

DOI: <http://doi.org/10.18352/ts.3447> TS •> #38, December 2015, p. 91-94.

Content is licensed under a Creative Commons Attribution 3.0 License. - © Maaïke Napolitano

Publisher: [www.uopenjournals.org](http://www.uopenjournals.org). Website: [www.tijdschriftstudies.nl](http://www.tijdschriftstudies.nl)

## *Ten Questions: Maaïke Napolitano, Delpher product manager*

### FOR WHOM IS DELPHER INTENDED?

Delpher is the full-text search engine of the (Dutch) National Digital Library. Delpher currently sees an average of 130,000 uses per month, by about 50,000 people. The largest group of users consists of amateur historians, journalists, and genealogists. There is also a large group of people who use Delpher incidentally, for instance to explore their own families or histories, or to find interesting articles to go with a particular publication. A smaller but not unimportant group is formed by the scholars, who, unlike the other user groups, use Delpher very intensively. The Delpher team is constantly mapping out the needs of the various target groups by means of regular user surveys, in order to be able to fulfil our users' needs and wishes as effectively as possible.

### WHO ARE THE PEOPLE BEHIND DELPHER?

In 2013, a collaborative effort between the university libraries of Amsterdam, Groningen, Leiden, and Utrecht, the Meertens Institute, and the National Library of the Netherlands (KB) gave rise to Delpher. The service is now managed and maintained by the KB, while the other partners are represented in the Advisory Board. A designated team is charged with the daily maintenance and improvement of Delpher, using the principles of "scrum", a development method leading to a new piece of software every three weeks. Every morning the core team meets for the so-called stand-up to discuss the current state of development. This team consists of back-end developers, front-end developers, UX designers, testers, someone known as the "scrum master", and myself as product owner. In addition, many other people are involved, such as those who make material available, collection specialists who deal with the content, marketers, a twitter editor, people carrying out the user surveys, and those who address questions of users. As product manager, I bring all these activities together. Based on all the input I receive, I chart the course we want to follow with Delpher.

### WHERE DOES THE MATERIAL SHOWN ON DELPHER COME FROM?

The collections shown on Delpher come from various heritage institutions, such as libraries, archives, and museums. The digitisation of these collections is the result of

various projects, including the historic newspaper project of the KB which digitised 8 million newspaper pages, the national ‘Erfgoed van de Oorlog’ (War Heritage) project, Early Dutch Books Online (EDBO), digitisation efforts by Google, and the Metamorfoze programme. The latter is currently responsible for the greatest influx of new material onto Delpher. Metamorfoze is the Dutch national programme for the preservation of paper heritage, aiming to ensure the digitisation of vulnerable Dutch materials. All books, newspapers, and magazines digitised via Metamorfoze will eventually be made accessible via Delpher.

WHICH TITLES WILL BE ADDED TO DELPHER IN THE NEAR FUTURE?

Strange as it might sound, we cannot predict with certainty which titles that will be. In the digitisation process, books can “drop out” for various reasons. Torn-out or mouldy pages may for instance prevent a book from being digitised in the first place. Complications may also arise when books are being processed, such as errors in file linking, or irregularities in metadata, which can cause the software to fail to process part or all of a batch of texts. Because Delpher operates on a bulk processing basis, the dropping out of individual books is not addressed right away; rather, they are included in a bulk batch again at a later processing date. This approach allows us to ensure that the digitisation proceeds rapidly.

In 2015, we added all books which were digitised and processed by the KB to Delpher. We are currently working on making more newspapers available, which will start their online life somewhere in the first six months of 2016. For the second half of 2016, we have scheduled the addition of more magazines to Delpher.

THE FULL TEXTS OF ARTICLES VARY IN QUALITY. ARE YOU WORKING ON IMPROVING THOSE WHICH ARE LESS THAN OPTIMAL?

Full texts are created using OCR software. OCR is good at dealing with printed texts but can have difficulty with special fonts or diagonal texts, such as is often used in advertisements. Research into the improvement of such software is constantly ongoing, nationally and internationally, and we keep abreast of the latest technology as much as we can. In cases where special technology is required which is not yet available, we are currently experimenting with manually processed texts. To this end, a crowd sourcing project has been set up – a collaboration with the Meertens Institute – which allows volunteers to manually enter seventeenth-century texts printed in a gothic font in a database. The first 100 newspapers processed in this manner have recently been added to Delpher.

WHAT ABOUT COPYRIGHT AND DELPHER?

The materials on Delpher are shown on the basis of licences and agreements. Copyright thus does not lie with the KB, but with the publisher or creator (i.e. the writer or photographer). Exactly in whom a copyright is vested depends on whether creators are identified or not, which in turn affects the term of validity. For example, if a newspaper

article does not indicate the name of its author, copyright belongs to the publisher and is effective for 70 years after publication. If the author's name is indicated, however, the copyright is his or hers and remains in force until 70 years after his or her death. As long as an article or photograph is copyrighted, one essentially needs permission from the copyright owner to use it. An exception is the right to quote, where the context in which a quote is included has a real and functional relation to the quoted work, for instance when a newspaper article is used to substantiate one's story. Still, in that case the source (and, if known, the author) of the quote must be clearly referred to.

#### CAN RESEARCHERS ALSO ACCESS THE DATA DIRECTLY?

Yes. Scholars and researchers are allowed access to all texts, images, and metadata on Delpher, for various purposes such as Text and Data Mining (TDM). Some data sets are freely available, such as the texts of the ANP (the Dutch news agency), and books published before 1800. For datasets which also contain copyright-protected materials, such as newspapers and magazines, advance permission is required. More information about gaining access to the data sets can be found on <https://www.kb.nl/dataservices>; one is also welcome to email [dataservices@kb.nl](mailto:dataservices@kb.nl).

#### WHY HAVE MAGAZINES NOT BEEN MADE AVAILABLE ON A PER-ARTICLE BASIS?

During digitisation, material is semi-automatically subdivided in various ways. When it comes to newspapers, for instance, a distinction has been made between advertisements, articles, family notices, and captioned illustrations. Because of budgetary restrictions at the time the decision had to be made, this has not been done for magazines. The software therefore treats a magazine on Delpher as one entity. That's a pity, of course, because it makes it impossible to search for separate articles or to include a link to a specific article when citing your sources. The KB is currently investigating the costs of retro-application of article segmentation to the magazine set. The results of that investigation will determine whether or not such segmentation will be implemented.

#### WHAT KIND OF NEW FUNCTIONALITY ON DELPHER CAN WE LOOK FORWARD TO?

Top priority currently lies with improving the search options. With every addition of new material, it becomes more and more important that the search function operates optimally, enabling users to rapidly find what they are looking for. For instance, we are working on optimising relevance scores, improving metadata, renewing indexes, and working with automatically generated selection lists, which show only those filters which actually apply to a particular search result. In 2016, we will also be working on a Delpher design for small mobile devices, such as smartphones, and we will be adding geolocation metadata to our newspaper set, making it possible to search by place name or the GPS of a phone.

FINALLY, WHAT IS YOUR PERSONAL FAVOURITE SOURCE ON DELPHER?

The first time I tested Delpher, I had to come up with a search word on the spot, and I chose my own maiden name. One of the first hits I found was an advert from the 1960s, a “help wanted-ad” from my grandmother, who needed a shop assistant for her *chocolaterie*. Because my grandmother’s shop closed before I was born, it was very special for me to find something so tangible from the past. As I continued my search, I found various adverts and articles about the *chocolaterie*, which I collected and shared with my family. They were amazed at the possibility of finding all that on Delpher.

•> MAAIKE NAPOLITANO *works at the National Library of the Netherlands as Delpher’s product manager.*