


CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Vocabulaire en Alignment Service

Hennie Brugman

Technisch coordinator CATCHPlus

Max-Planck-Institute for Psycholinguistics
Nederlands Instituut voor Beeld en Geluid

Werkgroep Nederlandstalige Erfgoedthesauri – 21 mei 2010

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Overview

- CATCH en CATCHPlus
  - Bijdrage aan infrastructuur voor digitaal erfgoed
- Vocabulary and Alignment Service
  - doelen en gebruiksmogelijkheden
- Wat hebben we al gebouwd?
  - Architectuur
  - Functionaliteit
  - Voorbeelden
  - Demos
- Status en beschikbaarheid
- Toekomstperspectief
- Conclusies en opmerkingen

CATC

- CATCH onderzoeksp
- CATCHPlus valorisat
  - 8 subprojecten bi
    - Levert (her)bruikbare tools en diensten
  - Verbonden door gemeenschappelijke diensten mbt
    - terminologie
    - annotaties
    - metadata (catalogus-informatie)
    - content
- CATCHPlus project bureau gehuisvest bij het Nederlands Instituut voor Beeld en Geluid (Hilversum)
- www.catchplus.nl

Rijksmuseum Amsterdam
Nationaal Archief
Koninklijke Bibliotheek
Beeld en Geluid
Gemeentemuseum Den Haag
Gemeentearchief Rotterdam
NCB-Naturalis
Rijksdienst voor het Cultureel Erfgoed

CATCH

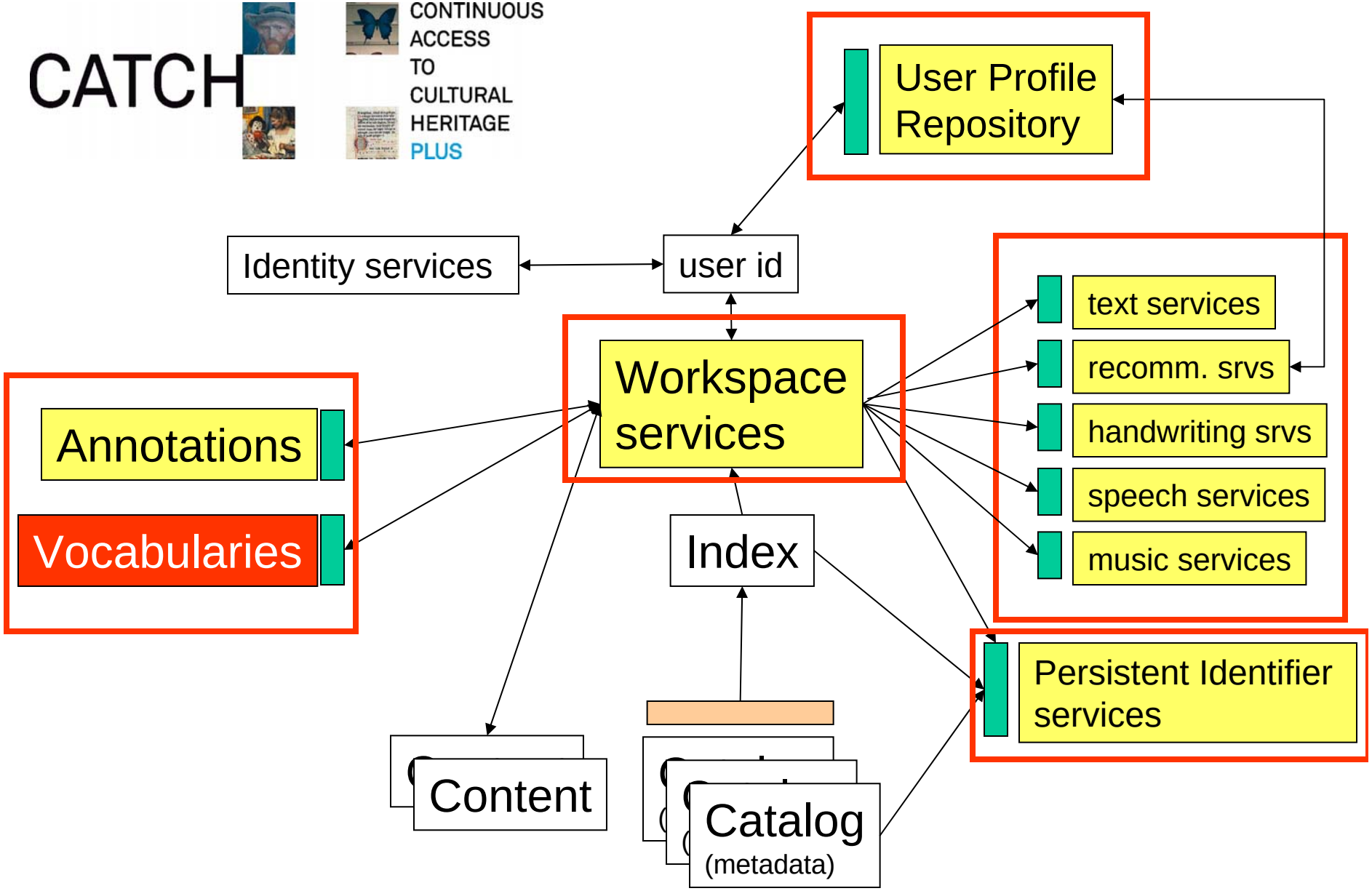
CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

CATCHPlus en infrastructuur voor digitaal erfgoed

Werkgroep Nederlandstalige Erfgoedthesauri – 21 mei 2010

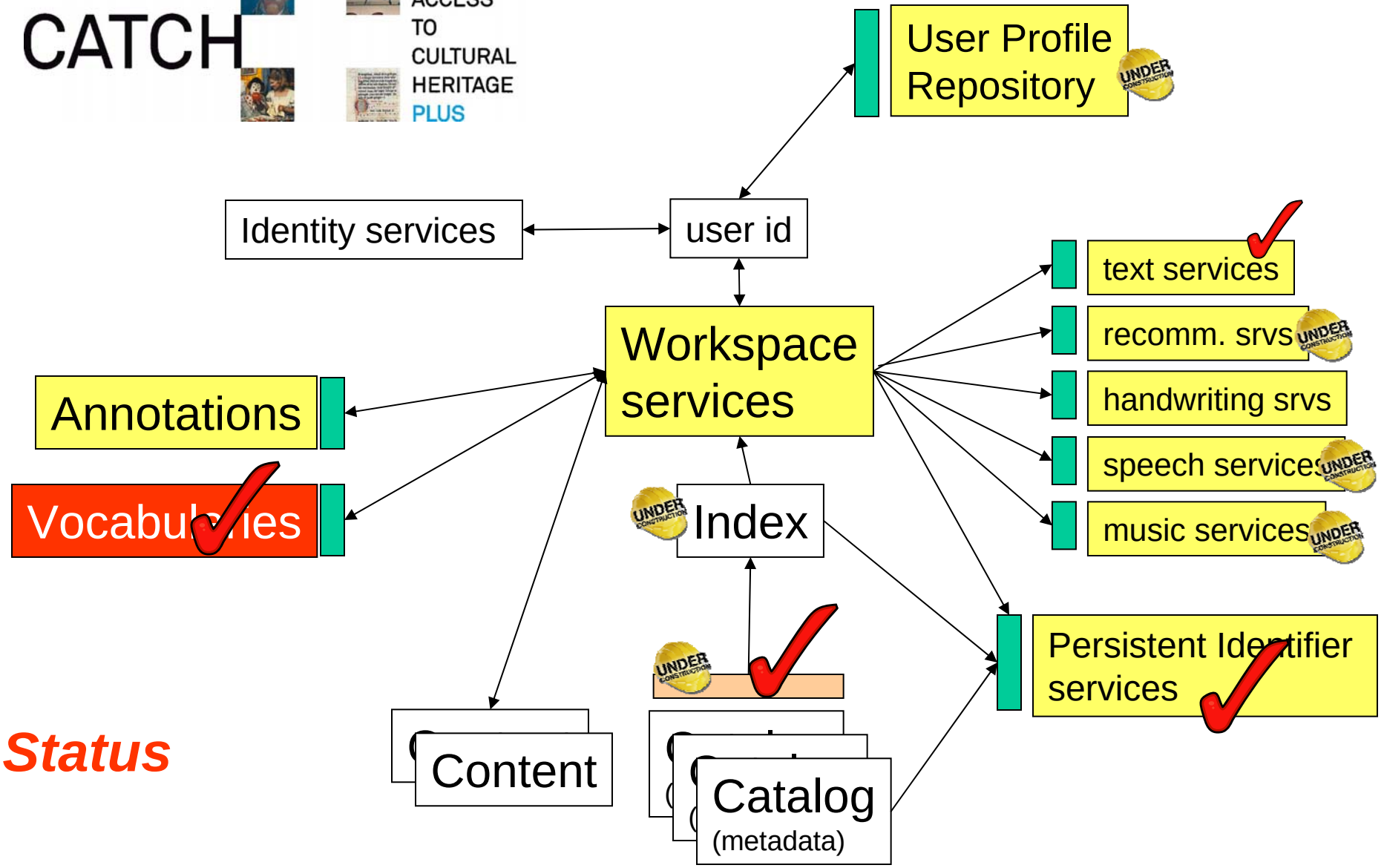
CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Status

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Vocabulary and Alignment Service – doelen en gebruiksmogelijkheden

Werkgroep Nederlandstalige Erfgoedthesauri – 21 mei 2010

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Belangrijke VAS doelen

- Standaard *formaat*
  - SKOS
- Web *publicatie* van vocabulaires
  - Als doorzoekbare en browse-bare dataset (mbv REST API)
  - Als Linked Open Data
  - Bruikbaar voor duurzame referenties naar concepten → persistent identifiers
- Verbeterde *semantische interoperabiliteit* door het ondersteunen van “alignments”
- Ontkoppelen thesaurus-aanbod en thesaurus-gebruik
- Mogelijk regeling van *licenties* per community

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Belangrijke VAS doelen

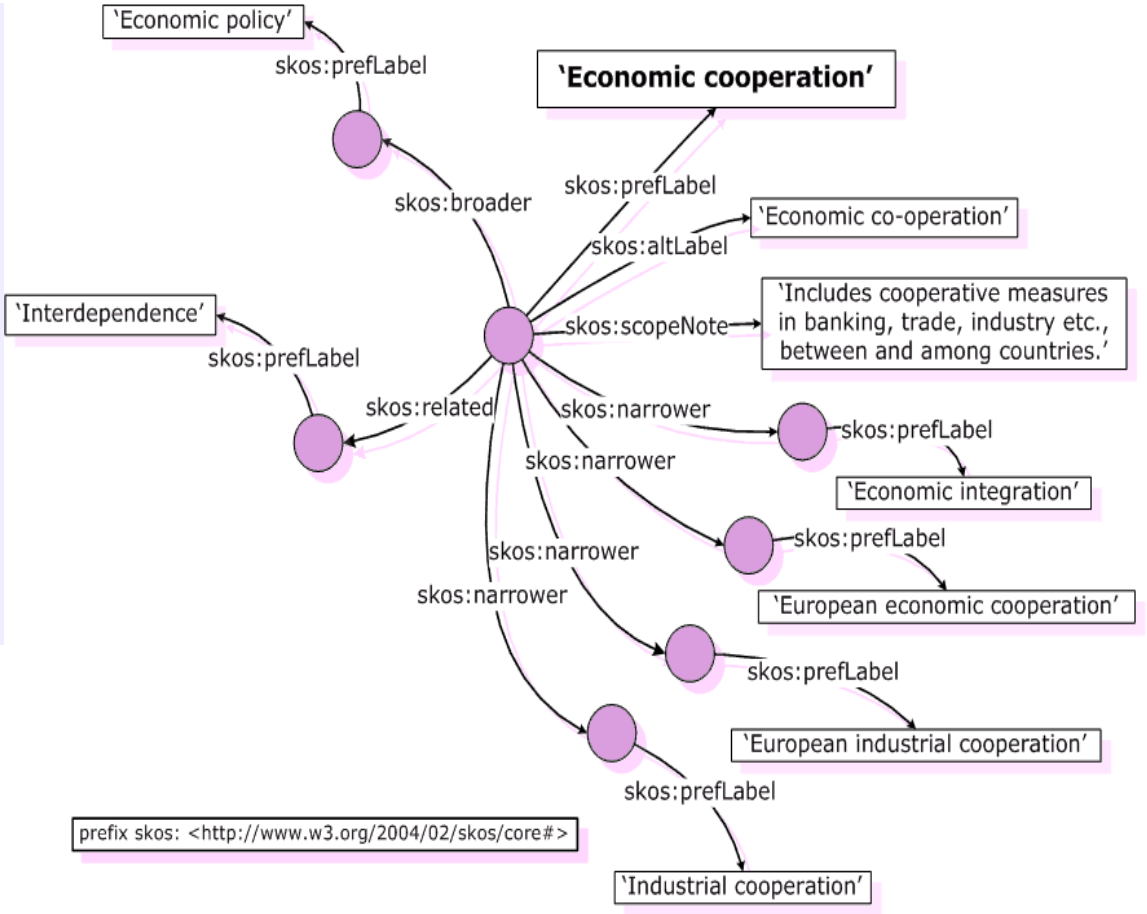
- Standaard *formaat*
  - **SKOS**
- Web *publicatie* van vocabulaires
  - Als doorzoekbare en browse-bare dataset (mbv REST API)
  - Als **Linked Open Data**
  - Bruikbaar voor duurzame referenties naar concepten → **Persistent Identifiers**
- Verbeterde *semantische interoperabiliteit* door het ondersteunen van “alignments”
- Ontkoppelen thesaurus-aanbod en thesaurus-gebruik
- Mogelijk regeling van *licenties* per community

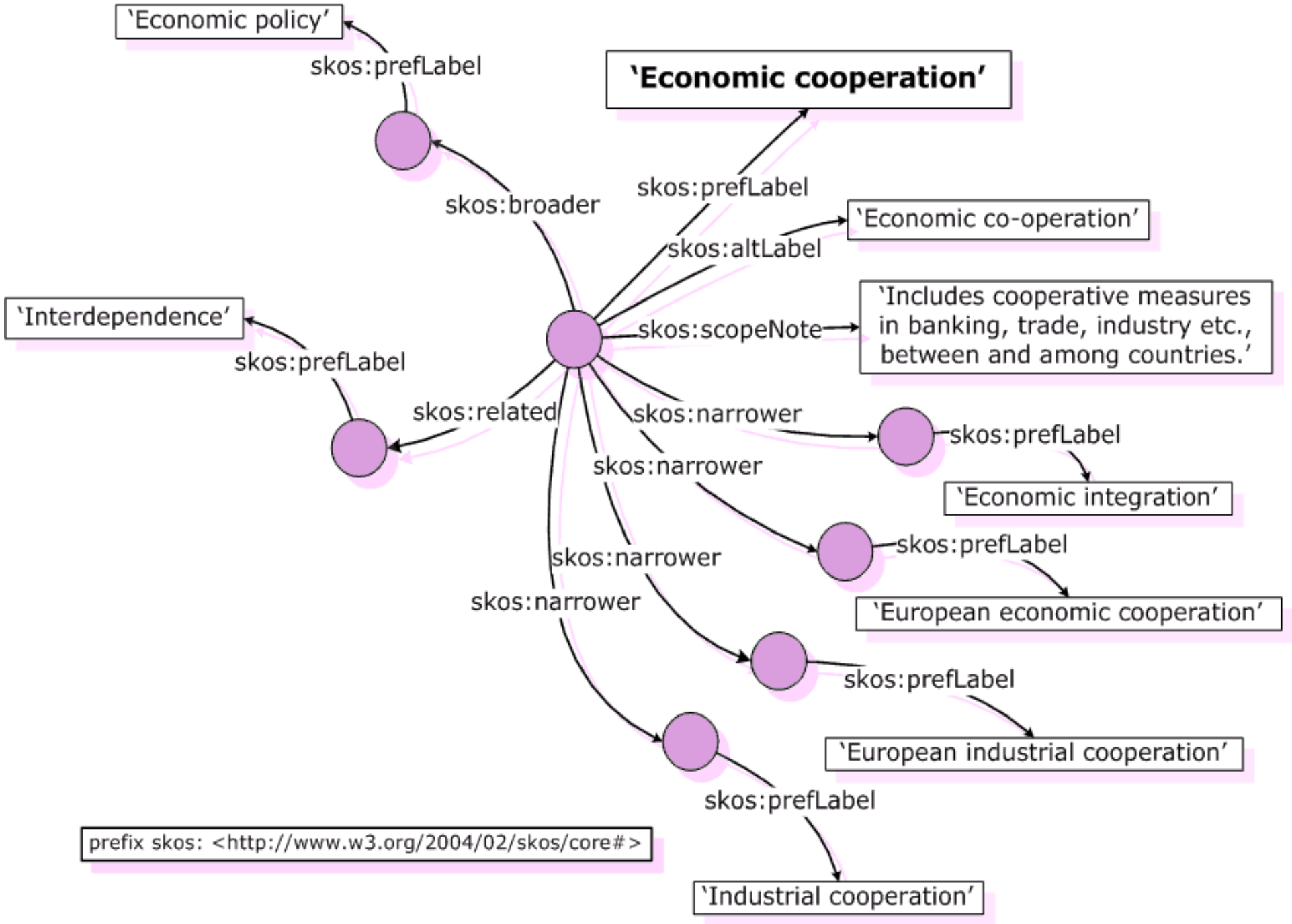
CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

SKOS

Term: Economic cooperation
Used For: Economic co-operation
Broader terms: Economic policy
Narrower terms: Economic integration,
European economic cooperation,
European industrial cooperation,
Industrial cooperation
Related terms: Interdependence
Scope Note: Includes cooperative measures in banking, trade, industry etc., between and among countries.




'Economic policy'
skos:prefLabel
'Economic cooperation'
skos:broader
skos:prefLabel
'Economic co-operation'
skos:altLabel
'Interdependence'
skos:prefLabel
skos:scopeNote
'Includes cooperative measures in banking, trade, industry etc., between and among countries.'
skos:related
skos:narrower
skos:prefLabel
'Economic integration'
skos:narrower
skos:prefLabel
'European economic cooperation'
skos:narrower
skos:narrower
skos:prefLabel
'European industrial cooperation'
skos:prefLabel
'Industrial cooperation'
prefix skos: <http://www.w3.org/2004/02/skos/core#>

CATCH

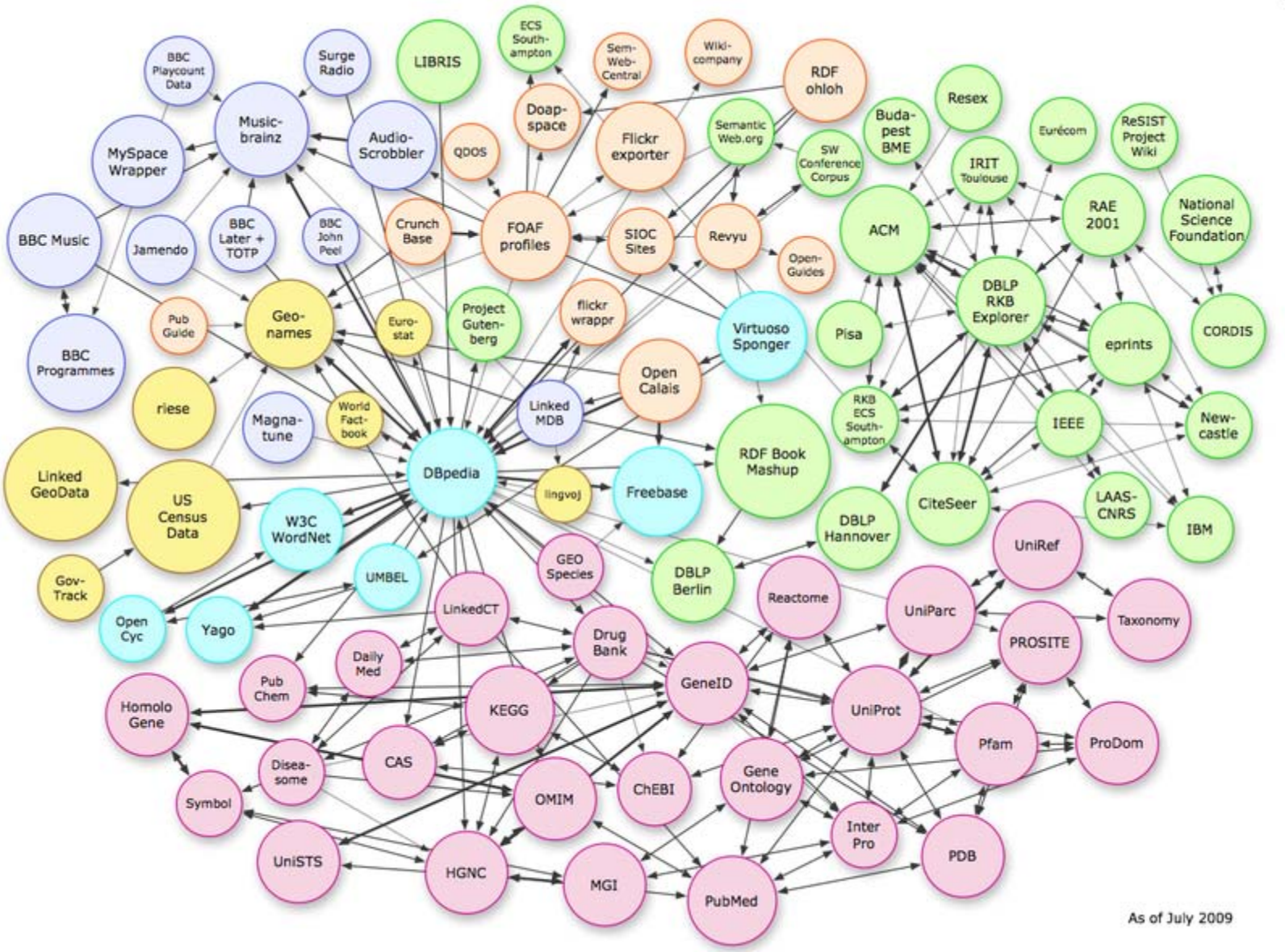
CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Linked Open Data

- Een methode om data te tonen, delen en verbinden via ‘dereferencable URIs’ op het Web.
- Principes (naar Tim Berners-Lee)
  - Gebruik URIs om dingen te identificeren.
  - Gebruik HTTP URIs zodat er naar deze dingen kan worden verwezen en dat ze kunnen worden opgezocht (“dereference”) door mensen en computer-programma’s.
  - Geef nuttige informatie over het ding als zijn URI wordt ‘ge-dereferenced’, met gebruikmaking van standaarden als RDF-XML.
  - Neem links naar andere, gerelateerde, open data op om het ontdekken van gerelateerde informatie op het web te verbeteren.

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL



As of July 2009

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Persistent Identifiers voor concepten

- Stabiele URIs per concept
  - Bv http://data.beeldengeluid.nl/gtaa/28181
- PID per “Concept Scheme” (sub-thesaurus)
  - 10574/<cs_identifier>
- Concepten mbv ‘templates’
  - 10574/<cs_identifier>-28181
  - Primair te koppelen aan stabiele URL

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Use cases

- Use cases uit CATCHPlus en Cultureel Erfgoed
  - Publiceer je thesaurus: importeer een SKOS vocabulaire, dan krijg je er REST toegang, tool support en Linked Data gratis bij
  - Selecteer het juiste concept om een object te beschrijven
  - Gebruik voor browsen en zoeken (naar terminologie en/of collectie-data)
    - VAS repository als een “topic map” voor erfgoed-collecties
  - Thesaurus-onderhouds-taken door online gemeenschappen
  - Vertalen, verfijnen, generaliseren van zoekvragen
  - Etc.

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Wat hebben we tot nu toe gebouwd?

Architectuur, functionaliteit, voorbeelden en demo's

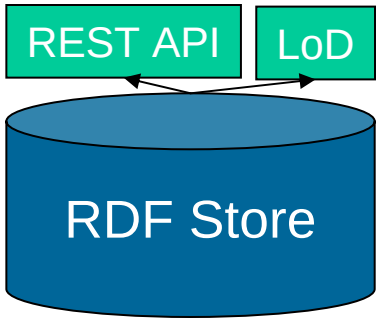
Werkgroep Nederlandstalige Erfgoedthesauri – 21 mei 2010

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Architectuur

- Repository voor SKOS data (inclusief alignment data)
  - RDF store (Virtuoso)
- REST API (search, autocomplete, upload, download)
- Linked Data interface



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

REST functionaliteit

- API basis functies
  - /find/concept
  - /find/conceptscheme
  - /find/conceptschemegroup
  - /find/relation

- http://catchplus.tuxic.nl/catchplus/serviceapi/1/

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

REST voorbeelden

(alle gtaa onderwerpen met prefLabel beginnend met ‘al’)

- http://catchplus.tuxic.nl:18080/vas/api
- /find/concept?
  - format=xml&
  - query=^al&
  - match_kind=regex&
  - limit=100&
  - cs=http://www.beeldengeluid.nl/Thesaurus/Onderwerpen
  - &sort=match_label&
  - match_language=nl

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

REST voorbeelden

- http://catchplus.tuxic.nl:18080/vas/api/find/concept?format=xml&query=^al&match_kind=regex&limit=100&format=xml&cs=http%3A%2F%2Fwww.beeldengeluid.nl%2FThesaurus%2FOnderwerpen&sort=match_label&match_language=nl
- http://catchplus.tuxic.nl:18080/vas/api/find/concept?format=xml&uri=http%3A%2F%2Fwww.beeldengeluid.nl%2FThesaurus%2F28181&info=skos_concept
- http://catchplus.tuxic.nl:18080/vas/api/find/concept?format=json&uri=http%3A%2F%2Fwww.beeldengeluid.nl%2FThesaurus%2F28181&info=skos_concept

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

LoD demo

(werk in uitvoering)

http://linkeddata.uriburner.com/ode

CATCH

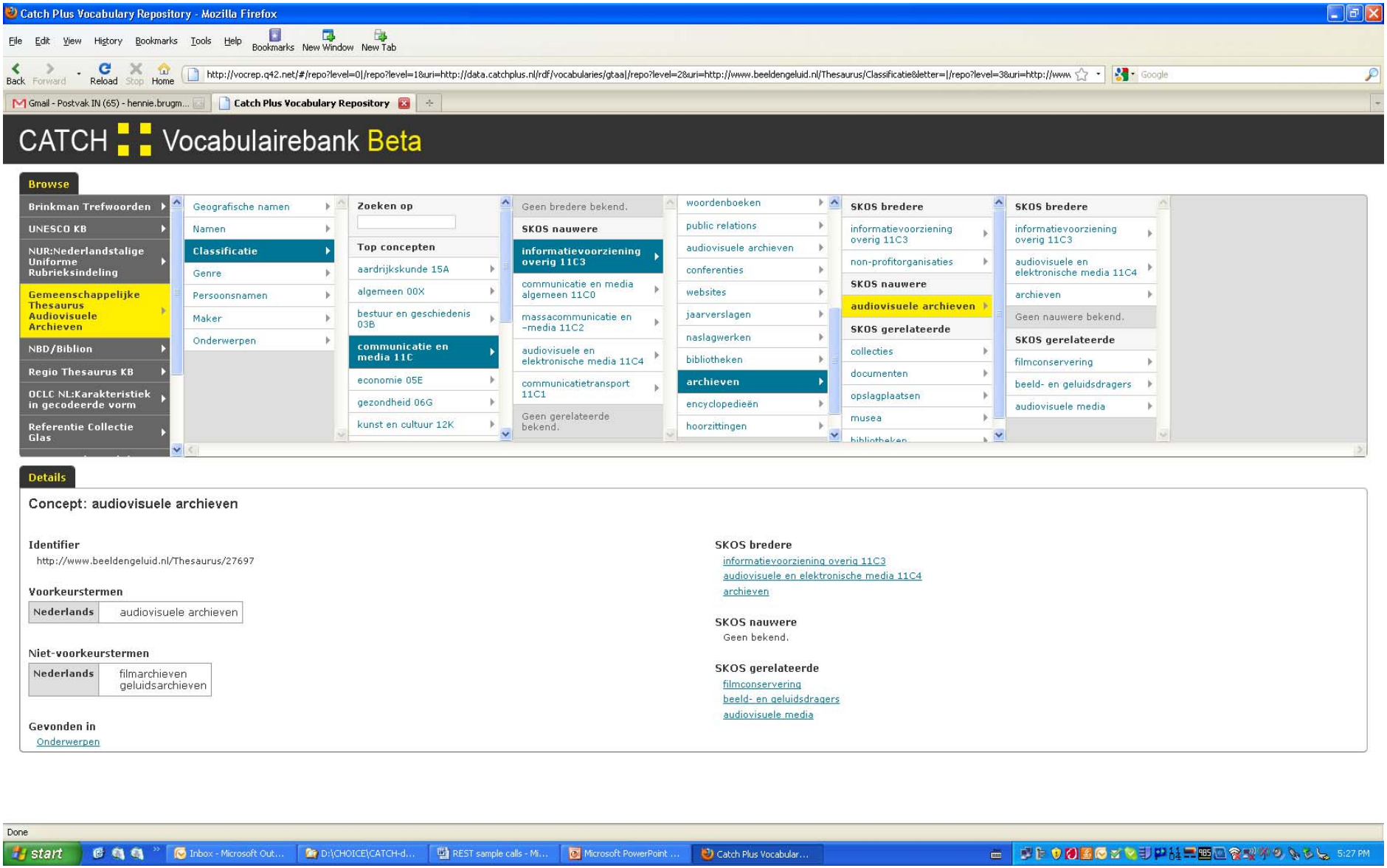
CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Client tools en diensten

(via de REST API)

- CATCHPlus cases (semantische annotatie, term ranking, art recommender, …)
- Commerciële collectie-beheer software bouwer gebruikt de API om thesaurus informatie te integreren
- Generieke browse en zoek webapplicatie

http://vocrep.q42.net/index.html



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

VAS SKOS guidelines

- SKOS, plus
  - ConceptSchemeGroups
    - Laden/verwijderen gaat momenteel per CSG
    - Concept counts, firstLanguage expliciet aan te geven
  - Conventies voor labels van concepten, csg en cs
  - Skos mapping, owl:sameAs, alignapi alignments voor het representeren van alignment data
  - Expliciet definiëren van top concepten voor concept schemes

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Status en beschikbaarheid

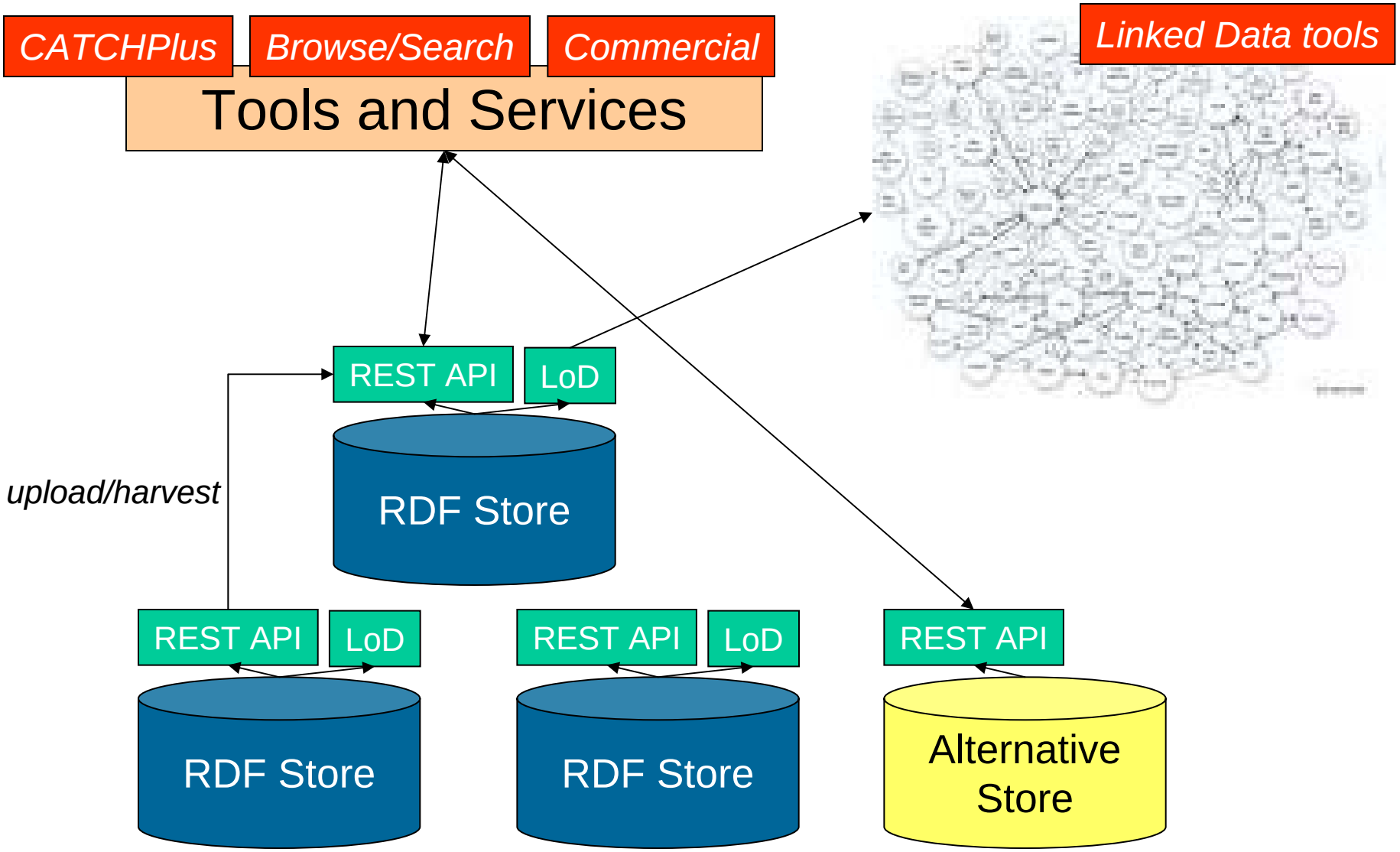
- Bevat momenteel 12 thesauri (meestal is de licentie-situatie niet geregeld)
- Browse/search tool (versie 2) is klaar
- Steeds bredere belangstelling van
  - Thesaurus aanbieders
    - VU, Wageningen SemWeb group, RKD, CLARIN-NL
  - Tool bouwers
    - Collectie-beheer software bouwers
  - Biedt een kans op harmonisatie van de API en/of de technologie
- Gebruikt voor samenwerking tussen Beeld en Geluid en Nationaal Archief mbt de GTAA thesaurus
- Doorontwikkeling via Open source model

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Toekomst

- Gedistribueerde operatie
- “live verbindingen” met thesaurus databases → automatische updates
- Linken met “DEN inventarisatie van terminologiebronnen”


CATCHPlus
Browse/Search
Commercial
Tools and Services
Linked Data tools
REST API
LoD
RDF Store
upload/harvest
REST API
LoD
RDF Store
REST API
LoD
RDF Store
REST API
Alternative Store

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Conclusies en opmerkingen

- Doorontwikkeling is nodig voor een langdurig beschikbare oplossing. Open source model, met bijdragen door diverse stakeholders.
- De/een API is essentieel. Minimaal is API harmonisatie tussen de diverse thesaurus-aanbieders en toolbouwers noodzakelijk.
- Hosting gebeurt momenteel via CATCHPlus, dus tijdelijk. Een meer duurzame hosting-oplossing moet nog gerealiseerd worden.
- De licentie-problematiek is nog grotendeels onopgelost. Promoten van ODbL (Open Database Licence) is zeer wenselijk.

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Dank u. Vragen?