



GEDE Workshop

Data Sharing a High Priority

Urgent Need to Act

14. April 2020

Online-Meeting

(Zoom Meeting – URL to be announced)

The RDA Europe 4 Project is coming to its end and with it also the funding support for RDA GEDE¹. It is therefore an excellent moment to reflect on what we did and on what a useful place for RDA GEDE in the evolving landscape could be which is dominated in Europe by discussions about the nature and purpose of EOSC. Our plan was to organise a face-2-face workshop in Brussels, but due to the Corona Virus epidemic we had to finally cancel this plan and turn to an online meeting. Of course, we will lack the social dinner and the more inspiring life interactions, but we should not postpone all activities.

Agenda and Keynote Speakers

The meeting will start at 9.00 and end at 16.00 CET with a lunch break around 12.30. We will have 4 sessions with well-known keynote speakers.

- In session 1 we will give a short look back on RDA GEDE
- In session 2 three domain scientists working at the cutting-edge in their domains (Matthias Scheffler – material science, Andre Dekker – medical science, Arturo H. Ariño – biodiversity science) will present their expectations about the development of data-driven science in the coming decade and their requirements with respect to an EOSC eco-system. A technologist Andreas Rauber (tbc) will reflect on these expectations and present his view on how technology will develop and support these wishes.
- In session 3 Karel Luyben (Chair EOSC EB) will discuss how EOSC will respond to the needs of cutting-edge science.
- In session 4 RDA Foundation's new RDA Europe entity and GEDE will present their views about their possible roles in the evolving landscape and how they intend to respond to the challenges.

¹ RDA GEDE received 20.000 € from the RDA Europe project to support its activities from 2017 on.

Goals of the Workshop

RDA Europe and RDA GEDE – each with a different focus – have served as useful interaction platforms for bringing together the DOers in advancing European data science. They have engaged in numerous discussions unimpeded by biases regarding funding opportunities. Thus, they have been able to contribute in constructive and impartial ways to the EOSC process. The goals for the workshop therefore are the following:

- contributing to shaping the core of EOSC to meet the foreseen challenges,
- identifying the role and importance of existing (research, e-) infrastructures to meet the challenges,
- understanding the evolving landscape of data infrastructures (incl. ERICs, EOSC LE, etc.) and positioning RDA EU and a follow-up to RDA GEDE in the next decade.

Relevant stakeholders from the ESFRI initiatives (ERICs, Clusters), e-Infrastructures, major research organisations, the EOSC process and related initiatives will be invited to participate and engage in cooperative discussions. Since these activities need to be closely synchronised with activities in other regions, we will invite participants from other continents as well.

Background and Motivation

In her speech at the Davos world economic forum, the new president of the EC, Ursula von der Leyen, presented “data” as a high priority for her organisation. This priority is warranted by the huge investments planned by the EC, by some member states and also in other regions of the world in the coming decade. This mission requires to step away from our usual activities, to think about innovative approaches and to be ready to take some radical decisions. The mission can best be described by the following high level questions

- ***Will European researchers be in the position to play a leading role in the exploitation of the data being generated for the benefits of society and economy?***
- ***Will European stakeholders be ready to deploy the right structures and basic sociological and technological mechanisms to enable researchers to carry out cutting-edge research?***

Billions of smart devices will create continuous high-resolution data streams leading to trillions of digital artefacts. Our understanding is growing that these quantities and their inherent complexity require radically innovative approaches. We can identify three major themes to be addressed: (1) the extraction of knowledge from combining data and methods in smart ways, (2) the smart combination of formally represented knowledge to derive conclusions and (3) the deployment of an eco-system of infrastructures to foster efficiency and stability of our work. There is a growing agreement that we cannot address these challenges by manual hunts for data, but we need to engage smart software agents that can interact with self-contained digital objects. Without ignoring the need to keep proven research paradigms, we still need to ask the question what is required to make necessary big changes happen in the coming decade. Obviously, the basis of a highly automated data domain will imply persistent identification and rich and semantically typed metadata, as recommended by the FAIR Principles and as suggested by various RDA group results.

Europe has already made enormous investments toward building awareness about the challenges of data-driven research. The ESFRI process, which was carried out mainly in the last decade, has already led to structuring the discussions in many disciplines, broadening awareness amongst researchers, improving the quality of data management, establishing community-inspired workflows and standards such as community-wide accepted metadata schemas, and much more. In this massive investment and engagement across the board, the scenario for future European research is far ahead of that in other regions. The e-Infrastructures helped us to extend knowledge about basic infrastructural mechanisms such as AAI, secure cloud services etc. It should be noted, though, that cloud technologies are not a

panacea. Some basic technologies will be replaced, for example, by quantum computing, and do not address the fundamental issue of improving interoperability of maintaining referential stability over many decades.

At the present moment, we are at a crossroads. We need to enter a more compelling phase of consensus building to maintain a leading position for European research. Short-term solutions to overcome cross-disciplinary boundaries, such as creating yet another aggregating portal or cloud will not be sufficient. We urgently need to converge on the radically innovative structures and mechanisms which are needed to shape the integrated and sustainable data and service landscape which is ultimately the goal of EOSC. Modernization only starts when the new structures and mechanisms are designed and implemented and add value to respond to the Global Societal challenges which is the mission of RDA GEDE. With earlier cross-disciplinary work on persistent identifiers, citation aspects, the role of repositories, the usefulness of blockchains in science and FAIR digital objects RDA GEDE could already contribute to this mission.