Dataset Register

Network Digital Heritage

nationaal archief

Ed de Heer – 6 september 2023 – Sound and Vision Hilversum

Dataset Register

For all heritage datasets!

The dataset register provides insight into the availability of datasets in the heritage field and thus encourages the use of these datasets.

For heritage institutions with datasets: →
add a dataset description

For users of heritage data: →
search **1877** dataset descriptions

Are you active with datasets at a heritage institution?
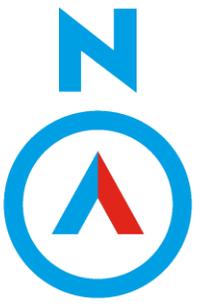
Are you looking for heritage datasets?

Do you develop software for heritage institutions?

Dataset Register

The purpose of the Dataset Register is to gain insight into heritage datasets.

Heritage institutions are encouraged to provide datasets from their system, describe these datasets and publish them online and to submit the URLs of dataset descriptions to the Dataset Register.

The Dataset Register retrieves the dataset descriptions, creating an overall picture of what is available.
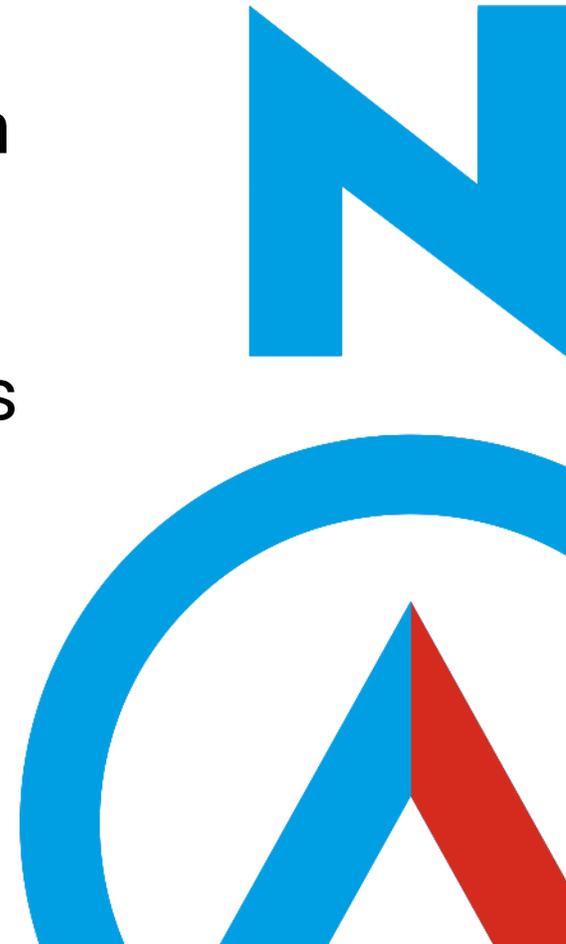
Datasets and descriptions

Dataset

A dataset is a collection of data. Data about heritage objects, such as a catalog, a set of museum objects or a collection of archives.
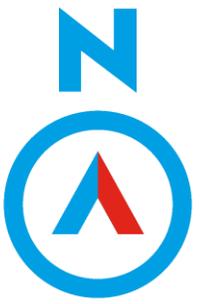
Dataset description

A dataset itself also needs to be provided with metadata. The dataset description is therefore data about the dataset. A rich dataset description in a standard way helps to make the dataset discoverable.

Datasets and descriptions

- The Dataset Register makes **no** datasets available. Data sets (in the form of data dump or APIs) are stored at the source (= the heritage institution) or on an open data platform that the institution uses.

- The Dataset Register makes **no** dataset descriptions available. Dataset descriptions are available at the source.

- You only provide a URL to the Dataset Register
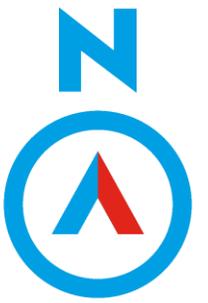
Description Requirements

Mandatory information

A dataset description contains information that describes the dataset, such as an identifier (URI), a name, a license, and a publisher.

Additional information

information can be given about the creation date, publication date, version, contact information, coverage in terms of place/area and time/period, keywords and genre.

Distribution

A data dump that can be downloaded in some format or an API that is queryable, for example a SPARQL endpoint. The information about distributions consists of a URL, a format and type, and optionally a license, description, language, publication date, modification date and file size.
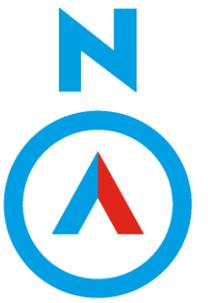
Validation

Enter a URL of a page containing a schema.org/Dataset or schema.org/DataCatalog (inline JSON-LD or direct RDF) to access it via the [Datasetregister API](#) . It is then checked that the found dataset description meets the [dataset requirements](#). The validation is performed on the basis of a SHACL file. If the dataset found on the URL is not satisfactory, the result of the SHACL validation is displayed.

Also

- The Datasetregister crawls along every URL and refreshes the information on a daily basis

- To search in the Datasetregister a dataset description use UI or Sparql.

- By analysing the Datasets we can build a Knowledge Graph on heritage data for better use.

- The Datasetregister can help software (Google) to find collections.

- Management for the Network by the National Archives

More information

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

https://netwerk-digitaal-erfgoed.github.io/requirements-datasets/

https://datasetregister.netwerkdigitaalerfgoed.nl/?lang=en

Questions?