

Linked Open Data: taking a step forward

Researchers want to ensure that others can use their research data – now and in the future. Linked Open Data offers a solution.

Linked Open Data (LOD) is a way to connect data sources and publish them as one dataset online. The connections are provided by Application Programming Interfaces (APIs). Thanks to Uniform Resource Identifiers (URIs) new data sources can also connect to the existing datasets. Through LOD, data are thus always linked to other relevant data sources.

Using the same standards

International challenges, embodied in programs such as the EU's H2020 program, increasingly underline the need for research communities to work together. This also implies an increase in the number of data sources and data formats. Obviously, the internet has become the standard platform on which all parties exchange knowledge and expertise, also in the form of LOD. In order to be able to apply LOD in practice, it is crucial that those involved use the same standards. The international web standards body, the World Wide Web Consortium (W3C), leads the way in the development of these standards. Connected information sources in Linked Open Data format are available.

Reusability tools

Meanwhile web APIs are increasingly being used as 'data-reuse tools' (over 10k). Mendeley, Figshare and other platforms provide useful APIs for researchers. As an example, everyone can use the Mendeley API to retrieve publication data and combine them with data from Figshare.

A world of APIs

Want to know which APIs exist? ProgrammableWeb offers a list of web-based APIs. It's good to realise that most APIs have been designed to answer questions like "Which publications have 'data' in their title?" The LOD format enables a different type of API, allowing the retrieval of raw data. For example "Give me all the data on X where X is the URI of a resource". Some sources offer a mix of different types of APIs. Which API is most appropriate depends on the use case. However, using a web API is almost always better than retrieving a dataset file.

International public data

Together we must answer the question which standard(s) we need to create an API. Then we can go and warm data providers to the idea of connecting their data with other datasets. International collaboration is important. Global sharing and linking of public data should be a priority for all of us, with the aim of enabling everyone to do research using the same contextualized data.

christophe.gueret@dans.knaw.nl
linkeddata.org
programmableweb.com
dev.mendeley.com

KNOWeSCAPE: visualizing to know

Beautiful visualizations, often based on Big Data, can be found everywhere: in research, in science communication, museums and art galleries. To be able to make them is not longer the privilege of a small group of experts. However, if it comes to libraries and archives visualizing their collections and the artefacts in them, is far from being standard.

Visualizations can have many uses. For digital humanities researchers they are a useful method in their quest for the past of our culture. For the libraries and archives providing

means for this quest, visualizations can be used to highlight special treasures in a collection, to support navigation through a collection and to monitor its use.

A European network of collaboration – KNOWeSCAPE – has been set up to foster the development and implementation of so-called knowledge maps for archives and libraries. The network has initiated all of the examples presented below. KNOWeSCAPE is a COST Action devoted to the analysis of the dynamics of knowledge spaces and map making for those spaces. It is

unique in establishing a dialogue between data scientists, digital humanities scholars and information scientists and professionals.

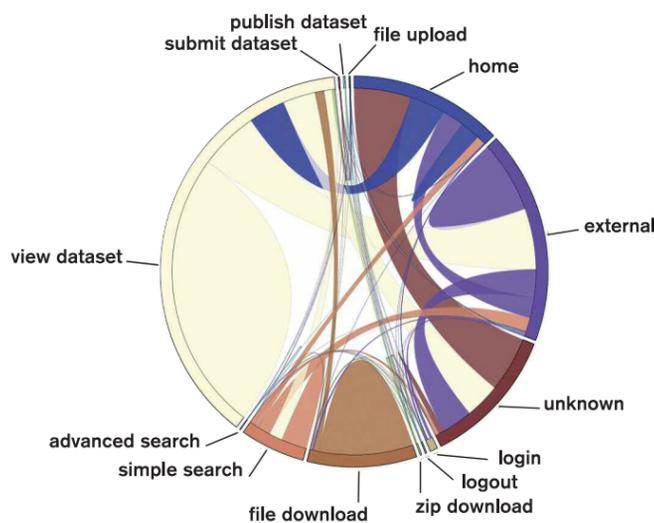
This page shows examples for three use cases of knowledge maps: **research, navigation and curation.**

The examples are taken from collections and services of DANS. Various visual elements are used: timelines, chord diagrams and complex network visualizations.

andrea.scharnhorst@dans.knaw.nl
www.knowescape.org

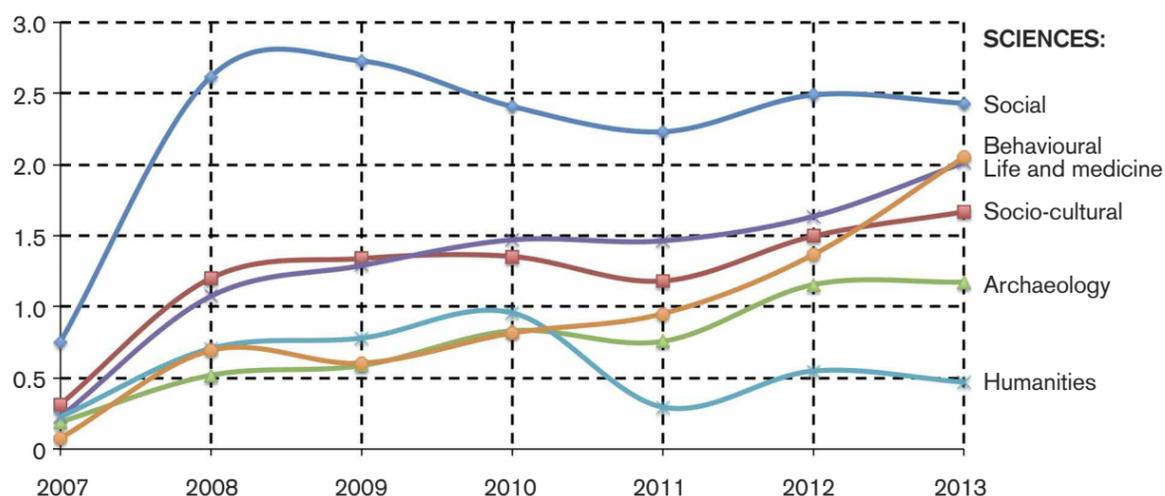
Curation

Each archive wishes to expand its collection for future use. At the same time, it also has to prove its present usefulness. DANS holds collections of digital research data in EASY (easy.dans.knaw.nl), and information about the Dutch research landscape in NARCIS (narcis.nl). Visualizations help to manage the services. Creating them is also a great opportunity to check the cleanness of metadata.



This chord diagram shows main paths of EASY users. Unsurprisingly, one sees a stream of transitions from external pages to views of datasets. What is remarkable is also that the advanced search option is hardly ever used. This kind of history of information seeking can be used to get insight into user behaviour as well as the user interface.

Ratio between # of downloads and # of available datasets



These timelines show that there is no direct correlation between demand and supply if it comes to archival information. Year after year more datasets have been deposited in EASY. They are tagged for different audiences, so we can arrange them into specific data set collections. In sheer numbers, archaeology (EDNA, www.edna.nl) has the biggest collection. But if we divide the yearly download of datasets by the num-

ber of available datasets, we see that smaller collections - tagged as social sciences or behavioral sciences - are in greater demand. Collection building should be based not only on current demand from scientific communities, but also on the experience and imagination of the archivists who anticipate future needs.