From content silos to an integrated digital library

Digital Libraries and Repositories

Prague 2013

Anders Conrad

Could have been:

- *From 400 years of print library to digital library: a practitioner's view*

- Digital is no longer just an add-on to the conventional library

- Digital collections are becoming a primary task

- We must provide digital infrastructure to support this change

Overview

- The Royal Library
- A bit of digitisation history
- Current challenges
- Our technical solution
- Library (technology) issues
- Conclusion

The Royal Libray

- National Library
  - Legal deposit of print or web publications, obligation to preserve
  - Special collections (manuscripts and private archives, newspapers, pictures, maps, music, theatre, judaica and orientalia)
- University Library for University of Copenhagen (UCPH)
  - Lending library
  - Licensed materials
  - Institutional repository functions (ETD's, student self-upload papers)
- Shared discovery solution (rex.kb.dk)

Digitisation history

- 1990'es:

- First website around 1994

- Digitised manuscripts, images, texts, sheet music

- Specialised applications and databases

- Rich resource sites and web exhibits

- Focus on low resolution images for web

- Master files on tapes, CD-ROMS, or ….

Digitisation history, contd.

- 2000'es:
- Goal of "mass digitisation"
- Larger projects: corpora instead of single objects
- Adl.dk, tidsskrift.dk, illustrerettidende.dk
- Content-driven dissemination sites
- Introduction of digitisation infrastructure
- Consolidation of 1st generation webapps, but
- Still a plethora of databases, repositories, data models, metadata templates, workflows, etc.

Main challenges

- Current technology not sustainable for continued digital collection building

- Unsatisfactory data management and metadata features

- Born-digital content
  - Legal deposit: files rather than books
  - Special collections: hard disks instead of sketches
  - New sources of materials, e.g. social media

- Open linked data

- User-supplied content and annotations

Strategy

- "The Digital Royal Library", official strategy since 2009

- Technology: one integrated digital library infrastructure for future digital collection building

- Organisation: National Library re-organizing into functional divisions
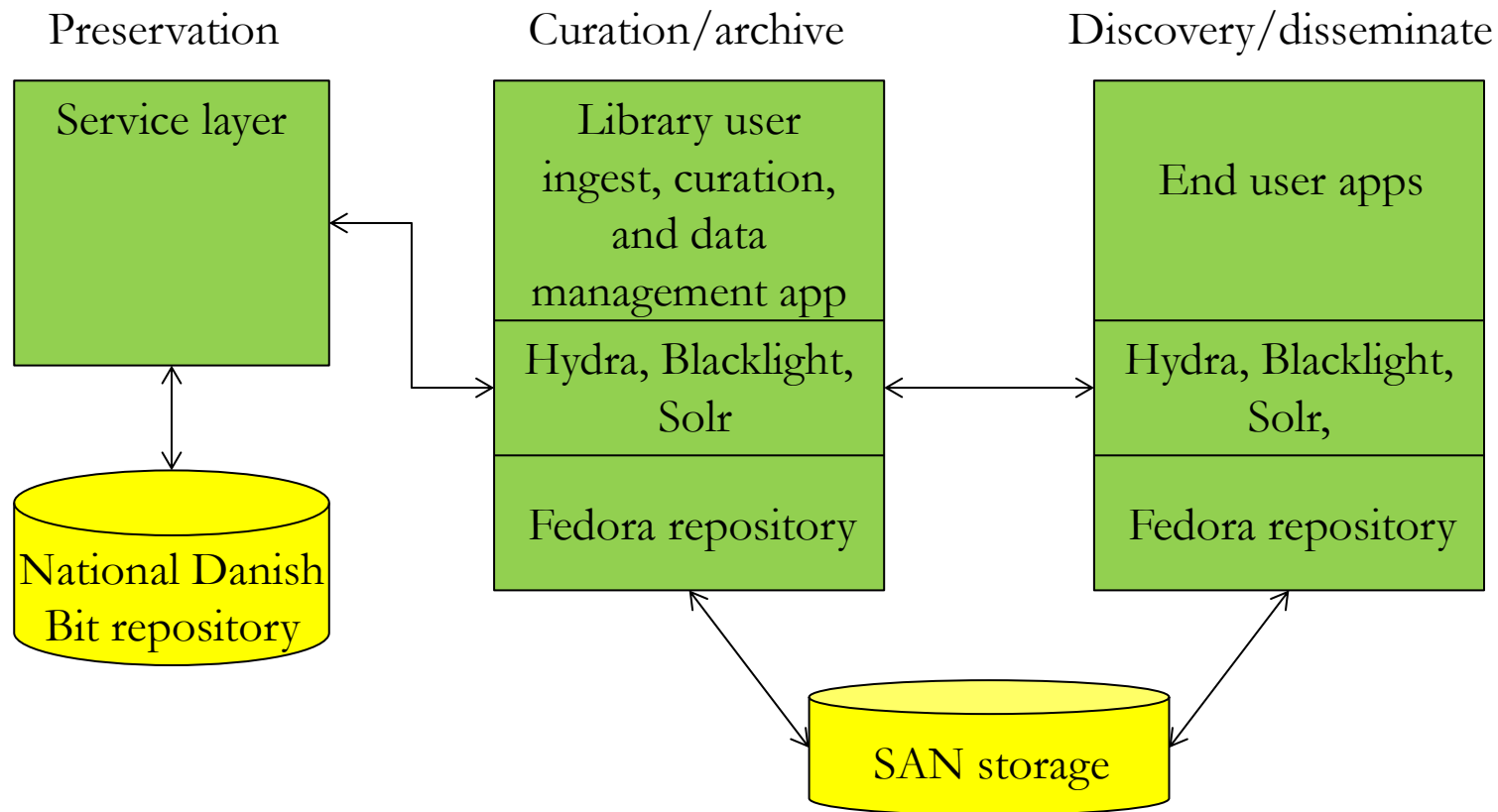
The technical process

- First steps: implementation project 2011-14
- Requirements process 2011
- Agile development, 3 month release cycle
- Trying to involve collection people as *product owners*
- Basic assumption: one repository will hold any kind of material and support all workflows
- Future applications and digital library solutions to be mapped into new infrastructure
- Infrastructure will be extensible after end of current project

Technical vision and architecture

Integrated infrastructure with three main parts:

Preservation	Curation/archive	Discovery/disseminate
Service layer	Library user ingest, curation, and data management app	End user apps
	Hydra, Blacklight, Solr	Hydra, Blacklight, Solr,
	Fedora repository	Fedora repository

National Danish Bit repository

SAN storage

Technology stack

- Fedora Commons – backend repository
  - http://www.fedora-commons.org/
- Apache Solr – state of the art search engine
  - lucene.apache.org/solr
- Hydra and Blacklight – Ruby application framework for digital libraries
  - Projecthydra.org
- The National Danish Bit Repository

Why Hydra 1?

- Open source framework to facilitate development in new directions

- Robust, state of the art technology

- Build specialised applications on top of one single repository ("One body, many heads")

- Existing code and solutions we can reuse

Our Dissemination Architecture
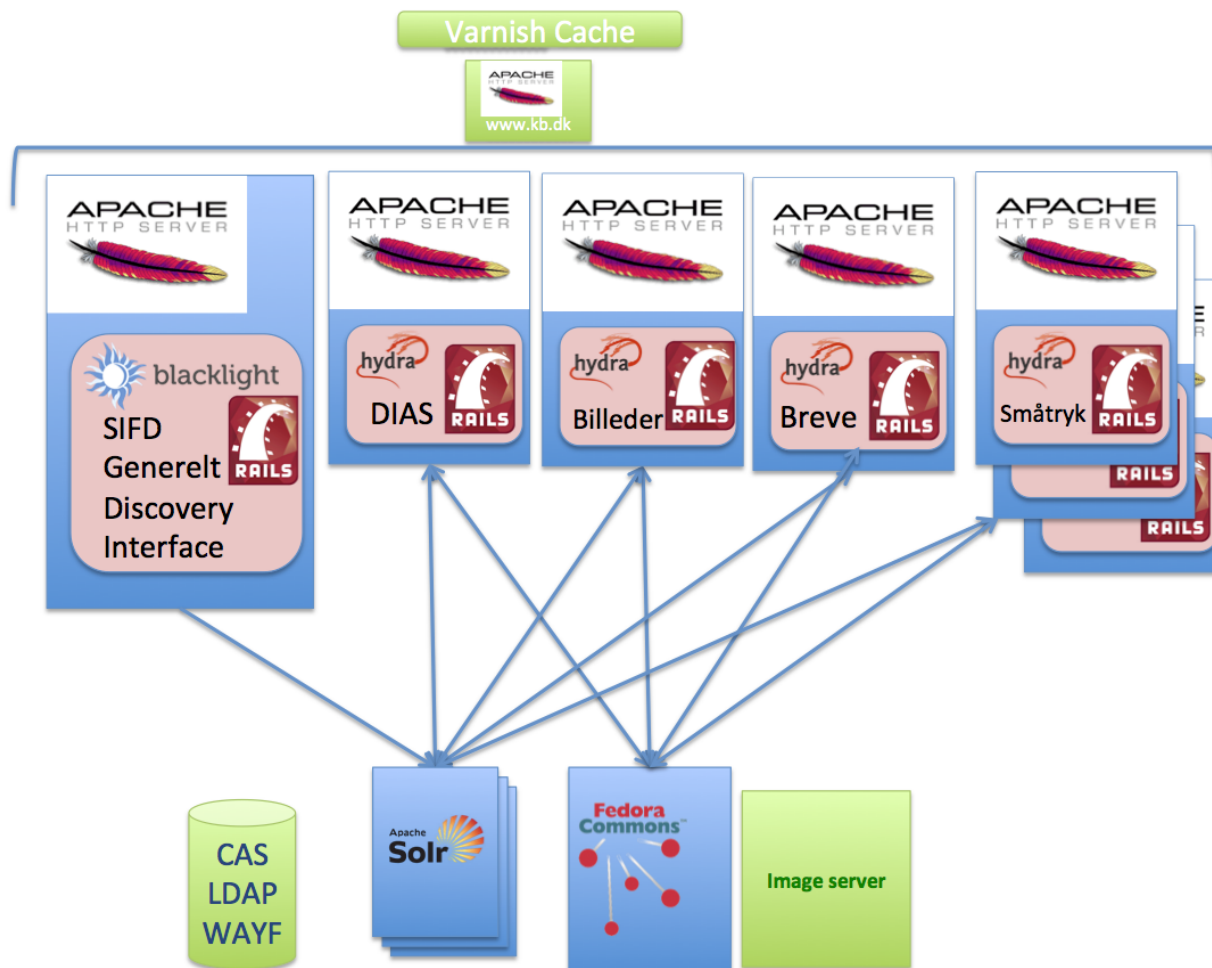
Why Hydra 2?

- Vibrant and growing community
- Embedded in Duraspace community
- A lot of expertise we can draw on
- 3 European partners, a few more adopters
- Space for more!

Hydra community focus just now

- Establish solution bundles/verticals
  - Institutional repository
  - Archival management
  - Image collections
  - Time-based media
  - Exhibits
  - Research data collections
- Fedora 4 development and testing
  - Steering group member
- RDF support for metadata modelling
  - Hydra Connect meeting January 2014

Library technology issues

- While storing everything in one repository:

- How to deal with different types of metadata?

- How to deal with different workflows?

- How to maintain professional standards and special features of various types of material?

# Analysis and strategy

- Support of strategic tasks, such as
  - TEL, Europeana
  - heterogeneous material
  - new business areas: research data, bit preservation
  - Open Access to our data
  - user supplied content
  - linked data
- Data management life cycle (DCC, DDI, OAIS)
- Reference models (FRBRoo)
- Standards (METS, MODS, PREMIS, …)

Metadata model

- Hope to be able to map everything to internal representation, possibly in RDF:
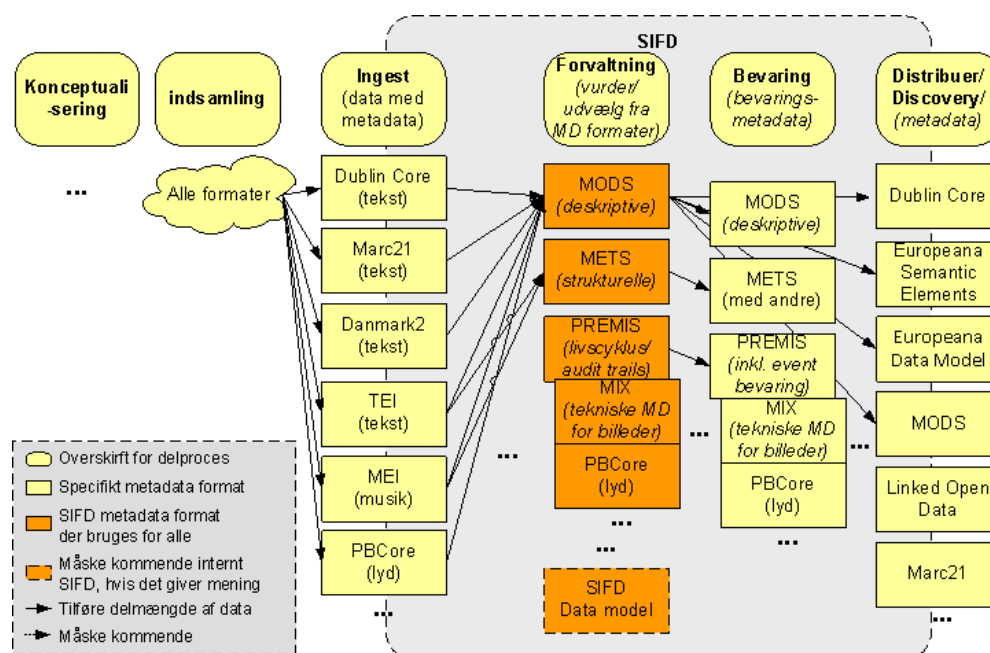
Conclusion

- A change project for the entire organisation
- Room for both organisational and technological development
- Align technological development with organisational change
- All involved parts must listen and learn
- Increasing interest and involvement from collection staff
- Challenging to move a centuries-old organisation!