Internal and external interoperability of books metadata using work concept and semantic web technologies

Pierre Boudigues
IT director, Electre.

Joëlle Aernoudt
Quality Director, Electre database.

Stéphane Bizeul
IT manager, Electre.

Gautier Poupeau
Data architect, Antidot

ABSTRACT:

Metadata is a key feature of book distribution workflows in general, and e-books in particular. Traditional players in the book industry have to take into account the production workflows, quality and scope of their metadata, in order to be able to keep a leading role in the digitization process unlike what happened for the music industry. This issue has to be addressed when designing the digital publishing workflow. Metadata management happens at every step of the process and involves every player (publishers, book- sellers, librarians...) each at their own level. Their successful collaboration relies on the use of standards, identifiers and vocabularies, all required in order to reach the necessary interoperability level for exchanging, linking, and using the data they produce.

1) Features of metadata in the book industry: the work concept
Traditional bibliographic records are no longer sufficient to provide useful metadata for the digital world. New models like FRBR, centered on the notion of Work and links between entities, are required. Use case: the Work concept, based on the FRBR specification, developed by Electre and the migration of the traditional bibliographic database.

2) Internal interoperability: Sharing and aggregating data in the book industry
Several types of metadata go beyond the traditional bibliographic description:
- metadata related to the audience and success of the book like critics (newspaper articles, awards, media events) and user reviews (social networks, comments)
- metadata related to the content (places, stories, characters...)
- metadata related to the author (biography, book signing events...).

The different types of metadata mentioned above are not created at the same time and rely on different producers at different steps in the workflow. Aggregating and linking these data requires the common use of standards and identifiers (ISBN, ISTC, ISNI...). The example of Electre's data warehouse using Semantic Web technologies shows the downfalls of current practice in the book industry in France and demonstrates the gap that has to be bridged in order to successfully combine and use the data at a global level.

3) External interoperability: Linking book data on the World Wide Web
It is important that the data thus aggregated and combined is shared on the web outside the book industry, so that external users can take advantage of highly structured information provided by the traditional players in the field. Then book data producers can link their data to other datasets already available on the Web in order to create new services with real added value.
We will demonstrate this topic with the external enrichment of our Data warehouse by collecting open data contents (RDF data from BnF's Website, data.bnf.fr, Dbpedia, Wikidata, Wikimedia Commons).
It is necessary to build a common ecosystem for producers and users in the digital publishing workflow, including a set of standards and identifiers. Semantic Web and Linked Data standards provide a suitable framework for linking successfully different types of data from different producers and silos. The business model for metadata distribution still has to be explored in this perspective.
"