Anders Hast, UU

There were two presentations from the eSSENCE community:

Anders Hast from UU, *Swedish eScience Education- a Graduate School in eScience*
Torbjörn Nordling, UU, *Identification of biomarkers and signatures in protein data*

Keynote Lectures



Hans-Peter Bunge, LMU

Hans-Peter Bunge, Department of Earth Sciences at the Ludwig-Maximilian University in Munich: *Data assimilation in global mantle flow models: theory, computational challenges and uncertainties in reconstructing global deep Earth structure back in time.*

Bunge does HPC modelling of problems in core, mantle and lithosphere dynamics. He has developed fluid dynamic inverse theory to reconstruct gross Earth structure back in time. The ability to extract the history of motion associated with large-scale structure that is imaged seismically in the Earth's interior, such as thermal plumes or cold, subducting oceanic slabs, is crucial to constrain the deformation processes of mantle convection. One can project gross Earth structure back in time by applying fluid dynamic inverse theory, based on a variational approach, in a global circulation model of the mantle. The uncertainties restrict the problem in practice, as our knowledge of deep Earth structure and its interpretation in terms of dynamic flow is necessarily limited.

Anton Frank, Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (LRZ), Garching bei München, Germany: *Environmental Computing – More than computing environmental models?*

Frank is the research coordinator at LRZ and supervisor of projects funded by third parties, including major infrastructure initiatives such as the PRACE joint supercomputing initiative. He coordinates LRZ's collaborations with earth and environmental scientists and is a member of the steering committee of several projects and initiatives in the environmental computing domain.

Most of the activities in environmental computing today rely – directly or indirectly – on modelling complex environmental phenomena and their interactions. Analysis of the modelling results on the scale of individuals, societies and regions is often uncovering previously hidden dependencies between theses systems, which need to be taken into account in planning, executing and sustaining different activities. Environmental computing can be seen as an emerging specialty that focuses on discovering these dependencies and producing *actionable knowledge* by optimally combining different models and data sources. This requires in depth analysis of the various feedback loops in the systems, and – equally importantly – presenting these relationships in a way that can be readily used both in different decision making processes as well as a foundation for further, cross-disciplinary research activities.

Sanna Sorvari, Science Coordinator, FMI, Nordic Atmospheric and Earth System Science RI Network: *European and International Collaboration in Climate and Environmental Research*

Sorvari has her scientific background in climate and Arctic research, paleoecology and limnology. In the recent years she has worked for establishing ESFRI Environmental Research Infrastructures at the national and European level (e.g. ICOS, ACTRIS, ANAEE. Sorvari has also facilitated the collaborative work of Enviromental RIs at the European level (Chair of the ENVRI stakeholders' Advisory Board) and at the international level (WP leader for COOPEUS-project, member of Belmont Forum's Env. data management action). She is also a member of H2020 Expert Advisory Board for RIs and e-RIs. In addition, she is working on Joint Programming, especially in JPI Climate as a co-chair of Working Group on Mowing towards Reliable Climate Predictions and as a member of the JPI Climate Management Board.

In her talk, Sorvari presented the work that is being done in the different Environmental Research Infrastructures (RI), which are defined as facilities, resources and services provided mainly from the scientific community to conduct top-level research. They are:

- Highly coordinated
- Open-access
- Long-term funded
- Single-sited, distributed and/or virtual
- Have well developed e-infrastructure components

Some examples are the European:
ICOS, IAGOS, EPOS, ESCAT-3D, EURO-Argo, EISCAT, and the Environmental RI cluster ENVRI (www.envriplus.eu) whose goals are sharing technologies and standards, sharing information technologies, developing policies, etc.

Julio Serje, UNISDR (United Nations Office for Disaster Risk Reduction): *Towards the use of super-computing resources in the Global Risk Model*

The UNISDR's mission is to:

- Coordinate
- Advocate
- Monitor Sendai FDRR (fSendai Framework for disaster risk reduction)
- Generate knowledge
- Inform the community
- Conduct platforms every two years
- Global assessment report (biannual)

Some of their activities:

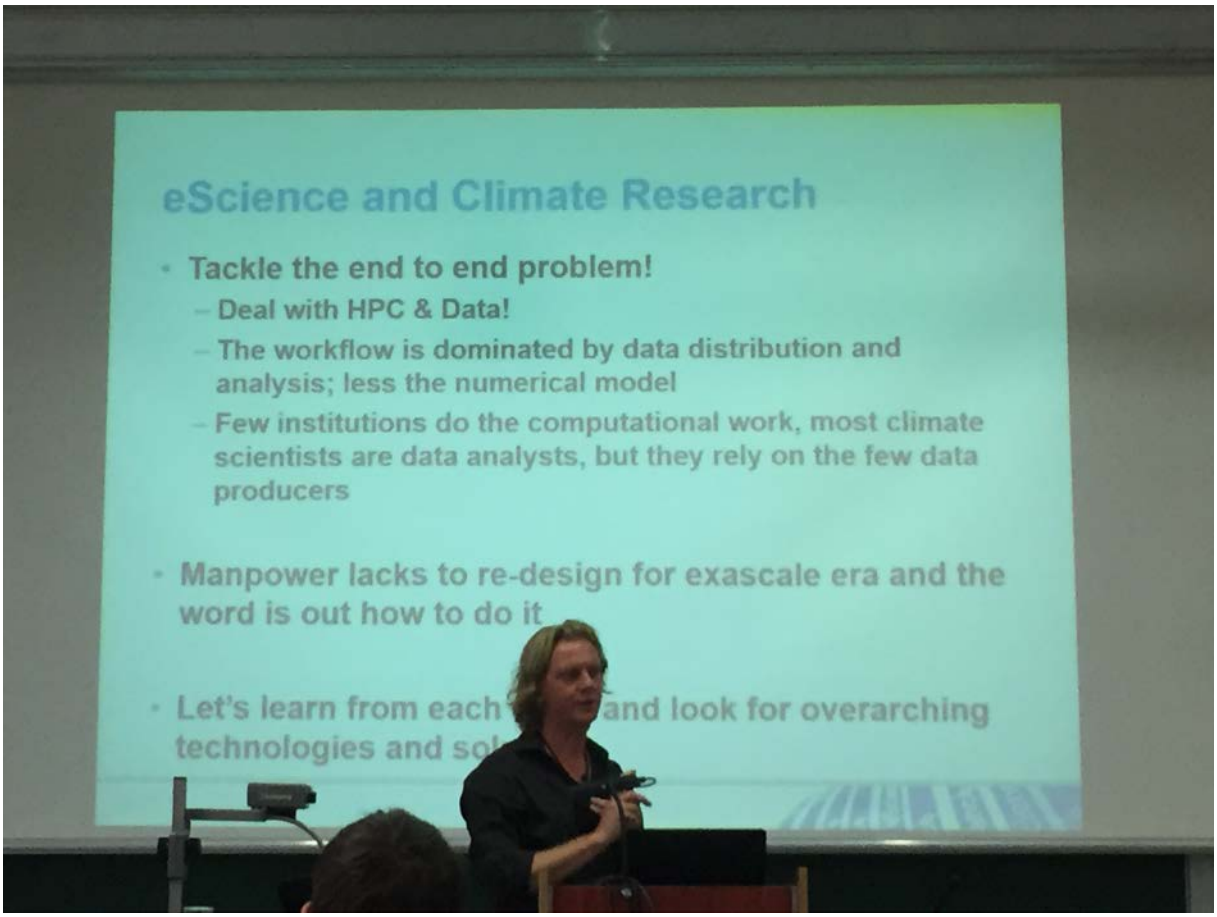
- Desinventar - Disaster information management system
- Global probabilistic risk assessment (probability of a certain level of losses)
  Components: hazard, exposure and vulnerability
- Hazard Models: Earthquake (1 million scenarios, 30km resolution), tsunami, flood (3 million scenarios at 1x1 km resolution), cyclone, storm surge, volcano (ash fall), drought.
- Exposure grids: 5x5 km grid with global coverage, 1x1 km grid for coastal areas
- Vulnerability modelling: susceptibility of exposed elements or assets to suffer damage or loss.

Prof. Dr. Wilco Hazeleger, Netherlands eScience Center (NLeSC), Amsterdam, NL/Wageningen University, Wageningen, NL/KNMI, De Bilt, NL, *Can e-Science transform climate science?*

Scientific understanding of current and past changes in the environment and predictions into the future, when human activities will severely impact the environment, involves extensive data research and computational studies.

Environmental research questions are characterized by processes which cover large ranges of spatial and temporal scales. Many relevant processes are described by known physical and chemical processes, often described by partial and ordinary differential equations. Although (numerical) solutions to these equations are often well known, the complexity and uncertainties in initial and boundary conditions, as well as interactions between subsystems, make environmental research extremely complex. Moreover, in many studies on recent and future changes, the behavioral human component becomes relevant which is not unequivocal described in mathematical formulations.

eScience technologies aid in addressing these scientific issues. New hardware architectures, which need to be energy efficient, are being used to make large scale computations at unprecedented detail possible. The large amounts of data allow for data-driven research



Wilco Hazeleger, NLeSC

approaches. In environmental studies traditional curve fitting and machine learning techniques are surpassed by advanced data assimilation techniques because of the dimension of the problem and the nonlinearities in environmental studies.

In his talk, Hazeleger illustrated these aspects with large scale computations of complex environmental systems, in particular the interaction between climate and the human activities at the global scale, the analysis of climatological and other environmental data and the fusion of the data in order to improve models and make skillful weather and climate predictions. He shows how the new eScience methodologies can transform environmental sciences beyond current capabilities.

Ewa Deelman, University of Southern California, Information Sciences Institute, *Challenges of Managing Scientific Workflows in High-Throughput and High-Performance Computing Environments*

Scientific workflows allow researchers to declaratively describe potentially complex applications that are composed of individual computational components. Workflows also include a description of the data and control dependencies between the components. The

Pegasus Workflow Management System developed at USC/ISI (http://pegasus.isi.edu) bridges the scientific domain and the execution environment by automatically mapping high-level workflow descriptions onto distributed resources. As part of this process, Pegasus may transform the workflow based on its properties and the target architecture. Optimizations and techniques have been developed to efficiently manage data and computations across heterogeneous computing environments. Pegasus can execute workflows on a laptop, a campus cluster, grids, and clouds. It can handle workflows with a single task or millions of tasks and has been used to manage workflows accessing and generating Terabytes of data. Upcoming, extreme-scale machines present challenges and opportunities to workflow management systems.

The eSSENCE delegation in Munich - Left to right: Anders Hast, Carmen Medina, Ingela Nyström and Torbjörn Nordling

Appendix 1 – Accepted papers

1. Alessio Sclocco, Henri E. Bal and Rob V. van Nieuwpoort.“Finding Pulsars in Real-Time”
2. Alexandru Uta, Andreea Sandu, Costache Stefania and Thilo Kielmann. “MemEFS: an Elastic In-Memory Runtime File System for eScience Applications”
3. Anders Hast, Michael Hanke and Hans Karlsson. “Swedish eScience Education- a Graduate School in eScience”
4. Anthony Nguyen, Andréa Matsunaga, Kohei Ichikawa, Susumu Date, Maurício Tsugawa and Jason Haga.“Deployment of a Multi-Site Cloud Environment for Molecular Virtual Screenings”
5. Ben van Werkhoven and Pieter Hijma “An integrated approach to porting large scientific applications”
6. Blesson Varghese, Javier Prades, Carlos Reano and Federico Silla. “Acceleration-as-a-Service: Exploiting Virtualised GPUs for a Financial Application”
7. Carlos Martinez-Ortiz, Marijn Koolen, Floor Buschenhenke and Karina van Dalen-Oskam. “Beyond the Book: Linking Books to Wikipedia”
8. Che Huang, Chawanat Nakasan, Kohei Ichikawa and Hajimu Iida. “A Multipath Controller for Accelerating GridFTP Transfer over SDN”
9. Chris Smowton, Georgiana Copil, Hong-Linh Truong, Crispin Miller and Wei Xing. “Genome Analysis in a Dynamically Scaled Hybrid Cloud”
10. Cristian Bonacic, Andres Neyem and Andrea Vasquez. “Live ANDES: Cloud-based application for citizen science and wildlife conservation”
11. David Wallom, Michel Drescher, Matteo Turilli, Diego Scardaci and Steven Newhouse. “Federating infrastructure as a service cloud computing systems to create a uniform e-infrastructure for research”
12. David Wallom, Timothy Booth, Andy Bowery, Ben Collier, Dawn Field, Philip Kershaw, Anurag Priyam and Yannick Wurm. “Facilitiating Environmental ‘Omics research through Desktop as a Service”
13. Eleni Mina, Mark Thompson, Kristina Hettne, Eelke van der Horst, Rajaram Kaliyaperumal, Katy Wolstencroft, Marco Roos, Barend Mons and Willeke van Roon-Mom.“Multidisciplinary collaboration to facilitate hypotheses generation in Huntington’s Disease”
14. Hrachya Astsatryan, Zaruhi Petrosyan, Rita Abrahamyan, Anna Shahnazaryan, Vladimir Sahakyan, Yuri Shoukourian and Hamlet Melkonyan. “WRF-ARWModel for the Prediction of High Temperatures in South and South East Regions of Armenia”
15. Jai Dayal, Karsten Schwan, Matthew Wolf, Greg Eisenhauer, Jay Lofstead, Hasan Abbasi and Scott Klasky. “SODA: Science-driven Orchestration of Data Analytics”

16. James Wetter, Ozgur Akgun, Adam Barker, Martin Dominik, Ian Miguel and Blesson Varghese. “Cloud-based e-Infrastructure for Scheduling Astronomical Observations”

17. Jane Hunter, Imran Azeezullah, Nigel Ward, Ross Barker, Tung-Kai Shyy, Robert Stimson, James Dentrinos, Gerson Galang, Stewart Mcadam-Wallace and Chris Pettit. “Enabling SDMX-based Retrieval and Spatio-statistical Analysis of National Census and Related Datasets”

18. Janneke Van Der Zwaan, Inger Leemans, Erika Kuijpers and Isa Maks “HEEM, a Complex Model for Mining Emotions in Historical Text”

19. Jeremy Cohen, Chris Cantwell, David Moxey, Jeremy Nowell, Peter Austing, Xu Guo, John Darlington and Spencer Sherwin.“TemPSS: A service providing software parameter templates and profiles for scientific HPC”

20. Jie Xie, Michael Towsey, Philip Eichinski, Jinglan Zhang and Paul Roe “Acoustic feature extraction using perceptual wavelet packet decomposition for frog call classification”

21. Jisk Attema, Bert Heusinkveld, Reinder Ronda, Gert-Jan Steeneveld and Bert Holtslag. “Summer in the City: Forecasting and Mapping Human Thermal Comfort in Urban Areas”

22. John Walsh and Jonathan Dukes. “Application Support for virtual-GPGPUs in Grid Infrastructures”

23. Karolina Vukojevic-Haupt, Florian Haupt, Frank Leymann and Lukas Reinfurt. “Bootstrapping Complex Workflow Middleware Systems into the Cloud”

24. Kary Ocaña, Daniel de Oliveira, Vítor Silva and Marta Mattoso. “Data Analytics in Bioinformatics: Data Science in Practice for Genomics Analysis Workflows”

25. Ketan Maheshwari, Justin Wozniak, Timothy Armstrong, Daniel S. Katz, T. Andrew Binkowski, Xiaoliang Zhong, Olle Heinonen, Dmitry Karpeyev and Michael Wilde. “Porting Ordinary Applications to Blue Gene/Q Supercomputers”

26. Kristopher Zarns, Travis Desell, Sergei Nechaev and Archana Dhasarathy. “Searching the Human Genome for Snail and Slug With DNA@Home”

27. Kyle Chard, Jim Pruyne, Ben Blaiszik, Rachana Ananthakrishnan, Steven Tuecke and Ian Foster. “Globus Data Publication as a Service: Lowering Barriers to Reproducible Science”

28. Kyle Goehner, Travis Desell, Susan Ellis-Felege, Rebecca Eckroad, Leila Mohsenian, Paul Burr, Nicholas Caswell and Alicia Andes. “A Comparison of Background Subtraction Algorithms for Detecting Avian Nesting Events in Uncontrolled Outdoor Video”

29. Malcolm Atkinson, Michele Carpene´, Emanuele Casarotti, Steffen Claus, Rosa Filgueira, Anton Frank, Michelle Galea, Andre Gemuend, Heiner Igel, Iraklis Klampanos, Amrey Krause, Lion Krischer, Siew Hoon Leong, Federica Magnoni, Jonas Matser, Alberto Michelini, Horst Schwichtenberg, Alessandro Spinuso and Jean-Pierre Vilotte. “VERCE delivers a productive e-Science environment for seismology research”

30. Natalia Villanueva-Rosales, Luis Garnica Chavira, Nicholas del Rio and Deana Pennington. “eScience through the Integration of Data and Models: A Biodiversity Scenario”

31. Nicholas Hazekamp, Joseph Sarro, Olivia Choudhury, Sandra Gesing, Scott Emrich and Douglas Thain. “Scaling Up Bioinformatics Workflows with Dynamic Job Expansion: A Case Study Using Galaxy and Makeflow”

32. Patrícia Cavoto, André Santanchè, Régine Vignes Lebbe and Victor Roth Cardoso. “FishGraph: A Network-Driven Data Analysis”

33. Richard Grunzke, Jens Krüger, Sandra Gesing, Sonja Herres-Pawlis, Alexander Hoffmann, Alvaro Aguilera and Wolfgang E. Nagel. “Managing Complexity in Distributed Data Life Cycles Enhancing Scientific Discovery”

34. Romulo Goncalves, Milena Ivanova, Foteini Alvanaki, Jason Maassen, Kostis Kyzirakos, Oscar Martinez-Rubi and Hannes Muhleisen. “A round table for multi-disciplinary research on Geospatial and Climate Data”

35. Rosa Filgueira, Amrey Krause, Malcolm Atkinson, Iraklis Angelos Klampanos, Alessandro Spinusso and Susana Sanchez-Exposito. “dispel4py: An Agile Framework for Data-Intensive eScience”

36. Rudolf Mayer and Andreas Rauber. “A quantitative study on the re-executability of publicly shared scientific Workflows”

37. Ryan Chard, Kyle Chard, Kris Bubendorfer, Ravi Madduri and Ian Foster. “Cost-Aware Resource Provisioning for Cloud-based Scientific Workflows”

38. Ryan McKenna, Vivek Pallipuram, Rodrigo Vargas and Michela Taufer. “From HPC Performance to Climate Modeling: Transforming Methods for HPC Predictions Into Models of Extreme Climate Conditions”

39. Samuel Barbosa, Roberto M. Cesar-Jr and Dan Cosley. “Using Text Similarity to Detect Social Interactions not Captured by Formal Reply Mechanisms”

40. Sarah Berenji Ardestani, Carl Johan Håkansson, Erwin Laure, Ilja Livenson, Pavel Straňák, Emanuel Dima, Dennis Blommesteijn and Mark van de Sanden. “B2SHARE: An Open eScience Data Sharing Platform”

41. Simon Woodman, Hugo Hiden, Mark Turner, Stephen Dowsland and Paul Watson. “Monitoring of Upper Limb Rehabilitation and Recovery after Stroke: an Architecture for a Cloud-based Therapy Platform”

42. Torbjörn Nordling, Narendra Padhan, Lena Claesson-Welsh and Sven Nelander. “Identification of biomarkers and signatures in protein data”

43. Yolanda Gil, Felix Michel, Varun Ratnakar, Matheus Hauder, Christopher Duffy and Paul Hanson. “A Task-Centered Framework for Computationally-Grounded Science Collaborations”

44. Yuwei Wang, Ze Luo, Yan Xiong and Baoping Yan.“Discovering loose group movement patterns from animal trajectories”

45. Zaoxing Liu, Nikita Ivkin, Lin Yang, Mark Neyrinck, Gerard Lemson, Alexander Szalay, Vladimir Braverman, Tamas Budavari, Randal Burns and Xin Wang. “Streaming Algorithms for Halo Finders”