Data.bnf.fr as a sandbox for FRBRization

Automated work creation in data.bnf.fr
Five entities...

Data.bnf.fr

Author (2M)

Place (117K)

Date (2K)

Concept (188K)

Work (263K)
...two faces

http://gallica
/ark:/12148
/btv1b90519196
The interface
The data
“Old works” at the BnF: a handcrafted artefact...

Validity control = persistence guarantee

https://catalogue.bnf.fr/ark:/12148/cb14473195c
Where to start?
We need ...
● a homogenic corpus of documents → the XXth century authors.
● an exhaustive collection of records from the legal deposit.
● A highly configurable robot which likes every kind of metadata…!

... and to keep it simple: no “aggregates” records!
Then, from titles clusters, generate the two faces...
The interface...
The data
...Calendar Information
First semester of 2019:
  ○ uploading computed works in the data.bnf.fr interface
  ○ Validation process

Second semester of 2019:
  ○ Uploading computed and validated works in the catalog
  ○ Attribution of permanent URIs
Concomitantly...

Evaluating the quality of the Main Catalog metadata:
- date: content and coherence
- title: content and structuration
- author: homonyms et function codes
- Language

Curation of the metadata in order to improve clustering performances
After works’ integration into the Main Catalog...

- **Side projects**
  - Non textual works
  - Foreign works
  - Before 1900 works
  - Expressions

- **“Benchmarking”**
  - Linking toward the ABES computed works to check validity of newly created works at the BnF
Thank you for your attention!