GEDE DO Participation Request

3.10.2018

Dimitris Koureas & Peter Wittenburg

Within the framework of the DO activities we are committed to write a few papers which may shed light on the usefulness of the DO concept. Here we give a short overview:

1. The first paper is the one written by Wittenburg & Strawn which already got quite some attraction (Common Patterns in Revolutionary Infrastructures and Data: <http://doi.org/10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0>)
2. The second one is currently being written by some C2CAMP and GO FAIR colleagues to look at DOs from a more computer science view.
3. **The third one will be a paper with views from scientific domains how they see the usefulness of the DO concept. Here Dimitris Koureas and Peter Wittenburg take the lead to kick it off. The hope is to have a paper until Christmas.**
4. A fourth one is currently being written by Erik Schulthes (GO FAIR) and Peter Wittenburg (RDA/C2CAMP/GEDE) about the relation between the DO concept and the FAIR principles. Erik already elaborated on this in his workshop talk.

What we have noticed in our various talks in the EC is that there is a great interest in this topic and that EC colleagues see now the potential of DO-based infrastructures to realise part of the EOSC, i.e. the purpose is to get the above mentioned papers as soon as possible without compromising on the quality of the content.

This is a request to interested scientific communities to describe their potential application of the DO concept within their infrastructure plans with the intention to integrate them where possible into the third paper to be written. Examples for such applications can be found in the GEDE Share[[1]](#footnote-1) to which you all should have access[[2]](#footnote-2). We will collect the contributions, probably discuss some questions with the proposers and then see how we can best integrate the contributions into the paper to be written. The contributions must make clear statements about the use case and the gain that are expected with the integration of the DO concept. In case of questions, don't hesitate to contact Peter (+49 15141858784).

Please submit a filled in template (see below) as is indicated below and provide references were possible. You can also attach papers to your contribution email, but these should not exceed 5 pages.

Please, submit the filled-in templates until 19. October 2018

Dimitris & Peter

GEDE DO Participation Template

## Name and Institution

SORIN HERMON, THE CYPRUS INSTITUTE

## Email and Telephone Number

[s.hermon@cyi.ac.cy](mailto:s.hermon@cyi.ac.cy) +357 22 208 654, +357 99 10 23 14

## Names of Collaborators

(if applicable)

FRANCO NICCOLUCCI

JOSEPH PADFIELD

## Research Community

(could be an ESFRI or another research community)

E-RIHS, ARIADNEPLUS

## Major Goals of the Ongoing Infrastructure work

(describe the major goals of your current infrastructure work with max. 5 lines)

E-RIHS is an infrastructure on Heritage Science, built around 5 modules: **molab** – a mobile laboratory for *in situ* investigations of works of art and heritage assets, **fixlab** – large-scale laboratories such as accelerators, **archlab** – access to physical archives and **digilab** – access to digital data and tools.

ARIADNEPLUS is dedicated to archaeological data, understood in the broad sense of data from sciences that contribute to a better understanding on the past. It includes 3D models, analytical measurements, plans, maps or aerial photos.

## Potential of the DO Concept for your Work

(please, describe your expected use of DOs in your work and the gain which you hope to achieve, max. 2 pages)

**Expected use of DO**

Archaeological data and in a broader sense Heritage Science data consist of a wide range of types deriving from a large variety of sources. It also relates to various purposes, such as research, education, public outreach or public administration. Within this context, research data is a specific type of data that is collected, or created, for purposes of analysis to produce original research results. It can be generated for different purposes and through different processes, and can be divided into different categories:

*  Observational: captured in real-time, often irreplaceable: sensor of survey data, or neurological images.
*  Experimental: from lab equipment, often reproducible, such as gene sequences or chromatograms.
*  Simulation: generated from test models such as climate models or economic models.
*  Derived or compiled: reproducible but expensive, such as text and data mining, 3D models.
*  Reference or canonical: a (static or organic) conglomeration or collection of smaller datasets, usually published and curated, such as gene sequence databanks, chemical structures, or spatial data portals.

Below some examples of research datasets:

* Text documents, spreadsheets, slides;
* Sets of measurements related to specific diagnostics protocols
* Text documents, spreadsheets, slides;
* Laboratory notebooks, field notebooks, diaries;
* Artifacts, specimens, samples;
* Database contents including video, audio, text, images, 3D models;
* Models, algorithms, scripts;
* Methodologies, standard operating procedures, protocols.

The expected uses of the DO are:

* Enable the correct identification and relationship between the physical or the virtual entity and their related cross-disciplinary datasets.
* Define an archaeological DO and a Heritage Science DO that would facilitate the implementation of the FAIR data principles

**Gain hoped to be achieved**

**Research Data Assessment**

Research data should be understandable to other researchers for purpose of reuse and validation. This means that data formats and metadata would best be assessed from the perspective of a user who has not worked with the data before. Can they find what they are looking for, can they gather the data they need, can they open the data, and can they understand the content? Metadata needs to be as complete as possible and fully transparent. If codes or variables are used, their explanation needs to be directly available. In addition, users need to be able to determine which files contain what kind of data. They need to be able to open the file format, as much as possible independently of the use of specific software or hardware.

**Data reliability (quality)**

A key element in using and re-using research results relies in the ability to assess quality of data. This concept relates to:

* How data has been created (data provenance)
* How data has been processed (data manipulation)
* How data has been interpreted (inference and argumentation models)

The gain hoped to be achieved by adopting and implementing the DO concept is an increase in the re-use of Heritage Science and Archaeological data once its quality has been correctly assessed.

1. <https://datashare.mpcdf.mpg.de/index.php/apps/files?dir=/GEDE/digital%20objects/GEDE-DO%20meetings/workshop-september-18&fileid=63958403> [↑](#footnote-ref-1)
2. Those who presented their plans already at the recent workshop do not have to submit these templates since we discussed the slides already. [↑](#footnote-ref-2)