Notes of the GEDE -DO meeting 15-11-2018

Editors: Zsuzsanna Szeredi, Peter Wittenburg

**Participants** (partly 29 colleagues participated):

Alexander Vasilenko, Peter Wittenburg, Andrzej  Sobecki, Christophe Blanchi, David Novák, Joe Padfield, Johannes Peterseil, Keith Jeffery, Kay Graf, Maggie Hellström, Roberto Aielli, Ruta Petrauskaité, Ulrich Schwardmann, Carsten Thiel, Heimo Müller,  Ingemar Haggstrom, Jose Lopez Coronado, Carlo Maria Zwölf, Wouter Addink, Eoghan o Carragain, Robertus Erdelyi, Rene van Horik, Cristiano Bozza, Kai Salas Rossenbach, Lucy Bastin

**Absent**: 24 colleagues apologised to not be able to join, but expressed their interest to follow the actions.

# 1. Updates

### 1.1 EUDON

The EUDON proposal did not make it and the major points of critique were: industry is missing (innovation), missing references to EDI/EOSC-pilot, EUDAT etc., SWOT/GAP analysis to be done beforehand, concrete examples from science missing, impacts & risks not detailed enough, country balance not optimal. But there were also very positive statements. The review report can be found on the GEDE share: <https://datashare.mpcdf.mpg.de/index.php/apps/files?dir=/GEDE/digital%20objects/EUDOn%20proposal&fileid=47600365>

A few statements were made during the discussion (and email interaction):

* We had comparatively short time to write the proposal and the DO issue was quite new. Only in March we presented it the first time at the EC to RJ Smits and his reaction was: why didn't we know this beforehand. We need to be honest in so far that we ourselves were in the process of clarification.
* We could not come with ready solutions for some aspects which the reviewers expected from us.
* It is normal that good proposals have been presented 2 to 3 times to COST.
* Obviously we now know where we have some weak areas of argumentation which will help us whatever proposal we will write.
* Some argue that we should write an updated proposal and then go for the April/March 2019 deadline. We need to quickly decide about that (see also 1.2).

***Actions:***

***Koenraad, Dimitris, Suzi, Peter: Make a suggestion wrt a renewed proposal, form a team and do the prep work.***

### 1.2 Proposal Preparation

Closely related to the EUDOn proposal issue is the question what we will exactly do wrt proposal writing. There was no doubt that EC officers expect from us that we come up with a "large" proposal, after they have seen the potential of the DO concept for interoperability at the data organisation/modelling layer for EOSC. They are also very positive about the broad involvement we have in GEDE DO. For many EC officers the first important phase - defining the EOSC Governance model and a FAIR implementation plan - seems to be over now and they will focus now on the question how to fill EOSC with life knowing that it is federating many heterogeneous repositories and systems. A few points were made during the discussion:

* We (Suzi, Edit and Peter) will arrange further meetings with responsible EC officers to fully understand the requirements and constraints for a number of options which were communicated to us and to also inform them about the progress.
* The interaction about this issue must be open within GEDE DO (and C2CAMP) and the papers and some documents such as DO Roadmap (see below) we are currently writing are all part of the preparation of a proposal. Everyone who contributes to the paper with serious use cases should be part of the formation process.
* The Blockchain technology is obviously a topic being discussed heavily within the EC. A link if possible would be good.
* Peter made clear that he does not have any stake in this except to get the proposal going.
* We need to include companies and Edit/Peter will follow their links to see what is possible. But everyone in GEDE DO with suggestions should send them. Since the matter is a bit delicate, please, send an email to Edit, Suzi and Peter first. For specific aspects companies have PID solutions and we need to see how this all fits together.
* We need to be in the EOSC WG on RoP to get DOIP into the discussions and link up with groups such as BDVA etc. EOSC/EDI is a potential market for companies making it perhaps easier to get them attracted.
* We have about 6-9 months to prepare a proper proposal.

***Actions:***

***Edit, Suzi, Peter: continue the interactions with the EC and start updating contacts with industry and report to the GEDE DO group.***

***all: contribute to papers (see below) and give ideas about companies***

### 1.2. Papers on DO and Use Cases

In addition to the Wittenburg/Strawn paper, three further papers are currently in the preparation. A paper together with GOFAIR colleagues about IT aspects of DOs is being finished currently. A paper comparing FAIR principles and DOs written by Erik Schultes (GOFAIR) and Wittenburg is also in the finishing phase showing complementarity which is very important for our EC colleagues. Another paper on the scientific aspects of DO is currently in preparation and all GEDE DO (C2CAMP) members are welcome to contribute with visions on use cases. This paper will be edited by Dimitris Koureas and Peter Wittenburg and they will take care that all (serious) use cases presented by GEDE DO members will be included in some way. A first skeleton of that paper is being discussed currently, but will depend also on the use cases submitted.

We have currently received 29 use cases which is excellent: 5 of them are on infrastructures and testbeds, 4 on fundamental designs how DOs can shape the domain of digital resources in disciplines, 8 on basic topics such as how to integrate a repository into the DO domain and 9 are about advanced cases such as using the DO concept in workflows, in the combination of laboratory notebooks with DOs, etc.[[1]](#footnote-0)

The use cases are on the share and Peter uploaded a first XLS overview: <https://datashare.mpcdf.mpg.de/index.php/apps/files?dir=/GEDE/digital%20objects/DO%20Participation%20Requests&fileid=68600770>

***Actions:***

***Suzy, Peter: start a second round and ask for DO use cases (visions) that could be integrated in the paper and evtl. later in a proposal.***

***Peter, Dimitris: analyse, categorise and summarise the different proposals***

***Dimitris, Peter: work on the paper skeleton and start writing.***

### 1.3 DOIP- specification

The DOIP 2.0 specifications have been officially published by CNRI and will be handed over to DONA. They will be free to use, publically available and DONA/CNRI will come up with reference implementations, i.e. code that can be reused.

The intention of these specs are to continue an intensive discussion. Already in the last months the many comments that came from GEDE DO and C2CAMP colleagues were very important inputs to create this new, much better version. GEDE DO and C2CAMP should again take the lead in commenting, discussing and implementing. Therefore two type of webinars are under discussion currently: one about the DO concept and the DOIP specs, the second about implementations. See below.

During the discussion it was discussed whether other PID systems than the Handle System could be used. C. Bianchi from DONA (one of the brains behind DOIP) stated clearly that DOIP is designed to be independent from the Handle System. Actually, it would be of interest to show that DOIP works with another PID system. However, some level of interoperability needs to be guaranteed. DOIP refers to the resolution of PIDs to state information (RDA: Kernel Information). It needs to be ensured that at that level interoperability is given as stated by ITU X.1255.

***Actions:***

***Suzi, Peter: organise webinars together with possible presenters (see below)***

### 1.4 Terminology Document

At the last meeting a lack of agreed terminology was recognised and it was agreed to have a document with term definitions. Peter started writing such a document mainly using RDA DFT definitions. We should not spend time to discuss this now, but people should comment on the paper so that all nuances can be seen and convergence can be achieved over time.

***Actions:***

***Suzi: make it a Google Doc so that everyone can contribute. This was done:*** [***https://docs.google.com/document/d/11tQR1NPqbc27QFVynBHxl56VG-p882ExI8khj7zn0vY/edit?usp=sharing***](https://docs.google.com/document/d/11tQR1NPqbc27QFVynBHxl56VG-p882ExI8khj7zn0vY/edit?usp=sharing)

***All: comment on the term definitions***

# 2. Intensive Discussions

### 2.1.DO Assertion Document

A first draft version has been written and already been commented on by several colleagues. We should make use of the same procedure as for the PID document produced last year in GEDE: 1) Bring in all statements relevant to DOs and DOIPs into this internal document and 2) write a paper that contains agreements and disagreements.

In the discussion the following points were mentioned:

* DOs can appear in various different contexts - not just in science. We need to be aware of these contexts.
* The question was raised whether the use of the DO concept and of DOIP makes sense without having a network connection. Two examples were given from industry where this point came up. AIRBUS industry wants to run a local information system and a large chemical company wants to run a mirror of a large database based on DOs, but add their own data and tools. Since DOs are representatives of data (for example) and not "transmissions" of data this should be possible. C. Bianchi made clear that there is no objection from his point of view to use DOs and DOIP locally. The issue one has to be aware of is that when you connect systems again via the net then one needs to be sure that the global handles are unique. Issues of serialisation, namespace consistency etc. need to be carefully looked at. In all these respects there are similarities to the OOP model. This should be commented in the assertion document.
* We should read the DOIP specs in depth and make sure that we state assertions which are compatible with specs unless we disagree.
* The definition of what a DO is in the DOIP document is much more minimal than the one defined by RDA DFT. The DOIP doc definition does not mention metadata, while for RDA DFT (and for FAIR objects) metadata is an essential point, since it opens the way to being F, A, I, R. The DOIP definition simply states that a DO has a PID implicitly assuming that the PID can be resolved to state information including different types of metadata. In our descriptions we need to make clear that there is no conflict between the two definitions, but that the RDA DFT definition requires a further action (PID resolution and attribute interpretation) to meet the expectations which are essential from a scientific point of view.
* A whole discussion was about "typing" of DOs. DOs should not have fixed types since they may change over time and may depend on the application case. DOIP leaves it open whether DOs (or its bit sequences) have fixed or dynamic types. The "type" is one specific metadata attribute and it can be changed. However, for many DOs mutability is dangerous wrt. consistency, reproducibility, stability. One has to be very careful in the design choices at application level. The question was raised whether the categorisation of "type" as being static or dynamic is needed, since it could simply be a variant of the type value.
* It becomes increasingly obvious that we need data type registries and also be aware of the linkage to Linked Data.
* Also versioning was discussed briefly. But this was discussed already in the previous GEDE PID document. People are free to use PIDs to implement versioning or to use metadata attributes. see also: <https://blog.zenodo.org/2017/05/30/doi-versioning-launched/>)
* It was kind of agreed that the final document should not be a document with too much technical specifications.

***Actions:***

***all: until end of November everyone interested should comment on the Google document and add "assertions". Wherever possible references should be given. (***[***https://docs.google.com/document/d/1Yb291ITgrqfEabmwGa4UYWi\_9t47-xVRSGbYLqeY0WA/edit?usp=sharing***](https://docs.google.com/document/d/1Yb291ITgrqfEabmwGa4UYWi_9t47-xVRSGbYLqeY0WA/edit?usp=sharing)***)***

***Koenraad, Dimitris, Peter: In December a new consolidated version should be created which is still a working assertion document (and not the final summary paper).***

***Peter, all: add references***

### 2.2. DO Roadmap Document

The ROADmap document is important as an internal document to understand in detail what is required to come to a DO-based infrastructure and thus to a decent proposal. ES (and other policy people) will always ask as a first question: which effort is required, how much do we need to change/adapt, what are the costs. We will need time to carefully discuss this, hoping that some colleagues will already setup and test parts etc.

In the first version uploaded and already commented by Tobias we focussed on infrastructural aspects, since the scientific use cases were not known yet. That section needs to be worked out much more carefully when the second round of use case submission has been done.

During the discussion the following points were made:

* We should put it up as Google doc so that everyone can make comments etc.
* At certain moments we should create consolidated versions.
* We should have authors, versions, date etc in the document.
* We have more time for this document and it is purely internal.
* Private sector use cases should be included in there as well.

***Actions:***

***Suzi: make a Google Doc version. Done:*** [***https://docs.google.com/document/d/1kGLe3zR8KzP50l1I-ma4bYf\_ax\_VmAriSf3C\_4Kknx4/edit?usp=sharing***](https://docs.google.com/document/d/1kGLe3zR8KzP50l1I-ma4bYf_ax_VmAriSf3C_4Kknx4/edit?usp=sharing)

***Peter: create consolidated versions***

***all: look for private sector use cases***

# 3. Next steps

### 3.1. Webinars

It was agreed to have

* 2 webinars about the DO concept and DOIP (December and January)
* 1 webinar about implementations (March)
* the target group for the first two webinars should include C2CAMP, GEDE DO, EX colleagues, RDA and eventually interested company people
* webinars should be recorded to allow others to listen

Actions:

Suzi, Peter: organise the Webinars and advertise them

### 3.2. Anything other business

It was agreed that the next GEDE DO meeting will be in Mid-January. An important topic will be the discussion of use cases based on the improved analysis (see above).

Christophe Bianchi offered to have telephone or skype chats with everyone interested about DOIP

Some had problems to access the share. Here is the link: <https://datashare.rzg.mpg.de/s/r2khPzQRBGTvL58> and Suzi can provide the password in case of questions again.

***Actions:***

***Suzi: make a Doodle for the next GEDE DO meeting***

***all: check access to the GEDE Share and contact Suzi.***

1. In a bilateral communication an "honest attitude" was requested in so far that an open discussion of use cases and visions requires trust that no one is misusing them. [↑](#footnote-ref-0)