A distributed network of digital heritage information

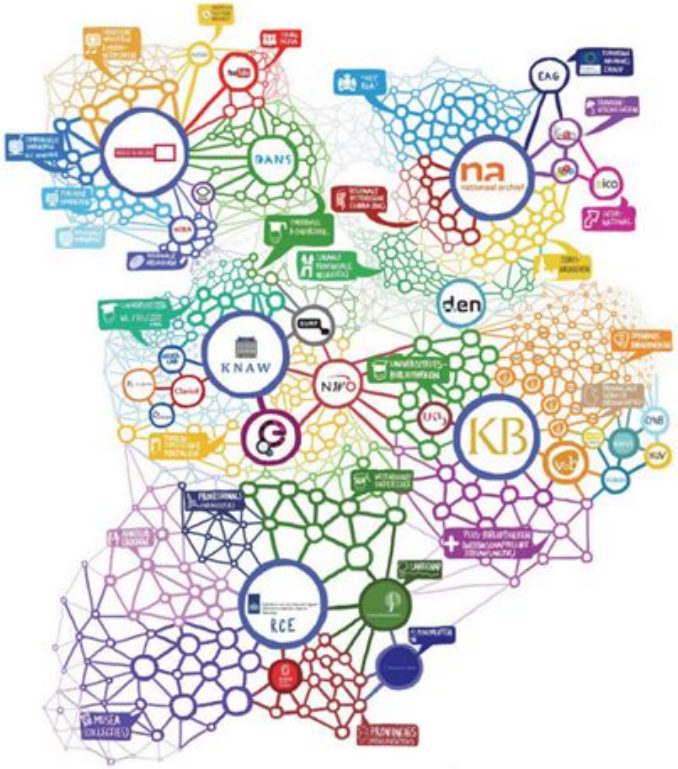
Enno Meijers - KB / Netwerk Digitaal Erfgoed

DANS Seminar - 1 May 2017

KB Koninklijke Bibliotheek
National Library of the Netherlands

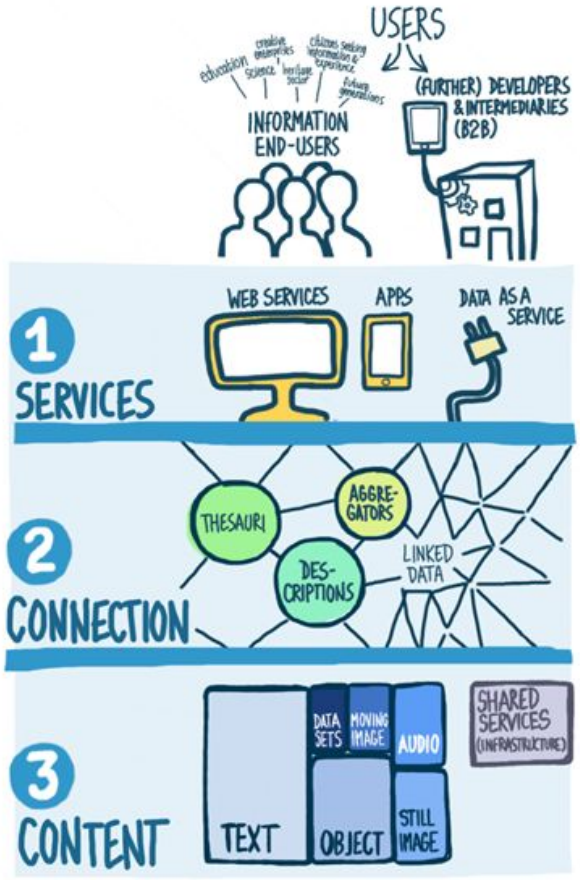
Contents

- The Dutch Digital Heritage Network
- Current aggregation practice
- A distributed web of cultural heritage information
- Design issues
- Roadmap



Netwerk Digital Erfgoed (NDE) aims at increasing the social value of the (digital) heritage information maintained by libraries, archives, museums and other cultural heritage institutions.

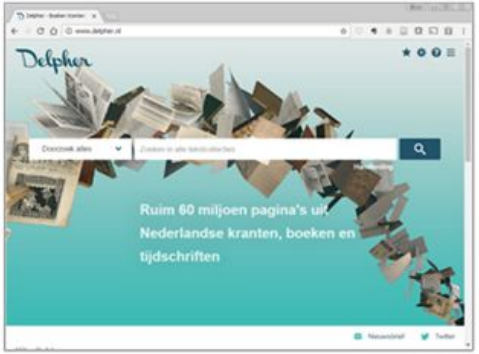
Long term cooperation between the government and the institutions on national, regional and local level. It's about information and people!



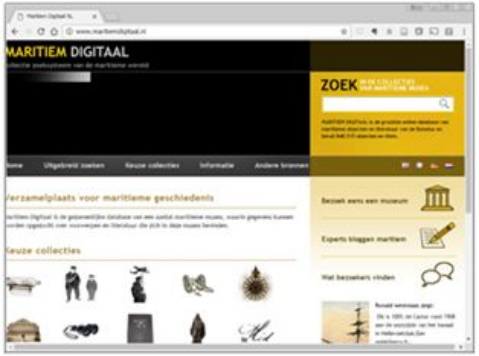
NETWERK DIGITAAL ERFGOED

Three layered approach for improving the sustainability, the usability and the visibility of digital heritage information.

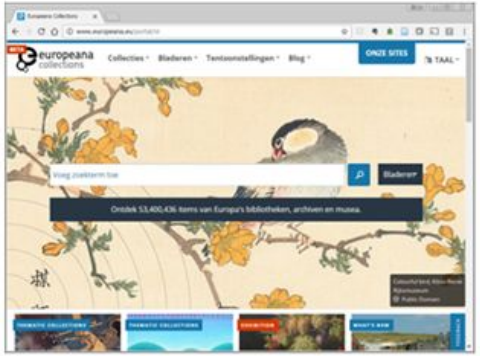

Collectie
Gelderland
zoeken
© CollectieGelderland


Delpher
Ruim 60 miljoen pagina's uit Nederlandse kranten, boeken en tijdschriften


Collectie Nederland: Musea, Monumenten en Archeologie
Doorzoek de meer dan 3419492 kunstwerken, monumenten en andere objecten uit Collectie Nederland.
Collecties van archieven, musea en andere cultuurinstellingen bijeen


MARITIEM DIGITAAL
ZOEK
Verzamelplaats voor maritieme geschiedenis
Keuze collecties


Ontdek Oneindig Noord-Holland
Beleef de geschiedenis van Noord Holland
Zoek in 17.236 verhalen, activiteiten en routes
Zoeken

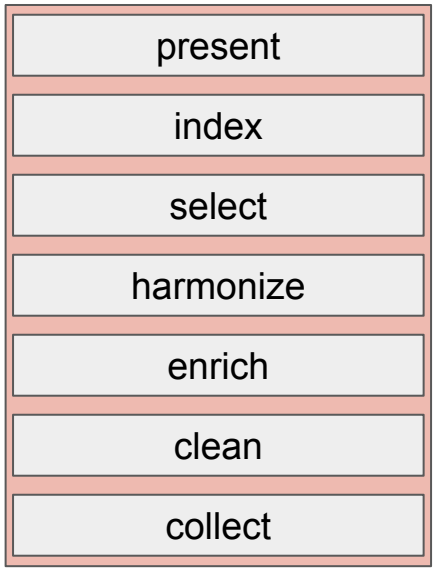

europeana collections
ONZE SITES
Ontdek 53,400,436 items van Europa's bibliotheken, archieven en musea.


BIOGRAFISCH
portaal van nederland
Nieuws
Zoeken
Bladeren


redbot
fries erfgoed hub
Home
RedBot
Projecten
Samenwerkingspartners
Nieuws
Contact
Een 21st century issue
Storyteller in het gevang


Doorzoek de collecties in Nederland over de Tweede Wereldoorlog
NETWERK
OORLOGS BRONNEN
ZOEK

Typical aggregation process



presentation at portal level => no feedback to the datasource

aimed at presentation => possible inefficient use of data

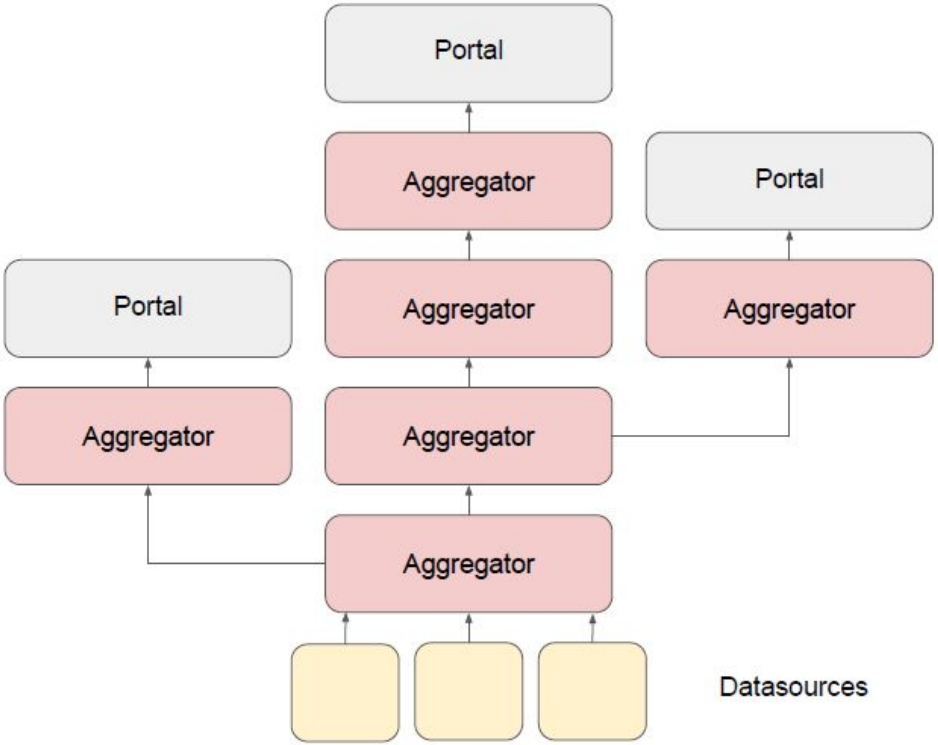
one size fits all => possible loss of details

isolated improvements => datasource stays unchanged

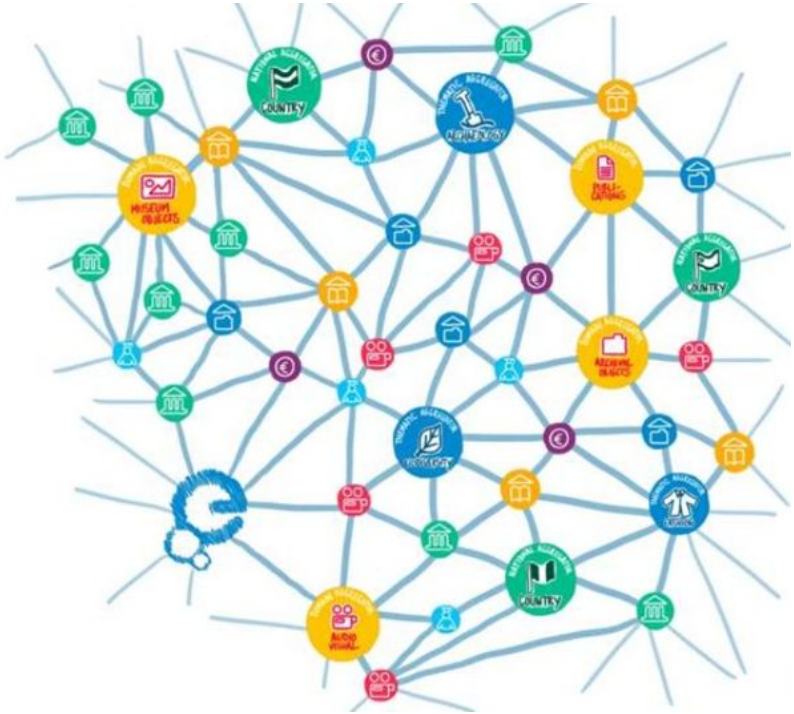
OAI-PMH based => data is copied

Based on a repository-centric approach...

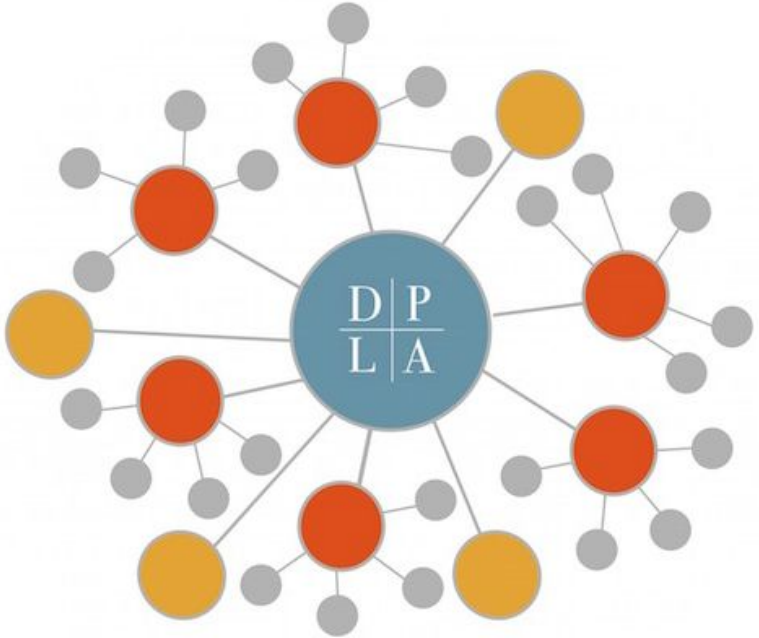
And we build networks of aggregators...



And it is a lot of work...



Europeana

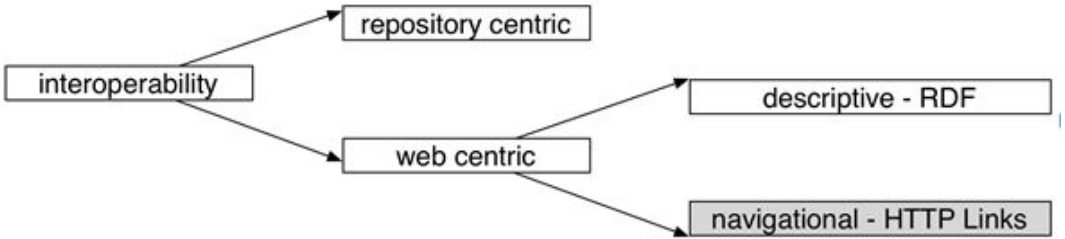


Digital Public Library of America

Web-Centric, Resource-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of the web

- **The resource at the center of the universe**
  - The notion of a repository, not even of a web server exists in the architecture of the web
- The tools of the interoperability trade are the primitives of the web



RDF - the Lingua Franca of Data Integration

- RDF is **simple**
- We can easily **encode and combine all kinds of data models** (relational, taxonomic, graphs, object-oriented, …)
- RDF supports **distributed data and schema**
- We can **seamlessly evolve** simple semantic representations (vocabularies) to more complex ones (e.g. ontologies)
- Small **representational units** (URI/IRIs, triples) facilitate mixing and mashing
- RDF can be viewed from **many perspectives**: facts, graphs, ER, logical axioms, graphs, objects
- RDF **integrates well with other formalisms** - HTML (RDFa), XML (RDF/XML), JSON (JSON-LD), CSV, …
- Linking and referencing between different knowledge bases, systems and platforms facilitates the creation of sustainable **data ecosystems**

Sören Auer (Fraunhofer): Enterprise Knowledge Graphs - 2016

Developing a resource-centric approach

First step: Implement Linked Data principles in the datasources

- use formal (sustainable!) URIs that describe the resources
  - specify a URI strategy for the Digital Heritage Network
  - implement persistent URIs (handles)
- provide resource URIs that deliver the data in RDF
- use of formal definitions for persons, places, concepts, events (URIs)
  - provide Linked and Open reference sources to link to (thesauri, taxonomies, etc)
  - build a Knowledge Graph for Dutch digital heritage
  - align and link related terms
- add support for cross-domain discovery (EDM, Schema.org,...)

=> but how can this help our ‘Googling’ users to find our information?

Design issues...

A tiny example...suppose a resource exists:

<http://data.museum1.nl/object1>
a nde:painting ;
dcterms:subject <http://knowledgegraph.nde.nl/term1> .

For ‘browsable Linked Data’ you would like to add the inverse relation:

<http://knowledgegraph.nde.nl/term1>
a skos:Concept ;
skos:prefLabel "Flowers" ;
skos:isSubjectOf <http://data.museum1.nl/object1> .

TBL about ‘browsable linked data’ [1]:
In practice, when data is stored in two documents, this means that any RDF statements which relate things in the two files must be repeated in each.

[1]: https://www.w3.org/DesignIssues/LinkedData.html

Possible approaches to make it work

1. Aggregate all the related Linked Data resources:
- collect datadumps and build a large triplestore
- impressive work: LOD Laundromat

2. Use federated querying:
- doesn't work very well using regular triplestores
- implement Linked Data Fragments?!
- realistic for a network with about 1500 nodes?
- use logic for predicting most relevant endpoint [1]?

[1]: Routing Memento Requests Using Binary Classifiers

Or should we build a registry?

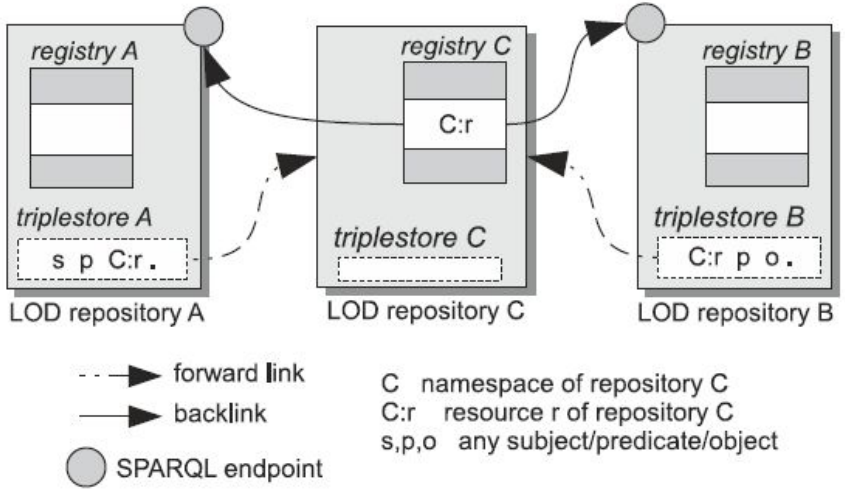
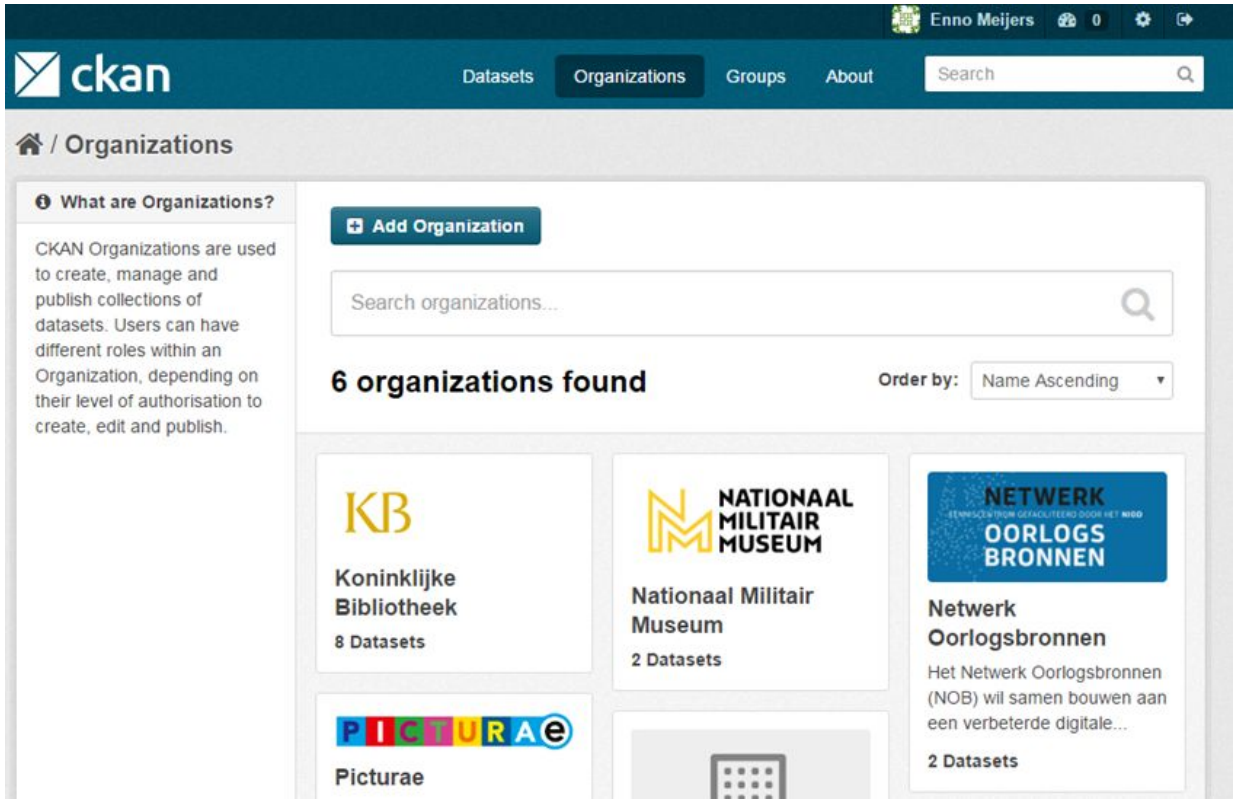


Figure 2: Interlinking LOD repositories via the Registry-based infrastructure

A decentralized infrastructure for the efficient management of resources in the web of data - Michalis Stefanidakis, Ioannis Papadakis - SWIM 2012

Register organisations, datasets and backlinks



using DCAT,
or DataID?

NDE’s strategy for resource discovery

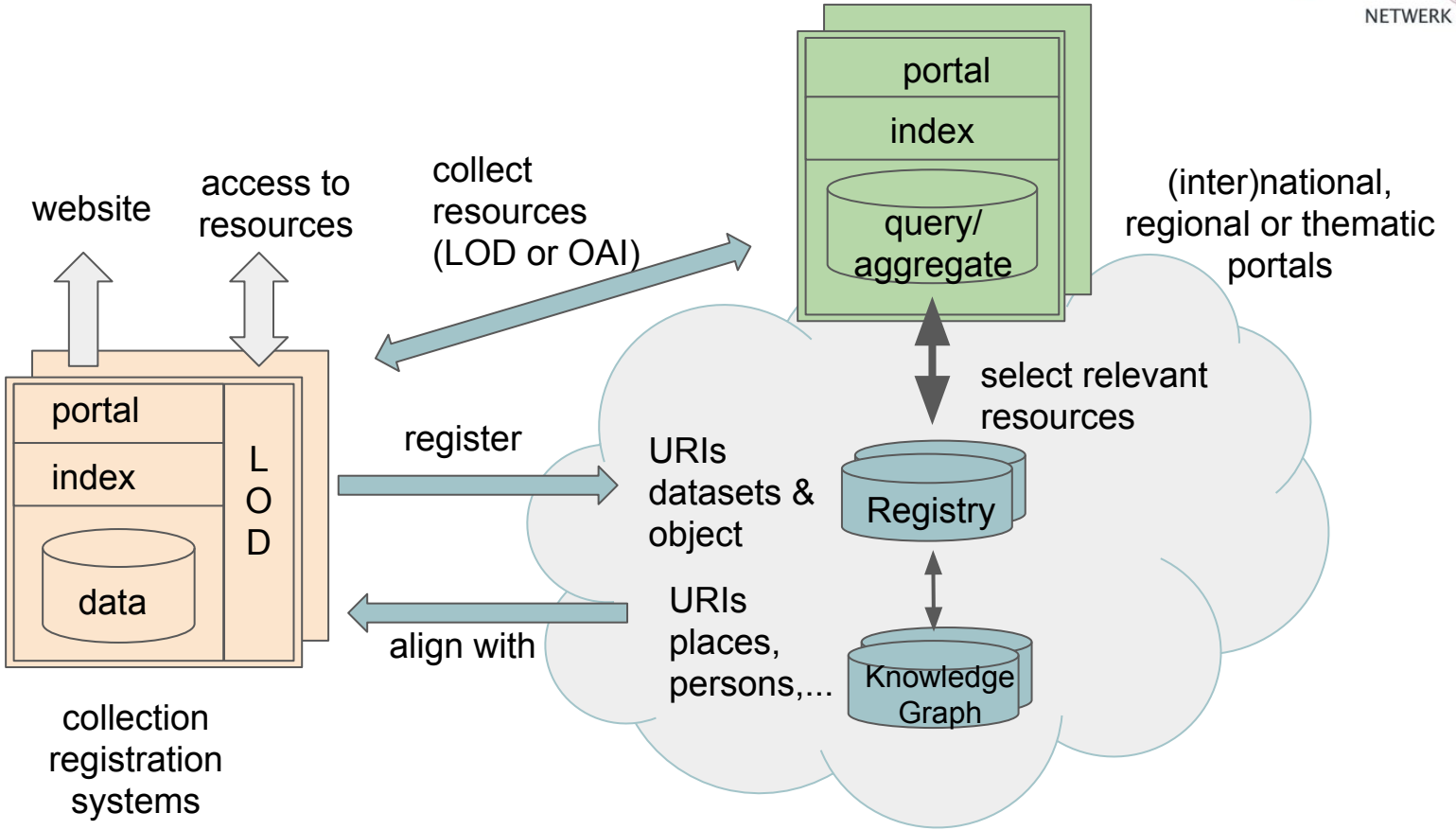
1. improve the usability of the data at the source:
   - implement linked data principles
   - align references for people, places, events, concepts (URIs!)

2. build a knowledge graph for Dutch digital heritage

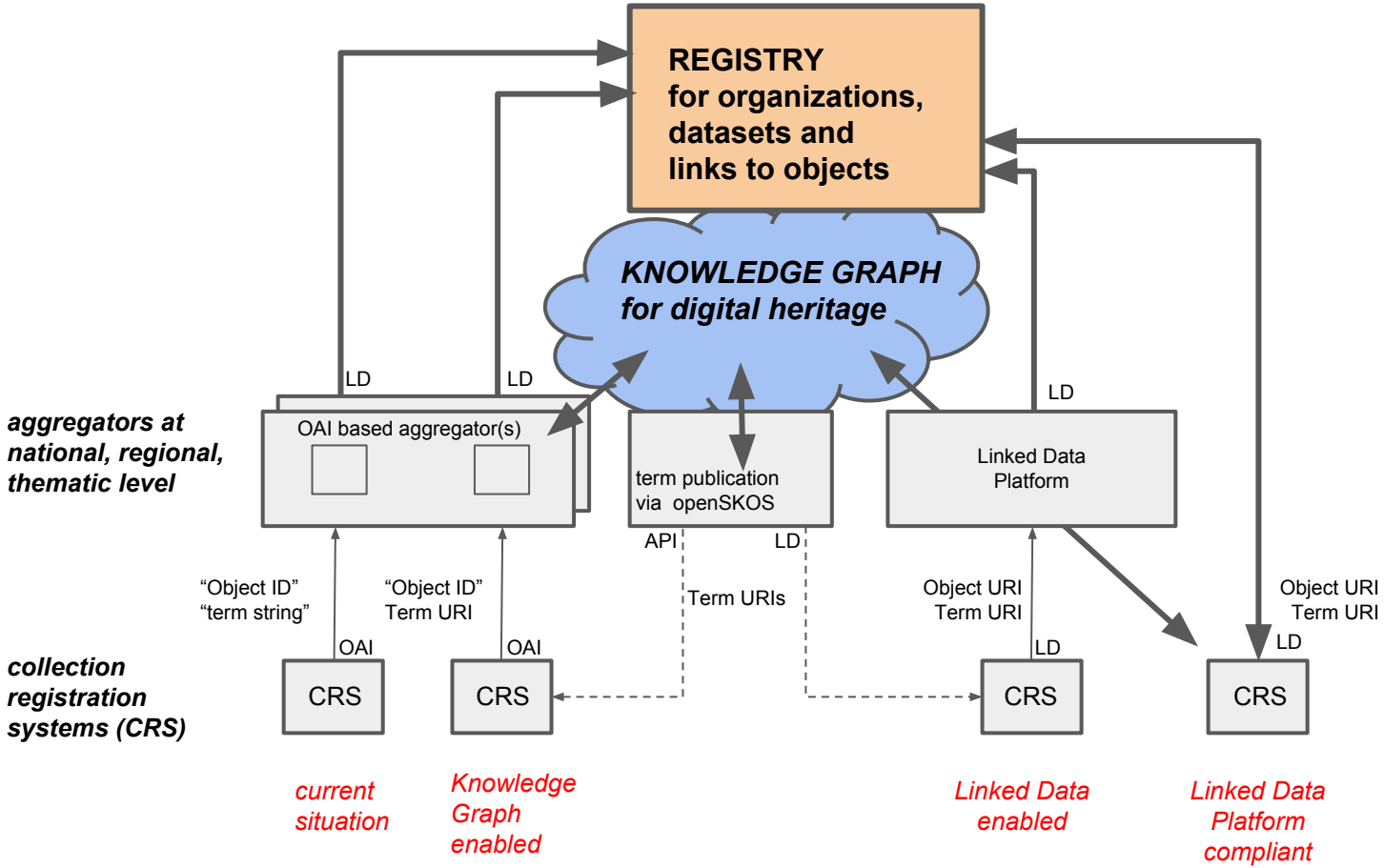
3. build a registry to link to the resources:
   - use formal definitions for organisations and datasets (DCAT/DataID)
   - send backlinks to registry (using Linked Data Notifications?)

4. support selective collection of data by federated querying (or harvesting):
   - use registry and knowledge graph for selecting the resources
   - use Linked Data Fragments for federated querying?

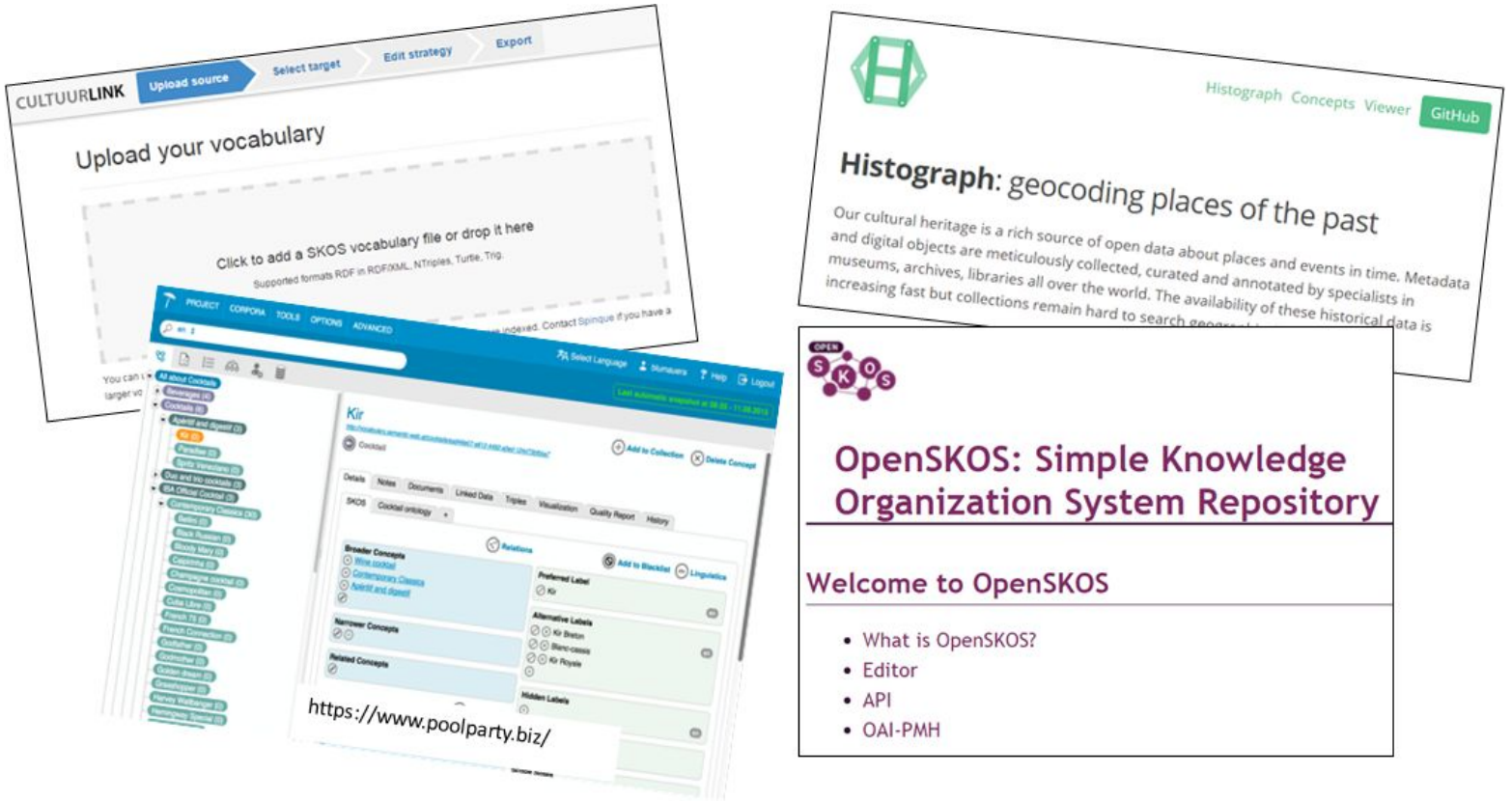
In a nutshell...



Roadmap from OAI-PMH to LOD



Building on previous work



Proof-of-Concept phase

Implementing (parts of) this vision in actual projects:

- Zuiderzeemuseum / Zuiderzee Collectie
- AdamNet
- Netwerk Oorlogsbronnen
- Modethesaurus
- Beeldbank Nederlands-Indië
- Groninger Archieven (webarchief)
- Nationaal Archief – project ‘Nadere toegangen’
- Nationaal Militair Museum
- RKD – RKD Artists

A distributed network of digital heritage information??

TBL on Decentralization:

This is a principle of the design of distributed systems, including societies. It points out that any single common point which is involved in any operation trends to limit the way the system scales, and produce a single point of complete failure.

Interesting work:

- Interoperability and FAIRness [1]
- IPFS and Linked Open Data [2]
- https://discuss.ipfs.io/t/whos-using-ipfs-in-libraries-archives-and-museums/130

[1]: Interoperability and FAIRness through a novel combination of Web technologies - Mark D. Wilkinson, Ruben Verborgh et al. - April 2017 - https://peerj.com/articles/cs-110/

[2]: Sharing Linked Open Data over Peer-to-Peer Distributed File Systems: The Case of IPFS - Miguel-Angel Sicilia, Salvador Sánchez-Alonso - November 2016 - DOI: 10.1007/978-3-319-49157-8_1

Thank you for your attention!

Please share your thoughts with me...
enno.meijers@kb.nl
twitter/slideshare: ennomeijers

Questions for the panel / audience

1. The ideal of the Web of Data seems to be unrealistic because links cannot be used in a bidirectional way. What are your expectations in regard to the future possibilities of real ‘browsable linked data’? Will a central infrastructure like a triplestore, registry or index always be necessary? Will an ideal implementation of the ‘read/write’ web solve this or a distributed web technology like IPFS?

2. Should web resources by default support a way to register backlinks similar to the mailbox idea of Linked Data Notifications?