

NETWORK DIGITAL HERITAGE

Towards a distributed network of heritage information

Position paper

March 2016

Authors:
Johan Oomen (Netherlands Institute for sound and vision)
Enno Meijers (KB Dutch National Library)
Wilbert Helmus (Network Digital Heritage)

Table of contents

Introduction	3
1. Aggregate and link	4
1.1 Issue outline	4
1.2 Starting points	5
2. Distributed network of heritage information	6
2.1 The concept	6
2.2 Linked Data, national registry and references	7
3. Next steps	9
3.1 Impact for heritage institutions	9
Annex 1 Overview of networks	10
At international level	10
National - central	10
National level – aggregation per field	11
Regional level	12
Thematic level	12

Introduction

In the position paper here presented, the Network for Digital Heritage (NDE) describes its vision for the future of digital heritage collections in the Netherlands. Naturally the NDE takes into account the national and international developments in the field of aggregation and linked data. The NDE is of the opinion that the future of a digital collection for the Netherlands lies in a *distributed network of heritage information*. However, there is still a long way to go before this network becomes a reality. Yet this vision of the future already provides a perspective on the scope of action for carrying out the National Strategy for Digital Heritage, in particular for the work package “Digital Heritage Usability”.

1. Aggregate and link

1.1 Issue outline

The National Strategy for Digital Heritage aims, among other things, to improve access to digital heritage. In the current situation, a large number of heritage institutions and (inter)national cooperation initiatives are actively collecting, transforming and enriching digital heritage information and make them accessible and available to the public. Many projects involve investments in and construction of various types of technical infrastructures that are difficult to connect. This approach, consisting in gathering and collectively presenting digital heritage, is inefficient, leads to unnecessary doubling of efforts and investments, and produces too little added value for the end user.

The current practice has a series of major flaws and is therefore not recommended to be continued:

- Metadata is still regularly copied and physically stored several times in different portals such as Europeana, APE, Digitale Collectie and/or DiMCoN. These portals harvest metadata from their original sources, copy them to their own storage spaces and create an index thereof. This procedure leads to unnecessary doubling of metadata¹, often resulting in older and out-dated information being used;
- In the course of this process, the original data is also often mapped into the data structure of a given aggregation platform. There is a risk of information being lost during these conversions that was initially present in the source data. Furthermore, these conversions must be updated regularly, which requires both money and time.
- The National Strategy puts the user first, which means that the user must be able to easily browse the available heritage data and compare and merge them with other information. These requirements, that also form the basis for DERA², are insufficiently met in the current situation: the digital collection of the Netherlands still poses major hurdles to the user and cannot be used in a transparent manner;
- Current practices drive heritage information providers into a web of portals and aggregators that each impose their own requirements and standards.
- Consequently, it becomes necessary to re-establish a vision for the future of the digital collection for The Netherlands.

The core of the solution lies in stimulating the publication of Linked Data³ straight from the source. Nevertheless, the vision expressed in this position paper does not exclude the potential necessity, with regard to particular goals, to aggregate and centrally process heritage information for the purpose of specific scientific research for example.⁴

¹ This is not the case with all aggregators; in the library sector for example, synchronisation is used to feed the common automated cataloguing system GGC.

² DERA stands for Digital Heritage Reference Architecture (NDE-OCW, 2016). The present paper is based on the content and key points of DERA.

³ The term Linked Data requires some further explanation. Not all data must (can) be 'open'. See <http://linkeddata.org/faq>: *Does all Linked Data need to be Linked Open Data? In a word, no, and it will likely never be so. The label "Linked Open Data" is widely used, but often to refer to Linked Data in general, rather than to Linked Data that is explicitly published under an open license. Not all Linked Data will be open, and not all Open Data will be linked. Therefore care should be taken to use the appropriate term, depending on the licensing terms of the data in question.*

⁴ For this reason, the NDE project "Usability" coordinates the aggregation needs of scientists with project 2 of the CLARIAH-network.

1.2 Starting points

Creating a nationwide distributed network of heritage information linked to a national network of reference sources takes time. During the transition towards a fully network-linked environment it remains vital to collect data centrally. It will be necessary to set up indexes based on aggregated information in order to be able to search and find heritage in various portals. In the long term, this procedure could increasingly take place within a distributed and linked network. In the meantime, during this transitional period, we are obliged to copy heritage information to collection points where we establish the links to reference sources, enrich the data and possibly publish it as linked data. However, we see possibilities to combine aggregation and linking of data already at this stage of the process, thereby enabling a gradual transition. The NDE will try to shorten the transitional period.

The Digital Heritage Reference Architecture (DERA) shall be used as a key starting point (framework requirement). The reference architecture is based on the strategic goal that consists of placing the user at the centre of the preoccupations in order to increase the usability and added value of digital heritage in the Netherlands. In order to implement these goals within the context of the NDE, six principles were established indicating how the heritage sector ought to facilitate these objectives:

1. evident authenticity of heritage information
2. clear availability of heritage information
3. user-friendly provision of clearly recognisable heritage information
4. unequivocal description of heritage information
5. heritage information refers
6. heritage information can be referred to

The DERA principles shall be further developed in the form of draft requirements and standards. The following chapter develops the vision of a distributed network of heritage information.

2. Distributed network of heritage information

2.1 The concept

The technical infrastructure for the digital collection for the Netherlands is developed within the NDE into a nationwide distributed network of heritage information. This network provides transparent access with a low threshold to heritage information that the users find relevant. In the long term, a dynamic connection between locally managed and presented heritage information and supra-local (thematic) intersections is created. The digital collection for the Netherlands thereby created can feed various portals containing direct references to the heritage information and related contexts. The disadvantages explained earlier with regard to current practices are thus minimised. The digital collection can be defined as a *register with references* to the cultural heritage as conserved by the institutions in question. By making the digital collection available as a common *finding aid* the heritage institutions have actually federated (along the lines of OAI federated archives⁵).

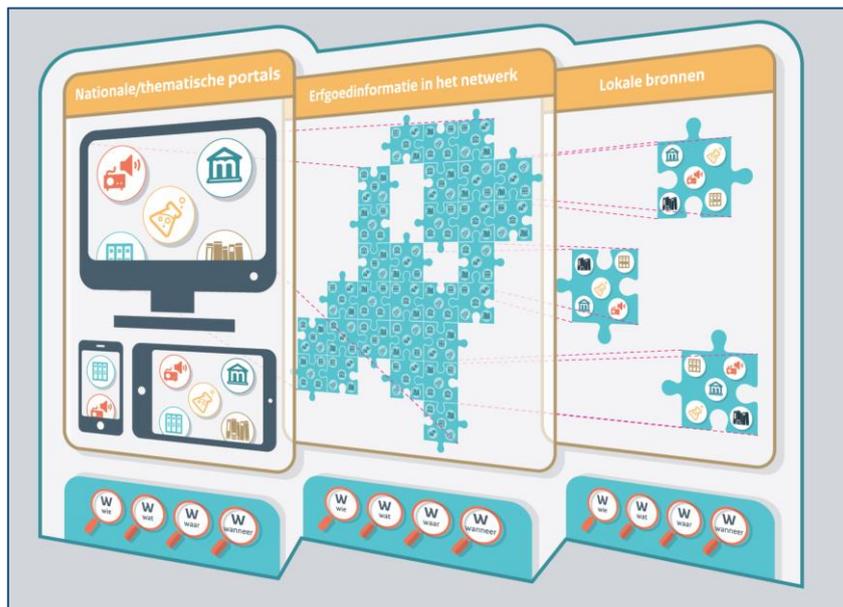


Figure 1: Concept distributed network of heritage information (infographic CreaBea)

The principles for implementing such a distributed network are based on DERA. The two most relevant principles are:

1. Heritage information refers: in the reference section, (inter)national terminology is used for the context (who, what, where, when) and for related information published by third parties;
2. Heritage information can be referred to: in the reference section, the information is published according to a common and standardised model applicable in the heritage sector, offering the possibility for third parties to refer to it.

There is a large consensus at national level about the pertinence of working with (inter)national references. The development of a network of references is in line with the prospect of renewing the

⁵ See also: http://wiki.dpconline.org/index.php?title=6.1.3_FEDERATED_ARCHIVES

technical infrastructure of the aggregation landscape. Various nodes in the Netherlands have been working hard in the past years to elaborate the prerequisites for implementing a network of heritage information: For example, projects in which SKOS⁶ has been used as a general data model to organise knowledge systems (that is to say terminology sources such as thesauri) and to compare definitions of concepts. The terminology sources can then be linked and compared by means of a harmonisation (alignment) service such as CultuurLINK⁷. Within the framework of the project Heritage & Location a tool has been developed, among others, in the past years that allows harmonisation of (historical and) geographical information. Terminology sources such as thesauri (i.e. checked terminology lists also indicating the relationships between terms) can be used to link metadata at the semantic level.

The (inter)national terminology and reference sources are linked in a sustainable and user-friendly manner in the form of a knowledge graph for heritage information thanks to the project Usability of Digital Heritage.⁸ In 2012, Google introduced its knowledge graph. The search engine of Google uses it to recognise information units such as people, places, companies, etc. and links them. The Knowledge Graph of Google is based on different sources, among which Google+, DBpedia/Wikipedia and Wikidata. In this way the search engine is able to better analyse the user intentions and the context of the search words he or she introduces. The knowledge graph provides an overview of the relationships between an increasing number of words (and entities)⁹.

Within the framework of the National Strategy, these separate components are developed into national tools available to heritage institutions as utilities for the publication and linking of their own heritage information. The NDE domain-nodes play a key role for institutions that want to introduce their collections in the register. There are different contact points according to the type of collection:

- Archives: National Archives
- Publications: Royal Library
- Museum objects; built and archaeological heritage : Cultural Heritage Agency (RCE)
- Audio/visual and new media heritage: Netherlands Institute for Sound and Vision (Beeld en Geluid)
- Immaterial cultural heritage, academic heritage and relevant datasets: Royal Netherlands Academy of Arts and Sciences (KNAW)

2.2 Linked Data, national registry and references

In an ideal world, local collection managers publish heritage information in a directly accessible manner (for example as linked data) and use links to national terminology sources for basic information: who, what, where and when.

In a national register, references to local URIs of source data are recorded when heritage information is published. At local level (source level), references to (inter)national terminology sources are also recorded in the form of URIs. In doing so, a two-way synchronisation takes place. At local level, the national references are used, and at national level, a *register of references in form of URIs to local collections and their items* is created. In environments in which supra-local presentations are made, this register of references can be used to carry out a selection and to gather data for indexes of spe-

⁶ Simple Knowledge Organisation System (SKOS) is an interchangeable data model for sharing and linking systems of knowledge representation throughout the Web. Examples thereof are thesauri, taxonomies, terminology dictionaries, classification schemes and systems for keyword allocation. The system is applied in one of the CATCH projects.

⁷ See: <http://cultuurlink.beeldengeluid.nl/app/#/>

⁸ See recommendation report on behalf of Netwerk Digitaal Erfgoed: A knowledge graph for the integration of heritage information (Spinque, 18 February 2016)

⁹ For more information about the Knowledge Graph of Google see also <http://www.frankwatching.com/archive/2014/04/09/het-semantic-web-zo-treed-je-t-vol-vertrouwentegemoet/> and a more in-depth article at: <http://www.seoskeptic.com/semantic-seo-making-shift-stringthings/>

cific portals.

This gives us the following image of a distributed network of heritage information:

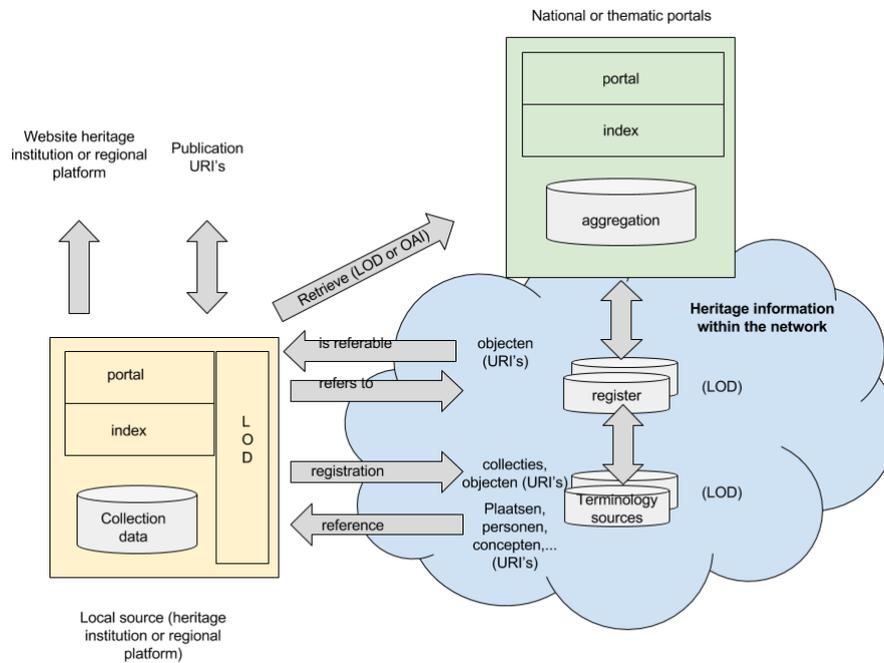


Fig. 2: distributed network of heritage information

The register of references to local collections and their items forms the backbone of the national infrastructure through which heritage information is linked to the network. The register does not contain the source data itself, but links to the locally available heritage information. This is a significant change from the current aggregation efforts, in which all information is copied and converted. Thus, when a local source publishes information as linked data in the future, this information can be accessed directly through the register.

Heritage institutions no longer provide and transfer their data, they simply make them accessible. They manage (meta)data in order to address their target group, to fulfil legal obligations, to carry out scientific work, for administration purposes, or for several of the aforementioned. These institutions generally use international data standards that are also included in the knowledge databases of DEN¹⁰. It is of key importance that heritage institutions continue to publish their data and make them available in accordance with international (and sometimes field specific) standards that fit the source in question.

The domain nodes play a central role for institutions that wish to have their collections listed in the national register. They are contact points depending on the type of collections: archives, publications, museum heritage, audio/visual and scientific heritage.

¹⁰ See: <http://www.den.nl/standaarden>

3. Next steps

With regard to the project “Usability”, the NDE considers that one of its main tasks is to further standardise metadata (models), to promote the use of linked data, to set up a national register and to promote the use of (inter)national terminology sources at local level. By encouraging co-operation within the network (among others with suppliers), the NDE intends to speed up, stimulate and facilitate these developments, both at the level of NDE nodes and at the level of local heritage institutions and providers of heritage software.

The technical framework requirements (how to make the technical infrastructure more sustainable), organisational aspects (how responsibilities are distributed at local, regional and national level) and financial implications (how do we bear the costs for creating a register and terminology sources and how do we share the potential added value) are further worked out in the project Usability of Digital Heritage.

The National Strategy aims at setting up a national infrastructure. One major reason for creating an interdisciplinary semantic infrastructure is the changing manner in which the large public uses digital heritage. For the first time in history, the public and heritage institutions meet on common ground: the digital world. In this world, the public does not care about the boundaries between sectors, institutions and collections, but expects cohesion between the many sources that can be found here. By bundling these sources and connecting them on the level of content, it becomes easier for the public to find this cohesion. In addition, heritage information becomes as accessible as possible for search engines such as Google.

The project Digital Collection, when finished, receives new impetus. Indeed, as of this year, the Digital Collection undergoes a technical overhaul based on DERA and the present position paper.

Another follow-up step concerns the suppliers of heritage software. The recently created node for suppliers has been asked by the NDE to adapt its software in order to make it as easy as possible for individual heritage institutions to publish their data sets in line with the present vision, thereby contributing to the development of the distributed network of heritage information.

3.1 Impact for heritage institutions

The goal of the NDE is to facilitate a progressive transition in which heritage institutions are supported and encouraged as much as possible to become a part of the growing network of distributed heritage information. At the national level, the NDE will invest in measures aimed at helping heritage institutions to publish their datasets as linked data, for example by setting up a register and national terminology sources for information such as “who, what, where and when” and a knowledge graph for the integration of heritage information. Many training opportunities support collaborators who wish to familiarise with new technologies and developments.

Annex 1 Overview of networks

Institutions that manage collections and end users must currently find their way in a scattered and complex ecosystem of aggregation. Due to the manifold of initiatives and the lack of an umbrella administration, an intricate landscape of sector-based, thematic, regional, national and international aggregators has developed, all of which handle their own instruments and communication channels. Collection managers are approached by various aggregators and sometimes supply to a range of different aggregators at the same time. The varying quality of the aggregated metadata and the manner of aggregation (lack of rich information) limit the possibility to create appealing views/presentations based on search functions for the aggregations.

At international level

Europeana gives access to tens of millions of heritage items. In recent years, the focus of the project was on quality. As of 2016, the so-called 'content re-use framework' is the core issue. In other words: heritage items are more or less prominently shown depending on the degree of "openness" of the metadata records that refer to it. For example, photographs available under a Creative Commons license are more often shown than archive items that can only be accessed in a password-protected environment.

Europeana aims at supporting a number of aggregators intended to develop into so-called 'expert hubs'. Examples thereof are Europeana Fashion, European Film Gateway, MICHAEL, TEL, CARARE, eFashion, IALHI and EUscreen. These hubs become responsible for standardisation and the exchange of knowledge within their respective networks. Europeana considers the national aggregation platforms as 'service hubs'. In mid-2018, a new aggregation infrastructure will be introduced to simplify the Europeana infrastructure.

New harvesting methods (based on crawling) and the linking of thesauri are also experimented with within Europeana. The partners of the work group Digital Collection take part in the Aggregator Forum on behalf of NDE and they closely follow the developments within Europeana.

However, the Europeana technology is not expected to be used at national level with regard to infrastructure, as the respective aims and operational contexts are too divergent.

National - central

Project Digital Collection (2011 - 2015)

The project Digital Collection was started as a national aggregator and completed in 2015. The aim was to establish a representative Dutch participation in Europeana. The Digital Collection comprises a national dataset with links to over 12 million heritage items. The infrastructure has a number of different components such as the servers of the Netherlands Institute for Sound and Vision, a harvester/validator, application management, technical support and various tools and services, for example data checker, data push, link checker and thumbnail service.

In the course of the project collaboration took place with nodes and initiatives such as Edurep, CLARIN, INNLE, Erfgoed & Locatie, various provincial portals, and third parties including DOKLAB and If Then Is Now. A publication service was set up that provides a range of functions: transformation of

data(models), data enriching with tools from Erfgoed & Locatie, an API access and a LOD demonstrator. The Digital Collection project was started as an aggregator without public interface.

NDE sequel – Digital Collection (2016 -)

The domain nodes are represented in the new working group Digital Collection. Within the framework of the National Strategy and its project Usability of Digital Heritage, the technical infrastructure and aggregation landscape is modernised based on the key points exposed in the present position paper.

National level – aggregation per field

Royal Library

The royal Library manages the National Library Catalogue (NBC+) in which all physical and digital library collections in the Netherlands are listed. Third parties can even develop individual interfaces based on the catalogue. For the Royal Library, the NBC will become their main platform providing access and visibility to metadata in the coming years. In addition to the NBC, the Royal Library manages Delpher, in which all content and full text sources from the heritage field are gathered (books, newspapers and periodicals) and made available for every person interested in history, and more specifically for scientists, journalists, students and genealogists. In 2015, an inventorying campaign took place in which advice was formulated for the best manners of making 515 library collections of 112 heritage institutions available and visible.

National Archives/APEF

The National Archives is one of the partners of Archives Portal Europe (APE). This international portal, created by a European consortium funded by the EU, provides access to information about the archival holdings conserved in some 30 European countries, plus information about archive institutions across the continent. The infrastructure of APE is used to provide access to the collections of the National Archives of the Netherlands. APE serves as archives aggregator for Europeana and can convert information about digital representations of archive items to the records system of the Europeana Data Model (EDM). The Archives Portal Europe Foundation (APEF) is responsible for managing the APE. The activities of the foundation are funded by the participating countries through a fee. Content providers associated with the APE do not pay a fee, but are given support by APEF if required.

Cultural Heritage Agency of the Netherlands (RCE)

The Cultural Heritage Agency of the Netherlands manages different portals, the main of which are the Digital Museum Collection of the Netherlands (DiMCoN) and the RCE image database for immovable heritage, archaeology and landscapes, and various knowledge portals. The RCE considers DiMCoN/Digital Collection as its main appraisal and presentation utility for all digital immovable and movable heritage collections in the Netherlands. The RCE develops this portal for professionals such as policy makers, curators and conservators, researchers and students but also for the large public. DiMCoN is currently being further developed, semantically expanded and fed with immovable heritage collections.

Netherlands Institute for Sound and Vision (Beeld en Geluid)

EUscreen is a network of archives, standardising bodies, research institutions and user groups providing multi-language and multi-cultural access to television heritage. EUscreen develops long-term solutions for legal issues and addresses various user needs, both in terms of service and content. EUscreen develops various scenarios for specific target groups active in fields such as: research, education, leisure and creative re-use. In addition, the content is complemented by context information.

Beeld en Geluid will use EUScreen infrastructure in order to integrate Dutch audio/visual collections. The aim is to make audio/visual heritage from all across Europe available through Europeana, for research purposes, for the creative industry and for home users. Funding is provided in the form of projects (ICTPSP regulation); the aggregator is included in the Europeana DSI (Digital Service Infrastructure, CEF regulation). EUScreen is a foundation funded by its members.

Royal Netherlands Academy of Arts and Sciences (KNAW)

The science sector has no national aggregator. The DANS aggregates information about scientific research through NARCIS.nl. In CLARIAH, Dutch human sciences institutions are gathered in order to make (heritage) data accessible in a centralised manner.

Regional level

In all provinces in the Netherlands, portals have been developed in the past ten years, often financed by the provinces themselves and managed by heritage institutions or regional history centres. Most portals will undergo modernisation, have already been renovated or are currently being renewed. Some of these portals are in a critical phase of their life cycle. Indeed, technological advances pave the way for new developments and innovative manners of connecting and presenting cultural heritage. A basic distinction can be made between portals geared towards making aggregated collections accessible and portals focussed on stories and activities, mostly making use of local collections and addressing communities and the large public. The possibilities to actively participate in collective projects or developments are limited by the lack of capacity and means of these provincial portals.

Thematic level

Thematic aggregation is basically centred on a particular topic and a form of organisational collaboration around that topic. For example, the War Sources Network aims at creating better digital access to the Dutch collections about the Second World War. The network wants to improve the usability and availability of the collection data of some 400 institutions conserving collections about the Second World War. One key aspect of this ambition is the development of a number of basic digital services that should enable the institution to better serve the public.

Brabant Heritage works together with heritage institutions from the Brabant area around the subject 'At home in Brabant', using the aggregation infrastructure Brabant Cloud among others. Brabant Heritage supports the professional collection management in small(er) heritage entities such as local history associations and initiatives, by providing, among others, training and risk analyses via the digital platform 'At home in Brabant'.

Musea with fashion collections have found together in the collaboration association 'Modemuze', in which Flemish fashion collections have now also been included. In Modemuze, the expert knowledge about fashion is among others used to feed the Art and Architecture Thesaurus.

One example of a knowledge centre of a new kind is the programme 'Digital Church Collection' of the St. Catherine's Convent Museum created in close collaboration with partners such as Stichting Kerkelijke Afgevaardigden, Landelijk Contact van Museumconsulenten, Cultural Heritage Agency of the Netherlands and many churches and dioceses throughout the country. St. Catherine's Convent has created a digital management environment based on the database and analogue church inventories of the (dissolved) Stichting Kerkelijk Kunstbezit in Nederland, in which church ownership holders

can register and manage their religious artworks. St. Catherine Convent helps churches (especially volunteers and parish counsels) through its digital management environment and makes use of its own knowledge in art history in the appraisal of the religious art collections of the Netherlands.