

Case Statement

Data Type registries #3

Cochairs

Ulrich Schwardmann, Larry Lannom

Background

The first two iterations of the DTR WG developed the concept, drafted a data model and expression for types, and then collected use cases to analyze the results. A number of data type registries are currently in use in prototype efforts, details of which can be found in the reports of WG sessions at prior plenaries, the last one of which was P11 in Berlin.

[<https://rd-alliance.org/wg-data-type-registries-1-2-rda-11th-plenary-meeting>]

The output of DTR Groups 1 & 2 was approved as an RDA endorsed recommendation and was then accepted as an ICT Technical specification.

[<http://dx.doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458>]

[https://ec.europa.eu/growth/industry/policy/ict-standardisation/ict-technical-specifications_en]

DTR WG activity is currently paused and the next step in DTR work is clearly governance.

Furthermore the topic of data type registries is under consideration by ISO as a potential standard. If and when an ISO group comes into focus, it will strongly influence the proper next steps for the RDA group.

Objectives

The third version of the DTR group to work on governance, either in collaboration with ISO or independent of it, is planned with the following objectives

Governance for DTRs

The goal of data typing is to provide a level of indirection to be able to impose semantics and this way to precise description of given data entities for human and machine processing.

The utility of this approach will follow the well-understood network effect. If every new data entity had a unique type the registries providing records of those types would still be useful. But the full potential power of the approach will come into play, when some set of types become widely used. As the use of common types becomes more widespread, software and services applicable to those types are more likely to be developed encouraging further use of the types and so on in a virtuous circle, greatly facilitating data interoperability and reuse.

Managing the growth of types and registries of types will be essential to this process. The availability and utility of existing types must be made known and data creators must be free to chose which types to use or to define their own types as needed. The findability of suitable existing types is key and combining existing types with new ones will become a common case.

Therefore the type descriptions in type registries must be consistent and interoperable, and related software and services must be readily discoverable. Relations between syntactically and/or semantically equal or similar types must be visible to data creators and users as a building block to

automatically define crosswalks between data descriptions.

To produce those results defining guidelines and policies most likely is the next step in RDA's data typing work. This process needs an organisational structure for all participating data type registries.

The Goal of Governance

Defining guidelines and policies needs input from stakeholders like working data type registries and possible use cases of communities.

A first goal of the working group would be to discuss how to engage the stakeholders in this process and to discuss different models of participation for possible interested and working data type registries.

Stakeholders and Participation

Stakeholders of an interoperability approach for DTRs are obviously the DTRs itself, but also communities will have an intrinsic interest to participate, and also standardisation body may have interest, because interoperability always needs some level of standards.

Possible models of participation range from formal federation to loosely coupled association of stakeholders. These possible approaches have to be described and discussed between the stakeholders. Which approach is most appropriate to reach the goal of interoperability between type definitions leaving enough room for innovation has to be evaluated. Different approaches for different classes of stakeholders, like DTRs, standardisation bodies or communities should be possible.

A first work item for the group would be to find and define different models for participation of the stakeholders and to provide an analysis of these models with pros/cons in a final recommendation.

The management of ontologies is an example of a self organized way to provide governance without central coordination. Despite the fact that there was no central or decentral kind of coordination, this approach has substantial impact nowadays. But on the other hand it took about ten years, or even longer, to become useful and it is not clear, whether its impact still increases.

Levels of Interoperability and Granularity

The type descriptions in type registries must be consistent and interoperable, because important common use cases like the findability of suitable existing types and the combination of existing types with new ones require these preconditions. But it is still unknown what level of interoperability and granularity is actually needed.

If for instance the use of a type describing the creator of a digital object is needed, is it sufficient to find a type definition that has the name creator, or does one need a description of the internal structure of the type like a list of prename surname couples, or even are the names of subtypes in such a structure important for a decision.

These discussions is very similar to those that were seen earlier in the PID Kernel Information WG and that certainly will remain in the PID Kernel Information and profile management WG.

To classify this kind of requirements and to provide a consistent view on the needed level of interoperability and granularity will be the second work item of this group.

Use Cases and Adoption

- which DTRs are currently running or are in a short term planned
- Which communities and use cases are covered?
- What are the topics that should be globally harmonized?

Related Work

- outcomes of DTR1&2
- ISO work
- PID Kernel Information profile management WG
- state of existing DTR
- ...

Possible Outcome of DTR3

Guidelines for governance

1. discuss how to engage the stakeholders in this process and to discuss different models of participation

Guidelines for interoperability as principles used in a governance structure

2. Describe level of type interoperability most useful. This may depend on the chosen governance model.
3. Define how interoperability can be technically achieved (metadata, interfaces, protocols)

suggestions for a consortium adopting the guidelines

4. Describe the way how the technical interoperability constraints can be implemented

Relation to other RDA efforts

As reflected in the interoperability objective, the group's work should be closely coordinated with the PID Kernel Information profile management WG as this group deals with a specific application case for DTR, and the PID KI and DTR both exist within the same digital object data ecosystem (one that builds upon, but is not limited to, the Handle system for PID resolution). The governance and management of Data Type Registries should be synergistic with the governance and management of PID KI profiles.

In addition, the use cases and stakeholder perspectives must be further taken in by connecting to the PIDs for Instruments WG, the emerging group on Interoperability of Observable Property Descriptions, the Biodiversity Data Integration IG and the Data Fabric IG. The latter is of specific importance as a forum to engage new adopters, both from use and added-value perspectives as well as infrastructure and service provisioning perspective.

Work plan and Deliverables

The WG will first assemble requirements for profiles and supporting services. Then, it will work towards the objectives in parallel.

In the end, the WG will deliver a recommendation for KI profile management, example profiles defined together with potential adopters, profile registry interface specification (if needed).

Milestones

Assuming an 18 month timeline, the work towards objective happens in parallel.

P14 (Helsinki): First working session of the renewed WG.

P15 (Australia): First draft of guidelines for governance and models of participation, interoperability principles and technical guidelines

P16: First draft of recommendation open for discussion

P17: Final recommendation presented

Initial supporters

Ulrich Schwardmann, GWDG

Larry Lannom, CNRI

Tobias Weigel, DKRZ