GEDE - EOSC Core Group

Carlo, Peter

Let's state very clearly again, as it was mentioned during the virtual meeting, to prevent misunderstandings:

***GEDE will not coordinate the ESFRIs/ERICs in their way interacting with the EOSC process, but we see a chance to use the GEDE umbrella to coordinate some actions of experts interested in specific topics. The Core Group will act in this coordinating role.***

## Task of the Core Group

* identify typical topics where GEDE can become active (and where not)
* identify a short list without excluding topics in the strict sense
* identify care-takers that take responsibility
* interact with the EOSC Sec project about modalities etc.[[1]](#footnote-1)
* helping to get the funding requests done

## List of possible Topics

A number of topics were mentioned which are listed here. These lists are not exclusive, but contain already many suggestions requiring an evaluation and rating.

### From the Co-Chairs' list:

* analysing the actual technologies being used by research infrastructures integrating RDA use cases and others (Technology Matrix), identifying so called common infrastructure components and abstracting from their individual flavours and describe their characteristics and interfaces using also knowledge from RDA groups, stimulating new RDA groups, etc.
* relate the outcome of specifications of reference architectures with work done within RIs etc.
* continue the intensive work on FAIR Digital Objects & continue the work by addressing other specific topics of interest
* organising surveys related to EOSC based on questionnaires & analysing the outcome of these surveys (current "Technology matrix" questionnaire & doc is just one example, could be others)
* organising webinars and workshops on topics related to EOSC

### Bert Meerman's list (from what he assumes about GEDE)

* PID Service ( Handle / DOI’s)
* EduGain service extension
* Digital Object model - FAIR DO
* Semantic Mapping exercise
* GDPR implications for data-exchange
* Research Workflow Automation use case or pilot.

### Topics mentioned in the interactions and the VM

* policy knowledge aggregation and abstraction
* Sensitive data - GDPR
* Repository Setups

## Evaluation Matrix

Here we comment on all suggestions with the goal to come to a first evaluation. Effort is categorised in days (D), weeks (W) and months (M). EOSC Relevance is categorised in low (L), medium (M) and High (H).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Topic | State | RDA Group | Effort | Care Takers | EOSC Relevance | Comment |
| Technology Matrix and Analysis | Questionnaire is ready, expect many reactions which requires effort for analysis | N | W |  | H | There is the hope that some analysis can be done automatically using nano publications, but experience shows that we need human analysis. |
| Analysing Reference Architectures | In Industry etc. reference architectures have been developed, they should be analysed on their relevance for EOSC | N | W |  | M | Industry (IDS, RAMI, IIC), DOA, EOSC Pilot all claim to have designed architectures that should be used for building complex infrastructures. One could investigate their potential, but this would require the involvement of real experts. |
| FAIR DO Related Work | GEDE has broad group working on this | Y | W-M | Peter, Dimitris, Koenraad | H | The effort depends on what will concretely be done. This needs to be checked with EOSC Sec (**see note 1**) |
| Organising surveys for EOSC across RIs | GEDE has tools and access to many RIs | N | D-W | Suzi, Peter | L-H | The topics of the survey would have to be defined. A first example is the Technology Matrix. Needs to be made concrete. |
| Formulating Statements on PID Services | GEDE has done an extensive analysis. We should summarise the results and make a suggestion | Y | D | Carlo, Peter | H | This should be done together with groups already giving services such as DONA, EUDAT, ePIC, DataCite, etc. Based on the GEDE report this can and should be done asap - probably without additional funds |
| Analysing AAI (eduGain, etc.) Service Gaps | Reoccurring issue of high relevance and a roadblock still | Y | W-M |  | H | The effort would depend on the formulation of the task. Within RDA there is a very active group on Federated Identity Management (**see note 2**) |
| Framework for Simple Semantic Mapping | Large ontologies are mostly underused and we still lack flexible frameworks for mapping | N | W-M |  | H | Semantic Interoperability is one of the key tasks in EOSC. At metadata level we are all used to map semantics for a while. At scientific annotation level semantic mapping will remain a highly challenging field where pragmatic approaches are urgently required. (**see note 3**) |
| Research Workflow Automation Pilot | Increased automation is one of the important aspects for future data analysis, still not used broadly | N | W-M |  | H | An increased automation of data analytics will be key for competitiveness in data intensive science, people are starting to use Jupyter and similar tools, but more could be done to facilitate creating workflows. (**see note 4**) |
| Sensitive Data and GDPR implications | important issue for some communities but underrepresented in EOSC discussions, GEDE work on blockchains could be integrated | ? | W-M | Jacques, Peter | H | there are special requirements for sensitive data which need to be pushed forward to the EOSC process, within GEDE only suggestions can be worked out how this can and need to be addressed; discussions about the use of blockchains could be part and current GEDE work can be integrated. |
| Policy Knowledge Aggregation | RIs have lots of experience in this respect making sense to aggregate this knowledge | N | W |  | H | This needs to be further specified to make a clear point. |
| Organising Webinars and Workshops | as the last virtual meeting showed it was very important to use GEDE as a platform for bi-directional exchange | N | D | Suzi, Carlo, Peter | M | This bi-directional interaction should be repeated bi-annually, a saturation effect will happen over time. It would only be very little money. |
| Repository Setup | repositories are key pillars, many would like to get guidance on how to set up FAIR compliant repositories; GEDE work can be integrated | Y | W | Michael, Peter, etc. | H | most of the RIs see repositories as key pillars and many asked for guidance and help, based on existing GEDE work and RDA work GEDE could be used as an umbrella to run hackathons to just show people how to do it |
| Citation Harmonisation | citation of data is diverse and GEDE and RDA work is there to harmonise suggestions | ? | W | Carlo | H | the many different suggestions for data citation are confusing and a harmonisation is urgently needed, GEDE could be used to express wishes and make suggestions |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

### 1. FAIR DO Related Work

The scope of work to be done is huge. Most urgent are test cases with the adaptation of repositories to DOIP. The existence of concrete software implementations to for example adapt fedora, D-Space or other concrete repositories will be crucial for acceptance. Many other DO topics could be taken up, but this seems to be the most straightforward topic to be addressed. We will discuss this within the GEDE DO group.

### 2. AAI

For cross-border (disciplines, repositories, countries) work in Europe it is still a nightmare that there is no accepted generic solution for the authentication side. It is not per se a technological aspect to be solved - there are suggestions that need to be analysed, the biggest challenge to be addressed is the policy harmonisation across countries. GEDE could be used to summarise the experiences, solutions being used and suggestions for change. This work needs to be synchronised with existing groups for example in RDA without losing the biggest advantage of RIS: they understand the practical needs.

Another unsolved aspect is the authorisation in distributed scenarios where the Finish IT providers in collaboration with EUDAT and ELIXIR worked out a solution. This solution would have to be analysed based on the RI requirements and suggestions could be made for adaptation and/or adoption.

### 3. Semantic Mapping

Semantic Mapping of metadata is what many people have done extensively between disciplines but also mapping to semantically loosely specified sets such as DC. Often these mappings are not explicitly registered, i.e. no one else can make use of them. Problematic is the mapping at scientific annotation level due to the well-known challenges due to semantic shifts, theory-bound definitions of concepts, etc. In some disciplines complex ontologies have been built, but in daily work when data is being used across narrow boundaries these ontologies are often underused or even ignored. What is needed is a simple to use framework where researchers can easily define mapping relations in a pragmatic way, store them and share them with others. CLARIN and perhaps others made early versions. These would have to be analysed and suggestions for a cross-disciplinary framework could be made. Implementing it would be difficult in this EOSC Sec context.

### 4. Automation

Automation is a big challenge and can only be achieved stepwise. In the RDA realm a training course has been organised for example which had as goal to integrate data and metadata in a simple repository and then use Jupyter notebook to execute simple analytics on the data. Other initiatives may have done similar concrete exercises helping young people to learn how to use these tools. Another aspect could be to create code snippets that are required in almost all FAIR compliant workflows such as: resolve a PID, register a PID including state information such as checksum etc., read existing metadata, create new metadata including provenance information, form packets of data & Metadata to be uploaded in repositories, etc. Having a library of such snippets would make it very easy for people to create workflows that create FAIR compliant data and store them in FAIR compliant repositories. Such code library could be made open and sharable - thus being FAIR compliant.

1. Proposals between 100 and 150 k€ seem to be appropriate. [↑](#footnote-ref-1)