

DATPRO: Surviving without a concept

Publications Office document 2015-05-21, Martin Scherbaum, A.2.001 Technical Support Contributions by Dora Gudlin, Maria Kardami, Marc Küster, Julia Pichon, Camilo Soares, Willem van Gemert

Business need

In the scope of exchange of publication data¹ between institutions and data providers/consumers using the IMMC protocol, it might happen that a sender (e.g. author) has to introduce new concepts for metadata elements, i.e. new codes for existing authority tables because e.g. a new corporate body was created. The time span between the (political) decision for the concept and the need to use it as metadata for an actual document might be very short, in many cases shorter than the complete procedure for inserting it into an authority table.

As the introduction of a new concept is part of the Metadata Registry's (MDR's) maintenance procedure, it is subject to possible delays until a new concept, once decided and implemented in the authority table, is available to all stakeholders via an MDR publication of the concerned Authority Table. But even then it is not assured that the new table version is available in all systems along the way as the moment of activation of a new table and thus availability on all systems can be detached from the moment of publication.

Indeed, a method is needed to be able for authors/stakeholders in the exchange of publication data to use provisionally the new concept without breaking the ability of processing systems for validation of the metadata. In order not to lose the additional information on the new concept, the new concept should be stored along with the other metadata.

Furthermore, such marked up data has to be possible to detect and recovered once the "real" concept is broadly implemented in all systems referring to it.

Introducing DATPRO

The method described above for specifying provisionally new concepts in metadata is done by the introduction of a generic concept for each authority table which was called DATPRO (acronym for provisional data). The presence of this value in a metadata element means that up to the moment of composing the transmission no specific concept was available to characterise the data.

The DATPRO value will then need special attention in all systems interpreting the values of metadata elements (e.g. newCERES, PlanJO, CELLAR, dissemination systems).

DATPRO from the authority tables and its corresponding concept OP_DATPRO in CELLAR is used since the beginning of CELLAR loading to make it possible to "fill" cardinalities where there was no proper value at hand. In CELLAR an OP_DATPRO value in a property of the Common Data Model (CDM) is

¹ "Publication data exchanges" encompasses all types of IMMC transmissions:

- manuscript contents (data) plus its related metadata, e.g. for OJ production,
- final works plus its related metadata, e.g. for EUR-Lex publication,
- and metadata exchanges only (e.g. for update).

always connected with an annotation referring to the context of the original value (if any). Like this it is indicated that "here is no proper value available, but in the original metadata the value 'xyz' which is not conform with the authority tables was specified" which makes a later recovery possible.

Explicitly, for the semantics of DATPRO this means: DATPRO is a provisional representative of a value that is NOT YET available in the corresponding table, it is a formal marker. DATPRO is NOT a category "other" on the semantic level.

How to get a code for a new concept

When a stakeholder in the publication process (e.g. author, contractor, production system, reception system) needs a new concept in an authority table, the MDR procedure for proposals applies.

In most cases, the stakeholder has already a specific proposal for the code to be created. Nevertheless, it is the MDR evaluation that either confirms the requested value or counter proposes a different (perhaps already existing) value.

If a fitting code already exists (meaning: available in a published version of the table and activated in all systems), the stakeholder can use the concept right away.

If the concept as proposed does not exist, then the MDR will evaluate its unicity and see if the proposed code is in conformity with the general code rules and will confirm to the stakeholder that the proposed value will most probably be accepted for insertion into the table. Or, the MDR will counter propose a different (new) code to be inserted into the table. In both cases, this reply authorises the stakeholder(s) to use the new code provisionally (see process below).

Using a provisional code

Once the MDR has attributed the provisional code, the stakeholder may use it in its transmissions. However, such occurrences of the code have to be marked by a specific code: DATPRO, which is available in all relevant tables. In IMMC v3, the authority table prefix has to be prepended to the DATPRO value, e.g. cob:DATPRO if the code refers to the corporate-body authority table.

Example:

A new corporate-body concept with the code "ABC" was requested by an institution. The MDR confirms that the code will be accepted for a new publication of the corporate-body table as "ABC" and might be used as provisional code. In the author field of an IMMC transmission this would then read as follows:

```
<cm:agent_work role="AUT" provisional_code="ABC">DATPRO</cm:agent_work> (IMMC v2)
```

```
<cm3:agent_work role="AUT" provisional_code="ABC">cob:DATPRO</cm3:agent_work> (IMMC v3)
```

The attribute `provisional_code` holds the attributed provisional value as a string. Thus, it will be validated by all parsers but can be used then e.g. to store it together with other metadata of the transmission.

Once the provisional value "ABC" is contained in an officially published version of the table, the stakeholder should replace the above mark-up with the "normal" values:

`<cm:agent_work role="AUT">ABC</cm:agent_work>` (IMMC v2)

`<cm3:agent_work role="AUT">cob:ABC</cm3:agent_work>` (IMMC v3)

Institutional stakeholders' systems

When the systems of the institutional stakeholders are mentioned, this refers to a potentially large number of highly differing systems. Thus, the following suggestion only characterises the general traits of a solution.

The document management system (DMS) of a stakeholder will use authority codes for conceptually coded entities (e.g. authors, countries, ...). When such a system should be able to deal with provisional codes, this has to be added as a functionality.

Suggestion 1: explicit conversion

First of all, the provisional code has to be entered into the systems configuration after attribution by the MDR. The provisional code has to be marked as such in the system: at this moment, the public version of the authority table will contain neither an authority code nor a mapping to the code.

Codes being marked as provisional will be translated into the above mentioned `provisional_code/DATPRO` construction in IMMC.

Once the provisional code is published via the MDR and all subsequent systems are able to handle the final code, the marker for "provisional" can be removed and the output will be produced as normal.

Suggestion 2: conversion in output filtering

The system does not directly use the authority tables for code selection and provisioning. Then the provisional code attributed by the MDR is just added as a "normal" code to the configuration.

If a transmitted document contains a code, which is not part of the current version of the authority table, this will be detected and the provisionality of this code be implied. In consequence the `provisional_code/DATPRO` construction will be used.

By installing the authority table containing the provisional code as final code (or mapping) the creation of the `provisional_code/DATPRO` construction will not be done anymore.