BHL-E Technology Workshop Agenda
Notes (Chris Freeland)
15 May 2009

WP2 Presentation - Bernard Scaife & Kai Stalmann
Big Picture
- Start with the deliverables, possibly restructure tasks
- Have to quickly deliver D2.1 Catalogue of content holder requirements
- Need to discuss D2.3 German UI
- Support a pragmatic approach to deliverables
- Need to discuss interfaces to EDIT & EUROPEANA
- Will engage content providers in standards selection process
    - Discussion:
        - Standards are shared across WP 2 & 3, so activities need to be in sync

Roadmap
D2.1
1. Circulated existing BHL specs for comment & review via BHL wiki
    a. Complete by 30 June 2009
D2.2
1. Use existing seriallist as prototype to develop method to ingest BHL-E partner catalogues' serials records.
    a. Define minimum fields to be supplied for MARC records & mappings to other MARC flavours. Produce standards to allow supply of metadata about whether digitized
        i. Complete: by 15 July 2009
    b. Acquire and Import MARC records (not deduped) to system (as hidden dataset?).
        i. Put in as 'hidden' because it might confuse users of existing system for production
        ii. Complete: 14 days (mid-Aug 2009)
        iii. TIME RISK
    c. Pseudocode of deduping algorithm (including manual verification option).
        i. Need code to be portable
        ii. Complete 1 month (end Aug 2009)
    d. Code and run the deduping and test/launch
        i. Complete: (end Oct 2009).
        ii. TIME RISK
    e. Create link with BHL portal to note what has been scanned
        i. Currently don't have this in place, but needed
        ii. Need considerations with CDL, other IA content
        iii. Complete: end July 2009
            1. Already begun
D2.3

1. Use existing seriallist as prototype for mono mashup with ability to bid. Acquire records at same time as for serials
   a. Complete: mid-August 2009
   b. TIME RISK
2. Investigate Goobi
   a. Do we implement cross BHL Goobi or just produce standards/advice do
   b. Complete: Sep? 2009
3. Implement Goobi or produce the standards
   a. Complete: Nov/Dec 2009
4. Integration with EDIT (ViTaL) for full text linkage from EDIT
   a. Complete: Ongoing, next 3 months for expansion to rest of EDIT community

D2.5
1. Migration to full system when architecture in place
   a. Complete: After yr 1

*DISCUSSION:
- Enhance questionnaire to include technical description of digitized content
- Invite Goobi to the upcoming meetings in Leiden and/or Prague

WP3 Presentation – Adrian Smales

Task 3.1
- Other WP depend on the standards/frameworks described in this task

3.1.1 – Management of the tech dev team
- D3.1 – Deliver composition of Tech Management board & Initial Mtg (M3)
  o DONE!
- D3.5 – Technical Architecture status & progress report with particular focus on the development of the German Prototype
- Team members
  o WP2, WP3, AIT, ATOS
- Time Line
  o D3.5 : Month 11 – 12
    ▪ Report that provides the funding for YR2 funding
- *DISCUSSION:
  o No more funding for this project
    ▪ Not Research, so needs a relatively low risk profile
    ▪ Need to contain & limit the exuberance of our partners
      • Scope creep
  o ATOS eager to work together, help

3.1.2 – Adaptation of EUROPEANA & BHL data model workflow
- D3.2 – Document agreed standards, best practice & components
- D3.4 – Implement plans for all components in Wp3
- Team members:
  o NHM, AIT, ATOS, MOBOT, Europeana
- Time Line
  o D3.2: Month 1 -2

- o D3.4: Month 5 – 7
- *DISCUSSION

3.1.3 – Definition of standards for images, metadata, harvesting
- D3.3 – Plan for managing interoperability issues
- Team members:
  - o NHM, AIT, ATOS, MOBOT, Europeanea, Tech Board
- Time Line
  - o Month 3 – 4

3.1.4 – Liaise with scanning centers of the
- No specific deliverable
- Team members:
  - o WP3 & WP2 Leaders
- Time Line:
  - o Month 4-5

3.1.5 – Development & adaptation of tools; name finding, OCR
- Deliverable post German Prototype
- Team members
  - o WP3 & WP2, Impact, EDIT, uBio
- Time Line:
  - o Month 18 – 24

Communication Methodology:
- PRINCE2, Project Server, BHL Wiki

Task 3.3
- Addressing distributed access & storage
- BHL-E is a catalyst for development of this data center, but outside true scope of project
- Needs a business model
- Team members:
  - o NHM, NMSI, V&A
- Time line:
  - o 24 – 30 months
- Communication Methodology
  - o PRINCE2, Project Server, BHL Wiki
- Data Center – "Darwin Repository"
  - o Green data center
  - o Multiple integrated repositories reducing the cost of storage, DR
  - o Storage & playout for major international projects
  - o A scanning center for high quality, low cost digitizations
  - o A dark library store for digitized library stock
  - o Suitable location found
  - o Need Economy of Scale
    - Provides additional revenue
    - DR & BC for all museums
  - o E-Infrastructure European initiative
    - Building digital repositories for Scientific Communities

- PESI (Biodiversity)
  - EU project to bring together big taxonomies
  o Proposed location
    - Swindon
    - Wroughton Science Museum
    - 15 aircraft hangars
      - 4,000 sq m each
        o put in 2nd floor, double space
      - 1 ½ available for use
  o Technology Partners
    - Many, discussions ongoing, may be at different stages
  o Plan B:
    - Imperial College near NHM

WP3 Presentation
Task 3.2 – Roger Essoh
- Review slides
- Consider Shibboleth
3.2.1. – Data management standards
- Need a framework for compliance with EUROPEANA & international requirements
- Team:
  o WP2 & WP3, EUROPEANA, Mobot
- Time Line:
  o M3
3.2.2. – Distributed data model to allow local control while enabling access through other portals
- Team:
  o WP2 & WP3, EUROPEANA, Mobot
- Time Line:
  o M4
3.2.3 – Prototype distributed data system for German language materials,
- integrate with BHL Portal
- Release candidate 1.0 of BHL-E Preservation & Archiving System
- Team:
  o WP2 & WP3, Mobot
- Time Line:
  o Report: M12
  o Prototype delivered: M18
*DISCUSSION:
- Use Mindomo.com
  o Online mindmapping
- Need to move quickly & manage risks

Task 3.4 – Walter Koch

- Key partners
  - AIT, ATOS, NHM, EDL, FUB-BGBM, NHMW, LANDOE, RMCA, MNHM

3.4.1 – Work with EUROPEANA to create multiple language access
- Liase with EUROPEANA-Connect on
  - Portal Language
  - Query Language
    - Eg www.dismarc.eu
    - Consider Z39.19 (2005 version)
  - Data Presentation Language
    - Native?
    - Live translation?
- Consider landing pages, search pages
- Results: User & Functional Requirements
  - Contribution to D3.3 & D3.4

3.4.2 – Create German prototype portal
- Based on requirements from 3.4.1, 3.2.3 and BHL portal specs
- D3.6 – BHL-Europe Portal
  - German instance – V1
- Contribution to D3.5
- *DISCUSSION:
  - DC probably insufficient for BHL needs; will need to review & suggest extensions
  - Disseminators
    - Subset of BHL will go to EUROPEANA

3.4.3 – Apply model to key European languages
- Should not consider list in proposal a complete list, but a priority list
- Based on BHL-E portal
  - German Instance V1, D#.9
- BHL-E Portal
  - German Instance V2
    - Contribution to the "Live System"

Example: DISMARC – Discovering Music Archives
- www.dismarc.eu
- Metadata store

*Suggestion: Use task brief; will be submitted to all partners

EUROPEANA
Technical Objectives & ESE specs
Julie Verleyen

WP4 Objectives
- Produce project website
- Set up open source infrastructure
- Ingest content
- Produce 2 major releases + 3 minor
  - V1.0 (Rhine) – July 2010

- - V1.2
  - V1.3 (Danube)
- EuropeanaLabs.eu
  - development space for Europeana
  - rapid prototype
  - access to collections for tests
  - Repository
    - All code + data
    - Download releases of software, dev stack, API
    - Partners/ext developers
      - Build components
      - Extend functionality
  - Sandbox (June 2009)
    - Environment for:
      - Test/validate content contribution
      - Test/validate technology contribution
    - Allocation to:
      - Partners of Europeana group of projects
      - Other projects, Open Source community
    - NOT a development platform
      - Validates results, **not** tools
    - Ingestion tool
      - Dashboard
      - Importing
      - Indexing
      - Caching of thumbnail images
      - Enabling/disabling collections
- Ingest content
  - Target by July 2010 – 10 million objects
  - BHL a contributor
- Ingest workflows
  - OAI-PMH
- Ingest planning
  - What & how big?
    - Size of datasets
  - By when?
- Tools & Docs for aggregators/content providers
  - Europeana Semantic Elements (ESE) Spec v3.1
  - Mapping & Normalisation guidelines for Europeana Prototype (draft)
  - ESE XML Schema
  - Sandbox (June 2009)
- Europeana Semantic Elements
  - "Schema" for the prototype
  - DC-based with Europeana-specific extensions
    - 39 elements
  - Elements

- **Country**
  - Country contributing the content
- **Provider**
  - Institution providing content
- **Language**
  - Language of provider's country
  - ISO639-1 language codes
  - "mul" multiple languages allowed
- **Type**
  - Type of the original object
    - 4 types
      - Image
      - Text
      - Sound
      - Video
- **Year**
  - Date of the original object

*DISCUSSION
- BHL-E needs a sandbox
- Probably best addressed in WP3
- Will non-European materials be submitted to Europeana?
  - Content from China?
  - What about all the content from US?
    - BHL
    - IA
    - ***It is approved

Time Line of Deliverables for first 12 months
Month 3
- D2.1 – Responsible: Bernard Scaife
- D3.1 – Responsible: Adrian Smales
Month 6
- D3.2 – Responsible: Walter Koch; Team: Bernard, Kai, Chris, Adrian, Roger Essoh
- M3.1
- D3.3 – Responsible: Adrian Smales; Team: Julie Verleyen, Kai Stalmann, Roger Essoh
  - Task 2.3.3
  - Task 3.2.1
  - Task 3.1.3
Month 9
- D3.4 – Responsible: Roger Essoh; Team: Adrian Smales, Chris Freeland, Kai Stalmann, Walter Koch
  - Task 3.1.5
  - Task 3.2.2
- D2.2 – Responsible: Tom Gilissen; Team: Bernard Scaife

- - Task 2.2.1
  - Task 2.1.2

Month 12

- D2.3 – Responsible: Kai Stalmann; Team: Bernard Scaife
  - Task 2.1.3
- D2.4 – Responsible: Kai Stalmann; Team: BHL-E
  - M3.2
  - Task 2.3.1
  - Task 2.3.3
  - Task 2.1.1
  - Task 3.2.1
- D3.5 – Responsible: Adrian Smales; Team: BHL-E
  - Task 3.1.2
  - Task 3.2.3
  - Task 3.1.5
  - Task 2.3.1
  - Task 2.3.3
  - Task 2.3.5
  - Task 3.4.1
  - Task 3.4.2
  - M3.2