From punch cards to online access to thousands of datasets

Fifty years of Dutch data archiving

Fifty years ago the first data archive in the Netherlands within the humanities and social sciences was established. Since then, a lot has changed, but the memories of the predecessors of DANS are still there.

1964 - Steinmetz Foundation

How to prevent valuable research data from being lost? Fifty years ago, the Steinmetz Foundation was set up to solve this issue. A few years later the Foundation was taken up by the Royal Dutch Academy of Sciences (KNAW). Marion Wittenberg, who works at the Steinmetz, remembers a story of a former colleague: "Punch files were taken to the University of Amsterdam's computer centre, SARA, on the back of a bike. There they were fed into a machine to be read, which sometimes went wrong. The cards would spray like a fountain from the card reader, warped or cracked, no longer reproducible. In the metadata, deviation rates were estimated. The punch cards were replaced with magnetic tapes, which were at first stored on SARA's mainframe computers, and later in the Steinmetz vault on Herengracht, with a back-up copy at SARA. In the early nineties, all data were copied to Optical Disks. The back-up CD-ROM was kept in Friesland, under the bed of one of the employees. If the bomb were to be dropped on Amsterdam, the data would not be lost. Ultimately, all of the data ended up in the EASY system at DANS."

She continues: "One of the aims of the Steinmetz Archive was that the collection could be reused. Director C.P. Middendorp was active in that respect. He replicated some 200 survey questions from 15 studies in his research on cultural changes in the Netherlands. This study from 1975 served as the basis for the biennial Cultural Changes survey by the Netherlands Institute for Social Research (SCP), which is still being held."

1989 - NHDA

In 1989, the Netherlands Historical Data Archive (NHDA) was established. Heiko Tjalsma was involved from day one: "The NHDA was founded in Leiden. In a way, we followed in the footsteps of the existing Steinmetz Archive, but for the historical sciences. From the start there were differences, however. Until 1995, when it became part of the Royal Academy, the NHDA was not an established institute but rather a project with an uncertain future. Funding would vary from year to year. As a consequence, the NHDA with its small staff engaged in any work related to 'history and computing,' even if only remotely. We had to, as we were almost entirely dependent on external project income. In particular, many digitizing projects were carried out. We did little data archiving as there was no prospect of keeping data longer than say a year. Sometime later we also played a leading role in a postgraduate programme funded by the European Social Fund set up to train graduates in historical information science. All in all the NHDA could best be described as a very creative, slightly anarchistic but always stimulating organisation in which multitasking was the normal work attitude." The NHDA was integrated in NIWI in 1997.

1994 - WSA

The Scientific Statistical Agency (WSA) was established by the Board of NWO in 1994 to improve the accessibility and availability of data files. The Agency played an intermediary role between researchers and data providers, and worked closely with other institutions mediating the unlocking of data for scientific research. Ron Dekker was head of WSA from 1997-2002. He remembers: "In 2001 WSA, NHDA and the Steinmetz Archive organized the international IASSIST data conference in Amsterdam. In one week we received more than 200 participants from 20+ countries, including (for the first time) from Eastern Europe. That week, there was a trip – yes, to the windmills!, there were workshops, plenary presentations and parallel sessions. This was in the pre-wifi era and the internet connections were literally laid through Amsterdam: from the UvA auditorium to the Doelenzaal and Singelkerk across the canal. We opened with a keynote from Paul Schnabel who quoted from the Odyssey. When Odysseus blinded the Cyclops, the latter asked who he was and Odysseus replied, "Nobody." Here we see the first instance of anonymized data because when the other Cyclops asked who had done it, the first one answered "Nobody" and Odysseus came out unscathed. Nero, no not the emperor but a Steinmetz employee, had bought the original soundtrack of the film. Later Kubrick decided to use

An Altamira Bison; this cave painting from 30,000 years ago is still accessible and readable today. Will this also be true for the digital data recorded by us? It's the core question of digital data archives.
credit http://en.wikipedia.org/wiki/Cave_painting#mediaviewer/

Computer use by historians in 1989, photo taken from the book 'Toverwoord informatie'
credit Maili Blauw

familiar music, but the soundtrack was already there – a nice bit of archiving! That collaboration of WSA, NHDA and Steinmetz anticipated the decisions made later. We were collaborating, but DANS did not come about until 2005. And the rest is history and, above all, a bright future."

2004 - EDNA

The e-depot for Dutch archaeology, EDNA, aims at the sustainable archiving and disclosing of archaeological research data and the dissemination of related knowledge. EDNA is a collaboration between DANS and the Cultural Heritage Agency of the Netherlands. Milco Wansleeben was involved in the creation of EDNA: "From the very beginning, when we had a small website with four datasets and a modified version of NIWI's Itor software, Dutch archaeologists could experience for themselves what exactly a data archive was, how it worked and what they could do with it. This visibility has contributed to the awareness that digital research will not naturally remain available, and it has ultimately been the key to the inclusion of a national deposit requirement in the quality standard for Dutch archaeology. Demonstrators and pilots did really help, and they still do. We went from 4, to 50, to 22,000 archaeological datasets in an e-depot that is now completely up and running. More and more archaeological organizations have left their initial reluctance about accessibility behind and open access seems to become the new standard for archaeologists."

2005 - DANS

In 2005, NWO and KNAW together founded Data Archiving and Networked Services (DANS). DANS took over the activities of the Steinmetz Archive, the NHDA, WSA and EDNA. Despite all the mergers, the data have been preserved over the last fifty years. Research files that started out as boxes full of punch cards in the early days of the Steinmetz Archive, can still be accessed through the EASY system today. Some of them are being downloaded regularly. Peter Doorn, director of DANS: "DANS promotes sustained access to digital research data. To this end, we encourage researchers to employ sustainable data archiving and reuse, for example through our online archiving system, EASY. And with NARCIS, for example, we provide access to thousands of scientific datasets, e-publications and other research information in the Netherlands. We never know what the future brings, but today we want to celebrate the success of 50 years of data archiving in the Netherlands."

Sharing data: good for science, good for you >>

3. Dutch historian Martijn Kleppe (EUR): "Such data sets tend to be very rich, there are so many articles you could write. Now that I've finished my thesis, my interest in other research areas is growing while my orginal data set still contains useful information. Sure, I can store it in my drawer or on my computer, but that's not going to be around five years from now. Or, alternatively, I can store it at DANS, so that people might extract relevant information."

4. Data that's worth analyzing, is worth storing. For yourself and for your successors And it's not hard, if you plan it properly, up front. The past is a great source of future knowledge. Data archives can be real treasure troves.

5. Quantitative data analyst Manfred te Grotenhuis (RU): "It feels like walking around in a candystore, just taking what could be of use to us. The access speed is a huge advantage, direct access without mediation, so you can start immediately, that's what I find enjoyable."

HISTORY