

CATCH



CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

Persistent Identifiers for Audiovisual Archives and Cultural Heritage

Hennie Brugman

Technical coordinator CATCHPlus

Max-Planck-Institute for Psycholinguistics
Netherlands Institute for Sound and Vision

EuropeanaConnect ERDS Meeting - Frankfurt, February 17-18 2010



Summary

- CATCH & CATCHPlus
- Requirements from CATCHPlus, CH and AV
- Our solution
 - Base technology
 - Identifier management
 - Organisational embedding
- Application to collections
- Concluding remarks



CATCH & CATCHPlus

- CATCH research program by NWO (14 projects)
- CATCHPlus valorisation project
 - 8 subprojects at large CH institutions
 - Connected by common services
 - Vocabulary services
 - Annotation services
 - Infrastructural: OAI-PMH, **persistent identifiers**
- Project bureau hosted by Netherlands Institute for Sound and Vision
- www.catchplus.nl



Requirements from CATCHPlus, Cultural Heritage and Audiovisual Archives



Requirements (1)

Software support

- Good resolving service available
- Proven technology, stable and 100% reliable
- Scalable, with respect to
 - Number of identifiers
 - Performance
- Globally working solution
- Redundant hosting and service providing
- Identification of parts of objects (AV, CH)
- Possibility to associate metadata with an identifier (AV, CH)
- “Actionable”: identifiers can be resolved using http URIs
- Support for identifier management tasks



Requirements (2)

Identifier management

- Identifier management should be independent of
 - System management
 - Web server management
 - Hosting of resolution services
- Can be done from the context of a collection management system
 - typically by a responsible collection manager
- Is efficient, powerful and simple



Requirements (3)

Organisation, policy

- What choices are made by partner institutions ?
(the fewer 'flavours', the better)
- Reliability and sustainability of the service providers
- Limited and controlable costs
- Freedom to switch between service providers



CATCHPlus solution

EuropeanaConnect ERDS Meeting - Frankfurt, February 17-18 2010



CATCHPlus solution: base technology

- Based on Handle technology
 - Best meets our requirements
 - <http://handle.net/>
- One Local Handle System and Handle prefix per participating Naming Authority
- Hosted by SARA, (eventually) mirrored by other EPIC partners (redundant hosting)
- Redundant resolving is inherent to Handle System



CATCHPlus solution: identifier management

REST web service

- For resolving, searching, creation and management of Handles (in one's own Naming Authority) over internet
- Also support for batch operations ("move collection")
- SARA has built the first version for CATCHPlus
- Available as Open source
- Ambition: uniform redundant service by EPIC

User interface will be developed (Q1-2, 2010)

- Prototype for evaluation by collection managers



CATCHPlus solution: organisational embedding

- **EPIC** (European Persistent Identifier Consortium)
 - SARA (Netherlands), CSC (Finland), GWDG (Germany)
 - Redundant and reliable PID services for eScience and eCulture in Europe
 - Based on Handles
 - European mirror of Global Handle Repository
 - Governance structure with technical board and board of stakeholders



Application to data sets



Collections and data sets

Currently assigning identifiers to:

- Concepts for the CATCHPlus Vocabulary Repository
- A subcollection of the Sound and Vision archive

Several Dutch cultural heritage institutions and projects expressed interest



Application to data sets

Some questions to answer first...

- What are the objects to assign persistent identifiers to? (versions, metadata records, formats, composite objects...)
- Is there a relation with already existing identifiers?
- What syntax to use? Include semantics in your PIDs?
- Where do your PIDs resolve to, especially for objects that do not have a web representation of their own?
- Who is responsible for identifier creation and management?
- What guarantees can be made with regard to persistence?
- Who does hosting? Who provides services?



Steps

- For existing objects
 - Determine your policies
 - Determine what URLs to resolve to
 - Create and publish PIDs for these URLs
 - Locally store association of URLs and proprietary identifiers
 - For all externally visible metadata: replace proprietary identifiers with PIDs
- For new objects
 - Ultimately, integrate PID creation and management in your collection management tools and workflows



BEELD EN GELUID

Sound and Vision pilot

- Objects:
 - metadata descriptions at level of broadcasts
 - Open data set: 'polygoon journaal'
- Existing identifiers: "task identifiers"
- Resolve to metadata record implies: resolve to dynamically created html page
- Persistent identifiers are published using OAI-PMH
 - Published metadata refers back to same dynamic web page
 - OAI data provider uses PID service to find handles for internal identifiers/URLs



Concluding remarks

- External accessibility of data and service depends on one resolver service: should be no single point of failure
- Identifier management is an extra task that explicitly has to be dealt with
- Explicit commitments with respect to persistency have to be made, and kept
- Identifier management requires tool support (otherwise too labour intensive and error prone)

- (Re)organizing your data internally becomes easier
- Publishing parts of your collections on the internet becomes easier, more consistent and more sustainable



Questions?