Poor and popular vs. Rich and lonely: can’t we achieve both?
Linked Data strategies at the BnF

data.bnf.fr
Bibliographic services in MARC

- BnF has served bibliographic records in (Inter/Uni)MARC formats for many years
- Records can be reused from the BnF
  - Via pick and choose « baskets » in the General catalogue (small sets)
  - Via Z3950 retrieval protocol (medium sets)
  - Via FTP protocol (bulk sets)
- Other French libraries use these records to avoid duplication efforts
  - Abes/Sudoc
  - Public libraries...
- International projects as well:
  - WorldCat, VIAF, European projects, etc.
OAI repositories in Dublin Core

- MARC formats proved too complex or specific to exchange our metadata in the framework of digital cooperation
  - Promoting contents from our digital library Gallica
  - Taking part in Europeana
  - And other digital library projects (World Digital Library...)

- BnF built an OAI-PMH repository
  - Widely used
  - Expands collaboration with other communities (archives, museums...)
  - But the DC profile loses too much of the bibliographic information in too many cases.
Since 2011, BnF is building the data.bnf.fr service to:
- Expose its bibliographic resources as linked and open data (with RDF & JSON dumps)
- Seek greater exposure and reuse
- As of today, 40% of the MARC records are available as RDF!

The data.bnf.fr is fit for standard reuse on the web.
- We lose some of our information but we give enough to many and have encouraging feedback as to reuse by communities we did not reach before
- But those are not libraries so far: the data.bnf.fr dumps are no acceptable substitutes for the MARC products
Challenges

☐ How to manage both ends?
  ■ Provide a simple and standard way to reuse our bibliographic data in an open, web environment
  ■ Maintain the possibility for other organization and communities with specific requirements to reuse our data in rich format but with modern, flexible means?
  ■ While avoiding the cost and burden of maintaining too many channels and data exchange protocols over time...
Let’s look at it! (1)

- Vintage linked data
  
  [http://catalogue.bnf.fr/ark:/12148/cb140202418](http://catalogue.bnf.fr/ark:/12148/cb140202418)

- Very fine grained information
- Library specific
- Targeted at finding / managing library assets
- Hard to extract reference information for 95% of the users
- References to the archives and manuscript catalogue
Let’s look at it! (2)

- The same in data.bnf.fr
  
  http://data.bnf.fr/ark:/12148/cb140202418

- MORE information at the same place
- Lightweight FRBR structure

- Enriched with more links (internal and external)
- Wider outreach

- LESS structured data (under the hood)
- Duplicated properties
Monster example

- Voltaire
  - 7646 bibliographic records in the « vintage », BnF General catalog
  - 62 authority records
  - 1 person authority record

- Summarized in **one page** on data.bnf.fr
- Enriched with new access points
  - (e.g: years / timeline)

[http://data.bnf.fr/11928669/voltaire](http://data.bnf.fr/11928669/voltaire)
Our roadmap

- **data.bnf.fr**
  - Keep its current datamodel unchanged
  - Keep adding more data from the vintage catalogues and other BnF sources
  - Add new services and links, in and out

- **alternative initiative** for library and specific communities needs
  - Finer grained information information in RDF
  - SPARQL as a new Z39.50
Fine-grained linked data: the goal

- Be closer to the standards’ state-of-the-art (FRBR, RDA)
- Lose no information
- Investigate fine-grained models (FRBRoo, PRESSoo)
- Joint project with French Higher Education bibliographic Agency

abes
agence bibliographique de l'enseignement supérieur
What could it do that data.bnf.fr does not do?

- Professionals often want fine-grained data (more than what we have in data.bnf):

- Professionals (librarians or NOT) may be interested in a very specific domain:

Some libraries may do not need fine grained data:

[https://demo.cubicweb.org/library](https://demo.cubicweb.org/library)
The future of bibliographic products?

- Most of patrons’ requirements and questions require a data extraction based on a request.

- SPARQL allows such queries to be done by the end user: pick up and chose the data you need – skip what you don’t need.
Articulation / needs / data:
Fine-grained linked data:

+ querying facilities to get any subset
What do we need for this?

USE

CASES
Use cases: why?

- Demonstrate the added value of fine-grained linked data in real life situations
- Tie the (complex, abstract) data model to concrete, specific needs
- Work step by step answering everyday needs rather than getting blocked in a modeling tunnel
- Lower the barrier to entry
Data flow: Identifiers to rule them all

- BnF Archives and manuscripts
- BnF General Catalogue
- ARK identifiers
- InterMarc
- RDF
- Fine-Grained linked data
- ARK identifiers

Alignments soon to be rolled back in the source data

- data.bnf.fr
- RDF HTML

EAD
Thank you for your attention!

gildas.illien
sebastien.peyrard
AT bnf.fr