

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



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

CATCH



CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

CATCH

- CATCH onderzoeksp
- CATCHPlus valorisat
 - 8 subprojecten b
 - 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

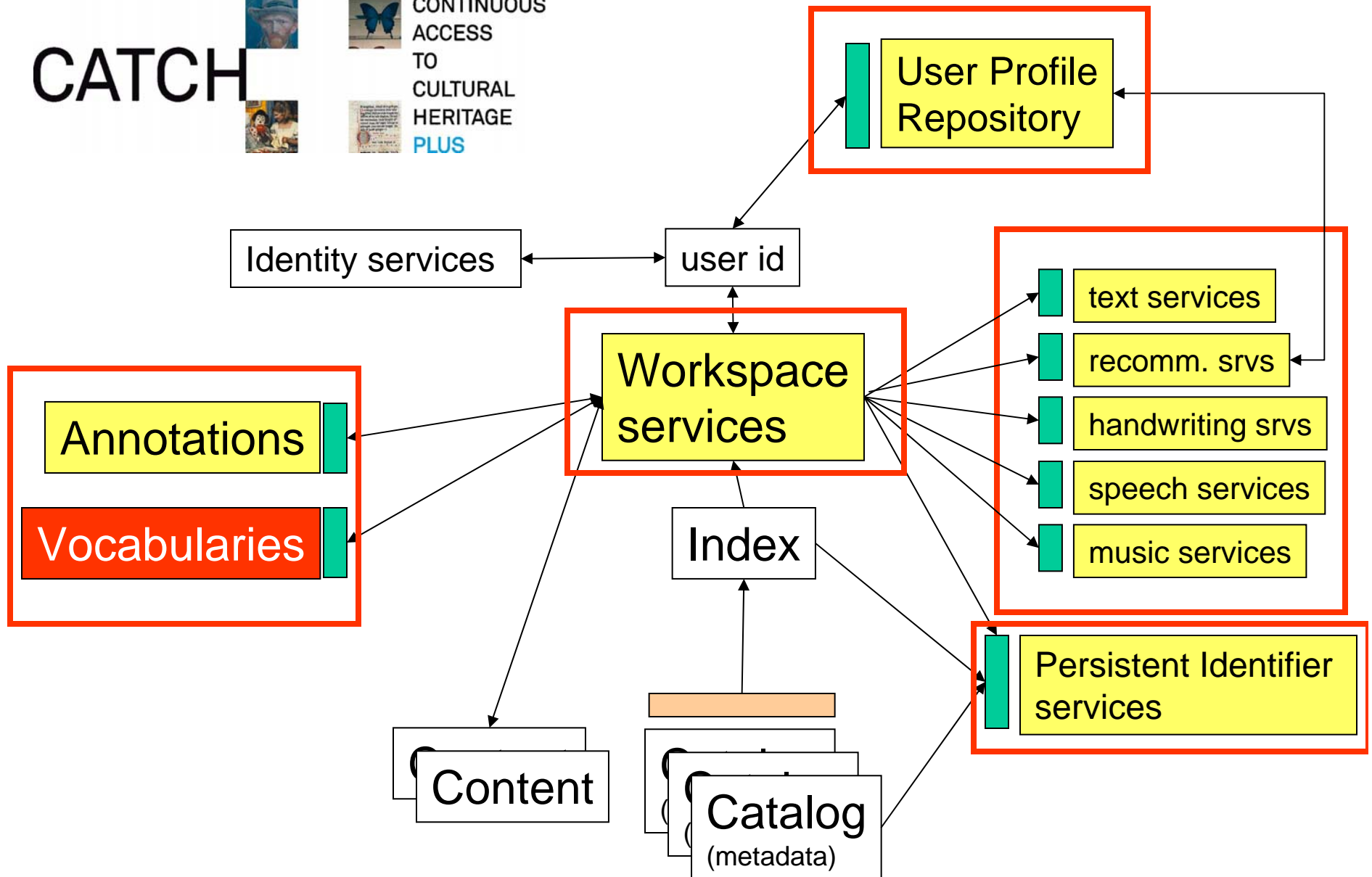


CATCHPlus en infrastructuur voor digitaal erfgoed

CATCH

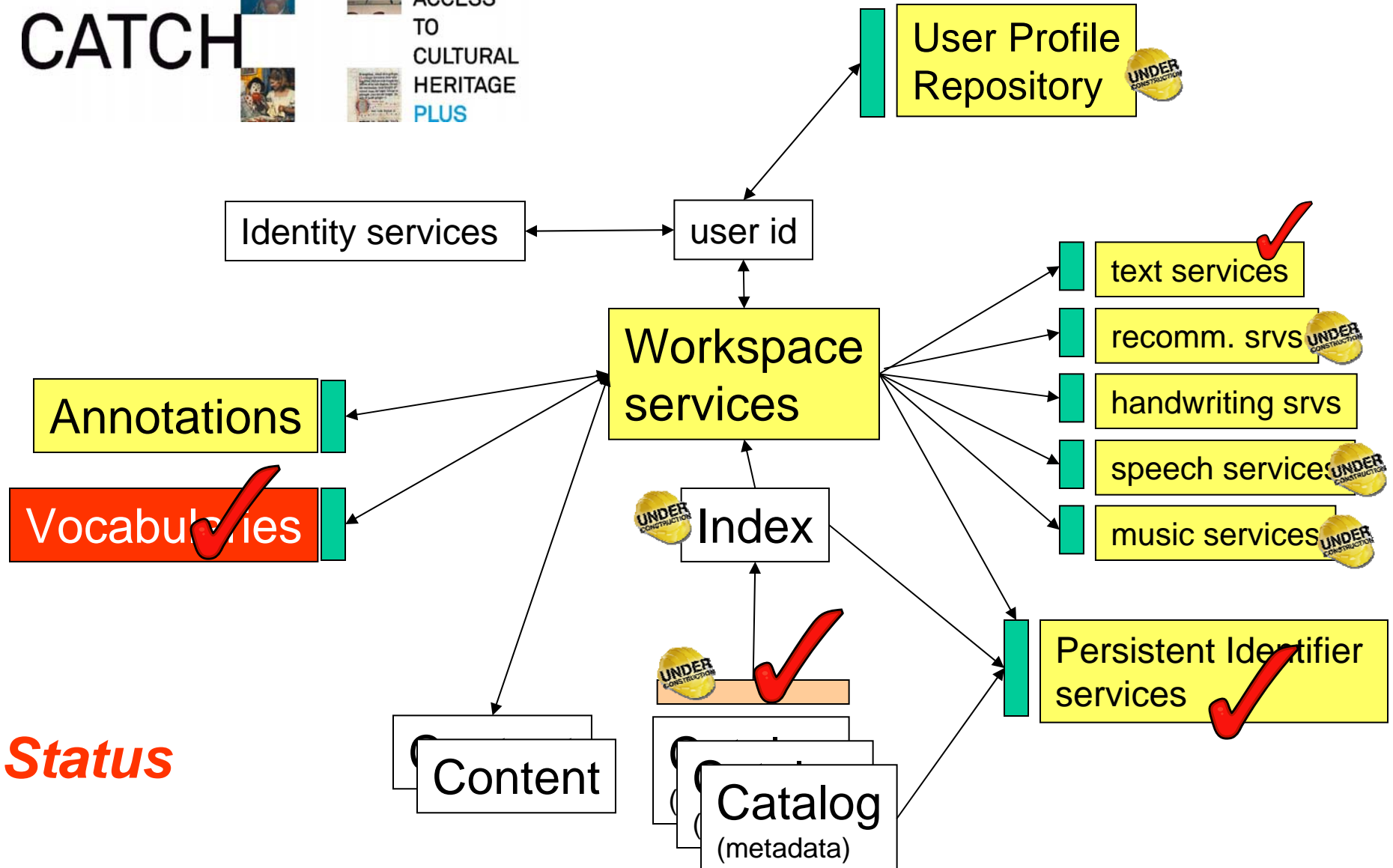


CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS



CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Status



Vocabulary and Alignment Service – doelen en gebruiksmogelijkheden



Belangrijke VAS doelen

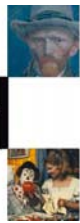
- Standaard *formaat*
 - SKOS
- Web *publicatie* van vocabulaires
 - Als doorzoekbare en browse-bare dataset (mbv REST API)
 - Als Linked Open Data
 - Bruikbaar voor duurzame referenties naar concepten → persistent identifiers
- Verbeterde *semantische interoperabiliteit* door het ondersteunen van “alignments”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



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

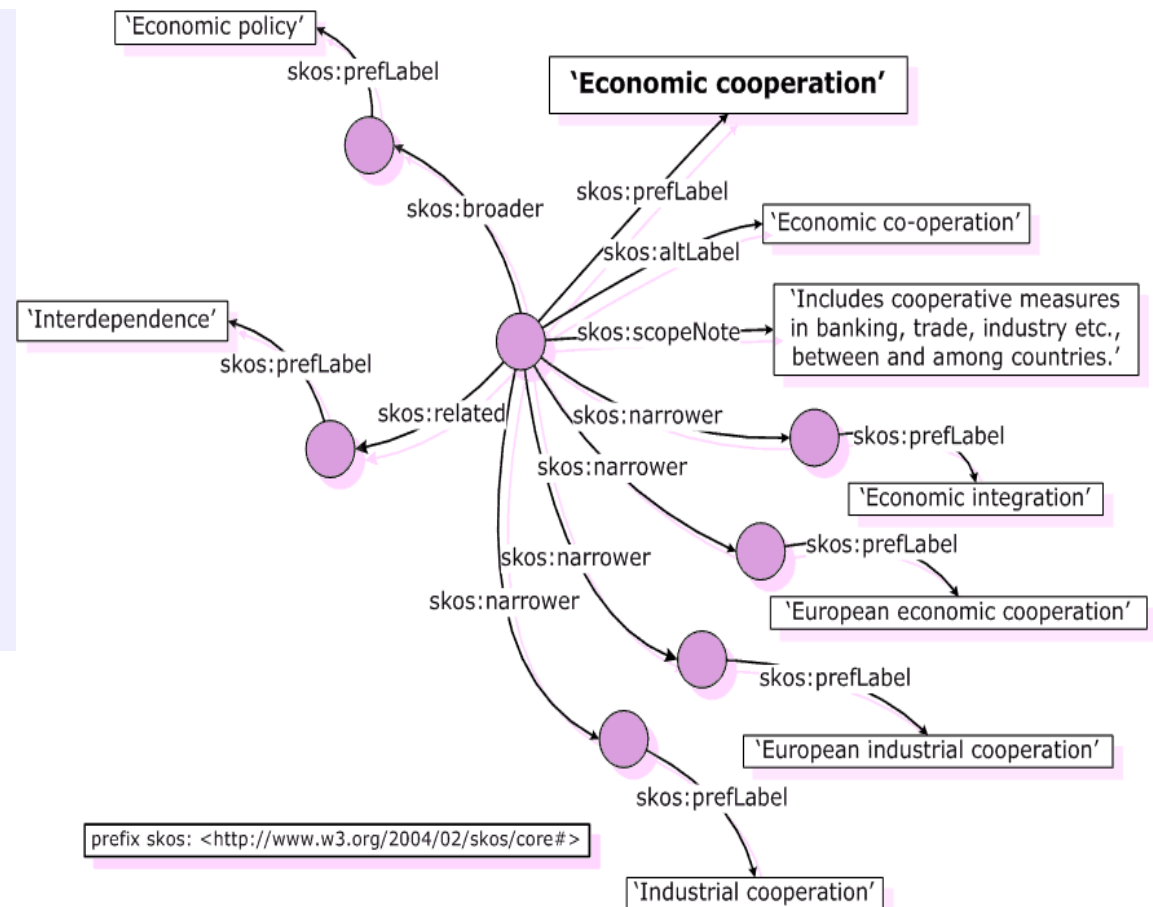
CATCH

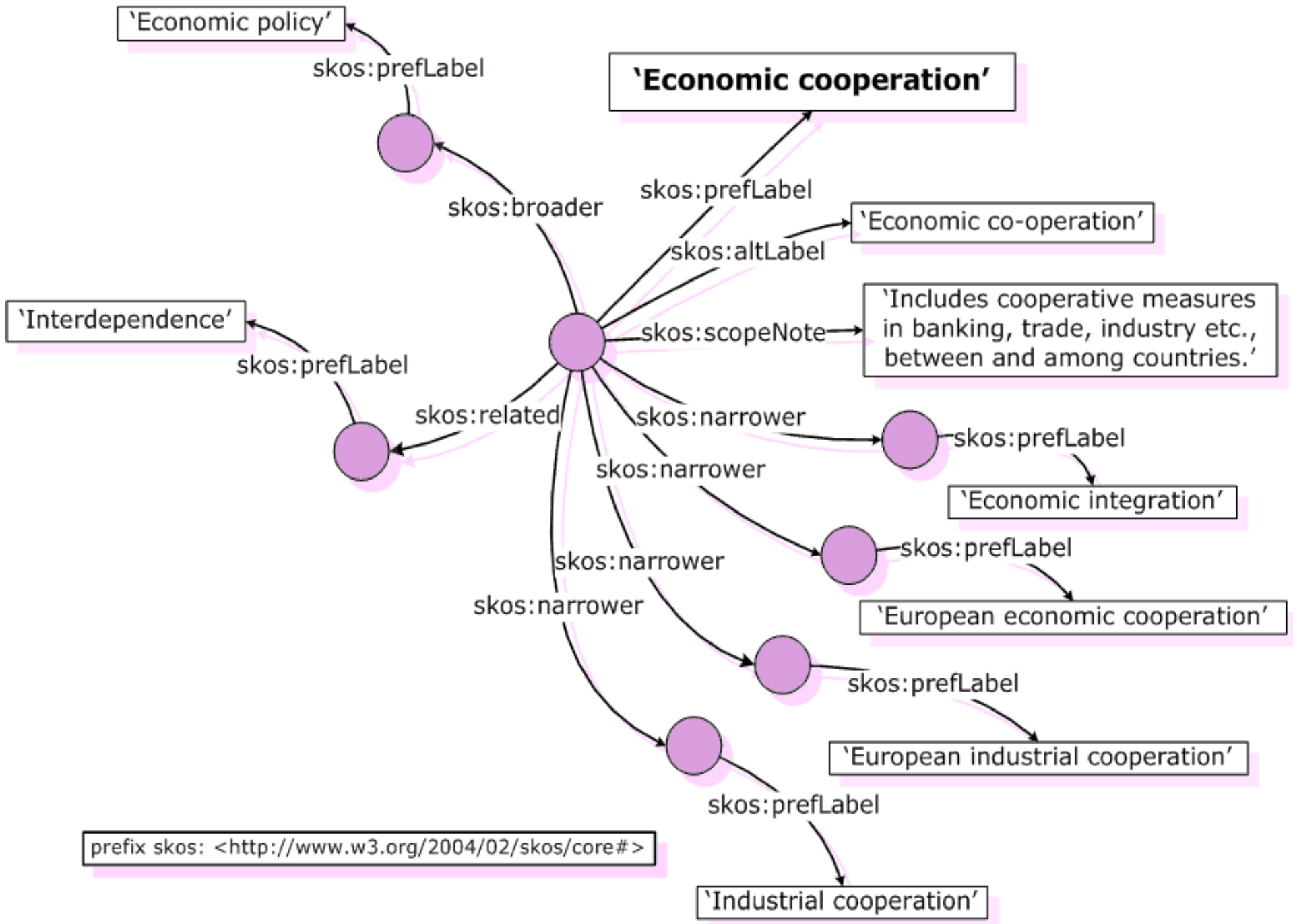


CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS

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.







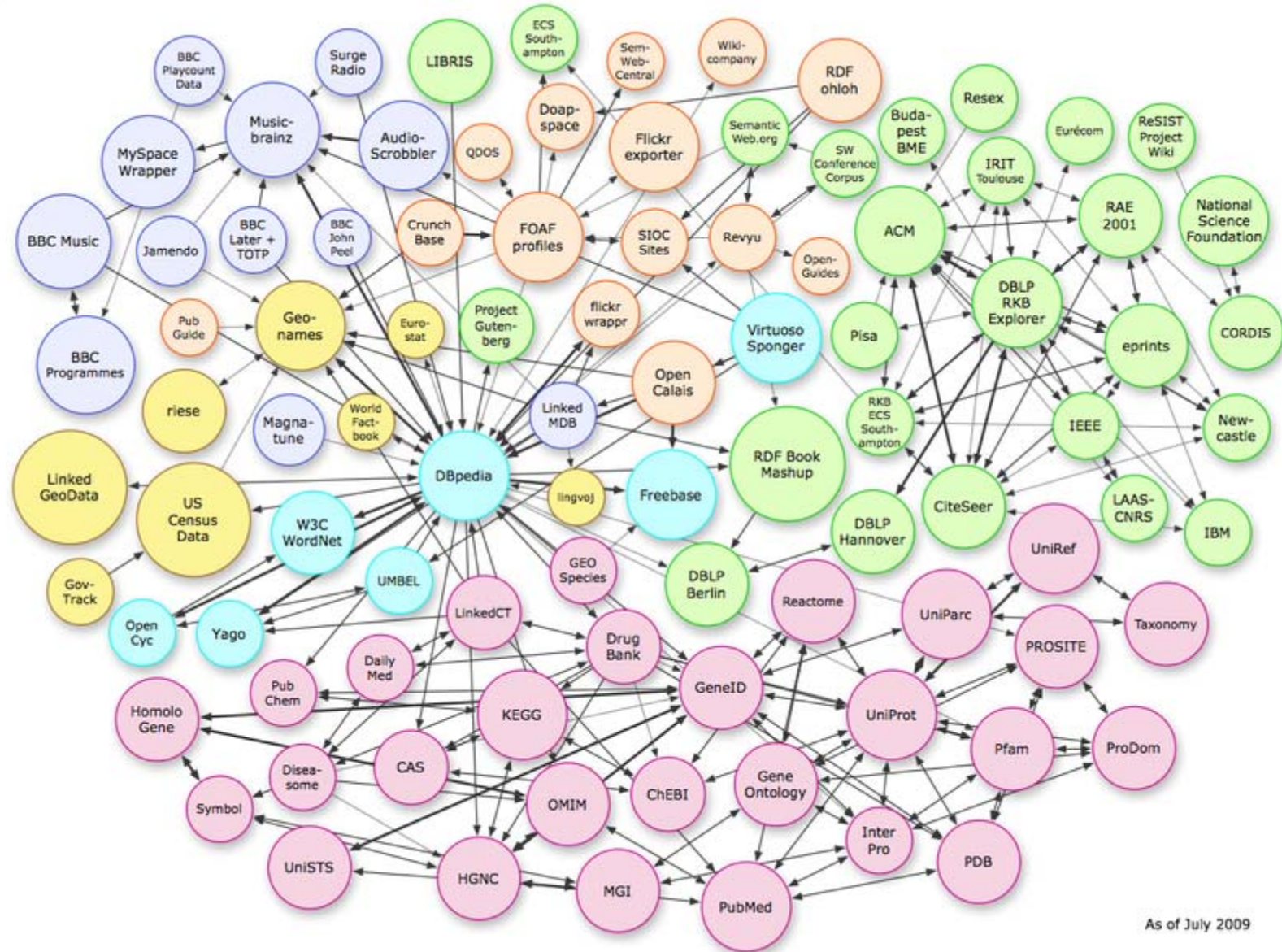
Linked Open Data

- 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 'gedereferenced', 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





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



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.

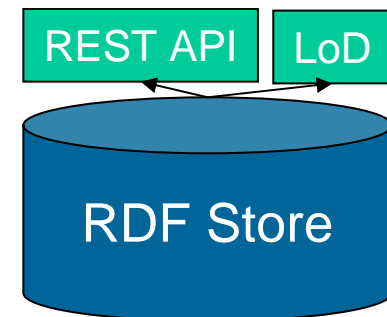


Wat hebben we tot nu toe gebouwd?

Architectuur, functionaliteit, voorbeelden en demo's

Architectuur

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





REST functionaliteit

- API basis functies
 - /find/concept
 - /find/conceptscheme
 - /find/conceptschemeigroup
 - /find/relation
- <http://catchplus.tuxic.nl/catchplus/serviceapi/1/>



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](#)



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



LoD demo

(werk in uitvoering)

<http://linkeddata.uriburner.com/ode>



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 Plus Vocabulary Repository - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home

http://vocrep.q42.net/#/repo?level=0|repo?level=1&uri=http://data.catchplus.nl/rdf/vocabularies/gtaal|repo?level=2&uri=http://www.beeldengeluid.nl/Thesaurus/Classificatie&letter=|repo?level=3&uri=http://www

Gmail - Postvak IN (65) - hennie.brugm... Catch Plus Vocabulary Repository

CATCH Vocabulairebank Beta

Browse

- Brinkman Trefwoorden
- UNESCO KB
- NUR:Nederlandstalige Uniforme Rubrieksindeling
- Gemeenschappelijke Thesaurus Audiovisuele Archieven**
- NBD/Biblion
- Regio Thesaurus KB
- OCLC NL:Karakteristiek in gecodeerde vorm
- Referentie Collectie Glas

Geografische namen

- Namen
- Classificatie**
- Genre
- Persoonsnamen
- Maker
- Onderwerpen

Zoeken op

Top concepten

- aardrijkskunde 15A
- algemeen 00X
- bestuur en geschiedenis 03B
- communicatie en media 11C**
- economie 05E
- gezondheid 06G
- kunst en cultuur 12K

Geen bredere bekend.

SKOS nauwere

- informatievoorziening overig 11C3**
- communicatie en media algemeen 11C0
- massacommunicatie en -media 11C2
- audiovisuele en elektronische media 11C4
- communicatietransport 11C1
- Geen gerelateerde bekend.

woordenboeken

- public relations
- audiovisuele archieven
- conferenties
- websites
- jaarverslagen
- naslagwerken
- bibliotheken
- archieven**
- encyclopedieën
- hoorzittingen

SKOS bredere

- informatievoorziening overig 11C3
- non-profitorganisaties
- SKOS nauwere**
- audiovisuele archieven**
- SKOS gerelateerde**
- collecties
- documenten
- opslagplaatsen
- musea
- bibliotheken

SKOS bredere

- informatievoorziening overig 11C3
- audiovisuele en elektronische media 11C4
- archieven
- Geen nauwere bekend.
- SKOS gerelateerde**
- filmconservering
- beeld- en geluidsdragers
- audiovisuele media

Details

Concept: audiovisuele archieven

Identifier

<http://www.beeldengeluid.nl/Thesaurus/27697>

Voorkeurstermen

Nederlands

Niet-voorkeurstermen

Nederlands

Gevonden in

[Onderwerpen](#)

SKOS bredere

- [informatievoorziening overig 11C3](#)
- [audiovisuele en elektronische media 11C4](#)
- [archieven](#)

SKOS nauwere

Geen bekend.

SKOS gerelateerde

- [filmconservering](#)
- [beeld- en geluidsdragers](#)
- [audiovisuele media](#)



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



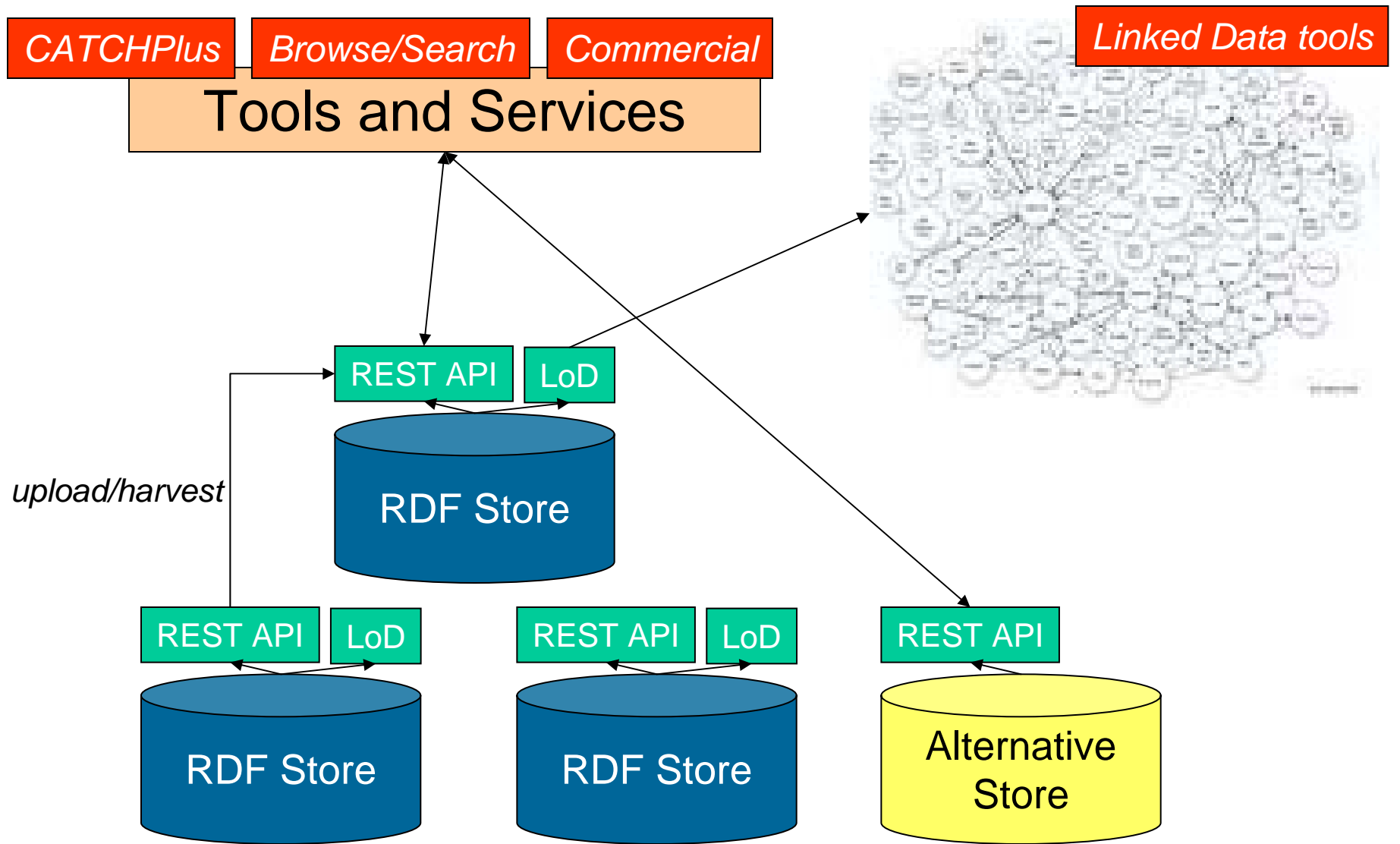
Status en beschikbaarheid

- 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



Toekomst

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





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.



Dank u. Vragen?