Report by the German National Library's Office for Library Standards, Office for Data Formats to the Bibframe Early Implementers Group

The German National Library has thankfully accepted the invitation to take part in the efforts of the Early Experimenters / Early Implementers Group during the initial phase of the Bibliographic Framework Transition Initiative.

After the first meeting in October 2012, we set up a small group of information specialists, with MARC 21 and Linked Data / Linked Open Data background. In addition, we shared our plans with two German groups, the Expertengruppe Datenformate (Expert Group for Data Formats), and the DINI AG KIM Group Titeldatenmodellierung (Modeling Bibliographic Data). From members of both groups we received valuable estimations and ideas.

During the testing phase, we studied the material that was and became available, mainly the draft report and the published report, the papers about annotations, the vocabularies, the mappings from MARC to Bibframe, as well as the visible data via id.loc.gov, and recently the source code for transformations. We discussed what we understood, and tried to decide what reasonable steps should be taken by us.

As a general remark, we have to admit that the tasks and timelines were challenging, and in some respect not perfectly suitable to what we in the German speaking part of the information community are used to. Comparing the migration from MAB to MARC on the one hand to the next step from MARC to Bibframe that now is taken, we struggled with the experimental approach. It is relatively easy to move from one well developed, established and documented standard to an even better one, it took time and defined efforts. But in this phase, neither the goal was and is clearly defined, nor was and is the path that has to be taken.

At first glance, the colleagues were critical about some parts of the Bibframe terminology. The Bibframe Instance seems to stem from a different context, and seeing the note "Instance of: Instance" in the test version of id.loc.gov may cause confusion.

The feeling of simplicity and incompleteness was one of the major impressions. We understand that in this early phase this had to be the case. But one of the values of the MARC format is its enormous richness, the complexity of elements, all of them with a need behind them, most of them still needed. If one of the Bibframe goals is to replace MARC, then there has to be an estimation how most of the MARC elements will find their way into the Bibframe model, in the core, or in MARC-based modules or additions. We would like to think about a structured way of local additions, in analogy to the MARC XX9/X9X/9XX fields.

We very much welcomed the distinction between Creative Work and Instance as core classes. By this, the line between the content and the carrier of a resource is clearly visible, and the comprehensive MARC record structure is broke up into reasonable levels. However, concerns were

discussed about the lightweight approach that was taken. Specifically from the RDA section, with a large project and commitment to implement RDA, there were critical voices about the "RDA-lite" view. The fact that a Bibframe Creative Work can be a FRBR/RDA Work, or a FRBR/RDA Expression, may cause rethinking the whole model of group 1 entities. In the Bibframe view, one Creative Work is the original work, and a second one is a translation of the work, or an adaptation. By relationships these Creative Works may be contextualized. We still see the need to separate the FRBR/RDA Work level from the FRBR/RDA Expression level. We don't think that it is sufficient to say that the one Creative Work that all relationships are pointing to is the FRBR/RDA Work. Instead, we can imagine and wish to see a clear designation (a "flag") in the model that one specific Creative Work is the equivalent to a FRBR/RDA Work, and that a Creative Work lacking this designation by implication is a FRBR/RDA Expression. Specifically when the community moves on from RDA Implementation Scenario 3 or 2 to 1 or something similar, this need may become a pressing one.

During experiments, one focus was given to Authorities as one of the four core classes, and the German authority file GND. A very simple mapping from MARC Authority (as we use it for the GND) to Bibframe (as we understood) was written, and based on this an XSLT transformation was coded, representing not more than a proof of concept. It was not possible to embed this piece into the productive DNB environment, for different reasons: Our data base is structured in Pica+, not in MARC 21; we don't use XSLT for ordinary transformations; and we were not able to invest more time, in order to do more productive approaches. A next step would be to write a full and detailed mapping for the GND from Pica+ to Bibframe, realize an interface by using Metamorph, or the usual Conversion Service (in the Pica fcvdef syntax, or Java), and prepare an implementation. One experience from the GND part of our experiments is the impression that the Bibframe is still at its very start, and a lot of elements are missing. We didn't invest into adding single elements to the Bibframe authority vocabulary, as we think that most (if not all) of the GND elements (as documented in the GND Ontology) deserve being added to the vocabulary. On the level of models, one limitation is that in Bibframe a Creative Work forms a core class on its own, whereas in the GND works are included as a type of entity (most of them from the Subject Headings file, some of them from descriptive cataloging by the German Musical Archive, with FRBR-like links into bibliographic records). An easier part may be the FRAD-based addition of the entity "family".

Experimenting with bibliographic records was more complicated for us. As one approach, we did a study in mapping of Types of Creative Works, and Types of Instances, aiming at resource categorization. This was done on the base of the types that were visible in the vocabularies. As these may have been selected according to criteria that we were not able to fully understand, and as we recognized that some types don't have a clear equivalent in MARC 21 (Leader, 007/00, 007/01, or RDA based 336/337/338), we decided not to follow this path. Instead, we began writing a mapping of single MARC elements to Bibframe. We discontinued this mainly for two reasons: A lot of elements that are needed are not yet defined in Bibframe, we hesitated to define "local" elements; and a very similar and detailed approach was taken by other partners, so that we could reasonably decide to wait, and benefit from their progresses. Here the same is valid as is said above about the GND => Authority experiments: For reasons of discussing the MARC => Bibframe step is needed, but for productive usages the Pica+ => Bibframe way would have to be taken. No line of code was written in this part of the experiments.

Nevertheless, some first observations may be worth sharing: In terms of part-whole relationships, there seem to be advantages in the decisions made a few years ago during the migration from MAB to MARC: Each physical volume gets a MARC record of its own, and a multivolume monograph gets one additional record. The (record for the) volume that lacks a title is linked via 773 $w to the (record for the) multivolume monograph, thus forming a tree. By this hierarchical description it is possible to distinguish both levels from each other, and to relate the parts to one other. One part of the discussion around the meaning of MARC field 505 may in this respect become irrelevant for

us. It is true that some other part-whole relationships below the physical level (e.g. collections, articles, audio tracks, etc.) do still have to be discussed.

The extensive usage of record control numbers for linking between records in the tradition of cataloging and systems in the German speaking countries seems to offer a lot of potential when it comes to Linked Data. Links inside of the GND are internally and physically available as record control numbers, and by the very clear definition of entity types they are easily to be expressed as a relationship from a representation of an entity to a representation of a second entity. In addition, the types of relationships are controlled by a list, admittedly most of them not yet controlled as MARC relators. The same is true for links from a Bibliographic record into the GND (1XX, 6XX, 7XX). But regarding inter-bibliographic links, things are more complicated. A 76X-78X $w, or an 8XX $w contains the record control number of a bibliographic record, so here the entity described isn't directly linked to. The link by itself doesn't contain information whether it is valid from one Creative Work to a second Creative Work, or, if it is valid from one Instance to a second Instance. In some cases, the field number may imply that the relationship is a Creative Work to Creative Work (765, 767), in some cases an Instance to Instance (775, 776), but some cases do need further analyses.

In each of these cases, we see the focus on links, and their types. We don't think that useful information that is available to a human being or a machine that follows the link has to be included into the syntax of the main information. No redundant information has to be given in the context of Linked Data. Translated back to MARC, e.g. a 100 with a $0 (and a $4 or $e) is sufficient, there is no urgent need to complete the field, as every kind of information is available if you follow the link.

Unfortunately, we didn't find the time to dive deeper into the Annotation topic. At first glance this class seemed to be a placeholder for "other" information that would not fall into the three core classes Creative Work, Instance, or Authority. With the two papers describing the Annotation class very thoroughly, and with reference to the Open Annotation model, we got the impression that this approach is a promising one. We do have a lot of information in an equivalent to MARC Holdings, mostly in the context of the German Union Catalog of Serials (ZDB). There, nearly 12 million holdings of thousands of libraries are contained, belonging to 1.6 million bibliographic records. In addition, licenses are contained, with specific challenges to modeling (resource, institutions, rights, time), and quite complicated format solutions. Apart from holdings, table of contents may form specific use cases for Annotations in our context.

On the level of strings, we are unsure whether it still makes sense to provide labels formulated according to ISBD, particularly using ISBD punctuation. A 260 field is to be deconstructed into one or more place(s), one or more institution(s), and one or more year(s). Both places and institutions in this example may even be linked/traced/controlled. Bibframe should not prescribe ways how to put together strings taken from different MARC fields and subfields. Wherever possible, data is better than strings, and MARC elements containing data (in coded form, or in otherwise controlled form) have to be taken over into the Bibframe model as data.

The German National Library is the only institution in the Early Implementers Group without an English speaking background. This may even pertain to syntactical issues of the Bibframe model. At this moment, every entity, every type, every element is expressed in terms which are derived from the English language. In some cases or branches of discussion, agreements are made based on language solutions. The abstract layer that a MARC tag = field number, an indicator or code position value, or a subfield code provides always adds to a language independent availability and acceptance of the standard. ONIX does provide most element names in both a long, English language based version, and a short version, comprised of one character and three digits. We can imagine considering that this helpful feature may be taken into account for the new Bibframe

standard. In any case, apart from the naming a clear definition is to be developed for each class, element, type, and controlled value.

As a lower priority, we see the need for a distinction between a part of a label that has to be included when it comes to presenting the label, and which at the same time has to be excluded when it comes to indexing, listing and sorting of labels. These non-sorting parts of labels (titles, names, etc.) have two MARC solutions at the moment (an indicator that counts positions, or control characters). We would like to see both of them, and the need behind them, covered in Bibframe. ONIX e.g. has taken a similar approach by defining the <TitlePrefix> = <b030> and <TitleWithoutPrefix> = <b031> elements.

We expect that Bibframe as an evolving standard will be and will remain to be open to the public and free of charge. We welcomed statements that go into this direction. This policy will add to the acceptance and persistence of Bibframe.

Summing up, the German National Library up to now has seen the chances and the challenges of defining the Bibframe. It is a great opportunity to actively take a role in this early stage of a new carrier standard's life cycle. We see and will continue to see our obligation to be partner in this effort for the times to come.