Darling, we need to talk
SETTING THE SCENE
From building to building
From Web Archiving to Linked (Meta)data
Leave me alone, I’m very busy saving the world.

Nobody understands me nor sees the beauty of my datamodel anyway, go away.
[he’s such a boring snob, isn’t he aware that catalogues are completely out?]

[How vulgar! I can’t believe she’s even collecting websites. Her repository is pure trash.]
The future is now and we’re getting short of money. We’re asked to demonstrate short term benefits to long term efforts.

We can no longer afford to maintain activities in silos. We need strategic and global visions, policies and organizations to improve the consistency and sustainability of our digital workflows.

We need to act in a global, online context. We ought to make the most of our existing networks and reach out new cooperations where needed.
OPENING QUESTIONS

- Preservationists and bibliographists, what do we have in common?
- Stuck for 7 days on a desert island (or in Lisbon?) what is it you would like to discuss?
PROPOSED AGENDA

- Governance, policies and advocacy
- Organizations, staff, roles and skills
- Ownership and identity in an open data world
- Some core issues for all of us: Identifiers, Vocabularies, Languages
1- GOVERNANCE, POLICIES, ADVOCACY
GOVERNANCE
POLICIES
ADVOCACY (local)

- Within organizations, growing needs for
  - Internal advocacy
  - Proper decision-making bodies
  - Transverse analysis of processes and workflows (rather than projects in silos)
  - Long term, high level policies

- Issues
  - Advocacy towards high-level management
  - Consistency of decisions
  - Policy means dealing with value
Advocacy:
External outreach might be our best internal advocacy.

Presidents Racine and Hollande at the Paris Bookfair, 2013

2013 SPIRL winners

- **Bibliothèque nationale de France**
  
  **Projects:** Gallica Library and Discovery Service
  
  The Gallica Library promotes French cultural heritage for scholars as well as a much wider public in a rapidly changing digital world. The Open Data of the BnF, Data, has integrated numerous BnF collections and made them available via an Open Data API. Together, both efforts drive a wider audience for the library’s content.

- **Biblioteca Virtual Miguel de Cervantes**
  
  Service oriented platform for the management of modern research digital libraries with a more open, user-oriented design and open-source development. Project submission. (See judges’ comments.)

Commendations of Merit

- **Griffith University**
  
  **Project:** Research Hub
  
  Research Hub addresses the need for a single, comprehensive, and accessible research platform that connects researchers, academics, and students across faculties and disciplines.
Advocacy: Metaphors and narratives are more effective than facts and figures.
Under the hood, Google is very complicated too. It might take complex datamodels and rich metadata to make an interface simple and intuitive.
Don’t be boring

- Focus on outcomes – service and marketing.
- Show what data can DO for your people, don’t even try to explain HOW.
- Datamodelling and bibliographic are only sexy to those who make them.

« What you need to do is to get people addicted. »

David Tarrant, OKF
Adjusting our internal governance structures: Trends at the BnF

- **Preservation**
  - Introducing digital preservation within an existing high level instance for analog conservation
  - Defining a proper level of ownership and planning for digital collections at middle-management level
  - Creating a product owner role for the digital repository and trying to organise decisions and planning

- **Metadata**
  - Turning the former bibliographic committee into a *metadata committee*
  - Change of scope: deal with *all* types of metadata
  - Creating a strategic instance at CEO level, which did not exist for these matters
  - Prioritizing the management of products & services rather than the discussion of cataloguing rules
Putting new stuff in old boxes?

- CEO board
- Conservation committee
- Metadata committee
- Legal deposit committee
- Middle management, product owners & experts
- Middle management, product owners & experts
- Middle management, product owners & experts

SPAR

data.bnf.fr
Policies: the hard piece

- It is, politically, very hard to define and quantify VALUE when it comes to Heritage or Research.
- Assessing priorities means assessing non-priorities as well.
- Any policy, if properly articulate, has a symbolic, social and economic impact on the overall organization. This message may be violent.
- It takes brains AND guts to have a long term and straightforward policy for metadata or preservation planning.
- Experts strongly depend on this strategic engagement.
Visions for the digital collection

- Policies come with a vision of the digital collection.
- Digital means: think the collection globally from the beginning to the end of the workflow (digital document lifecycle) – not in functional silos.

Collect

Conserve

Curation

Signal

Communicate
Visions for the digital collection (2)

A shift in paradigm?

- Before: collections → data
- Now: data → collections
- The same bulk dataset can be the basis of many collection views/facets/services
- Data is tangible. Collection will vary, adjust to context, policy, service and usage.
- Should we discuss data policy rather than, or together with, collection policy?
GOVERNANCE ADVOCACY POLICIES (global)

- Between organizations
  - Standardization (how to promote it?)
  - Publication and co-maintenance of vocabularies and registries
  - Institutional initiatives / crowdsourcing?

- International and local policies
  - Good knowledge of the international context helps defining your own added value...
  - ...including your digital facilities or skills (it is no longer only about contents distribution)
Digital communities: - motivations - intersections

GA in Ljubljana, 2013
Next IIPC GA: PARIS
May 19-23

Portuguese Web Archive joined IIPC this summer!
2- ORGANIZATIONS, STAFF, ROLES, SKILLS
Manual / Automated processes
- Preservation started with automation but is now requiring more human intervention
- Bibliographic started manually but is now requiring more automation

Opportunity to share best practises in:
- Processes & workflows
- Quality
- Organisation & skills

Explore computational curation
STAFF, ROLES & SKILLS

- Talents and knowledge at risk
  - Demography, generations: those who know, whether young or old, will soon go
  - Lonely networks: a risk of confinement

- New roles, new borders (no borders?)
  - IT vs. curators – operations, expertise, management: need to work differently together
  - Agile methods and project-based communities

- Tip: advocate and anticipate!
  - Empower your people
  - Make them move before they move
future-ex preservation metadata expert

ex-web archiving preservation expert

future LOD metadata manager

New web archiving lead

MAKE THEM GROW, MAKE THEM SWAP
Incubators and forums to discuss digital issues?

The ORHION initiative: BnF’s « Digital Anonymous »
BACK TO THE FUTURE

☐ It’s not only about getting senior staff interested in the new stuff.

☐ Junior staff also need to look back at the past (physical conservation, old bibliographic wars and accomplishments). Different tools, same problems.
3- OWNERSHIP AND IDENTITY IN AN OPEN DATA WORLD
OPEN DATA
OWNERSHIP
INCOME

☐ Key issues
  ■ open source code sustainability
  ■ open (meta)data licences

☐ Shared paradoxes
  ■ Request to generate new income
  ■ Request to open the code and the data

☐ Alternative business models?
  ■ Sell services, not data?
  ■ Demonstrate benefits rather than income?
Some ideas
Some fears

- **With preservation**
  - Selling service to third party vs. dealing with your own stuff? (but what if this actually pays to deal with your stuff?)
  - Open software co-development: are institutions fit and ready for this?

- **With metadata**
  - Is it OK for all to use public money to help companies making business out of it?
  - If I stop selling my metadata, does it mean nobody will try to sell them back to me?
  - Are we fit for selling our expertise?
Open Licences:
The French context

Freely reusable data...

... with obligation of source citation.
A NOTE ON PROVENANCE

- Is different from ownership
- Is very much needed for internal management, including rights
- Is very much needed for external reuse, citation, sourcing and data qualification
- Crowdsourcing is acceptable in metadata only if you have provenance information
Use case: bringing together web archive and linked data

☐ A case for local collaboration between preservationists and bibliographists:
  - For legal reasons our Web archive is only accessible on the Library premises.
  - For economic reasons, our Web archive isn’t accessible from our catalogue so far.
  - We’re testing the possibility to expose some archived websites metadata (websites used as sources for authority data) as part of our LOD service data.bnf.fr

☐ Demo: www.data.bnf.fr
4- SOME CORE ISSUES: IDENTIFIERS, VOCABULARIES, LANGUAGES
IDENTIFIERS

- Common needs for unique, trustful and persistent identifiers
  - For internal management purposes
  - For external, sustainable discoverability and cooperation

- Joint issues
  - Interoperability (global identifiers)
  - Application domains
  - Governance, cost and accessibility

Identifiers are the bridges between data silos. A bridge needs to be robust: identifiers require policy, governance, good management.
Growing need for open vocabularies
Growing need for open registries

- Ontologies
- File formats
- Environments...

How specific, how generic?
- Poor & popular
- Rich & lonely
- Mixed strategies & mapping
Today, shared languages...
- XML...
- RDF...
- Sparql...

But above all and over time, it is the language (and standards) of the Web (and whatever will come next) which should be our commun tongue
- Avoid communautarism
- But don’t throw away your added value for the sake of being generic
In conclusion...
I always felt our paths would meet some day.

We’re equally weird and smart. There’s so much more for us to share.
Did you hear about this joint conference in Lisbon?
keep in touch:
gildas.illien[at] bnf.fr