Minutes: DTR WG Call: 1 Feb 2017

During this one-hour call we heard project updates from four groups and had one email report. These minutes will summarize those reports, with conclusions and next steps at the end.

**Report #1** - Ulrich Schwardmann - GWDG: Ulrich approaches the DTR issues from the point of view of PID Info Types, as he is deeply involved in ePIC, the European PID service. For a summary of his work up to that point see his presentation on Building and Maintaining a Registry for PID Info Types at P8:

<https://www.rd-alliance.org/sites/default/files/attachment/2016-09-17-Building_and_Maintaining_a_Registry_for_PID_Info_Types_-Schwardmann.pdf>

He brought us up-to-date on this call. He is running two type registries, one for types in preparation, one for final candidates and prototyping, with over 200 types in total defined across the two. Currently working on a process for transferring types from preparation into candidate state. Such a process will include both technical and organizational components. He is also increasing automated aspects of registration, including schema creation. Most of the defined types are basic types, e.g., types of dates, types of locations and so on, including many derived types. The candidate registry can be examined at <http://dtr.pidconsortium.eu:8081>. An important use case for ePIC is transferring digital objects with a given handle PID into the DOI world, which requires a specific metadata set. A search of the candidate DTR for DOI will show some of the current type definitions.

**Report #2** - Jim Duncan and Mike Finnegan - Vermont Monitoring Cooperative (VMC): VMC monitors forest health and maintains a data archive used for research and management. Their DTR use cases are data set discovery and traversal, standardization, and multidisciplinary application. A presentation at P8

<https://rd-alliance.org/sites/default/files/attachment/RDA-P8_VMC_MFinnegan_16Sept16.pptx>

gave their considerable progress to that point in time, with 250 defined types associated with over 2000 fields in their data store. Many of these are basic measurement types. Since that P8 report they have moved on to utilize types in a data set recommendation system whereby users of a given data set can be pointed at similar data sets based on type similarity and other metadata elements. Their work in this area is continuing, in collaboration with CNRI. Going forward on DTR use per se they are looking at expanding their use of types from the attribute level up to the data set level.

**Report #3** - Tobias Weigel - Deutsches Klimarechenzentrum (DKRZ): Tobias, who is one of the co-chairs of the DTR WG, is heavily involved in the World Climate Research Program CMIP6 project. DTR work figures in this project both in PID profile work and typing the results of their data collection and processing work. See his presentation at P8

[https://rd-alliance.org/sites/default/files/attachment/DTR2 P8 - Type registry scenario for climate data-TW.pptx](https://rd-alliance.org/sites/default/files/attachment/DTR2%20P8%20-%20Type%20registry%20scenario%20for%20climate%20data-TW.pptx)

for an overview of that work. In this call Tobias reported on new project ideas that would use types to bind objects to processing and/or analytics services. This led to an interesting discussion on the possibilities of using types in the orchestration of various kinds of services and workflows. The idea of using registered types to connect registered objects to registered services has been present in the DTR work from the beginning but with no visible work or progress to date. These topics have now resurfaced in this group and in the Data Fabric Group, especially in the interactions with the Brokering Group. It seems likely that this topic will be addressed at P9 in multiple fora.

**Report #4** - Andres Ferreyra - Ag Connections, Inc: Andres, who made contact with the DTR group at P8, is working with Ag Gateway, a consortium of over 240 companies in the agricultural sector and Andres is involved a DTR-like project. He was unable to make the call, but sent in a brief note on what he would have discussed had he been able to attend.

*"Ag Gateway, an industry consortium of 240+ companies in the agriculture sector, has set up a variable type registry (aka The ContextItem System) to source geopolitical-context-dependent variable data definitions used by the industry as part of its SPADE, PAIL and ADAPT interoperability projects. It's implemented as a RESTful API, and documented at* [*https://api.contextitem.org/swagger*](https://api.contextitem.org/swagger)*. We are putting in place an ISO 19135-based governance process for the resources to be delivered by the API."*

We look forward to more contact with Andres and his project.

**Report #5** - Giridhar Manepalli - CNRI - Giridhar has taken the lead, with encouragement from the U.S. National Institute of Standards and Technology (NIST), to work with ISO to determine if data type record standardization warrants a new activity at ISO. ISO wanted first to see if any existing standard already covered this issue. A new ISO Study Group formed June 2016, with delegates from China, UK, US, and Japan. Giridhar is the group lead. The goal of the group is to see if metadata fields needed for type records are already covered by one or more existing ISO Standards. The next virtual meeting is scheduled for March. During that meeting the study group members will discuss whether the likely metadata record structure is already covered elsewhere. Giridhar has the task of proposing that structure for that call, a task for which this DTR call has proved useful. Note that this is not a list of types but a structure to describe registered types. In June, ISO will make the decision as to whether a DTR record structure is a new standard of work or not.

**Discussion and Conclusions**: Larry started by announcing that our proposed P9 Breakout was approved and scheduled for Thursday afternoon, pretty much in the middle of the W-F Plenary. There will be at least the usual level, if not more than the usual level, of conflicts as each breakout slot has to accommodate eight or nine sessions.

Tobias then kicked off the rest of the session with a number of discussion topics. The first was the potential use of software design patterns in various kinds of type-related processing, e.g., service orchestration, or perhaps even to the designing of type records. None of this would be intrinsic to types and their registration but the creation of well-designed patterns that could lead to common re-use would encourage best practices and make the adoption of types and type-based processing considerably easier.

His second topic was the creation of a type catalog, not as a permanent service but as an exercise to analyze the current state of play. There are now hundreds of types in multiple registries with many likely overlaps. Before we fully address the technical issue of federation and the organizational issue of governance it would be useful to get a snapshot of the current collection of types. This was seen as a good and reasonable goal for P9, with Tobias taking the lead.

This led naturally to the third topic - federation and organization/governance. Lots of input here with no specific proposed actions as yet, although this is clearly a key topic going forward, especially in terms of uptake. To use types from a given DTR one would want some kind of assurance that those types would continue to be available in a formal and registered form. Ulrich commented on the need for agreements by or across organizations maintain types or certain aspects of registration. Tobias pointed to the EUDAT type service - this will have to be maintained over time. How is that supported? There will always be a cost. Margareta Hellstrom pointed out the existing difficulty of proselytizing type usage to those who see ontology work and format registries as providing all of the need infrastructure for data interpretation, making a reliable type infrastructure all that more important. And Giridhar pointed out that the usual network effect will apply to DTR uptake and warned that unless type overlap and duplication were managed somehow, the result would be confusion - which version of a type to use?

This ended an hour of very productive discussion. Should we have a second call between now and P9? The conclusion was that we would wait for these notes and any reaction to them as well as Tobias' plan of pulling together a current-state catalog and see if the result of that exercise warranted a second call.

Thanks to all participants and we look forward to seeing many of you in Barcelona.

Larry