



**UvA-DARE (Digital Academic Repository)**

**Archaeology of the Amsterdam digital city; why digital data are dynamic and should be treated accordingly**

Alberts, G.; Went, M.; Jansma, R.

*Published in:*  
Internet Histories: Digital Technology, Culture and Society

*DOI:*  
[10.1080/24701475.2017.1309852](https://doi.org/10.1080/24701475.2017.1309852)

[Link to publication](#)

*Citation for published version (APA):*  
Alberts, G., Went, M., & Jansma, R. (2017). Archaeology of the Amsterdam digital city; why digital data are dynamic and should be treated accordingly. *Internet Histories: Digital Technology, Culture and Society*, 1(1-2), 146-159. <https://doi.org/10.1080/24701475.2017.1309852>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Archaeology of the Amsterdam digital city; why digital data are dynamic and should be treated accordingly

Gerard Alberts<sup>a</sup>, Marc Went <sup>b</sup> and Robert Jansma <sup>b</sup>

<sup>a</sup>Korteweg – de Vries Institute, University of Amsterdam, Amsterdam, The Netherlands; <sup>b</sup>Graduate School of Computer Sciences, Vrije Universiteit/Universiteit van Amsterdam, Amsterdam, The Netherlands

### ABSTRACT

One of the major initiatives in The Netherlands promoting the use of the Internet by private individuals was De Digitale Stad (DDS), which is the Amsterdam digital city. DDS was launched in January 1994 and soon evolved from an elementary bulletin-board-like system to a full blown virtual city with squares, houses, post-offices, cafés and a metro. Archaeology of the digital city makes it clear that there is no beaten track for preserving and, after two decades, unwrapping “born digital” material. During the research to reconstruct the digital city two routes were tried, one emulating the old system, another replicating it. The outcome, together with the harvest of two working systems, is a lesson, a concern and an appeal. From the experience of reconstructing digital heritage, we draw pragmatic lessons. Tools for digital archaeology are tried and contemplated. The lessons, however, do not unequivocally support the use of the notion “archaeology.” The concern is one of the social responsibilities. Web archaeology, being part of contemporary history, confronts the researcher with such issues as privacy and the ethics of “young” data. A case is made for treating digital data dynamically.

### ARTICLE HISTORY

Received 5 December 2016  
Accepted 19 March 2017

### KEYWORDS

Web archaeology; De Digitale Stad; emulation; replica; research ethics; privacy; digital heritage; dynamic treatment of digital archives

## Introduction

Participants in the digital city had an avatar. DDS, De Digitale Stad, that is the Amsterdam digital city, was much more than an Internet server, if only because the community shaping it had not settled for a definite meaning. The systems conveyed a sense of community building, and although there was not one but many communities, there was this one basic sense of being a virtual “citizen” expressed by the avatar. The DDS-team was self-conscious enough to adorn its second anniversary with a full backup of the system “to be studied by archaeologists in a distant future”: The FREEZE, 1996.<sup>1</sup>

Two decades do not create a great distance. However, we do embark on the archaeology of DDS. Following the actors of 1996, the effort to read and resuscitate vintage digital material may be called “archaeology.” The metaphorical expression does not come without repercussions. First, the lack of distance poses problems of contemporary history, hardly associated with the notion of archaeology. Second, notions like digital archaeology

or web archaeology suggest a kinship to a media archaeology or to an archaeology of knowledge in the sense of Foucault (1969) or media theory (Ernst, 2013; Presner, Shepard, & Kawano, 2014, p. 84ff), which is hardly explored here. The present contribution is about digital, but very material, old tapes. Getting hold of the vintage tapes is one thing, reading them and making sense of the content quite another. Different tools were tried and developed. In the perspective of accessing the data for historical research and possibly museum presentation, major issues arise. Ethical issues, not unusual for contemporary history, hit the digital archaeologist in the face. More specific questions related to the technological aspect of DDS impose themselves, issues of security and integrity of the data. The question of privacy, perhaps not strictly related to technology, poses itself with new urgency.

### *Archaeologies, emulation vs. replica*

The effort to reconstruct the digital city and have it in operation almost naturally proceeds along two routes: emulation and a replica (simulation). The two notions of emulation and replica have a long evolution of shifting meanings in art history. Here they are taken after their, equally unstable, meaning in computer science (Smith & Nair, 2005, Chapter 2). The resulting contradistinction adds a nuance, informed by computer science, to the discussion on emulation in electronic art (Jones, 2004).

Emulation is the effort to run the original code on a new platform. This does not come without compromise. The DDS-system as preserved does not easily “un-freeze”. It will not run, primarily because the physical systems supporting it are not readily available. Migration to present day systems is tedious, but feasible. Getting the legacy software operational requires adaptations. The authenticity-question, as to what system one is “actually” running, always remains.

Replica, the look-alike remake by present day means, seemingly discards the authenticity-question – one is obviously not running the real system – and focuses on presence to the user, rather than on the system.

The difference is that the mimicking is on a different level and from a different perspective. The original sense of emulation is that of one artist out of admiration mimicking an other. Here one computer system is thought to mimic the other. The agency is with the system; it is placed on level with the artist – in the original sense of emulation. The perspective is that of the system; the boundary is between hardware and software, or between layers of software. In a replica, the perspective is that of the user. Irrespective of what happens under the hood of the system, its surface performance for the senses is what counts. The boundary is between the user and the system. No sharp distinction is assumed between user and system, or between hardware and software. The crucial point here is that considering something either an emulation or a replica is a matter of perspective and comes with diverging expectations. Criteria vary from reconstructing the operation of the system to recreating the user experience. Recreation implies making the experience present, and is in that historiographical sense presentism. Does the archaeologist in her interpretation identify (with) the code or with the user? It seems that archaeologists by default choose the first; we are preoccupied with the authenticity of the code. We are, but in fact we show the alternative route as well.

What approach to choose and which tools to develop is contingent upon the goals set for such a project, which in turn depends on the context. Heritage has a community on

the donating end, a group strong and dedicated enough to not throw away the material remnants, and on the receiving end, a group feeling strongly about the value of these materials. In fact there are several groups caring for the DDS heritage and their goals vary. For the joy of a one-time replay a system may be fired up and run with loose ends. For public access, by contrast, for example a museum exhibit, integrity and security of the system and privacy of personal data are key issues. The mere consideration of the latter purpose was one of the motives inspiring the alternative approach of building a new system from scratch with the same functionality, a replica.

### *Dynamic approach*

If preserving and reconstructing data may be called web archaeology, what are the “scoops” and “brushes” in the digital practice? For the most part tools of digital archaeology require manual calibration and application; automated procedures are in their infancy. Working with legacy digital material brings home one crucial insight: whether the unearthed objects are data, scripts or full blown software; their archaeology involves getting the code to work. Born digital material is dynamic. Executing a script may yield different outputs each time it runs. Static material does not react, let alone react differently. The archaeologist will not be satisfied with images or screenshots.

This is in stark contrast to the existing practices of Web Archiving, viz. to preserve snippets of the Internet as pages, as snapshots. The maturing of the field of Web Archiving has been well captured in the volume edited by Julien Masanès (2006). The pioneering work of the Internet Archive and its Wayback Machine from the late 1990s has been broadened and institutionalised on a national scale in many countries. National Archives and Media Archives have automated their harvesting with crawlers and filters. The materials they gather, however, are static, or rather, are treated statically. Even if websites contain code, they are saved simply as pictures of pages. Today, the Internet Archive does more. It replays the harvested pages as much as it can. Our plea is to reinstate the pages as they were born, not starting from the resulting page, but from the server. Given the dynamic character of born digital material the archaeological approach should be dynamic. To put it differently: such material is called “born digital” to emphasise that its symbols are not just text. The text is considered as working code. And for working code, emulation and replica offer themselves as feasible approaches, each with their own tools.

### **De Digitale Stad: a local history**

De Digitale Stad, the Amsterdam digital city, exemplified the electronic social network. It facilitated the exploration of all the possibilities to connect, including almost incidentally access to the Internet. It was designed with the city metaphor in mind. On 15 January 1994, the digital city opened its gates. Beyond the practice of earlier FreeNets, De Digitale Stad appealed to its users to adopt the metaphor and create a true community. It allowed the users to be “citizens” or “netizens” and enter the unknown world of the Internet.

The project was initially funded for 10 weeks by the city of Amsterdam, on the assumption of bridging the gap between local politics and the ordinary citizens. The number of

subscriptions skyrocketed. After the first 10 weeks, with the project clearly growing bigger than anyone had anticipated, DDS acquired further funding to continue beyond the initial experiment (Castells, 2001; Lovink, 2002; Rustema, 2001, pp. 42–67).

The first version of DDS was a bulletin board system (BBS), a static menu offering the user a choice of line numbers to continue towards further pages. Imagination was an essential asset for the user to walk the streets of the digital city.

The second version made a major step to change from the Gopher communication protocol, used in version 1, towards the newer HTTP protocol still in use today on modern web pages. Within weeks, yet another version of DDS was released. Through a major overhaul this version 3 had become a truly interactive system, embodying the metaphor of a city. The overarching metaphor was further detailed by such facilities as “post office”, “city square” and “café.” These allowed the “citizens” to navigate the city more intuitively. Users could fetch their mail at an email facility called “post office”, they could set-up their own homepages called “houses” that were reachable by traveling across “squares”, those were web pages linking between each other, or they could hang out in a “café”, which we nowadays see as a chat room. The city metaphor was introduced and promoted the Internet as a common, a public space, which it hardly was in 1994/1995, when network services were mostly available in universities, libraries and as private facilities in large companies.

DDS grew amidst optimistic expectations, expectations of technology having a democratisation effect. In the sense of spreading the technology itself among larger section of society it certainly had this impact. Hope that the technical facilities by themselves would bridge the gap between politics and public, and thus solve the representation and legitimatisation problems of politics soon evaporated. Such hopes were certainly played out in acquiring the initial subvention of DDS by the city of Amsterdam. The idea of a push-button direct democracy is a recurring dream, also in DDS circles. In the same vein but stronger and more specific for DDS were the expectations of an emerging community and a new kind of sociability. Conceiving DDS as a commons, as a public sphere, deeply motivated many of the early actors. Their motivation was even strengthened when in those very years in the US the Internet rapidly evolved in a commercial connectivity with a new kind of economy (Aspray & Ceruzzi, 2008). Lovink (2002) observes that not one but many communities shaped around the digital city and created their own niches subcultures. From recent research into the themes of the cafés and lists in DDS, one may infer that the electronic facilities did in cases serve as vehicles of emancipation.<sup>2</sup> Dennis Beckers and Peter van den Besselaar have in a series of publications shown the dynamics of the various groups involved in DDS-initiatives (Beckers & Besselaar, 1998; Besselaar & Beckers, 1998). Thus, while Lovink (2002) characterised the divergence between the communities as cultural differences, Beckers and Besselaar were more political in their analysis by pointing at divergence through conflicting interests.

At a more fundamental level studies in sociology, STS and media students have hinted at new kinds of sociability emerging in such connected commons as DDS. Castells, famous for his trilogy on the Information Age (1996), in *The Internet Galaxy* (2001) presented DDS and many other case studies as vistas on new forms of society. The more radical approach in this direction was Howard Rheingold’s (1993) effort to continue where Dürkheim and Weber have left us with “Gemeinschaft” and “Gesellschaft.” Reinder Rustema (2001) emulates Rheingold’s search for a “society” beyond *Gesellschaft* by the example of *De Digitale*

Stad. Sociology has not yet come to conclusions, but without any doubt DDS is part of the empirical material to be reflected upon. Media studies on their part show that the industrial society is superseded by the platform society (Van Dijck, Poell, & Waal, 2016) – of which then DDS is not a forerunner or prime example.

### Local frost and defrost

On 15–16 January 1996, the DDS servers were down for most of the night to allow for a full backup of De Digitale Stad. A full 1-on-1 disk copy of all the servers running DDS was created on 3 Digital Linear Tapes (DLT). DDS congratulated itself with a city frozen in time, preserved “to be studied by archaeologists in a distant future”: the FREEZE. Further heritage material was gathered at “gravediggers parties” and the Amsterdam Museum installed a small exhibit on DDS.

In restoring old data, it soon came to light that the package would not simply unwrap, or defrost. The DLT Tapes holding the FREEZE did not easily render their content. After a good deal of searching for auxiliary hardware, the tapes had been read and converted into the more common format of compressed gzipped tarball (.tar.gz). Initial attempts to extract this tarball of 10 Gb failed because, for no apparent reason, it exceeded the available storage. After several tries, each time with more storage available, the files were finally extracted to a network attached storage (NAS) with 12 Tb of free space. The size of the completed extraction revealed why earlier attempts failed: the data filled little over 2.2 Tb of storage space. When searching for the cause why a 10 Gb .tgz file extracted to over 2.2 Tb (220x its own size), we detected four corrupted files, each over 500 Gb in size – quite possibly the effect of a “decompression bomb.” Omitting these four files, the extraction returned to reasonable size, approximately 35 Gb. The project “DDS 3.0 operational” worked from these cleaned files.

### Sockets

However, in an effort to understand what had gone awry in handing down the legacy files, and to make sure that no major parts of the original files were missing, we made a detour going back to the original servers, still extant in the Amsterdam Museum. Other than the FREEZE the content of these servers was not strictly dated, let alone of the same date as the FREEZE. Because of its historical significance, the original, but not necessarily operational, hardware is preserved at the store of the Amsterdam Museum. For the purpose of our research project, we were granted access. For just this one occasion the original hard drives were retrieved from these servers – and put back.

In order to read the 20-year-old hard drives, vintage equipment was needed, with the sockets of the cables as major obstacles. The original server had eight hard drives, connected using three different cable sockets. The interface was SCSI, which is a parallel interface subject to different standards. The solution was found with the help of a former system administrator digging up the fitting connectors from an old drawer. To read out the content of the hard disks a Linux live USB was set-up. The disks were connected one at a time. A full-disk-image was made – and preserved carefully.

This sidestep of the project greatly improved our understanding of the legacy material being studied, specifically the hardware. The newly retrieved content read from the original servers, showed sufficient overlap with the FREEZE of 1996 to confirm the adequacy of the cleaned files, but being of different date, was not included in our current reconstruction project.

### *Avatar generator*

In January 2015, after the defrost and clean-up of the FREEZE had been achieved, first forays into the data were started. Since there was little indication of the structure of the stored file, except that the system was an old Sun SPARC system, investigations began with mapping the folder structures and sizes, and listing the installed software. The first observation was that the system was not as systematic as one might naively assume. In particular, there were no systematic locations for source code, if preserved at all, going with the installed code. Harsh lesson for the archaeologist: with the programs installed and running in 1996, there will not even be an indication of source code being preserved and, if so, where it might be stored in the system. Another harsh lesson, even the non-techies have to learn some jargon.

Software appears in various modes, basically “source code” and “binary.” Programmes are written by the programmer – with the help of a whole factory of tools – written in a programming language. This written version is called source code. This is the version one usually refers to when discussing programs. Source code is just text, it does not work by itself. A program in source code is translated to fit on a system, in our case on the Sun SPARC station. The result of translation is an “executable” or “binary” code. This is the version that does the work. Hence, the digital archaeologist looks for the “binaries” to see what a system can do, which engines are available.

The tool doing the translation from source code to binary is called translator or “compiler.” In some special cases and only for specific languages, a tool exists which can do the reverse translation, back from binary to source code, a “de-compiler”.

### *Memories*

In exploring the files, it was of great help that former “citizens” vividly remembered the avatars. This clue from oral history set the challenge of locating the relevant software. With the introduction of DDS 3.0, the system had become truly interactive and logging in became more than having access. Every user would now have an avatar representing him or her whilst “walking” through the digital city. A program would generate a small icon-like image representing the user, an avatar. Somewhere on the server there must be such a piece of software performing that function, the avatar generator. In 1995, for lack of memory space and operational speed, the system would not allow for pictures or other complex images to be inserted. Therefore, the avatar generator created simple but effectively distinctive images for every user, varying on a pattern inspired by a character from the Muppet show, Beaker.



Exploration of the frozen data yielded no such file as avatar or avatar generator. A further clue was to search for the programs governing registration to the system, because that was the procedure including the creation of an avatar. This led to locating Apache and its original configuration file, from there the original registration page was traced. The registration scripts had been written in CGI PERL, easily readable, and thus the steps of registration could be traced.

Further recollections from oral history suggested that these icon images were never called avatars, but DoDoS, in an apparent play on the name of the system DDS. And “Dodo” is the extinct bird from Mauritius. So, the avatar generator would revive the extinct bird. The script, now located, had the name Dodo.cgi. This shows the development of terminology of the past 20 years. Where nowadays “avatar” is the appropriate jargon, it shows that this term was not as pervasive back in the day.

Soft lesson for the archaeologist: follow the challenges set by oral history.

## The reconstruction project

With the avatar generator resuscitated in 2015, and several other chunks of software brought to life in 2016, we could not resist the temptation of trying to get the whole system back into operation, project “DDS 3.0 Operational.” The project group soon split up and worked in two opposite directions. One part of the research focused on what seemed to be the most obvious thing to do, viz. to try and run the original software again. The other part focused on the idea of reinstating the user experience, regardless of the machinery behind the screen. This second route, to replicate the digital city, leads to rebuilding the system from scratch and working per modern technology and standards.

## Preservation of the digital city

The goal of the FREEZE had been to preserve De Digitale Stad as it had existed and run two years after its introduction, to ensure for future generations the possibility to experience and to study the early days of the web. However, simply backing up one’s data does not automatically result in preservation. A 1-on-1 copy may be historically accurate, but such a copy loses much of its attraction if it cannot be run, if the context of the original production server is missing. DDS ran on the Solaris SPARC computer architecture. Because Sun SPARC has become proprietary software and hardware from Adobe, with prohibitively high licensing costs, virtualisation seemed the road towards reconstructing



an adequate rempalcant context. It proved to be not that easy, since emulation of the SPARC architecture on a different system has usually low performance, particularly on the most common architecture in the 2010s, the x64 architecture.

## Emulation

The one research direction was to revive De Digitale Stad in its original state, from the perspective of the software running it. Like a hyena going for the innards, the software archaeologist goes looking for the software that was run, the software that was executed to create the performance of the system. The search is for the executable files, or “binaries.” Being familiar with the operating system, mostly Unix, and knowing how to identify and search for these files, binaries, archaeologists use such instructions as “grep” and “find” to trace the location of the executables in the FREEZE.

Lesson of exclusion: only close familiarity with the legacy system and its operating system will allow one to do the work of a web-archaeologist. To a large degree this is tacit knowledge.

Knowing which programs did in fact do the job is not enough to run them again. Binaries are executable only in a specific context, in this case the SPARC context.

At this point the web archaeologist, in general the software archaeologist, bereft from straightforward automatic tools like a Virtual Machine doing the work, must revert to more subtle and more individual methods. For the DDS 3.0 Operational project, a pragmatic decision was made with far reaching consequences. In spite of the dependencies – i.e. the points where the programmer had created constructions particular to the specific machine – we chose to emulate on an x64 architecture. The consequence being that the binaries on the old system (Sparc) are of no use and one has to create new binaries for the new system (x64). We had to go back to the source codes, the programmes as written, and would in that mode of the programme have to deal with the dependencies each individually. The task at hand was now to find the original versions of the software running in 1996 and compile these programs anew for the operating systems coming with the x64 architecture. Fortunately for this project a good deal of the source code, even if not systematically preserved, was retrieved scattered throughout the FREEZE. For future archaeology there is a use for systematic tools of de-compilation, in cases where there is no source code is available at all.

The positive side of emulating on an x64 architecture that this system is so common that one may expect it to survive for the foreseeable future. We are good, not for archaeological stretches of time, but at least for one or two generations. One may hope that the present work of getting a version of DDS running needs not be repeated from scratch in 20 years from now.

The downside is that corners of DDS with the heaviest dependencies are hard to restore. In particular, the system for authentication, logging in, was most specific for the SPARC architecture and has therefore been left out of the present emulation. By consequence the avatar generator, Dodo.cgi, was found and brought back into operation, but the programs constituting its context, logging in, are not. So, we can play and make avatar-wallpaper today, but we cannot have our DoDo walk through the system for us. When the former systems administrator was consulted on how to emulate this feature – which he had in fact programmed – his answer was: “consider not to.” In terms of DDS 3.0 as it ran in 1996, one can only visit the city

in the “guest” mode. In other parts of the DDS program functionality was restored by pragmatic patches. Thus, this emulation comes with compromise.

Along the route of emulation, the digital city has been restored as close to the original project as possible. In as far as it functions, it does revive with a feel of authenticity – so say DDS’s former inhabitants. As an extra benefit the emulation preserves code of historical interest and allows comparison to modern standards. It reveals the challenges as the DDS developers perceived them and the answers they chose when designing the early pieces of Internet. It makes the look and feel of born digital media accessible, including its inner workings and some of the thoughts behind it.

## **Replica**

The other approach was to revive the digital city from the point of view of the end user, replicating the original system as closely as possible using current technologies of 2016. This was feasible because, compared to today’s projects, the size of the digital city was relatively small and quite straightforward. Although providing static images only, the Way-back Machine did show the appearance of DDS to the end user. The programming for the replica was done in parallel to, and strongly inspired by, the emulation. While missing out on historical accuracy in the back-end software, the replica proves the feasibility of creating a user interface very near to the original, and in practice indistinguishable from the emulation – so say again DDS’s former inhabitants. The user will “walk” through the city without noticing it is not the original software. In that sense the experience is effectively preserved.

Major advantages of the replica are, beyond its technical maintainability and sustainability, its security. If one were to consider the creation of a publicly accessible version of DDS 3.0, for example as a museum exhibit, a replica would offer a doable solution; whether feasible in terms of museum practice, remains to be seen. As long as maintenance is kept up a replica can be secured and it could be filled with part of the legacy content upon authorisation by the, former, users.

## **Tools**

### ***The archaeologist’s brushes***

If preserving and reconstructing data may be called web archaeology, what are its brushes and spades? Digging up the digital city has in large part been a manual labour. In fact relied heavily on the tacit knowledge typical of craftsmanship. It required familiarity with Unix and other operating systems. Key element in the tacit knowledge is practical insight in the way server systems were usually built up, preferably joined with expertise in today’s systems. It took the joint forces of former system administrators who had not forgotten their trade and were willing to share, and archaeologists who are able to absorb. The latter are talented computer science graduates using some of their academic lessons and heavily relying on their experience of working as system administrators to pay for their studies. Their concerted effort has allowed for a “handmade reconstruction” of DDS.

But in a digital environment, should one not expect “systematic and automated tools”? Some tools do exist in the realm of software archaeology. For example, source code is

compiled into executable code. For some situations, e.g. programs written in high level programming languages, tools do exist for the reverse process: decompilation to reconstruct source code from executable files. Further automated tools are dearly wanted, like excavation tools to help to recognise the various types of files.

### *Beyond the archaeological metaphor*

As far as craftsmanship goes, the metaphor of archaeology sounds attractive and serves well. But in fact, what we are describing here has little to do with archaeology proper, but all with handing over, i.e. heritage through living tradition. It was not Pompeii but Amsterdam 1996 with its inhabitants still around today.

Once access is gained, from 2015 through unfreezing the FREEZE and now by its dynamic reconstruction, research is not solely focused on software. The floor is open for an analysis of the content. Approach and tools are quite different from the above. Whether historical, sociological, anthropological or phrased by media studies, the further research questions involve a completely different set of tools. Technologies of searching through data of filtering and of visualisation are available in the computer sciences and the data sciences. Some of the more sophisticated tools are being developed in a branch of data science going under the name of forensics.

In every part of the research on DDS, even in the most technical niches of the reconstruction process, the “archaeological” research is mingled with dialogue. Oral history helps. More than that, the intermingling reminds the researcher that in fact the historical approach is the umbrella underneath which it all makes sense: the dusting and scooping and fitting fragments together. Our archaeology of the web has all the benefits and the pitfalls of contemporary history. The takeaway message is that in spite of what the reconstructed operability of the software may suggest to some, the historical distance remains. It is in the very process of reconstruction that the archaeologist is reminded of the ineradicable historical distance.

### **Personal data**

Manipulating such “young data” on people acting as inhabitants of DDS two decades ago, and existing as fellow citizens today, may well produce a moment of shivers. The archaeologist finds herself swimming in a pond of personal information. Applied to the study of “young data” the metaphor of archaeology is brutally misleading. The work is social science or contemporary history and carries with it the social responsibilities tied to such sciences. Accessing young data poses major privacy issues.

In 1994 a major shift occurred that allowed people to share intimate experiences with the click of a button. Privacy was a difficult question and people did not, nor could they, predict the long-term implications of posting their information. Google, Facebook, Amazon and many other tech companies use these data to their advantage. Where users used to be proud “citizens” or “stakeholders in the commons”, today they are seen as natural resources for those companies having evolved into platforms. The meaning of “data” has shifted dramatically in the past 20 years, from information bearing, to monetary gain. The notion of privacy has changed even more.

### *Purpose limitation*

As a researcher in web archaeology, one will quickly find oneself dealing with legal and ethical concerns regarding the privacy of historical subjects. The web was first and foremost a communication medium and as such is filled with personal information of its users. Personal information is not just the information directly linked to one individual person. Any piece of information, that could possibly be linked to an individual, counts as personal information. Linking data to a username would link all these data to an individual, once that username has been linked to a person. By consequence, most of the web is subject to privacy laws, as most of the content is about, or created by, individuals – regardless of whether they can be directly identified through this content.

The ensuing legal considerations may vary greatly depending on national or state laws. And since the Internet hardly stops at national borders, legal considerations are complicated even further. Inside of the Netherlands, when handling personal data, the concept of “purpose limitation” (Dutch: *doelbinding*) is a core concept of the privacy legislation. Its purport is that personal data should only be used for the purpose for which they were acquired (Ketelaar, 2000). Medical data are to be used for health care only and not by insurance companies; income data gathered for taxes should not be shared beyond tax administration, etc. An exception to this rule is made if the purpose is historical research. However, leaning on this exception will only turn the legal consideration into an ethical one, which it already was.

By publishing web archaeological finds, the researcher will encroach on the privacy of the group being studied. And due to the freshness of the sources in the FREEZE, chances are that the group being studied is still alive and might object to the personal data being spread. Therefore, researchers must weigh the potential harm their research might cause, against the potential gain for society of including the personal data under consideration. The more so while, as researchers, we are inherently biased towards publishing. Therefore, other researchers must be consulted, and in an ideal world ethical committees should be installed overseeing projects of web-archaeology (Markham & Buchanan, 2012). In the DDS case, our provisional measure has been that any access to the retrieved data for the purpose of research is given under an agreement of confidentiality between the researcher and the Amsterdam Museum, procuring the source material.

### *Security*

If for historical research, privacy issues may be addressed in similar ways as other research, with an extra caveat, because of the rapid changes in Internet practices, the discourse takes a different turn in museum context. Suppose the purpose were to create a public exhibit out of legacy web sources, not only should the above reticence towards the publication of personal information be taken into account, but technical matters need to be considered as well.

Systems built in the 1990s were not created with today's practices of collecting data in mind. Technologies of protection have evolved accordingly. The legacy systems, even if one wanted dearly, could not possibly be made secure and safe to the standards of the twenty-first century. This thought has considerably strengthened the inspiration to build a replica next to the emulation, which in terms of safety and security must be judged hopeless.

## Treating digital heritage dynamically

The established practice of archiving the legacy of the web is to store “snapshots”, that is take a momentary image of a website and download it. This can be done automatically by so-called crawlers and yields enormous haystacks of information on the history of the Internet. Not only the *Internet Archive* and the *Wayback Machine* operate like this, national libraries and archives have adopted and standardized this approach (Mason 2007). The *Internet Archive* will replay the pages thus harvested. In doing so, they are reducing websites to mere pages. But the Internet is not a book, its sites are generated dynamically, be it at a pace of once per day or ten thousand times per second. Upon return the visitor will find a new thing. Michel Serres (2015) reminds us what a parochial way of organising our knowledge it is to put it page by page, now that we could liberate ourselves from that format thanks to the very Internet. As early as 2003 Helen Tibbo (2003, p. 16) observed the dynamic nature of the web:

A related problem is the Web’s dynamic nature. Web archiving initiatives can only preserve “snapshots” of sites or domains at the expense of their dynamism, rather like insects trapped in amber. Once snapshots of Web content are located outside the active Web, it is arguably missing one of its most characteristic properties.

The appeal was picked up by Michael Day in (2006, p. 193). We urge to take consequence of the dynamic nature and call for an adequate, dynamic approach. Far from incidental, the complexity and dynamism of the web reflect its digital nature. And the way we conceive of this heritage should change accordingly. The web should be seen not just as text and image. It is working text: code or software. Preserving the web, thus conceived, may well burst the frame of archiving. In that sense web archaeology is a truly new field, an extension of archiving proper.

The dynamic character of a website’s content expresses the underlying code. Whether simple CGI scripts, markups or complicated software, executable files lend the web its dynamic character. This code is the “working text.” Without it the documents as the user sees them are different, static. The dynamic approach to web heritage implies to take the text on the surface inclusive of the underlying code, its context. To be able to contain the context of a web page one cannot assume static snapshots as a solution. To properly preserve web content for future research, it must be stored with its dynamic elements in mind.

## Notes

1. In Dutch “gedeponeerd in een archief ter bestudering voor archeologen in een verre toekomst”, in “De digitale stad bestaat (bijna) twee jaar”. Post on the DDS, 1996. <https://hart.amsterdam/nl/page/37138/gevondenfreeze> <http://web.archive.org/web/20100830120819/> <http://www.alme dia.nl/DDS/Nieuws/freeze.html>
2. Unpublished reports of student work.

## Acknowledgments

“History of Digital Cultures” is a regular graduate course by Gerard Alberts in the joint MSc Computer Sciences programme of University of Amsterdam and Free University Amsterdam. “DDS 3.0 operational” was a special course in 2016 taken by Marc Went, Robert Jansma, Ronald Bethlehem, Tim

Veenman, Kishan Nirghin, Millen Mortier and Thomas Koch. Reports on this work are available on [Re:DDS 1.0]. The authors further extend their gratitude to Tjarda de Haan and the team at Amsterdam Museum for support throughout the project and hosting the presentation, to Waag Society for facilitating our meetings, to Jesse de Vos and Dennis Beckers for guest lectures, to Theun van den Doel for tireless moral support, and most of all to former system administrators Michael van Eeden and Paul Vogel for immediate help and for sharing their implicit knowledge of the systems. We gratefully acknowledge the Digital Preservation Coalition for distinguishing *The Digital City Revives*, of which our *DDS 3.0 operational* is a part, with the 2016 National Archives Award for Safeguarding the Digital Heritage.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors



**Gerard Alberts** is an associate professor for history of computing and history of mathematics at the Korteweg - de Vries Institute for Mathematics at the University of Amsterdam. He is editor of the Springer Series History of Computing and a member of the editorial board of IEEE Annals of the History of Computing and of this journal.




**Marc Went** is a master student at the joint MSc Programs Computing Sciences and Information Sciences of the Vrije Universiteit Amsterdam and University of Amsterdam.



**Robert Jansma** is a master student at the joint MSc Program Computing Sciences of the Vrije Universiteit Amsterdam and University of Amsterdam. He is also a research assistant at the Amsterdam Museum involved in sustainably preserving De Digitale Stad in the project “DDS herleeft”.

## ORCID

Marc Went  <http://orcid.org/0000-0002-1133-5805>

Robert Jansma  <http://orcid.org/0000-0003-1121-4189>

## References

Aspray, W., & Ceruzzi, P.E. (Eds.). (2008). *The internet and American business*. Cambridge, MA: MIT Press.

- Beckers, D., & Besselaar, P. (1998). Sociale interactie in een virtuele omgeving: De Digitale Stad [Social interaction in a virtual environment: The Digital City]. In *Informatie & informatiebeleid* 16-4.
- Besselaar, P., & Beckers, D. (1998). Demographics and sociographics of the digital city. In T. Ishida (Ed.), *Community computing and support systems — Lecture notes in computer science* (Vol. 1519, pp. 108–125). Berlin: Springer Verlag.
- Castells, M. (1996). *The information age: Economy, society and culture* (Vols. 3). Cambridge, MA: Blackwell.
- Castells, M. (2001). *The internet galaxy: Reflections on the internet, business, and society*. Oxford: Oxford University Press.
- Day, M. (2006). *The long-term preservation of web content*. In J. Masanès (Ed.), *Web archiving* (pp. 177–199). Berlin: Springer Verlag.
- Ernst, W. (2013). *Digital memory and the archive*. Minneapolis: University of Minnesota Press.
- Foucault, M. (1969). *L'archéologie du savoir* [The archaeology of knowledge]. Paris: Gallimard.
- Jones, C. (2004, June). *Seeing double: Emulation in theory and practice. TheEerl King case study*. Paper presented at the electronic media group. Annual Meeting of the American Institute for Conservation of Historic and Artistic Works. Portland, OR.
- Ketelaar, F.C.J. (2000). Elke handeling telt. archiefdiensten en de wet bescherming persoonsgegevens [Every act counts. Archival agencies and the law on the protection of personal data]. *Nederlands Archiefblad*, 104, 18–23.
- Lovink, G. (2002). *Dark fiber: Tracking critical internet culture*. Cambridge, MA: MIT Press.
- Markham, A., & Buchanan, E. (2012). *Ethical decision-making and internet research. Recommendations from the AoIR ethics working committee (version 2.0)*. AoIR. Retrieved from <https://aoir.org/reports/ethics2.pdf>
- Masanès, J. (Ed.). (2006). *Web archiving*. Berlin: Springer Verlag.
- Mason, I. (2007). Virtual preservation: How has digital culture influenced our ideas about permanence? Changing practice in a national legal deposit library. In M.V. Cloonan & R. Harvey (Eds.), *Preserving cultural heritage, Library trends*, 56-1 (summer 2007) (pp. 198–215) [Special issue].
- Presner, T., Shepard, D., & Kawano, Y. (2014). *HyperCities. Thick mapping in the digital humanities*. Cambridge, MA: Harvard University Press.
- Rheingold, H. (1993). *Virtual community, homesteading on the electric frontier*. Reading, MA: Edison Wesley.
- Rustema, R. (2001). *The Rise and fall of DDS: Evaluating the ambitions of Amsterdam's digital city* (Unpublished master's thesis). University of Amsterdam, Amsterdam. Retrieved from [http://reinder.rustema.nl/dds/rise\\_and\\_fall\\_dds.pdf](http://reinder.rustema.nl/dds/rise_and_fall_dds.pdf)
- Serres, M. (2015). *Thumbelina: The culture and technology of millennials*. London: Rowman & Littlefield International.
- Smith, J., & Nair, R. (2005). *Virtual machines: Versatile platforms for systems and processes*. San Francisco, CA: Elsevier.
- Tibbo, H.R. (2003). On the nature and importance of archiving in the digital age. *Advances in Computers*, 57, 1–67.
- Van Dijck, J., Poell, T., & Waal, M. (2016). *De platformsamenleving. strijd om publieke waarden in een online wereld* [The platform society. The struggle on public values in an online world]. Amsterdam: Amsterdam University Press.