Policy Brief

Community engagement: The 'last mile' challenge for European research e-infrastructures

Dimitrios Koureas[‡], Christos Arvanitidis[§], Lee Belbin[|], Walter G. Berendsohn[¶], Christian Damgaard[#], Quentin John Groom[¤], Anton Güntsch[¶], Gregor Hagedorn[«], Alex Hardisty[»], Donald Hobern[^], Arnald Marcer[ˇ], Daniel Mietchen[¡], David R Morse[ˀ], Matthias Obst[ˁ], Lyubomir Penev[₵], Lars B Pettersson[ℓ], Soraya Sierra[ᔆ,ᴾ], Vincent Stuart Smith[‡], Rutger Aldo Vos[ᔆ]

‡ Natural History Museum, London, United Kingdom
§ Hellenic Center for Marine Reaearch (HCMR), Heraklion Crete, Greece
| The Atlas of Living Australia, Carlton, Australia
¶ Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Berlin, Germany
# Aarhus University, Silkeborg, Denmark
¤ Agentschap Plantentuin Meise, Meise, Belgium
« Museum für Naturkunde Berlin, Berlin, Germany
» Cardiff University, Cardiff, United Kingdom
^ GBIF, Copenhagen, Denmark
ˇ CREAF, Cerdanyola del Vallès, Spain
¡ EvoMRI Communications, Jena, Germany
ˀ School of Computing and Communications, The Open University, Milton Keynes, United Kingdom
ˁ Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden
₵ Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria
ℓ Biodiversity Unit, Department of Biology, Lund University, Lund, Sweden
ᔆ Naturalis Biodiversity Center, Leiden, Netherlands
ᴾ Tint Hue, Amsterdam, Netherlands

Abstract

Europe is building its Open Science Cloud; a set of robust and interoperable e-infrastructures with the capacity to provide data and computational solutions through cloud-based services. The development and sustainable operation of such e-infrastructures are at the forefront of European funding priorities. The research community, however, is still

reluctant to engage at the scale required to signal a Europe-wide change in the mode of operation of scientific practices. The striking differences in uptake rates between researchers from different scientific domains indicate that communities do not equally share the benefits of the above European investments. We highlight the need to support research communities in organically engaging with the European Open Science Cloud through the development of trustworthy and interoperable Virtual Research Environments. These domain-specific solutions can support communities in gradually bridging technical and socio-cultural gaps between traditional and open digital science practice, better diffusing the benefits of European e-infrastructures.

Keywords

e-infrastructure, Virtual Research Environment, Open Science Cloud, Europe, user engagement, research infrastructure

The era of open digital science

We are entering a new era of Open Digital Science where e-infrastructures, Web-based services and the globalisation of the scientific community are paving the way towards scientific progress founded on collaborative and data-intensive research. These technologies facilitate dynamic, open, transparent, democratic and replicable research. Europe strives to address societal challenges in sectors including Health, Energy and the Environment, and has acknowledged that an efficient way of doing so is by investing in innovative scientific research and by linking the outcomes to industry, policy making and society. The European roadmap for tackling these challenges includes, amongst others, strengthening of the European Research Area (European Commission 2012) and promoting a Digital Agenda (European Commission 2010) for Europe. It is, however, at the intersection of these two continental-scale efforts where the much needed cross-disciplinary innovation can emerge.

Through strategic funding schemes (e.g. the Research Infrastructures part of the Horizon 2020 work programmes), Europe invests in research e-infrastructures that give researchers access to the European Open Science Cloud (EOSC). The objective is to promote innovation and integration of a still highly fragmented European research environment. During 2014 and 2015, the European Commission (EC) invested over €170 million through Horizon 2020 (the European Union Research and Innovation programme 2014-2020) on development and integration of e-infrastructures. That is on top of €572 million invested between 2007 and 2013 (33.35% of the FP7 budget for Research Infrastructures) (European Commission 2015b). A question arises about the extent to which these investments have impacted the modus operandi of scientific practice across Europe and across different research communities.

Researchers are still reluctant to engage

At the heart of European e-infrastructure developments is the provision of robust, reliable and interoperable services that generate global solutions for data sharing and preservation, high performance and cloud computing, user-authorisation and authentication. This set of core services forms the backbone that supports high-throughput, collaborative and data-driven scientific research. Despite European investments, researchers still find it difficult to discover and use these services. Many services are too technical, do not provide easy-to-use interfaces and cannot easily be integrated into the majority of day-to-day research practices. Researchers often need to switch between different digital environments and rely on manual work to structure and transform data to conform to the specifications of each service. Furthermore, research is a global enterprise involving researchers and infrastructure elements from all parts of the world. The problem cannot be solved by European infrastructures alone. The solution requires investing in consistent and efficient service interfaces, internationally coordinated specialization and large scale cooperation.

Not all research communities are 'created equal'

The lack of a seamless framework supporting the entire digital research lifecycle hinders adoption of existing digital solutions and reduces their enabling benefit to science. A recent survey (European Commission 2015a) showed that the main barriers (with >80% totally and partially agreeing) for researchers to engage with practices of 'Science 2.0' (The term 'Science 2.0' has been broadly replaced in recent EC policy documents with the term 'Open Science'), are uncertainties about: (i) quality assurance, (ii) attribution (receiving credit for work), (iii) integration between different infrastructure components, and (iv) limited awareness of 'Science 2.0' and its implications for research. Usage statistics from developed science-wide e-infrastructures show that the above barriers are equally preventing uptake of e-infrastructures by different communities of practice. For example, the European Grid Infrastructure (a flagship initiative that delivers integrated computing services to European researchers) announced (Dec 2015) a user base of 35,959 (European Grid Infrastructure 2015), with scientists from the physical sciences accounting for ca. 47.2%, scientists from the biological sciences ca. 4.3%, earth scientists 1.4% and humanities ca. 3.6% of the total user base. These striking discrepancies in uptake rates among researchers from different disciplines, suggest that different e-infrastructure audiences require different approaches. They need approaches that enable different users to realise the possibilities inherent in e-infrastructures and build trust relationships between scientific communities and e-infrastructure providers.

There is a significant disconnect between the rates of technological progress in the development of research e-infrastructures and uptake by researchers. To mitigate this risk, it is imperative that e-infrastructure services are accessible through consistent easy-to-use interfaces, which provide integrated and ubiquitous access. These interfaces should have the same simplicity and maturity as the consumer-oriented Web applications we are

already familiar with. Intuitive user interface experience, seamless data ingestion, and collaboration capabilities are among the features that could empower users to better engage with provided services. For the investments in technological development to achieve their full potential, however, communities need to address significant challenges, also from a socio-cultural aspect. Investing in both formal and professional training across science disciplines would improve the capacity of communities to engage with provided services.

The need to respect diversity and continuously developing needs

Issues relating to accessibility of data, data annotation, collaboration or even publishing norms are often perceived in completely different ways within different disciplines. For instance, researchers working on genomics, physics or astronomy have long appreciated the value of data sharing. By nurturing a culture of shared physical and computational infrastructure, open-source software and open data, they have embraced the principles of open science. Other disciplines have less open traditions and require social impulses as well as technological collaboration environments to stimulate the adoption of open practices. Well-implemented services have improved data sharing in communities that traditionally were lagging behind. The development and support of Dryad for example, has provided a robust and trusted solution for sharing datasets in natural history, botany, zoology and ecology (to illustrate a few). It has enabled the development of a new generation of data publishing scientific journals (e.g. Scientific Data, GigaScience and the Biodiversity Data Journal). Despite the above discrepancies, all communities recognise that data quantities are exploding and that in order to fully exploit the potential associated with this data wave, a gradual shift in their traditional scientific practices is needed.

The Science Europe association in its response to the Science 2.0 European Commission consultation recommends that Europe needs to: "Recognise that research communities are developing Science 2.0 practices organically and that they are best placed to explore which of these contribute to the advancement of their discipline". This recommendation underlines the need to continue supporting different scientific communities in developing the required technical and socio-cultural research environments, including adaptation to generic e-infrastructures as community-driven initiatives.

The 'last mile' challenge for research e-infrastructures

To capitalise on earlier investments, it is crucial that we incentivise and support research communities to better understand the benefits and to explore the opportunities presented by e-infrastructures. The challenge starts with identifying how professionals work within their research communities and understanding the processes that lead to innovation becoming embedded into common practice (May and Finch 2009). Providing e-infrastructures that seamlessly couple with the work practices of a particular profession

requires layers that abstract from a technical level and use the language of each profession. Such layers are typically web-based applications that address elements across the lifecycle of data and research, i.e. data mobilisation and discovery, experimentation, analyses, publication, and open collaboration. Such "Virtual Research Environments" (VREs) should act as intuitive and responsive interventions between researchers and core services. VREs should maintain domain specificity of data, standards and workflows created by the relevant communities. These components are needed for the proper operation of their professional activities and for harnessing the underlying capabilities and capacities.

VREs have to be offered in combination with processes to help implement new practices that are aligned with Open Digital Science and to foster interdisciplinarity. In the long run, VREs can grow into trustworthy discipline-specific 'commons' that provide technical, social and governance frameworks. These discipline-specific commons need to be compatible with each other and ultimately should lead to the gradual formulation of a science-wide accepted e-infrastructure commons, as described by the e-Infrastructures Reflection Group (e-IRG) (e-Infrastructures Reflection Group 2013). As such, the role of the VREs is not to replace or replicate the backbone European e-infrastructure, but rather to build on top of it to complete the research infrastructures value chain.

VREs have already proved in principle that they can drive and underpin a sustained paradigm shift in the way research communities manage, compute and publish data in open collaborative environments. For instance, the Biodiversity community (a traditionally reserved community regarding aspects of e-science) has more than 7,000 researchers actively engaging with virtual services through the efforts of EU-funded projects (CORDIS 2014).

The importance of VREs to the challenge of engaging researchers with backbone e-infrastructure services is analogous to the 'last mile' challenge in telecommunications or transportation, where the marginal cost and complexity of 'connecting' end-users to the backbone (core) e-infrastructures (e.g. cloud high-performance computing or data services) is high when compared to the core infrastructure itself. These costs vary based on the distance of end-users from the backbone. The technical and socio-cultural 'distance' of different research-communities from the core e-infrastructures determines the level of investment that is required to bridge the 'last mile'. As such, this 'last mile', is the critical section which needs to be bridged in order to disrupt the current mode of science functioning and its daily practice, since it lowers the barriers for accessing computational capacity, and improves transparency and efficiency. Thus, 'last mile' investments (i.e. VREs) are as integral to the development of research e-infrastructures as the operation of the European Open Science Cloud. Indeed, without VRE's the value of the European Open Science Cloud cannot be realised for most research communities.

The role of research infrastructure funding policies

In a report from the Research Data Alliance (RDA) Europe (RDA Europe 2014), it is argued that for a "truly effective data-sharing system", 5% of the total global research budgets would be required. That can be calculated at over €10 billion a year. It should be expected that a significant portion of this funding needs to be invested in developing, supporting and sustaining cross-domain user-engagement mechanisms. For the ecosystem of digital research services to be fully effective across the scientific communities, it is imperative that e-infrastructure operators and funders continue to invest in the development of VREs. To achieve maximum return on investment, European (European Commission and national) funding programmes need to promote a balance between the backbone and discipline-specific data e-infrastructures. In the past, VREs have been supported through both national (e.g. JISC in the UK, SURF in the Netherlands) and European (e.g. Framework Programmes) resources. In the absence of a common European e-infrastructure backbone, VREs were previously developed with limited access to persistent backbone services. The latest advances at both the technical and governance level of European core infrastructures have completely reformed the European e-infrastructure landscape, providing opportunities for more efficient and parallelized development of the required domain-specific virtual environments. To efficiently develop the next generation of VREs, it is crucial that funders, VRE operators and user communities work together in support of a balanced model between core infrastructure development and domain-specific solutions.

User communities need to be able to: (i) articulate and communicate their community-specific needs in regards to data and services, and (ii) translate these needs into clear functional requirements that will drive the development of VREs. VRE operators need to: (i) develop VREs looking beyond the ephemeral timeframes of project-based approaches, (ii) invest early in building public-public and public-private partnerships that ensure sustainability and, (iii) robustly link VREs with existing underlying e-infrastructure, building on top of available backbone services.

Funders need to (i) further acknowledge the pivotal role of VREs in support of user community engagement and, (ii) develop, with a particular eye to long-term sustainability, dedicated VRE funding programmes with targeted calls to discipline-specific communities. The definition and observance of key indicators will facilitate continuous assessment of the communities' progress towards the sustainable uptake of e-infrastructure services. These indicators should be informed by metrics such as (a) overall user buy-in that is taking into consideration quantitative (number of users) and qualitative (best practices) aspects, (b) level of integration of domain-specific VREs with European core e-infrastructures, and (c) proven capacity to develop and sustain domain solutions.

The Digital Agenda for Europe is setting out ambitious goals, which aim, among others, to transform science, making research open, global and collaborative. For the European Research Area, however, to fully benefit from investments in e-infrastructures, it is critical that no community of practice falls behind. Though previous practices for developing Virtual Research Environments need to be revisited (to better align with the overarching

implementation strategy for the Digital Agenda for Europe), we hereby highlight their integral role in the development of a robust ecosystem of e-infrastructures and services in support of the Europe 2020 strategy.

References

- CORDIS (2014) VIBRANT: New virtual research communities to create and share data on biodiversity. http://cordis.europa.eu/result/rcn/150383_en.html
- e-Infrastructures Reflection Group (2013) White Paper. 9. http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf/ce8a2253-aebd-4cbe-9a93-4709a1166214
- European Commission (2010) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A Digital Agenda for Europe. http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52010DC0245
- European Commission (2012) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. http://ec.europa.eu/research/era/pdf/era-communication/era-communication_en.pdf
- European Commission (2015a) Validation of the results of the public consultation on Science 2.0: Science in Transition. http://www.eesc.europa.eu/resources/docs/validation-of-the-results-of-the-public-consultation-on-science-20.pdf
- European Commission (2015b) FP7 Research Infrastructures. https://ec.europa.eu/research/fp7/index_en.cfm?pg=infra. Accession date: 2016 7 12.
- European Grid Infrastructure (2015) Discipline Metrics Report. URL: http://operations-portal.egi.eu/metrics/disciplineMetricsReports/discipline/2015-12?disciplineId=1
- May C, Finch T (2009) Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory. Sociology 43 (3): 535-554. DOI: 10.1177/0038038509103208
- RDA Europe (2014) The Data Harvest Report – sharing data for knowledge, jobs and growth. https://rd-alliance.org/data-harvest-report-sharing-data-knowledge-jobs-and-growth.html