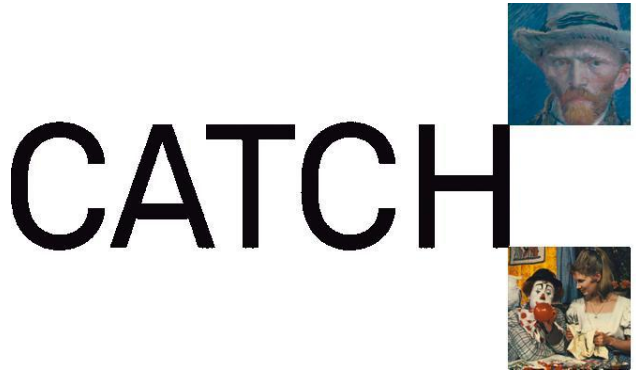



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# PIDs in the CATCHPlus project

Hennie Brugman

Technical coordinator CATCHPlus

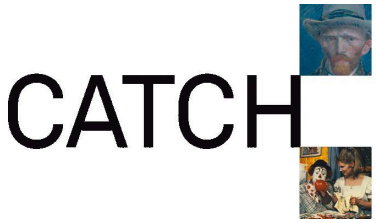
Meertens Institute

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Summary

- CATCH & CATCHPlus
- Initial requirements from CATCHPlus and CH
- Progress report
  - Base technology
  - Identifier management (API, application case)
  - Organisational embedding
- Applications of the PID service
- Lessons learned
- Plans




C

National Archive
National Library of the Netherlands
Netherlands Institute for Sound and Vision
Gemeentemuseum Den Haag
Rotterdam Municipal Archives
Naturalis (National Museum of Natural History)
Rijksdienst voor het Cultureel Erfgoed
Meertens Institute

- CATCH
- CATCH
  - 8 subprojects at large CH institutions
  - Connected by common services
    - Vocabularies, Workspaces, Annotations, User Profiles
    - Infrastructural: OAI-PMH, **persistent identifiers**
- Project bureau hosted by Meertens Institute
- www.catchplus.nl

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Initial requirements from CATCHPlus and Cultural Heritage

EPIC User Forum 2011 - Amsterdam, April 12-13, 2011

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# 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
- Distributed hosting and service providing possible
- Identification of parts of objects
- Possibility to associate metadata with an identifier
- “Actionable”: identifiers can be resolved using http URI

# 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
- Is secure

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Requirements (3)

## Organisation, policy

- What choices are made by partner institutions ?
  (the fewer ‘flavours’, the better)
- Reliability and sustainability of the service providers
- Quality of Service: redundancy, high availability, performance, capacity to scale up
- Limited and controlable costs
- Freedom to switch between service providers
- Control by user community

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

# Where are we today?

CATCH

CONTI
ACCES
TO
CULTU
HERITA
PLUS

# Requ
S

## Local Handle Systems

- 1 per participating Naming Authority
- Hosted by SARA

- ✓ Good resolving service <u>available</u>
- ✓ Proven technology, stable and 100% reliable
- ✓ Scalable, with respect to
  - ✓ Number of identifiers
  - ✓ Performance
- ✓ <u>Globally</u> working solution
- • <u>Distributed</u> hosting and service providing possible
- • Identification of <u>parts</u> of objects
- ✓ Possibility to associate metadata with an identifier
- ✓ “Actionable”: identifiers can be resolved using http URI

CATCH

CONTINUOUS ACCESS TO CULTU HERITA PLUS

• Mirrored by EPIC

# 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
- ✓ Distributed hosting and service providing possible
- • Identification of parts of objects
- ✓ Possibility to associate metadata with an identifier
- ✓ “Actionable”: identifiers can be resolved using http URI

CATCH

CONTINUOUS ACCESS TO CULTU HERITA PLUS

• Requires Handle software update by SARA/EPIC?

# 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
- ✓ Distributed hosting and service providing possible
- • Identification of parts of objects
- ✓ Possibility to associate metadata with an identifier
- ✓ “Actionable”: identifiers can be resolved using http URI

CA

CATCHPlus RESTful web service

- For searching, creation and management of Handles
- SARA has built the first version for CATCHPlus
- Currently operational and used for a full scale collection
- Available as Open source

✓ 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

✓ Is secure

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

- Side effect of collection management
- ✓ Side effect of collection publication

# 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
- ✓ Is secure

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE

- Stakehold • To some extend covered by EPIC consortium
  - Organi
  - 6 large cultural heritage institutions participated
    - Plus DEN, CATCHPlus, SARA and CLARIN-NL
  - Important topics:
    - What is the business case?
    - Business models
    - What formal agreements are necessary? Contracts?

- Limited and controllable costs
- ✓ Freedom to switch between service providers
- Control by user community

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

# Applications of the PID service

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Methods for identifier management

- Side effect of collection management
  - Action in collection management system triggers PID management REST call
- Side effect of collection publication
  - Update of your public data set (e.g. OAI data provider's internal database) triggers PID management REST call

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Beeld en Geluid

- Local OAI indexes maintained with ‘inbox mechanism’
- ‘Create’, ‘update’, ‘delete’ events on inbox trigger PID management REST calls
- 1.4 million Handles created and maintained this way
- Subset (polygoon) published using OAI-PMH

# beng:Expressie:422163

| | |
|---|---|
| **Herkomst** | Nederlands Instituut voor Beeld en Geluid |
| **Titel** | Polygoon Hollands Nieuws<br>Weeknummer 53-41<br>BEZOEK VAN HM DE KONINGIN AAN DUIVELAND |
| **Auteur** | Bloemendal, Philip (commentaar) |
| **Onderwerp** | Juliana (koningin Nederland)<br>Watersnoodramp 1953 |
| **Omschrijving** | Bioscoopjournaals waarin Nederlandse onderwerpen van een bepaalde week worden gepresenteerd.<br>Koningin Juliana bezoekt het gebied van de Watersnoodramp op Schouwen-Duiveland. Ze arriveert per helikopter, reist verder per boot en per jeep naar onder meer Nieuwerkerk en Ouderkerk; bekijkt de aftermath; spreekt met bevolking en hulpverleners; deelt geschenken uit. |
| **Uitgever** | Polygoon |
| **Datum** | 05-10-1953 |
| **Type** | bioscoop |
| **Formaat** | WEEKNUMMER534-HRE0000E1E0 |
| **Identifier** | /12068/12070/12070/422163<br>518888<br>http://hdl.handle.net/10574/CA92F810879A11DFA0CB001D0911E44A |
| **Bron** | Bewegend beeld/Polygoon/Hollands nieuws |
| **Taal** | nl |

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Other CH tests and pilots

- Meertens Institute
- Rijksdienst voor het Cultureel Erfgoed (intended)
- Several other interested organisations

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

# Lessons learned

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Some lessons learned

- Technical
  - Maintainance and publication via OAI-PHM feasible
  - Current API works well
  - Handle URIs can be used for Linked Data
- Organisational
  - Chicken and egg problem?
  - Business cases are not clear – good/bad examples needed
  - In need of a good business model
  - In need of contract models

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

# Plans

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Plans

- API version 2
  - Collect feedback
  - Freeze API
  - Implementation
  - Again: freely available, open source
- General PID inspection and management web application
- Pilots, tests and applications at CH institutions
- Make progress on business case, businessmodels and contracts

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

# Questions?

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

EPIC User Forum 2011 - Amsterdam, April 12-13, 2011


CATCH


CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS


# 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 garantees can be made with regard to persistence?
- Who does hosting? Who provides services?

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# 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

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

BEELD EN GELUID

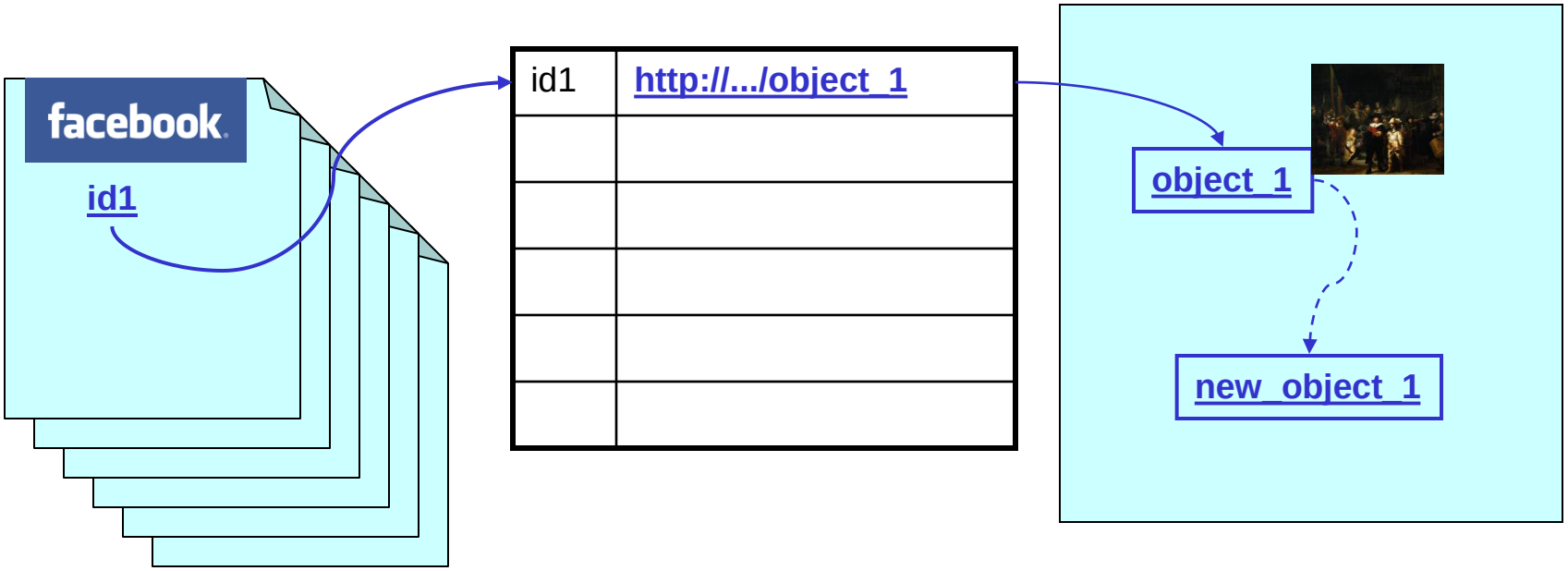
# Sound and Vision pilot

- 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

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

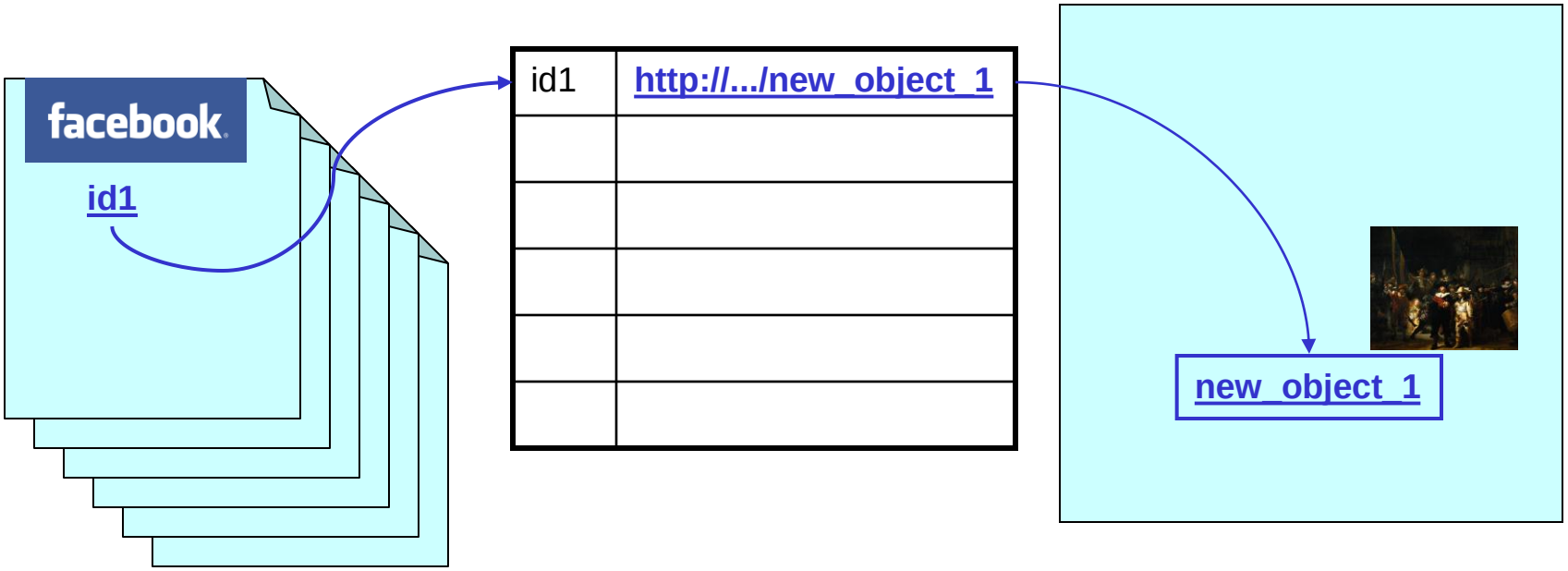
# Basisoplossing



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

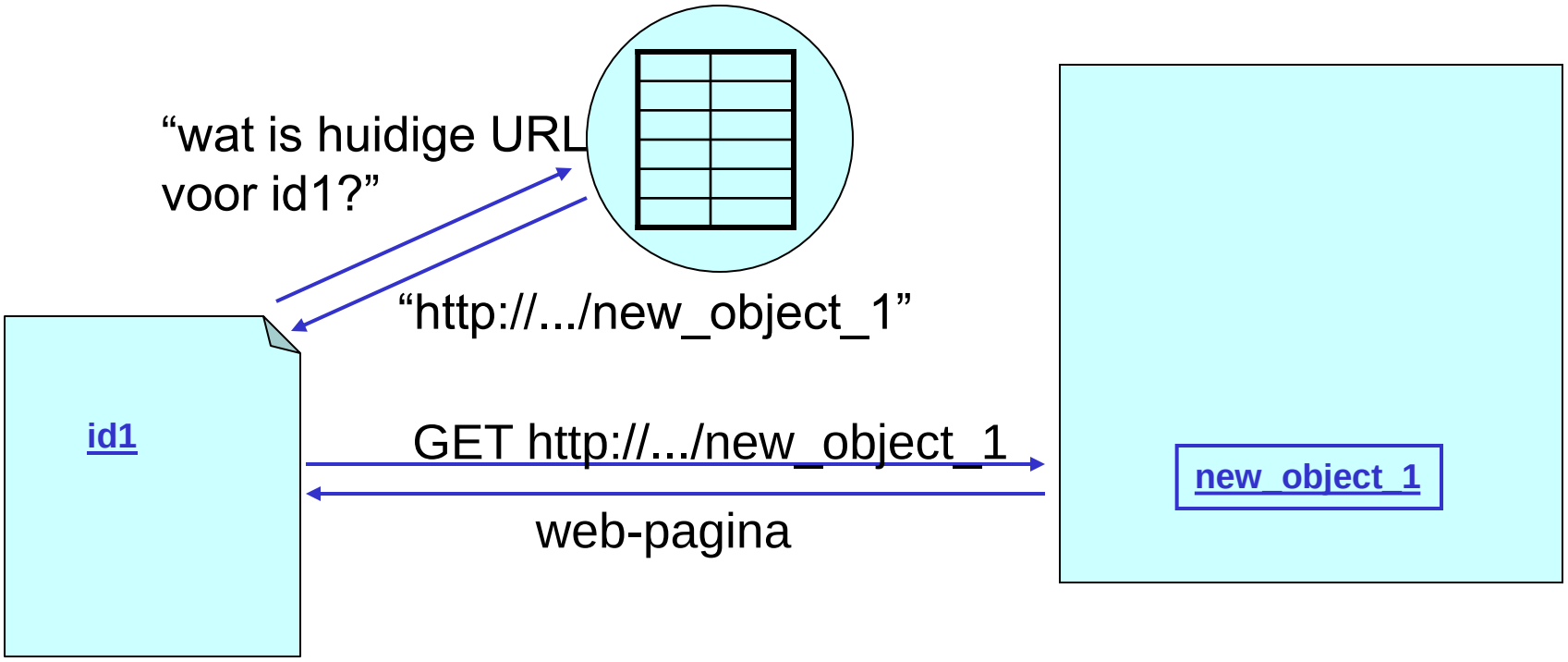
# Basisoplossing



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Resolver dienst



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Naming Authority

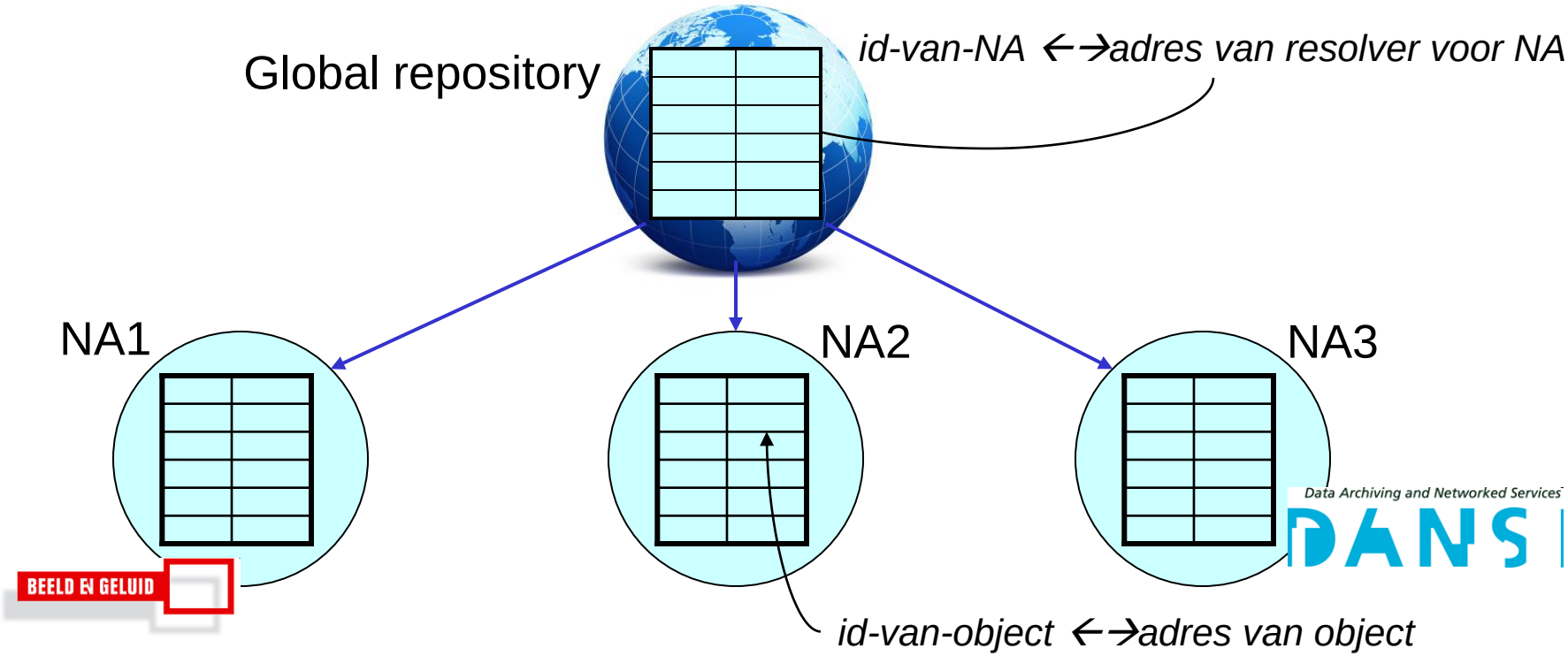
- Naming Authority: beheerder van de inhoud van de tabel
- NA heeft ook een unieke persistente identifier
- Globaal geregistreerd waar de resolver voor de NA is te vinden.
- Veel soorten persistente identifiers hebben de basisvorm:

  *<unieke-id-van-NA><unieke-locale-id>*

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

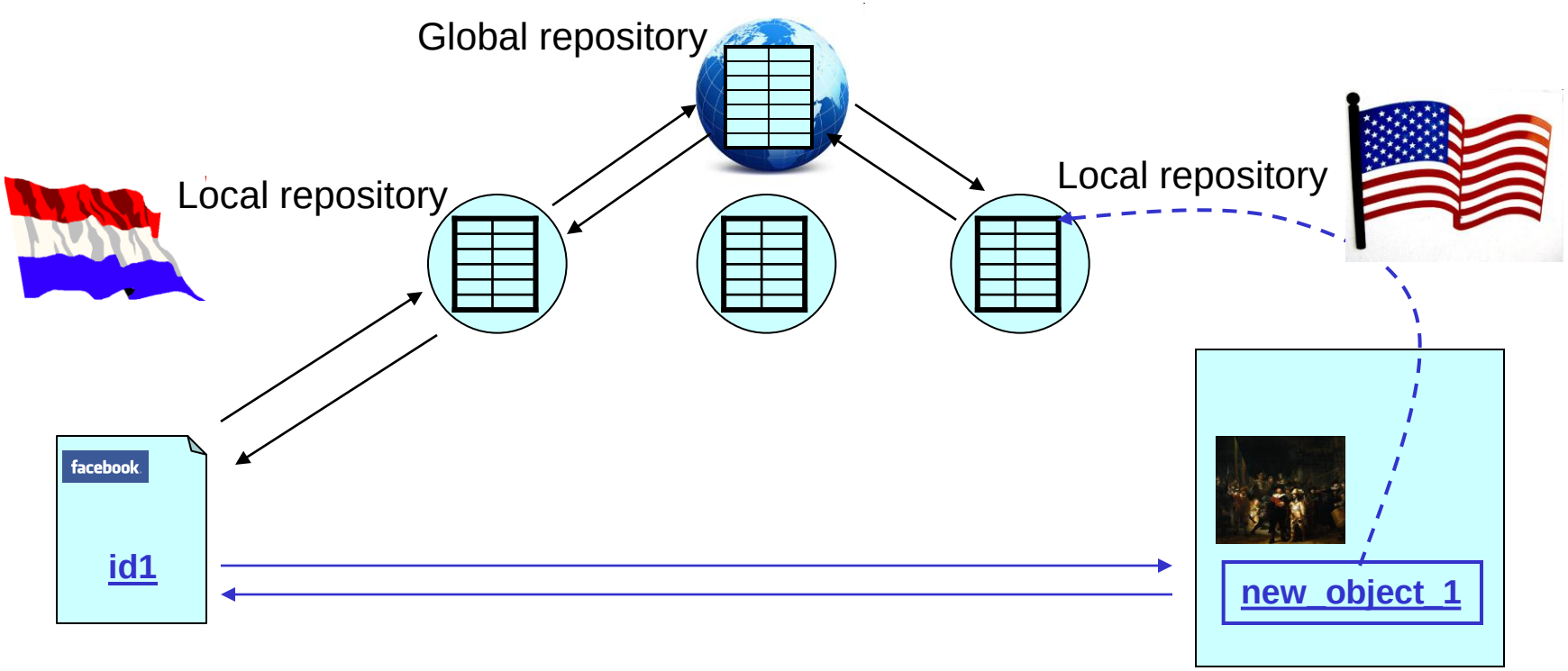
# Naming Authority



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Global resolving



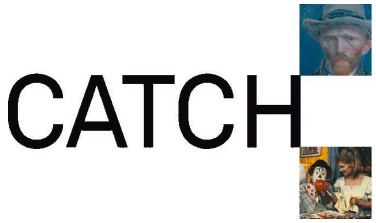
CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

# Voorbeelden van PID oplossingen

- Er bestaan een aantal verschillende oplossingen
- Binnen erfgoed in NL spelen een rol

  - Gewoon URLs gebruiken
  - URN-NBN
  - Handles
  - DOI
  - ARK
  - PURL




CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

http URL

- http://identifiers.erfgoed.nl/local_id_1821

URN-NBN

- urn:nbn:nl-local_id_1821

Handles

- 10574/local_id_1821
- http://hdl.handle.net/10574/local_id_1821

DOI

- doi:10.1594/PANGAEA.726855
- http://dx.doi.org/10.1594/PANGAEA.726855

ARK

- http://identifiers.erfgoed.nl/ark:/128014/local_id_1821

PURL

- http://purl.org/vocabularies/iconclass/concept1821

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

HOST

NA

LOCAL ID

X http URL

- http://identifiers.erfgoed.nl/local_id_1821

X URN-NBN

- urn:nbn:nl-local_id_1821

Handles

- 10574/local_id_1821
- http://hdl.handle.net/10574/local_id_1821

DOI

- doi:10.1594/PANGAEA.726855
- http://dx.doi.org/10.1594/PANGAEA.726855

ARK

- http://identifiers.erfgoed.nl/ark:/128014/local_id_1821

X PURL

- http://purl.org/vocabularies/iconclass/concept1821

- Part identifiers - some use cases
- Selections from audio-visual objects
- Thesaurus concepts
- Lexicon entries

- How: rewrite rules per Naming Authority and/or per PID, built into the resolver

- *"Part identifiers"* preferred over "*many PIDs"* when:
- potentially infinite number of pids (AV)
- impractically large numbers of PIDs (~$10^7$ concepts)

- Possible criterium:
- Managed as one coherent (web) resource (1 base URL)