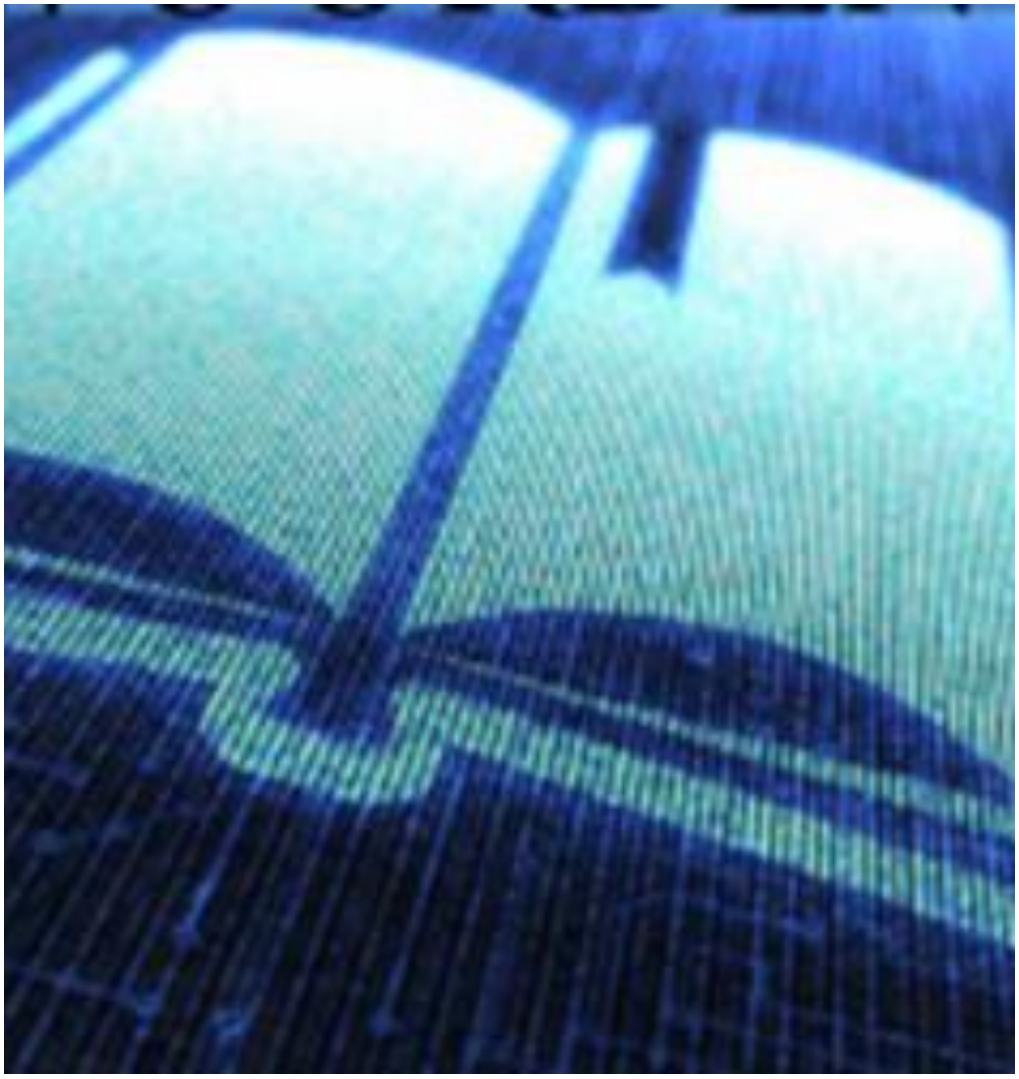


Content strategy

What are our plans?

2019-2022

KB national library of the netherlands

CONTENT STRATEGY

What are our plans?

1 Preserving the written word

The KB collects and stores Dutch books, newspapers, magazines, scientific publications, blogs and websites. In other words: we preserve the written word, whether it is on paper or digital, so that a version remains available forever; one which we want to make both accessible and easy to find one's way around for everyone.

We would, of course, like our archives to be as complete as possible, but we have to make choices. We therefore need to be clear about which publications we will and will not keep in the coming years.

2 What do we keep?

In the past, the written word was always on in a physical form, primarily paper: e.g. in a newspaper, in a book or on a pamphlet. A publication was a tangible object, something you could actually handle. That changed with the advent of computers and the internet. There is no physical version at all of some publications. For example, articles appear on the websites of newspapers that are never published in the papers themselves. Or take De Correspondent: this is a journalistic initiative that only appears online.

So what are we going to keep and what are we not going to keep? The definition of "a publication" we use is deliberately broad: 'a coherent text unit made public by a person or organisation'. Additional conditions are therefore required. With the advent of the Internet the number of publications has become infinite, but our budget is not.

3 Core profile

In the coming years our focus will be on the NETHERLANDS COLLECTION: publications from and/or about the Netherlands. We want to collect and store these as completely as possible, within our financial, legal and practical constraints. We will focus on digital copies and will only include physical copies in our collection if there is no digital alternative available. In practice, this means, for example, that where possible only the ebook of a new Dutch novel will be included.

International publications and special collections, for example medieval manuscripts, are also included, provided that they contain written words and that they are about the Netherlands or come from the Netherlands.

This focus does have particular consequences for our international e-repository for scientific articles. As part of the SAFE PLACES NETWORK, we are going to focus on sustainable access to home-grown publications: the researcher, publisher or university must come from the Netherlands.

4 The written word - Putting it to work

Not only do we want to preserve the written word, we also want to make it easily accessible to everyone. We have several ways of doing this.

Through our DELPHER and DBNL services, we make digitised heritage available to anyone interested in the language and culture of the Netherlands. These websites are freely accessible, no membership is required.

For ebooks that are still available for purchase, we enter into agreements with publishers. We then try to make these works available to all members of public libraries, for example via the ONLINE LIBRARY app.

Finally, digital scientific sources (ebooks and e-journals) are accessible to members of the KB via the KB website.

5 A larger and more diverse facility for readers

Together with public libraries, we ensure that the range of online libraries on offer continues to grow. In addition to ebooks, we offer an increasing number of audio books. In the coming years we want to expand our range further to include e-time-books and e-newspapers.

6 Supporting scientific research

The KB wants all Dutch people to have easy access to digital scientific information. Our members therefore have access to a selection of licensed scientific sources and publications. We also guide them to scientific information that is available elsewhere.

We are helping researchers by making a large number of texts accessible within the framework of TEXT & DATA MINING. We publish our METADATA as LINKED OPEN DATA. This means that machines can read our data and automatically link to it. This improves the online accessibility of our heritage.

Glossary

DELPHER

Delpher is a website of digitised historical Dutch newspapers, books and magazines. The name Delpher echoes the Dutch verb ‘delven’: evoking mining and digging up precious resources. It is also based on the famous Greek oracle at Delphi. On Delpher you can now view 900,000 books, 1,850,000 newspapers and 156,712 magazines. And more are added every year.

DBNL

The Digital Library for Dutch Literature (DBNL) is a digital collection of texts that belong to the national literature, linguistics and cultural history of the Netherlands from the earliest times to the present. The collection represents all Dutch language areas; not only the Netherlands, but also Belgium, Suriname and the Caribbean. It includes texts from the Low Countries or variants of these, such as regional languages and slide lectures. On 31 December 2019 there was a total of about 17,000 titles available in the DBNL, with an aggregate volume of over 4.8 million pages. The collection includes both books and magazines.

LINKED OPEN DATA (LOD)

Linked Open Data is a term for data (texts, for example) that are open and linked, structured and published according to the principles of Linked Data. The data is interconnected, freely accessible and divisible.

METADATA

Metadata is actually data about data. Metadata for a publication consist, for example, of the name of the author, the date of composition, the publisher, the number of pages, the language in which the publication was written, and so on. We call the recording of metadata ‘metadataing’. We do this in order to make the works easier to find, for example via online search engines. Part of the metadata is created manually, part of it with the aid of technology. Would you like to know more about this enrichment process? See below for more information.

NETHERLANDS COLLECTION

The Netherlands Collection comprises all publications from

and/or about the Netherlands: mainly works published by a publisher in the Netherlands, but also works published abroad and about the Netherlands.

ONLINE LIBRARY

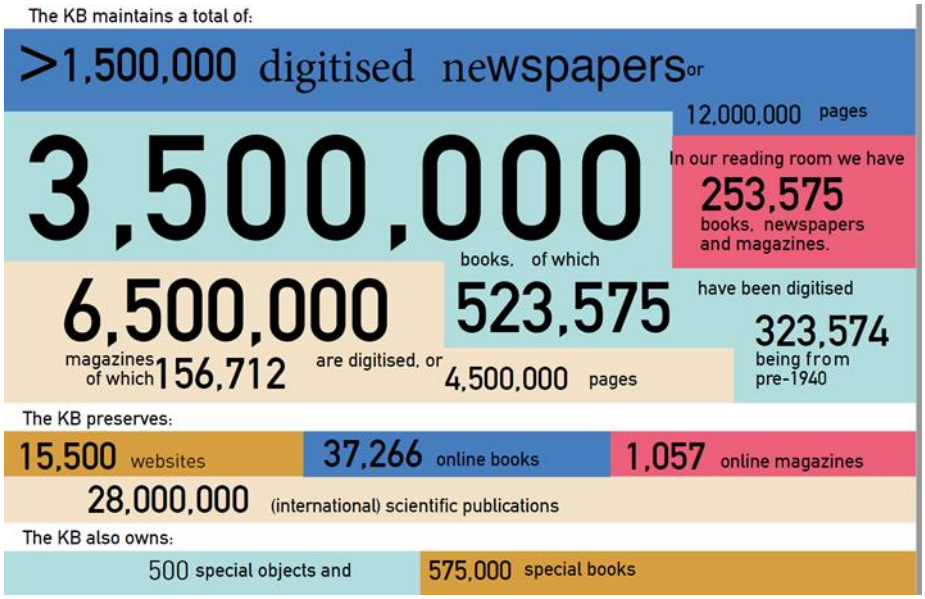
As a library member you also have access to ebooks and audiobooks via the Online Library. The Online Library has more than 29,000 ebooks and over 4,200 audiobooks in its collection. The collection changes regularly, so there are always new and interesting titles to be found. We enter into agreements with publishers to make online lending possible.

SAFE PLACES NETWORK

The safe places network is an international body of organisations that ensures the sustainable preservation of texts. The organisations involved are, for example, national libraries or archive institutions that preserve the material of their country.

TEXT & DATA MINING (TDM)

TDM is used to search for links or patterns by mining into a large mountain of data or texts.



MORE INFORMATION ON METADATA

Enrichment of digital content

Research into improving the searchability and use of our digital content is one of the core activities of the KB Research department. One of the ways we try to achieve this is by enriching our collections with extracted or related information, from both internal and external sources. These enrichments can be of many different sorts: from a derived genre or sentiment, to geographical coordinates or a related video on the web. Our current focus is on enriching our historical newspaper collection with linked named entities, i.e. names of persons, locations and organisations in the newspaper articles linked to descriptions in international knowledge bases such as DBpedia, Wikidata and VIAF.

Enrichment infrastructure

We have set up a generic enrichment infrastructure, consisting of an enrichment database and a number of services, in which any type of enrichment of any object in the KB collections can be stored, without changing the original data. Generally speaking, the enrichment database contains links between identifiers of objects from our collections and related identifiers. In the case of linked named

entities, there are links from identifiers of newspaper articles to records that represent an entity, combining the links known to us to descriptions of that entity in a thesauri or knowledge base. If available, we add metadata about, for example, the origin and reliability of the links.

The entity linking process

Named entities are automatically recognised in articles using special software. By searching these names in an index based on DBpedia dumps we generate a collection of potential links. For each ‘candidate link’, a number of characteristics are determined on the basis of properties of the name itself and from context information, such as date of birth and occupation. A machine learning model trained on a collection of manually annotated articles then selects the best link. Although our software now has an accuracy of 85%, we invite users to correct remaining errors and add missing links. This user feedback also serves as additional training data for the entity linking software.

Semantic search

If an article is requested for indexing or presentation purposes, the corresponding enrichments can be retrieved from the enrichment database. Presentation software can provide links to the descriptions or show relevant information from those descriptions, such as a summary or an image. Indexing the identifiers of recognised names together with the newspaper articles offers new (semantic) search possibilities. For example, users can search for articles containing entities that possess certain (combinations of) properties, such as articles about Roman emperors. Our software then extracts the identifiers of entities with the property of having been a Roman emperor from Wikidata in the background and uses these in a search of the enriched newspaper index.

Demonstration

To demonstrate this functionality, we offer an online research portal (external link), in which users can view the available enrichments and experiment with semantic searches in the newspaper index. This portal supports full SPARQL searches in Wikidata, but also offers a number of more user-friendly options for semantic search, for example by generating a "best guess" SPARQL query based on a conventional query. Links to additional services are also provided, including a page to remove and add enrichment to each article.

KB national library of the netherlands