

FACTSHEET:

WEB ARCHAEOLOGY

Data is the new clay, scripts are the new spades and the World Wide Web is the youngest layer that we are going to dig up. Web archaeology is a new area in e-culture where we excavate and reconstruct relatively new (born-digital) objects, which were lost not so long ago, using new digital tools. Both the archaeological finds and the methods of unearthing and reconstructing our digital past are very recent and still in development.

In the *DIY Handbook for Web Archaeology, intended for heritage professionals*, we get started at once, bit by bit, byte by byte. We not only show how digital archaeology can be used to ‘dig up’ a digital heritage object, we also show how you can safeguard your archaeological finds and preserve them for future generations. The DIY Handbook is a step-by-step guide with practical tools for setting up and carrying out a web archaeology project. To give substance to the new discipline of web archaeology, we immediately translate our theory into practice using the case study DDS Herleeft. The DDS case study is a running theme throughout the DIY Handbook. We will answer the following questions: How do you dig up the Digital City (DDS) and transform it from a virtual Atlantis into a virtual Pompeii? How do you reconstruct, store, unlock and present your DDS material in a sustainable way for future generations?

A web archaeology project involves the stages of planning, digging, reconstructing, unlocking and knowledge sharing. See the sneak preview of highlighted examples from the DIY Handbook.

1. PLANNING

Activities: Write a plan, describe the objectives, research the sources, describe the digital object and its relevance, make an inventory of what you need, draw up a plan of action, make a budget, set up a communication platform.

HIGHLIGHT: Describe the digital object and its relevance

Provide arguments based on which a digital-heritage object should be included in the collection, such as its artistic or technological value, the uniqueness of its user experience, its provenance history, and include owner/contact details for more information. Explain your arguments: Why is it important to ‘dig up’ this object, to reconstruct it and include it in the collection? Also describe how this object relates to the mission and vision of your organisation.

HIGHLIGHT: THE DIGITAL CITY (De Digitale Stad - DDS)

Twenty-three years ago, a unique city made of computers, modems and telephone cables emerged. On 15 January 1994, De Digitale Stad (literally: The Digital City), the first virtual city in the world, opened its virtual gates in Amsterdam. But like so many other cities in world history, it collapsed. In 2001, DDS was taken offline and the Digital City disappeared like a virtual Atlantis.

DDS was the first Dutch online community. It played an important part in the Internet history of Amsterdam and the Netherlands. For the very first time, the Internet was accessible to the public – for free. The virtual city introduced many people to the possibilities of the Internet. The virtual city and its inhabitants produced objects, ideas and traditions in new digital (hybrid) formats, such as web pages, news groups, chats, audio and video. DDS is an important and unique historical source about the early years of the Internet in the Netherlands and a treasure trove of digital heritage.

The virtual city was not just a website. It was a complex, dynamic information system which over time developed several applications. DDS always followed new developments closely, which resulted in a number of unique 'cityscapes' (interfaces):

- 1994-1995: DDS 1.0 - 15 January 1994 (telnet/command line interface). The city metaphor was reflected in the organisation of the interface. It had a post office (for email), public forums to meet other visitors, a town hall and a central station (gateway to the Internet).
- 1994-1995: DDS 2.0 - 15 October 1994 (website: static HTML). First DDS website with a graphical interface, hyperlinks, text and images.
- 1995-2001: DDS 3.0 - 10 June 1995 (web system: static HTML and interactive web pages). In the multi-square interface, each square had its own theme and character, and served as a meeting place for people who were interested in a specific theme. 'Residents' (users) could build their own 'house' (web page), send and receive emails worldwide(!), participate in discussion groups, chat in cafés, walk around and help build the 'Metro'.

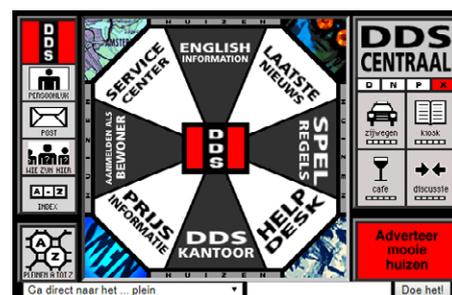
DDS 1.0 15 January 1994



DDS 2.0 1 October 1994



DDS 3.0 10 June 1995



The three 'cityscapes' of DDS

TIP: Set up a strategy for a sustainable digital collection.

The score model: <https://www.scoremodel.org/>

Catalogue of sustainability policies: <http://wiki.ncdd.nl/>

2. DIGGING

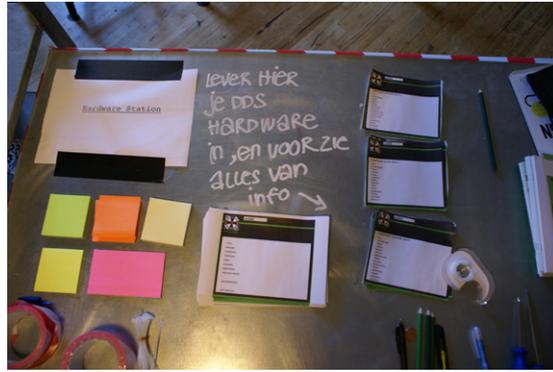
Activities: Make an inventory of the objects that need to be included, publish your plan, create info cards, bring people, stories and objects together, guarantee the data integrity of your finds, secure your finds.

HIGHLIGHT: Bring people, stories and objects together

Organise meetings to reunite people and collect lost objects and memories. Bring together original builders, creators, visitors and users of the digital object, and your team of web archaeologists. This promotes knowledge transfer and creates a better understanding of your object. Digital heritage tends to be complex and requires experience and expertise, mainly technical. What is essential is not just contemporary knowledge, but particularly knowledge of historical data, software and hardware. Together we are smarter. Moreover, you might end up finding lost artefacts. That's the power of crowdsourcing!

HIGHLIGHT: Grave Diggers Party

“Help us dig up this unique city, witness the kick-off of our web excavation work and contribute to the restoration of this historic monument! Look in your attic and search your hard drives. Bring all the servers, modem banks, VT100 terminals, freezes, hard drives, scripts, zip files, floppy discs, tapes, backups, log files, videos, photos and screenshots you can find and tell us your memories and stories!” That is how we invited former residents, former DDS employees and kindred spirits to the ‘Grave Diggers Party’.

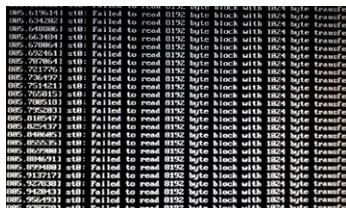


Web archaeologists and info cards at the Grave Diggers Party

Armed with digital tools, web archaeologists researched the supplied historical DDS materials to study the history and use of early computer architecture, peripherals, operating systems, programming languages, system management tools, et cetera. The tools included computers (digital excavators), storage devices (buckets), USB flash drives (find bags), UNIX commands (spades, picks, trowels) and scripts (metal detectors). The web archaeologists collected the objects, original remnants of the lost virtual city, on a hard disk, including the metadata. The origins of the retrieved artefacts were documented on info cards: who contributed the material, where did it come from, how was it originally used, what is the current state of the objects and who can you approach for more information?

HIGHLIGHT: Freeze!

After the Grave Diggers Party, we received one more unique find: the FREEZE. The Digital City was ‘frozen’ on 15 January 1996. This snapshot with “exactly what the city has to offer” was saved to disk and “deposited in archives to be studied by archaeologists in a distant future.” After a one-year search, we found the technician who was able to read the data, so they could be used as building blocks for the reconstruction.



The FREEZE tapes and the reading of the data

HIGHLIGHT: Transferring and securing data

Follow three procedures to transfer all the recovered data and secure them for future use: 1. Transfer the data from the data carrier: Create images. 2. Validate the data: Create and compare checksums of the images. 3. Create multiple backups of all data! Also, always work with copies of the data only.

TIP: Always keep the metadata (information about the data)

<http://wiki.ncdd.nl/index.php?title=Duurzaamheidsbeleid/Beleidsuitgangspunt-Metadatas>

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

3. RECONSTRUCTING

Activities: Describe the starting points, research the ethical issues, start the preparatory work, reconstruct (emulate or restore?), interpret the data.

HIGHLIGHT: Describe the starting points

Our principles:

- Restoration, conservation, preservation and presentation of digital objects are new and important museum tasks. Reconstructing and unlocking born-digital objects involves different requirements than preserving physical objects. How do you handle this difference? How do you ensure that endangered or lost digital objects remain available to museums?
- Making the digital object operational again enables the audience to walk through the Digital City once more and experience or re-experience the evolution of this unique virtual project and the early days of the World Wide Web. How do you restore the original user experience? How do you repair dynamic web objects that are interactive, networked, process-oriented and context-dependent? How do you reconstruct historical data when there is tension between 'historically true' (authentic) and 'feasible' (quick & dirty)?
- As a web archaeologist, you are responsible for the legal and ethical aspects. How do you deal with users' privacy (note that the World Wide Web is still young and that most of the people involved are still alive)?

HIGHLIGHT: Reconstruction – emulate or restore?

Which digital preservation strategies (migration, conversion, emulation or reinterpretation of the data) can or should be used? Getting an original system working again has great historical value and significance. How 'historically faithful' the restored digital object can be depends on your recovered data, hardware and software. Is emulation possible? Is restoration necessary? Is replication desirable? An emulation equals the original, bit by bit and byte by byte. A restoration makes as much use as possible of the original programs and data and adjusts the components (or data) to get the digital object working again. A replication is a completely new object with the same functionality as the original, with or without the original data.

During the reconstruction work, two variants of DDS 3.0, based on the FREEZE data, were reconstructed. The 'Replica', the museum variant, only contains the information that was accessible without logging in and does not include privacy-sensitive and non-public data, such as account details and emails. The 'Emulation', the scientific variant, does contain privacy-sensitive data and has outdated (historical) security problems. If, in the future, interested parties want to view the Emulation and the other data found, they may submit a request to the DDS Digital Heritage Foundation, which consists of the current project partners who supervise the re-use of DDS data. The Foundation assesses all applications made by researchers.



Students presenting the Replica (left) and the Emulation (right)

TIP: Search the digital haystack

Use the metaphorical dustpan and brush of the web archaeologist: 'find', 'grep', metadata dating, DeNISTing, DeNISTing-NG (Next Generation; <https://github.com/tifkap/>) or Bitcurator, a free digital forensics and digital preservation tool (<http://www.bitcurator.net/>).

4. UNLOCKING

Activities: Archive (statically or dynamically?), find a partner, describe the most favourable preservation scenarios, give access to target groups, document, document, document.

HIGHLIGHT: The three preservation scenarios of 'DDS Herleeft'

- **Bit-preserving** (no public access, read-only). Only storage of the data (bits and bytes). You save all the data that you have excavated during your project, both in original and processed condition. This is about saving the ones and zeros, without sustainability actions. This backup can be used to make a more complete reconstruction in the future.
- **Sustainable dynamic storage** (no public access, read-only, master). To preserve the extensive and diverse DDS collections with dynamic (working) data permanently, it was decided to store both reconstructions, the 'Emulation' and the 'Replication', as 'virtual machines'. These machines can mimic the original software and hardware and execute the reconstructed objects. Virtual machines are the perfect archiving format for dynamic (working) digital objects. To ensure the future functioning of data, the archived digital object must be monitored regularly, so that risks and changes that may affect the digital collection are noticed in good time and the right measures can be taken (preservation planning). Both the technology and the services for this new method of dynamic preservation are currently being developed at the Netherlands Institute for Sound and Vision.
- **Web archiving** (public access, read-only). This involves recording Web ARChive (WARC) files of the 'Replication'. WARC is a widely used archiving format for storing web crawls.

HIGHLIGHT: Use the OAIS model

There are several processes to be considered when you want to preserve digital heritage in a sustainable way and make it accessible in an e-depot. Both the provider and the recipient (e-depot) of the digital object are involved. You can use the Open Archive Information System (OAIS) model. We used this model as a guideline in our project.

- *Ingest*: Receiving the data (digital objects and metadata). How are DDS data read into the archive and what agreements can be made?
- *Archival storage*: Storing the digital objects. How are the DDS data to be stored?
- *Data management*: Managing the digital objects' metadata to enable search, retrieval and management features and performing checks. How should the DDS data be managed and which metadata does this require?
- *Administration*: Managing the e-depot, coordinating activities (services and functionalities).
- *Preservation planning*: Monitoring the archive. Planning the sustainable management of digital objects. How will the DDS data remain accessible in the future and how do we manage them? This is the key to sustainable storage and preserving accessibility for future generations of the digital object.
- *Access*: Making digital objects or information available to users. How can interested parties get to the DDS data?

TIP: Use the OAIS model

<http://journals.sagepub.com/doi/10.1177/0955749017725930>

<http://www.ncdd.nl/news/artikel-duurzaamheidsbeleid-od-magazine/>

5. KNOWLEDGE SHARING

Activities: *Let the Bytes Free! Describe and share your results and findings and give recommendations for future web archaeology research.*

HIGHLIGHT: Share your results and your findings

Find our results and more information at: <http://hart.amsterdam/freeze>

HIGHLIGHT: CALL TO ACTION!

We invite all current and future web archaeologists to get started and share their knowledge and experiences with us. Dig in and Do It Yourself! Plan, excavate, reconstruct, preserve, unlock, present and share your knowledge. What web archaeology research are you working on? Let us know and share your knowledge and experiences.



DDS on permanent display at the Amsterdam Museum

Powered by...

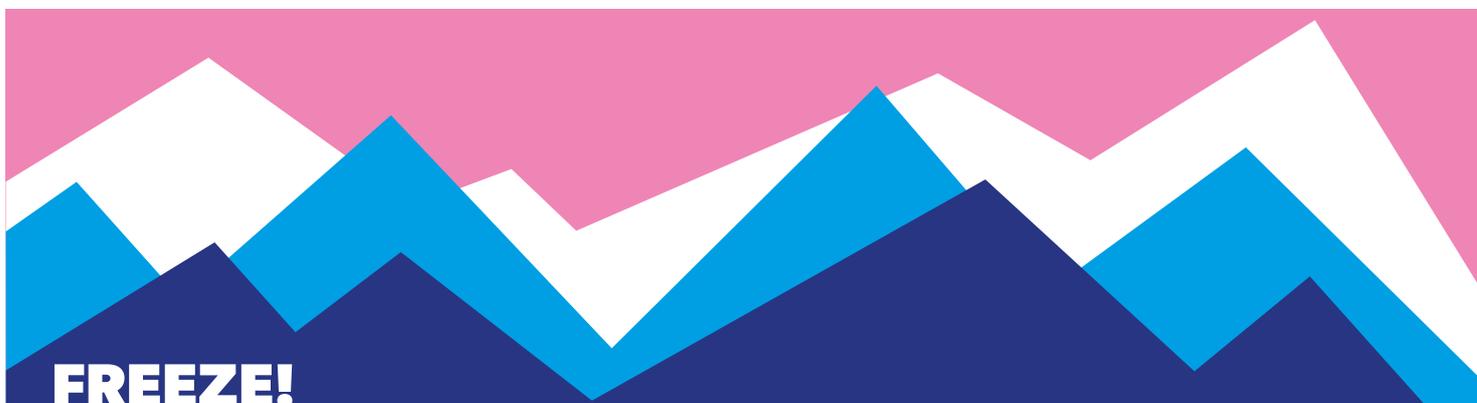


stimulerings
fonds
creatieve
industrie

The project received the international Digital Preservation Award 2016

Credits

This factsheet is based on 'DIY Handboek voor Webarcheologie. Do It Yourself: Plan, graaf, reconstrueer en ontsluit!'; written by Tjarda de Haan, Robert Jansma and Paul Vogel for the 'DDS Herleeft' project. The project was carried out in 2016-2017 by the Amsterdam Museum, the Netherlands Institute for Sound and Vision, the University of Amsterdam, the Waag Society, and was realised with financial support from the Mondriaan Fund, the Netherlands Coalition for Digital Preservation, the Digital Heritage Network, the Prince Bernhard Culture Fund, and the Creative Industries Fund NL.



A manifesto for safeguarding and preserving born-digital heritage

Aims.

Finding ways to preserve born-digital heritage has become a matter of urgency and growing concern. Websites, games and interactive documentaries each bring specific challenges that need to be addressed. It takes three to tango: Ensuring that our digital lives and digital creativity are not lost to future generations requires a joint effort by the principal players: creators, heritage professionals and policy makers. This manifesto lays out the actions they need to take today to safeguard born-digital heritage.

Creators.

Digital products are at risk of being lost from the moment they are created. Creators are therefore part of the preservation process - whether by writing code, editing digital content or by creating some other form of digital expression. We encourage creators as follows:

- Invest time to describe your work carefully, whatever platform you use to store and manage your work. Provide at least a minimal set of metadata (who, what, where, when). Always include versioning data and information about the rights status of the work.
- Document your work as copiously as possible. Documentation enables future users to understand and reuse your work more easily. Describe the technical specifications of your work, for example the hardware and software used to create the work.
- If possible, assign open licenses (such as Creative Commons) to your work. This enables content to be reused. Reuse will help to ensure the longevity of your work.
- Where possible, use open-source software and open-source hardware. Your work will withstand the test of time better, since open means: independent of proprietary technology and vendor lock-in, and transparent availability of the source code and building blocks of your work.

Heritage Professionals.

Digital material presents several challenges for heritage professionals. For instance, the sheer amount of material created, dispersed among diverse platforms,

hardware and domains makes selection a daunting task. There is little standardization of file formats and environments that supports these. Norms for describing and managing this complexity are inadequately developed. The tasks involved in collecting, preserving and making digital materials accessible fall into three categories. We encourage heritage professionals as follows:

Policy

- Identify vulnerable digital heritage in your area of activity and find out which forms of digital heritage your organisation develops, manages or intends to manage (in line with collection policy plans). Create a convergent digital landscape by harmonising collection policies with other institutions. To ensure success, avoid overlaps and gaps in the combined collections.
- Develop policies for acquiring and keeping born-digital material accessible sustainably. Use existing models, as described in the 'DIY Handbook of Web Archaeology'.
- Obtain legal advice regarding storage and reuse. Act responsibly when using, managing and making personal data or information accessible.

Implementation:

- Where possible, cooperate with (fellow) institutions and industrial partners to find collective solutions. Choose robust –preferably open - technical infrastructures and operating systems.
- Assume that your current technology will need to be updated regularly. So prepare your exit strategy: can you move data from system A to system B easily?
- Use well-documented, open standards, e.g. for storage formats and exchange protocols. Non-dependence on suppliers ensures your archive material remains interchangeable in the future.
- Agree clear guidelines for delivery of acquired and transferred born-digital material: when, why and under what terms. Outline the rights and obligations before and after material is transferred. If accessibility is an objective, organise this when the acquisition is realised: lay down terms for accessing the collections.

- Ensure copious metadata records are kept of digital objects so that the context in which these were created is clear for future users. E.g. record the hardware and software environment in which objects function. Document data in the form of descriptions, photos, screenshots, screencasts, videos etc, and establish conservation procedures.
- Ensure collections can be used and reused, that digital objects can be found, accessed, interoperated, reused and stored in a sustainable manner. Use and reuse by a large group of users increases awareness of the importance and need for preservation.

Knowledge sharing:

- Exploit the power of the community. Introduce your team to the original creators, inventors and users. Organise meetings to share expertise.
- Keep track of developments in amateur communities involved with digital works. Much can be learned from bottom-up initiatives by amateurs who keep older digital cultures alive.
- Invest in your co-workers' increasing expertise and keep track of developments by following blogs, seminars and participating in domestic and international expert communities, such as NDE (Digital Heritage Network Netherlands) or iPRES.
- New born-digital products require new instruments and new research queries. To keep pace with rapid changes in technology, be prepared for new ways of working and new ways of cooperating.

Policy makers.

You hold the key to creating a sustainable policy that ensures sustainable digital heritage. We encourage policy makers to do the following:

- Stimulate cross-domain collaboration and use of collaborative instruments. Value hands-on expertise. Encourage the experts who love tracking down bit rot, annotating ancient code and building emulators. Put them in a position to share their knowledge and obtain recognition for their contribution.
- Bring the need to develop sustainability policy and preservation policy to the attention of institutions.
- Stimulate the emergence and use of open and collective services to ensure that as many heritage institutions in the Netherlands as possible will be able to guarantee long-term access to the digital collections they manage. Encourage collaboration within the Digital Heritage Network, based on the National Strategy for Digital Heritage.
- Encourage copyright reform to facilitate the preservation, availability and reuse of born-digital heritage.

Invitation

There are many ways to advance the aims outlined here. We are the steering committee of the DDS project,

realised in 2016 and 2017 with support from Mondriaan Fund, National Coalition Digital Preservation, Digital Heritage Network, Prince Bernhard Culture Fund, Creative Industries Fund NL. We welcome your feedback regarding this version of our manifesto and are keen to learn about other activities designed to pursue these goals. We encourage others to join us to help further the overarching objective of preserving born-digital heritage.

Judikje Kiers Amsterdam Museum
 Julia Noordegraaf University of Amsterdam
 Johan Oomen Netherlands Institute for Sound and Vision

Marleen Stikker Waag Society

More information: <http://hart.amsterdam/freeze>

This material is licensed under Creative-Commons-Licence BY 4.0 International: <http://creativecommons.org/licenses/by/4.0/>

