Dataflow for Research Data at the BGBM

David Fichtmueller

Freie Universität Berlin

Botanic Garden and Botanical Museum Berlin-Dahlem

Funded by

Software for Management and Curation of Data

- **JACQ** (as presented yesterday)

- reBiND (special datasets)
- EDIT (taxonomic data)

Dataflow for Multimedia Data

- mediastorage.bgbm.org -> FSI Server

- Users upload images

- File names based on convention: containing collection numbers

- Nightly script checks for new images to link to corresponding datasets

- Hosting of multimedia data for other institutions as well (e.g. DSMZ)
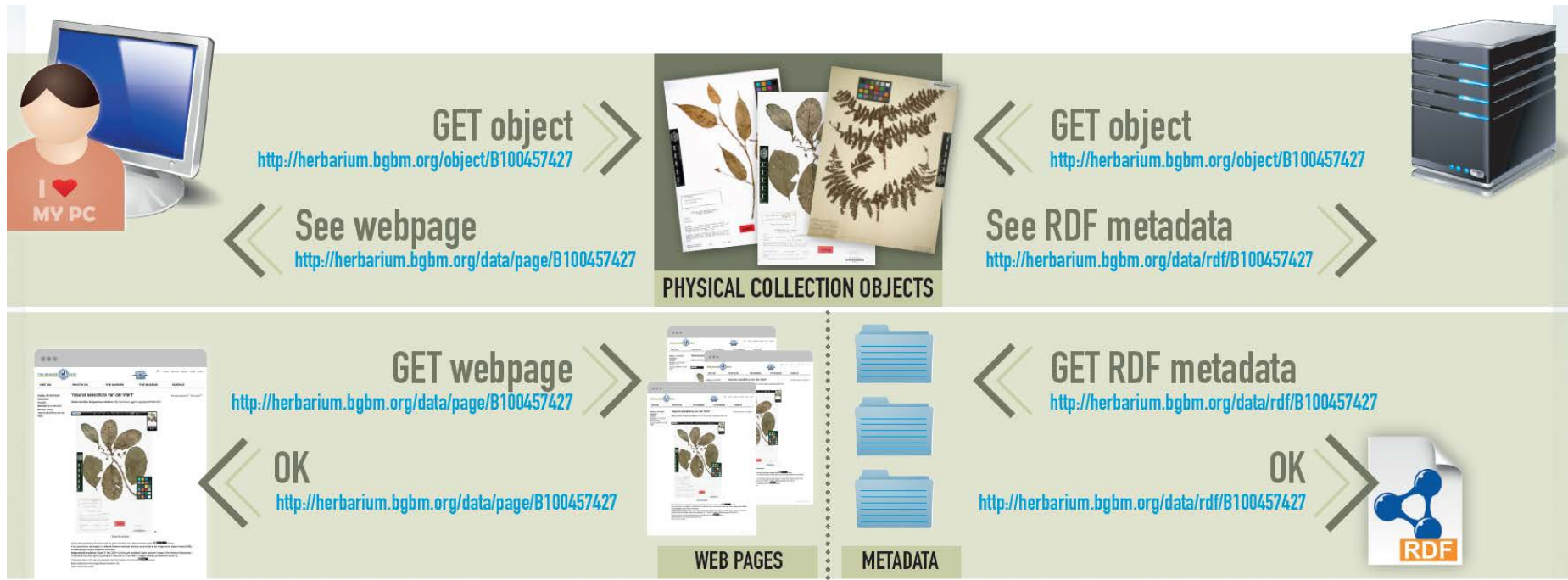
- Generate Views in Database

- Mapping in Biocase

- Register endpoint to be indexed/harvested

- BioCASe: 3.5.2 (current)

- ABCD: 2.06b (default) and 2.1 (BiNHum)

CC BY SA Steven Byles

Usage of 'datasources' and 'datasets'

- Usually one data source (endpoint) for each dataset

- Exception:
  - reBiND: one endpoint, many datasets

- Maybe in the future:
  - One dedicated GFBio endpoint, many datasets?

Usage of Identifiers

- Stable Identifiers according to the CETAF-ISTC Suggestion
  - Resolvable HTTP-URIs
  - Content Negotiation based on HTTP Accept-Header
    - Website with Image Viewer for Humans
    - RDF Data for Machines

- Collection number is part of the URL
  - http://herbarium.bgbm.org/object/B100457427

- Default
    - Metadata: CC-0
    - Multimedia: CC-BY

- Individual data sources can have different licenses