27. 06 GEDE-DO Meeting Notes

Participants: Peter Wittenburg, Alex Hardisty, Alexander Vasilenko, Attila Kreiter, Birger Jerlehag, Brigitte Hausstein, Carole Goble, Christoph Quix, Daniel Bangert, David Novak, Dimitris Koureas, Eoghan O Carragain, Eva Soroli, Holly Wright, Joe Padfield, Jose Lopez Coronado, Judit Gardos, Karolj Skala, Lucy Bastin, Maggie Hellström, Manfred Kohler, Markus Fiebig, Milena Stefanova, Nevena Veljkovic, Nick Juty, Oya Beyan, Karl Presser, Raphael Ritz, Robert Pergl, Robertus Erdelyi, Stian Soiland-Reyes, Ulrich Schwardmann, Wolfgang Kuchinke, Wouter Addink, Zsuzsanna Szeredi, Frank Toussaint

GEDE was set up by RDA Europe to include the ESFRI practitioners in relevant discussions, since they are often not heard. The work on the usage of PIDs was good try-out case and it was finished in 2017 (<https://zenodo.org/record/1116189>) . GEDE is a voluntary bottom-up platform to interact and the co-chairs are pace makers and moderators.

(Please, also have a look at the attached chat log)

1. Purpose of the meeting:

This meeting was intended as a warming up to start the real content interaction in September. Basis is the EUDON proposal for a COST Action, where we expect a decision to be made in November. Independent of the COST decision, we will start the interaction on DOs, since there is a lot of ongoing work in various initiatives, pressure on research infrastructures to make advancements in their infrastructure work and code is being developed requiring proper specifications. We need to understand what the current practice in the different Research Infrastructures is and how DOs may fit into their work. Also the plans for EOSC put pressure on all of us to come up with suggestions.

2. Agenda

The agenda was accepted as suggested (see slides).

3. Admininstration Update

Please check the presentation slides for information on email lists and the document share. RDAEU4 is the new RDA EU project since 1st of March and although GEDE is mentioned in the DoA we still lack a clear statement on funding support for GEDE. Nevertheless, we will go on and are happy for the help from V&V, here in particular of Zsuzsanna, and MPCDF.

GEDE-Share - DO-Documents

This should be used to store all documents that are relevant for our discussions about DOs. This could be strategy notes, white papers, specification notes, position papers or so. Everyone should feel free to send it to Zsuzsanna and/or Peter and we will upload it to the Share. We should not accept PR papers for concrete technology, but can refer to it.

If EUDON will be accepted we will see how we have to change the whole setup.

People should regularly have a look into the share for new documents, we will from time to time send an email to point to new material.

4. EUDOn overview

COST projects are meant to establish elaboration forums and foster capacity building.

EUDON 3 pillars:

1. Global discussion on DO brought in a form more useful for RIs
2. Within and beyond Europe there are several organizations and forums where these issues are being discussed. We need to synchronize activities and see how to go forward.
3. Hard to find money for networking, to work based on consensus. EUDOn (and GEDE) are platforms to do so.

Submitted: last cut of date.

Evaluation: mid November

Success rates: only about 5%, but we have a good proposal

The proposal revolves around the main challenges of bringing all stakeholders to the table

1. Research coordination objectives:
   1. Collectively carry out SWAT/GAP analysis
   2. Propose specific approaches
   3. work on DO architecture
   4. Stimulating and contribute capacity for the specification work
2. Capacity building objectives
   1. Dissemination of the results to get wider engagement.
   2. Organizing workshops and trainings.
   3. Exchange of personnel between stakeholders

Network:

* 94 proposers
* 31 come from inclusiveness target countries
* Heavily supported by the communities of scientific practice: many disciplines. We present EUDOn as the impact it could have on how research infrastructures could deliver more and create a common framework around RIs
* 27 countries, Russia, Australia associate countries
* If we are successful: opportunity to put together working group and plan first management community meeting. 4 years of action, with 2019 first-quarter start.

Secondary proposers: indicative of the support of the action, but does not mean more people won’t be in it if we succeed.

We now added all experts to the email lists who we could not integrate any more during the last days before submission. We have now a list of 124 interested experts.

5. Motivation for some of us in RDA to take actions on DO

(see the slides for some background information)

Some RDA activities contributed to the forming of ideas and concepts:

* Data Foundation & Terminology WG
* PID Info Type WG
* Practical Policy WG
* Data Type Registry WG
* Data Fabric IG
* Collections WG
* PID Kernel Information WG

DFIG finally focused on discussing virtualisation concepts with globally resolvable PIDs including binding information and DOs in its core and the potential of using types associated with DOs to automate processes.

A spin-off of the DFIG work is the C2CAMP collaboration with a few institutions who want to implement DO centric testbeds. Also C2CAMP is open but a certain level of commitment in contributing to the goals is required. C2CAMP is member of the GO FAIR implementation network which is another relevant initiative complementary to RDA.

The next priority for C2CAMP is to specify a generic DO Interface/Access Protocol which will be central in a DO based infrastructure. Also GEDE-DO should participate in the discussions about DOIP, i.e. C2CAMP will take care that the information will be flowing in both directions.

6. Next steps:

* All GEDE-DO participants should send relevant information about relevant and related initiatives to Peter/Zsuzsanna ([zsuzsanna.szeredi@gmail.com](mailto:zsuzsanna.szeredi@gmail.com) ) and we will add it to the share. Everyone is welcome to study all documents until September to extract essential assertions, draw relations etc. Peter certainly will work in this direction.
* It is also of interest to get more ideas from various scientific communities about the potential scientific relevance of DOs. In the EUDON proposal we just included four examples, but having more by receiving positions notes or so would be welcome.
* We may need a website, but probably should wait until the decision about EUDOn.
* Prepare September meeting: give opportunity to people to present, all recommendations & comments welcome.

Discussion:

Wolfgang Kuchinke: GEDE to be present at the DI4R in Lisbon?

(Peter Wittenburg: we will check this oportunity)

Question: To what extent is DOAP/DOIP different from SOAP and SOA architectures? I feel it would be more data-centric than operation centric.

This question needs to be answered until the September discussion.

Carole Goble points to the workshop on Research Objects during the IEEE eScience meeting which would give us a chance for a face to face meeting on the topics of interest. There seems to be funding for students and researchers to attend the meetings. Carol will send a short email to everyone about the opportunities. Obviously C2CAMP will contribute in some way.

Peter Wittenburg: Since GEDE is about RIs, it would be excellent to see various contributions in the workshop.

Wide agreement that we should make use of this workshop.

Question: How will DOs relate to or act as the virtual representation for physical objects that are being studied or discussed?

Peter Wittenburg: Physical Objects will have a digital representation which can be of different kind either as a metadata description or as a documentation of a complete configuration in case of an "accelerator" for example that has a different configuration for each experiment.

Dimitris Koureas: Indeed very relevant to CISCO, we looked into specificities how we can use DOs to represent specimen in the digital world. We are trying to solve this through the DISCO infrastructure architecture.

Stian Soiland-Reyes: should see DOs not just as passive objects

Alex Hardisty: This is exactly the the view within DISSCO.

Peter Wittenburg: That's exactly how it was discussed in DFIG and C2CAMP. DOs have types and with help of Data Type Registries (<https://www.rd-alliance.org/groups/data-type-registries-wg.html>) one can relate types with procedures.

Eva Soroli: is there a full list of participants/countries/people included in the EUDOn proposal? And what are the research domains included ?

Maggie Hellström: uploaded the full proposal to the share

Dimitris Koureas: not exhaustive list of domains but only give an idea of breadth of the proposal (since we could not include various colleagues in the proposal at the very end). New domains should be included as well. Wrt case stories we just had 4 from Neuroscience, Linguistics, Material Science and Biodiversity.

Peter Wittenburg: We should have more case stories from other disciplines. This is exactly one of the priorities now in GEDE DO.

Eva Soroli: someone in the domain should then send his documents on what they are doing.

Peter W.: Indeed make a note on what your institutions is doing in relation to DOs. Send it to Peter/ Zsuzsanna.

Karl Presser: Coming from the area of food science, data on consumption, analysis of food. Is your proposal going into the area of matching terminologies? There is a challenge to match between objects to see whether we talk about the same thing?

Peter Wittenburg: If I understand the question correctly your point is more about ontologies and mapping between different conceptual spaces and terminologies. "Semantic mapping" is indeed a topic that is being addressed in almost all scientific domains. In RDA there is for example a group in the agriculture area working on standardizing formats, mapping between terminologies etc. in collaboration with FAO and other global organisations. This is not really where the work on DOs is about, although DOs will facilitate the mapping task.

Dimitri Koureas: DOs are to be seen as an architectural pillar of the interoperability of RIs in the future. We need to be able to demonstrate how the implementation of DO in different fields can be transformed into a better infrastructure: it's about the conceptual framework and the underlying technical platform. If the DO concept is a reasonable one, we need to be very serious about moving forward in a more consistent way on how to implement DOs in the context of EOSC. There are plenty of opportunities in H2020 to integrate such work. Design studies are the first step to allow us to work on a more sustainable infrastructure. Should go for a design study for the next call.

Eoghan Ó Carragá: Agree c.f. RDA Research Data Repository Interoperability Group final recommendations which basically recommends a "research object"/"data packaging format" as the lowest common denominator for repository interoperability.

Joe P.: Would we be looking at analytical data equivalent to how [http://iiif.io](http://iiif.io/) is trying to standardise the sharing of images?

Chat log from GEDE DO focus group telco, June 27 2018

=====================================================

Alexander Vasilenko, MIRRI

Eoghan Ó Carragáin, University College Cork

Robert Pergl from Czech Technical University in Prague

Joe Padfield from National Gallery, London

David Novák, Institute of Archaeology of the Czech Academy of Science

Jose Lopez Coronado from the Spanish Type Culture Collection (CECT). Spain

Wouter Addink from DiSSCo/Naturalis

Carole Goble(ResearchObject.org, ELIXIR, U Manchester, SSI)

Carole Goble: submissions for the workshop on RO expected: http://www.researchobject.org/ro2018/

Dimitris Koureas: It would be great to submit something asap to this workshop! Is the most relevant space for us to bring this discussion forward

Carole Goble: Indeed - its a discussion forum so light abstracts are encouraged.

Stian Soiland-Reyes: We can also propose a discussion or collaboration break-out for the workshop

Dimitris Koureas: It would be great to submit something asap to this workshop! Is the most relevant space for us to bring this discussion forward

Carole Goble: Indeed - its a discussion forum so light abstracts are encouraged.

Stian Soiland-Reyes: We can also propose a discussion or collaboration break-out for the workshop

Joe Padfield: Have people already been added to the email list?

Zsuzsanna Szeredi: Yes, we have about 30 people who we are awaiting on to confirm, but we already have 124 on the list.

Stian Soiland-Reyes: https://post.rzg.mpg.de/mailman/listinfo/rdaeu-gede-do to subscribe yourself to the list

Zsuzsanna Szeredi: please use this link to the share: https://datashare.rzg.mpg.de/s/lrPcA4kXsNUUYRY

Zsuzsanna Szeredi: password:GEDE

Zsuzsanna Szeredi: I will correct the slides in the share.

Wolfgang Kuchinke: Will GEDE be present at this year's DI4R in Lisbon? The call is still open for presentations.

Alex Hardisty: Dimitris is reading objectives from the EUDon propopsal. There's no slide with those on.

Carole Goble: Nick Juty leads the identifiers WG of ELIXIR by the way and is part of the identifers.org team ...

Alex Hardisty: Summary of the Data Fabric IG on 'Global Digital Object Network' here: https://www.rd-alliance.org/sites/default/files/Summary\_of\_Virtual\_Layer\_Recommendations-20180516.pdf

Carole Goble: https://www.go-fair.org/implementation-networks/overview/c2camp-in/

Stian Soiland-Reyes: just a thought.. To what extent is DOAP/DOIP different from SOAP and SOA architectures? I feel it would be more data-centric than operation centric.

Joe Padfield: Just curious how DOs will relate to or act as the virtual representation for physical objects that are being studied or discussed?

Stian Soiland-Reyes: that's a very good point - and not just as passive object descriptors like on the Semantic Web, but more active, e.g. that you request for access to study a specimem in a museum or ask for a new scan of something

Alex Hardisty: @Joe: This is exactly what we're thinking about in DiSSCo / ICEDIG for natural history specimens.

Eva Soroli: About the EUDOn proposal: is there a full list of participants/countries/people included in the proposal ? And what are the 7 research domains included ? Neuroscience, Linguistics...

Maggie Hellström: Hi, I just uploaded the full proposal PDF into the EUDOn proposal folder of the GEDE share

Carole Goble: https://rd-alliance.org/groups/physical-samples-and-collections-research-data-ecosystem-ig for those of you interested in physical objects

Paul Finglas: My name is Karl Presser

Eoghan Ó Carragáin: Agree c.f. RDA Research Data Repository Interoperability Group final recommendations which basically recommends a "research object"/"data packaging format" as the lowest common denominator for repository interoperability

Eoghan Ó Carragáin: https://www.rd-alliance.org/groups/research-data-repository-interoperability-wg.html

Joe Padfield: Would we be looking at analytical data equivalent to how http://iiif.io is trying to standardise the sharing of images?