Gunilla Jonsson
National library of Sweden
Leader of working group 4   2003-04-26

Janne Andresoo, from the National Library of Estonia has put a specific question to the conference:

"The bibliographic unit in digital context – how to define it?"

The discussion of the bibliographic unit as a general issue will be included in working group 4, which deals with multi-part structures. It seems appropriate, however, to make a specific comment concerning the digital context. My comment builds on the experience we have for some years back in the National Library of Sweden.

Firstly, I want to develop Janne Andresoo's question into three parts:

1. What is a bibliographic unit?

2. Is there always a one to one relationship between a bibliographic record and a bibliographic unit?

3. Do we need descriptions at a deeper level of granularity for the digital environment?

What is a bibliographic unit?

I suggest it is any publication or part of a publication that can be distinguished as some sort of independent unit. Clearly, you can carry that principle very far, if you wish. A single paragraph, line, or even a single sentence or phrase could be distinguished that way. I would even argue that all these cases are covered by the entities defined by FRBR:

works
expressions
manifestations
items

How far we carry the distinctions in practice is a function of the economical balance between the user tasks (find, identify, select, obtain) and what libraries are likely to be able to achieve, given the ever-growing output of intellectual creations.

Is there always a one to one relationship between a bibliographic record and a bibliographic unit?

The bibliographic record, basically, describes manifestations with the work-expression levels integrated, but the work-expression perception is decisive as to which manifestations or parts of manifestations to describe. The assumed user interest/need decides to what level of granularity we carry the cataloguing work. – Always counteracted by economic constraints, I

can't help adding! (Maybe, it should have been put the other way around: Economic constraints decide at what level of granularity we can meet the assumed user need.)

Our record structure is flexible, and we find examples where one record describes several manifestations (the different parts of a multi-part publication described in the same record), examples where independent works are disregarded (the different contributions in a festschrift or magazine articles, e.g., do not necessarily get any recording at all) or we can find the opposite, so called analytical records for contributions in festschrifts, anthologies, or journals. The latter, however, require a record for the host publication, to which reference is made in the analytical record.
Items may be accounted for in a designated part of the bibliographic record or in separate records linked to the bibliographic records.

The conclusion is that, presently, we choose the bibliographic units to describe, the level of granularity, rather pragmatically, and that different levels are accommodated in the same catalogues without difficulty.

Do we need descriptions at a deeper level of granularity for the digital environment?

My preliminary answer would be both yes and no. To explain that answer I will give a short account of our experiences with digital deliveries and a digital archive.

We have pdf files that comprise complete novels, like common paper books, just digital instead. The relationship is one work-one manifestation, and we make one record. – No problem.

We also have works split on several files, pdf and other files. We may have one file for the table of contents, one for the summary, one for the main text, one for images and one for statistical tables. The relationship is one work- one *reader* manifestation - one item in the database in several parts. This is actually the same situation as with multi-part publications on paper. The difference is that we get smaller, and consequently many more, parts in the digital environment, and we are not used to having table of contents or summaries as separate entities. We haven't yet had separate files for separate chapters of a book, but I expect that may come as well – we all know that it isn't practical to work with a substantial text in one single file! – Problems? Not necessarily for the traditional bibliographic record. We may, as for the paper publications, choose the level of granularity according to the perceived user need, and for the example I have given it would definitely be one record for all the parts making up that work. For the digital archive, the needs are different, and that is why my answer was yes and no. Further, the item level and especially the "obtain" functionality are affected. I will come back to those circumstances later.

Finally, we will soon have xml files with links to images and tables, which come as separate files, comprising one work. In this case we also have style sheets, which may be used for many different xml files, realizing different works. Even in this case the obvious solution seems to be one record for the manifestation making up the work. Again the relationship is one work- one *reader* manifestation – one item in several parts. – Problems? Yes, because the manifestation level is split in a way we have never seen before. We are used to handling manifestations where the layout, the typographical arrangement, is inextricably integrated.

Here it comes as a separate layer, and it is exchangeable. Typically, there may be alternative style sheets between which the user can choose, and they are also likely to change over time.

We think we know pretty well how we shall handle this in the digital archive, but we are not yet quite sure of how we shall handle it in the bibliographic record that is designed for the common user. The problems, however, lay in the descriptive area, and, perhaps we can leave discussion of changes to description for the meeting and our working group discussions.

Coming back to the digital archive, we do need archival records with a certain amount of preservation data about our files, but that will be handled automatically and will not affect the common cataloguer. What is very clear is, that for the archive, we need a record at some level for each manifestation, or more precisely each file making up the item of a certain manifestation, and that the archive, primarily, is not concerned with works. For the bibliographic database, we do not need separate records for all these separate physical parts of the item stored in our database, but we do need records which make it possible to find works-expressions at the level of granularity we perceive necessary, and we need to be able to select the manifestations and obtain items (single items or aggregates of items and even sometimes parts of items) from the archive. In order to do this, we must establish links between our traditional bibliographic records and the archival records. Presently, it seems as if the easiest way to achieve it is some kind of "hook", by which we can establish that link. Identifiers present themselves as better suited to act as such hooks than URLs. In the future, maybe, all information needed for different purposes can be present in the same database and different style sheets can be used to display the information needed in different user situations.

To sum up:
we will have more items per work in the digital archive than in the paper format. Every single file potentially constitutes a bibliographic unit, which needs a technically detailed item record, but that does not mean that we must have a separate bibliographic record for each part of an item held in the archive. We may have more expressions per work than what is common today (there is some evidence to that effect in our archive), and there may be a greater need for uniform titles, or authority records for works. Much of the manifestation information will probably be automatically generated, but it will need updating more frequently than today, and we will have to find some automatic solution for that. The linking system between the authority records, the bibliographic records, and the item records in the archive will be crucial.

The bibliographic unit in the digital context is not different from the bibliographic unit in the analog environment. It comes in different constellations and it has different physical properties, and we must build a new infrastructure to be able to obtain items from the digital stacks, the archive. There may be possibilities to produce bibliographic records for a deeper level of granularity than what we achieve for publications on paper, due to economical constraints. It will, however, require that text be structured according to specified standards to allow automatic extraction, or generation, of metadata. Cataloguing guidelines need to be extended to the field of document type definitions and the building of XML schemas, and cataloguing experts need to understand the new technical environment.