A distributed network of digital heritage information

SWIB17

Enno Meijers / 6 December 2017 / Hamburg

Contents

1. Introduction to Dutch Digital Heritage Network

2. The current digital heritage infrastructure

3. Strategies for improvement

4. Building a distributed network for digital heritage information

1. Introduction to Dutch Digital Heritage Network

●○○○○ vodafone NL ⊙   Koninklijke Bibliothe
●○○○○ vodafone NL ⊙   gahetNA | gahetNA is
●○○○○ vodafone NL ⊙   Home (English) | Nederlan
●○○○○ vodafone NL ⊙   https://cultureelerfgoed.nl
●○○○○ vodafone NL ⊙   The Academy institutes — K   ✕
10:06   ∗ 85% ▮

knaw.nl/en/institutes

NETWERK DIGITAAL ERFGOED

KB Koninklijke Biblioth
National Library
of the Netherlands

na
nationaal archie

Exhibition   Home

Star
in ou

Search for

Choose c

News

'Novelty bo
children's b

› A history o
with interc

One of the
manuscript
Eleanor of

› Peruse the

Renewed KI

› Visit our di

Open Prese
appoints ne

› KB's Barba
the founda

› More news

Discover our

News   Calendar

The world of the Du

Photos show the di

More news

› Big Panorama Mesdag
› The National Archives
   the Year Award 2015
› All At Sea: The Prize P
› Marens Engelhard new
› More news

gahetNA is a website of the

BEELD & GELUID

SOUND AND VISION
TREASURES

OSCAR® NOMINATED
AND WINNING FILMS

WINTER
SCHOOL
for audiovisual archiving

SOUND AND VISION
25 - 27 January 2017

Pauzeer slideshow

Zorg voor het e

RSS nieuwsberichten   RSS pers

Home   Erfgoed   Kenni

Nieuws

Aanvragen
herbestemmingss
vanaf 1 oktober 2
– 1 oktober 2017

Aanmelden
Monumententalen
tot 18 oktober
– 28 september 2017

Ontginningen Voo

SITE MAP   CONTACT   RSS   NEDERLANDS

Zoeken...   🔍

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

HOME |NEWS |ABOUT US |MEMBERS |INSTITUTES |ADVISORY WORK |AWARDS |INTERNATIONAL |TOPICS

> Home

THE ACADEMY INSTITUTES

Dutch scientists
build colon cancer
progression model

Press release
Hubrecht
Institute/UMC
Utrecht

The Academy is responsible for fifteen research institutes. The institutes are active in the humanities, the social sciences and the life sciences; one institute works at the interface of science and government policy. The Academy institutes are meant to play a leading role in Dutch and international research. They serve as national centres of expertise, manage unique infrastructures, and provide access to their collections, many of which are world-famous.

Policy

The institutes are meant to play a leading role in the Netherlands as agenda-setting organisations. They should also serve as international points of contact. In order to become (or remain) an Academy institute, they must strive to achieve excellence in their work. For more information about the Academy's policy regarding its institutes.

Knowledge transfer

Knowledge transfer is an important point of attention at the Academy institutes. To support the institutes and individual Academy researchers in this area, the Academy has set up its own Knowledge Transfer Office (KTO). The KTO helps institutes and researchers seek the expertise needed for knowledge transfer and supports them in attracting funding. More information about the KTO

All institutes of the Academy
Humanities and Social Sciences

› Data Archiving and Networked Services (DANS)

› Fryske Akademy (FA)

› Huygens ING

› International Institute of Social History (IISH)

› Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV)

› Meertens Institute

› NIOD Institute for War, Holocaust and Genocide studies

› Netherlands Interdisciplinary Demographic Institute (NIDI)

› Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS)

Life Sciences

› Hubrecht Institute for Developmental Biology and Stem Cell Research

› Netherlands Institute for Neuroscience

› Netherlands Institute of Ecology (NIOO)

› Spinoza Centre for Neuroimaging

› Westerdijk Fungal Biodiversity Institute

Programming & Social Debate

National Digital Heritage strategic plan (2015)

NETWERK DIGITAAL ERFGOED
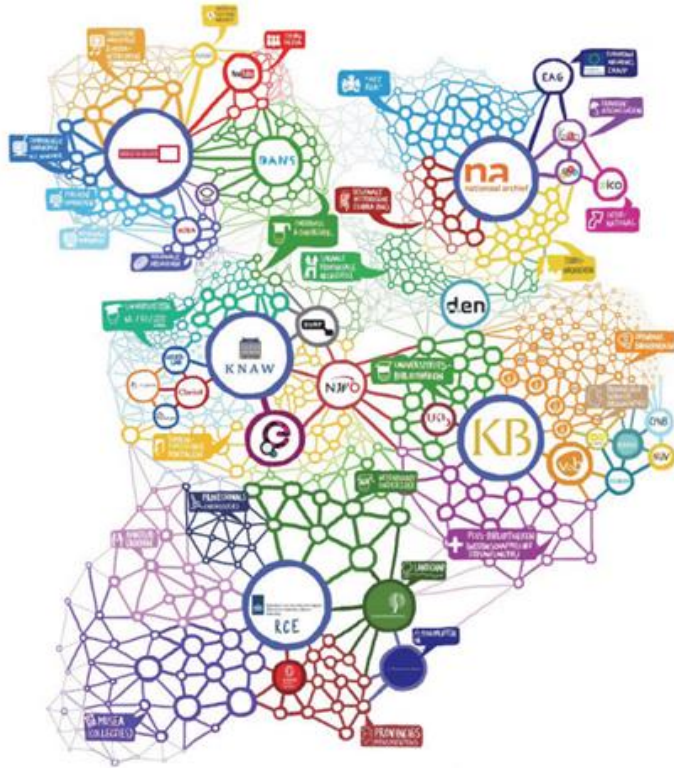
The Digital Heritage Network (NDE) aims at increasing the social value of the heritage information maintained by libraries, archives, museums and other cultural heritage institutions.

This strategy offers a perspective on developing a national, cross-sector infrastructure of digital heritage facilities.

It focuses on long term cooperation between the government and the institutions on national, regional and local level. It is about organizing the network of people and information!

Thinking from the user's perspective

Thinking from the user's perspective also means seeking out the digital platforms and work environments where potential users can already be found.

The attractiveness of information to a certain user group is not determined only by the nature of the information, but also by the method and location through which that information is offered.
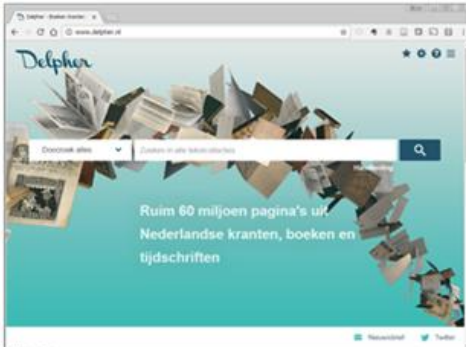
visible

usable
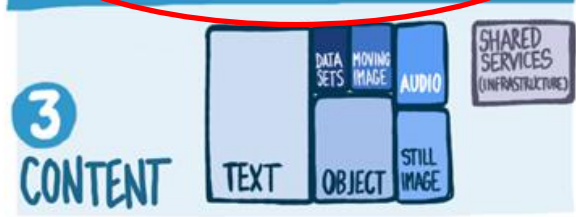
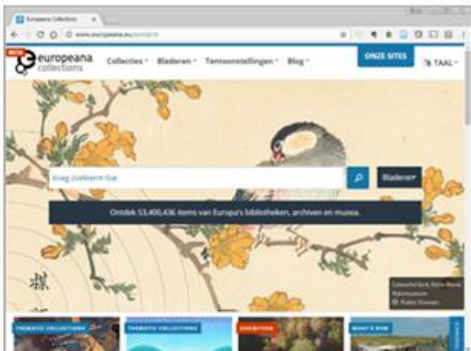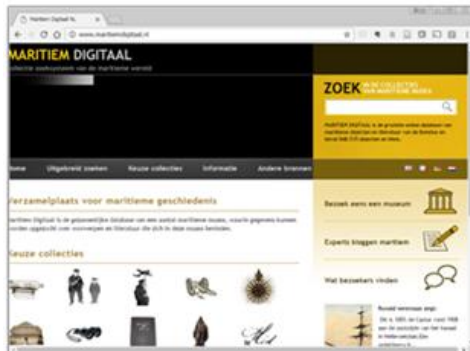**sustainable**

The Digital Heritage Network is developing a three-layered approach for improving the sustainability, the usability and the visibility of digital heritage information.
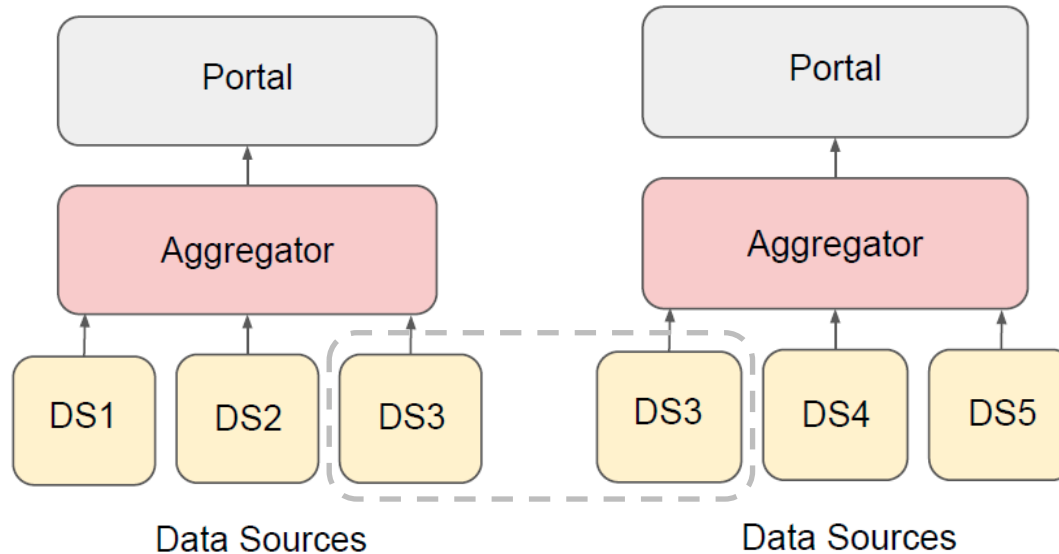
2. The current digital heritage infrastructure

NETWERK DIGITAAL ERFGOED

# General setup of digital heritage portals

Heritage information consisting of GLAM datasets and science collections

Evaluating current approach (1)

Positive results so far:

- many sources available through OAI-PMH protocol
- powerful and smart protocol for metadata synchronization
- opened up data silos
- created the need for aligning data models
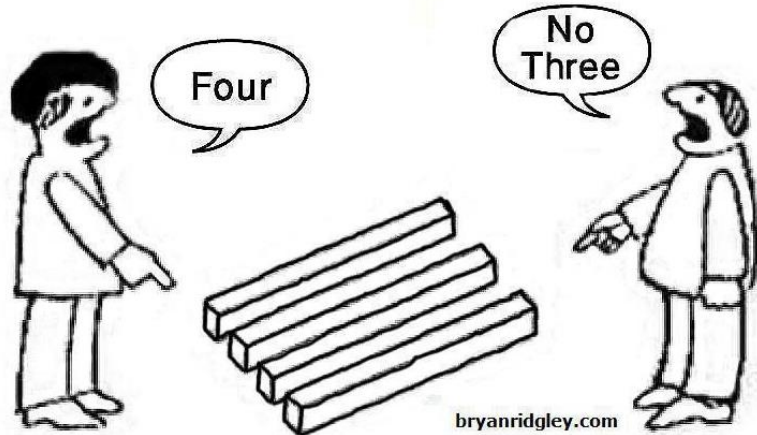- made cross-collection and cross-domain discovery possible (e.g. Europeana)

Networks of aggregators...
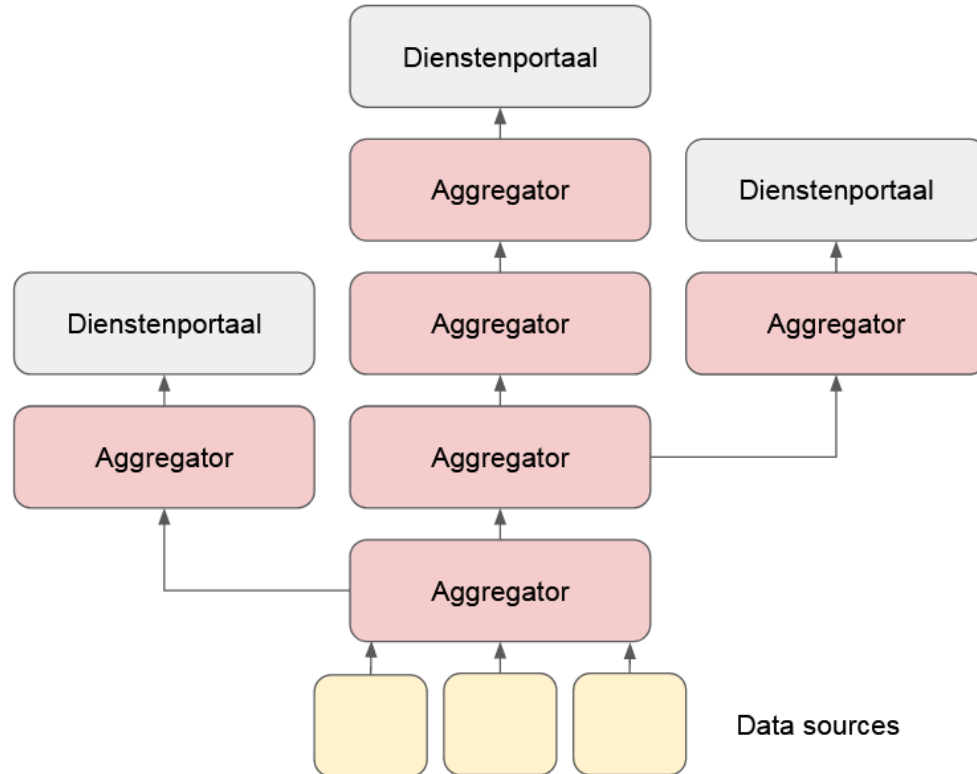
Evaluating current approach (2)

But there are two main problems areas:

- poor semantic alignment

- inefficient data integration

See also:
Miel Vander Sande et al. , Towards sustainable publishing and querying of distributed Linked Data archives - Journal of Documentation (2017)
Herbert Van de Sompel - Reminiscing About 15 Years of Interoperability Efforts - D-lib Magazine - December (2015)

3. Strategies for improvement

Design principles for a discovery infrastructure

- build service portals as views based on a common data layer
- minimize the intermediate layers
- support decentralized discovery
- refer to the source instead of copying
- maximize the usability of data at the source
- develop a sustainable, 'web-centric' solution
- use HTTP, RDF and RESTful APIs as building blocks

=> implement the Linked Data principals

Inspired by the work of Ruben Verborgh, Herbert Van de Sompel and colleagues:
See for example: Miel Vander Sande et al. , Towards sustainable publishing and querying of distributed
Linked Data archives - Journal of Documentation (2017)

Implementing Linked Data principles (1)

At the data source level:

- use sustainable URIs to identify the resources
- use formal definitions for persons, places, concepts, events
- use domain data models to describe the data
- add support for cross-domain discovery (Europeana Data Model, Schema.org,...)
- publish the collection information as Linked Data

=> Work with the IT suppliers as strategic partners for the implementation!

Implementing Linked Data principles (2)

At the network level:

- create a 'network of terms' for shared terminology
- provide tools for alignment and linking
- create alignments and links between different terminology sources
- provide easy access to shared terminology for collection management systems (API)

=> Provide open and cross-domain solutions at the network level!

Building on previous work

https://www.poolparty.biz/

Ok, but how will our Linked Data be found?

The Semantic Web is still a dream… #1

lod-cloud.net · 120% · Zoeken

Insight · OÉ Gaillimh NUI Galway · HPI Hasso Plattner Institut · UNIVERSITY OF MANNHEIM
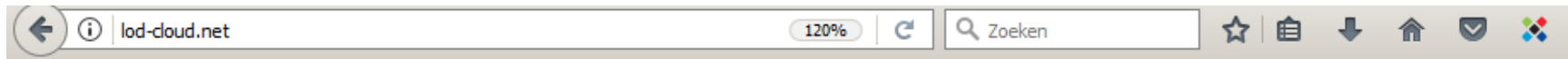
The Linking Open Data cloud diagram

This web page is the home of the *LOD cloud diagram*. This image shows datasets that have been published in Linked Data format, by contributors to the Linking Open Data community project and other individuals and organisations. It is based on metadata collected and curated by contributors to the Data Hub . Clicking the image will take you to an interactive SVG version, where each dataset is a hyperlink to its entry in Datahub.

The diagram is maintained by Andrejs Abele and John McCrae (Insight Centre for Data Analytics at NUI Galway). For any questions and comments, please email andrejs.abele@insight-centre.org and John.McCrae@insight-centre.org. The original version was developed by Richard Cyganiak and Anja Jentzsch.

Last updated: 2017-02-20

⇒ So discovery of Linked Data requires registering datasets?!

Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics

Collaboration in the Web of Data

- DBpedia's mission is to
  - build a hierarchical-decentralised network
  - serve as an access point for data
  - facilitate collaboration
  - disseminate data on a global scale
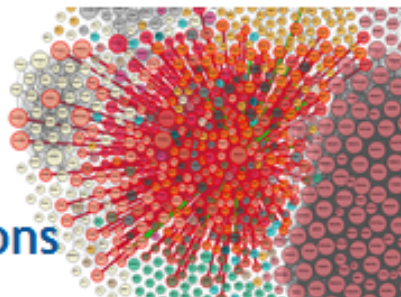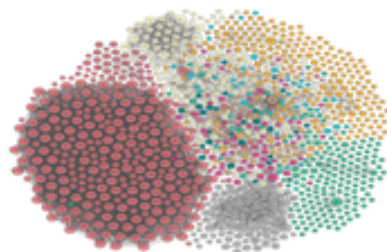
Connecting data is about connecting people and organisations

The Semantic Web is still a dream… #2

A tiny example...suppose a resource is defined as:

```
museum_X:object1
    a              nde:painting ;
    dct:subject    aat:windmill .
```


"Windmill"

The Semantic Web is still a dream… #2

A tiny example...suppose a resource is defined as:

```
museum_X:object1
    a               nde:painting ;
    dct:subject   aat:windmill .
```


→ *"Windmill"*

For 'browsable Linked Data' you should(!) add the inverse relation [1],[2]:

```
aat:windmill
    a                   skos:Concept ;
    skos:prefLabel      "Windmill"@en ;
    dct:isSubjectOf     museum_X:object1 .
```

"Windmill" →

[1]: Tim Berner's Lee on 'browsable linked data' (2006)   [2]: Tom Heath and Christian Bizer on 'Incoming Links' (2011)

The Semantic Web is still a dream… #2

A tiny example...suppose a resource is defined as:

```
museum_X:object1
    a                nde:painting ;
    dct:subject   aat:windmill .
```


$\longrightarrow$ *"Windmill"*

For 'browsable Linked Data' you should(!) add the inverse relation [1],[2]:

```
aat:windmill
    a                    skos:Concept ;
    skos:prefLabel       "Windmill"@en ;
    dct:isSubjectOf      museum_X:object1 .
```

"Windmill" $\longrightarrow$

=> a Linked Data integration problem, the lack of "backlinks"

[1]: Tim Berner's Lee on 'browsable linked data' (2006)   [2]: Tom Heath and Christian Bizer on 'Incoming Links' (2011)
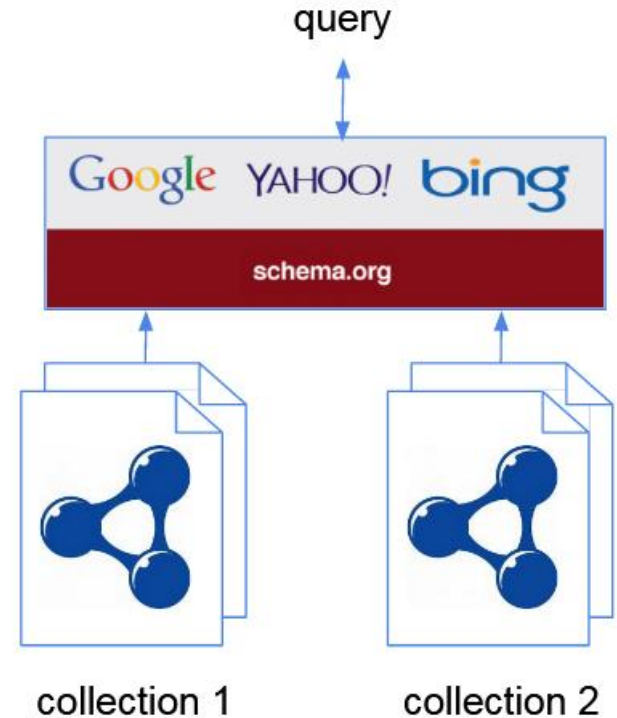
1. Semantic integration only

Actions:

- implement schema.org
- let search engines 'infer' the relations
- query the search engines

Outcome:

- is the data interesting enough for Google?
- what about special thematic or regional views?
- can we reuse the results of the integration? (NO!)

query

Google YAHOO! bing

schema.org
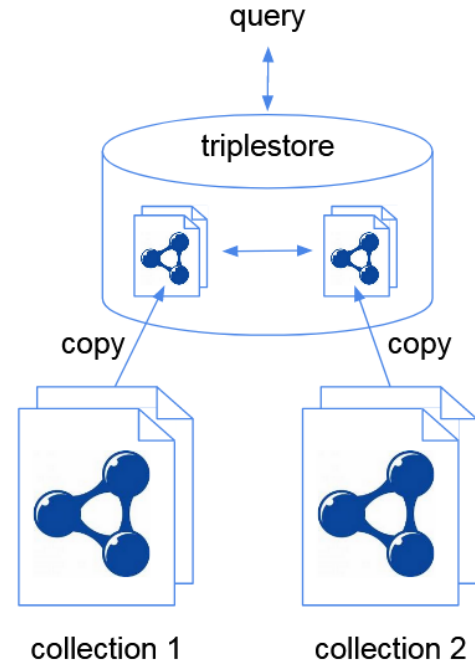
collection 1   collection 2

2. Physical integration of linked data

Actions:

- aggregate all the related Linked Data sources
- build large triplestore and infer the relations
- query the aggregated data

Outcome:

- approach still based on copying
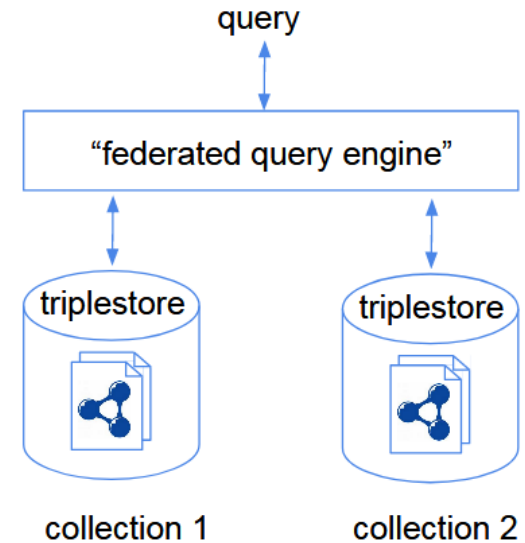- same problems as traditional aggregation!

3. Virtual integration - standard approach

Actions:

- publish Linked Data through triplestore with SPARQL endpoint
- build a central query engine to integrate the results

Outcome:

- implementing a triplestore is hard for small data providers
- federated querying over multiple triplestores performs poorly

4. Virtual integration - using LDF

Actions:

- publish Linked Data using [Linked Data Fragments (LDF)](#) technology
- build a central LDF based query engine to integrate the results

Outcome:

- easy implementation for small data providers
- federated querying is supported
- more difficult to process the result
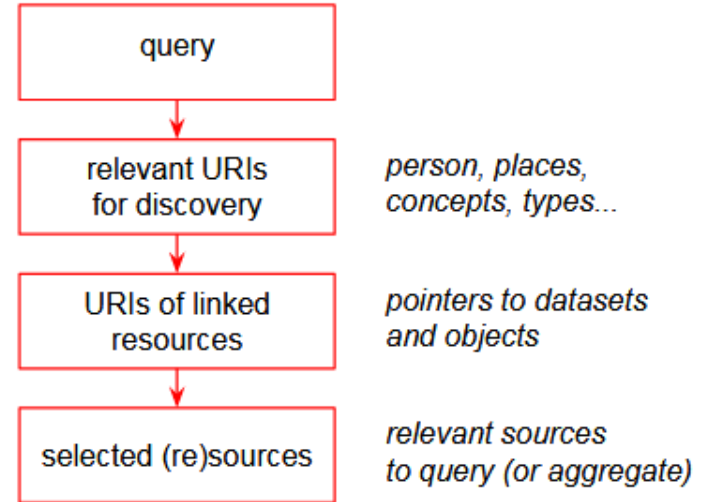- possible support for time-based versions ([Memento](#))

See also: *Miel Vander Sande et al. , (2017) Towards sustainable publishing and querying of distributed Linked Data archives* - Journal of Documentation
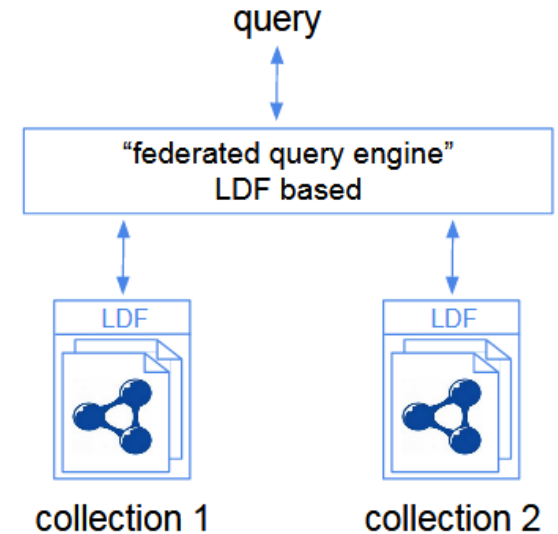
But federation needs selection of sources…

Problem:

- query many data sources at the same time is not realistic…

Solution:

- build a Knowledge Base with backlinks to support the discovery process
- select relevant sources for querying based on the Knowledge Base

query

↓

| relevant URIs for discovery | *person, places, concepts, types...* |

↓

| URIs of linked resources | *pointers to datasets and objects* |

↓

| selected (re)sources | *relevant sources to query (or aggregate)* |

***More advanced**: data source profiling or dataset summaries*

See also: *Miel Vander Sande et al. (2016) Hypermedia-Based Discovery for Source Selection Using Low-Cost Linked Data Interfaces (IJSWIS) 12(3) 79–110*

4. Building the distributed network of Dutch Digital Heritage information

Strategy for our distributed network

1. *build a service for shared terminology for Dutch digital heritage*

2. *improve the usability of the data source:*
   - align object descriptions with shared terminology
   - publish data as Linked Data

semantic
alignment

Strategy for our distributed network

1. *build a service for shared terminology for Dutch digital heritage*

2. *improve the usability of the data source:*
   - align object descriptions with shared terminology
   - publish data as Linked Data

3. *build a discovery infrastructure:*
   - register organizations and datasets in a (automated) registry
   - build knowledge graph to support discovery ("backlinks")

4. *implement **virtual** data integration technology :*
   - use registry and knowledge graph for selecting the resources
   - support federated querying (or selective aggregation)

semantic
alignment

data
integration

High-level design of our discovery infrastructure
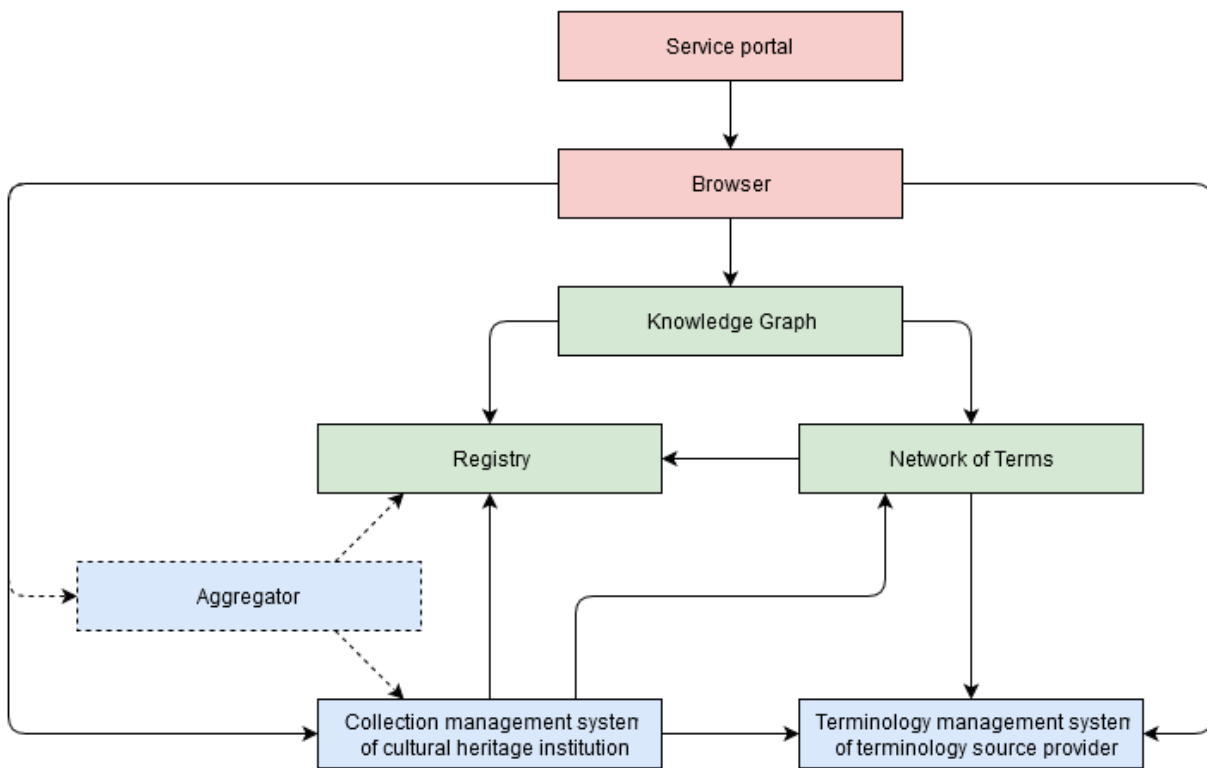
Roadmap (1)

Phase 1 – functional design / developing partnerships:

- design of supporting cross-domain functionality

- develop partnerships with IT suppliers and specialists

- develop domain and cross domain strategies

Phase 2 – enrich the current (OAI-PMH based) infrastructure:

- build a network of terms to provide shared terminology for discovery

- upgrade object descriptions with formal definitions (URIs)

- build an (automated) registry for organizations and datasets

Roadmap (2)

Phase 3: implement Linked Data technology at the network level

- make aggregators Linked Data compliant

- build a knowledge graph with backlinks for discovery

- support federated querying (or selective harvesting)

Phase 4: realize the distributed network of heritage information

- make collection management systems Linked Data compliant

- transform aggregators to service portals for discovery

Work in progress…

- Working with many institutions in multiple projects

- Prototyping infrastructure components

- Making terminology sources available and findable (OpenSKOS)

- Developing showcases for adoption of the Linked Data principles

- Adamnet: adamlink.nl (@lukask,…)

- Zuiderzeecollectie.nl

Thank you for your attention!

please share your thoughts with us...

email: enno.meijers at kb.nl
twitter, slideshare: ennomeijers

https://github.com/netwerk-digitaal-erfgoed

NETWERK DIGITAAL ERFGOED