Digital First? Saving Digital Worlds, Artefacts and Inhabitants

GERNOT HAUSAR, University of Vienna, Austria

This paper offers observations on the state of *"digital first"* preservation efforts in the *"digital ecosystem"* and sheds light on their place in the context of the digital Sciences and Humanities. It showcases approaches taken by different actors, identifies important but neglected challenges and offers some ideas for future "digital first" preservation approaches.

In the context of this paper *"digital ecosystem"* encompasses beings ("users"), material things and content related to "digital" aspects and is defined by the author as "all infrastructure (hard- and software), especially the critical components, used in the operation of the global systems of interconnected computer networks that users interact with and the signals that can be sent, stored and /or exchanged through it".

"Digital First" in the title refers to digital information and artefacts in the digital ecosystem that "originate in and have characteristics that can only fully exist in the context of a digital environment".

INTRODUCTION

While digital tools are widely used in the digital Sciences and Humanities (e.g. [Hausar 2016]), the "Digital" as an ecosystem where purely digital artefacts including the context and the "worlds" they exist in need to be experienced, recorded, reconstructed and reactivated for posterity is a minority view. With parts of the digital ecosystem not clearly defined but in some cases treated like state-territory, where purely digital artefacts can be created and also be stolen – with real-life criminal law repercussions – the question of preservation of those artefacts is paramount for digital archaeology and history. The sheer amount of data produced every year (1,7 MB of data every second [Domo 2017]), the different actors involved and the fast advances in digital technologies make it a daunting task.

Even using bridging technologies like (3D-)printing, a digital artefact can seldom be recreated in a physical form that closely resembles the digital, as sometimes it is a mere copy of the form without the (full) functionality. These "digital first" objects can best be recreated in a digital environment.

It only gets more complex if we think about whole digital spheres like web-portals, social networks or – even more complex – "Massive Multiplayer Online Games" (MMOs). The recreation process needs the emulation of the necessary hardware, the server-side software and (emulation of) the input devices. It further needs all the historical user interaction data with each other and the software. Only then can an artefact be adequately recreated and researched. Finally access to this recreated artefact is reliant on preserving or emulating the digital infrastructure. As preservation efforts often operate under constraints, emulations using modern technologies (e.g. emulating the DOS operating system in DOSbox, a software running on modern computers) offer an adequate compromise in some cases.

While there seems to be consensus on *what* is needed to preserve digital artefacts, the question of *who* should finance, coordinate and lead the preservation effort is still contested. While national libraries in principle would have the legal foundations to archive digital content (e.g. [Austrian National Library 2019]), they do lack financing and a clear political commitment. That leaves preservation efforts for now mostly in the hands of private enterprise (e.g.

Author's address: Gernot Hausar, Department of History, University of Vienna, Austria, 1010, Universitätsring 1, Wien; email: gernot.hausar AT univie.ac.at / gernot.hausar AT gmail.com

Google Books[1]), non-profit organizations (e.g. Internet Archive[2]; Project Runeberg[3]; Project Gutenberg[4]; eLibrary Projekt[5]), museums (e.g. Computerspielemuseum[6]), universities and educational institutions (e.g. alo – Austrian literature online[7]; phaidra[8]) and individual archivists. While there is some cooperation especially in the case of digitalization of corpora, there is still not enough done on preserving "digital first" artefacts.

DIGITAL FIRST ARCHIVES: WHY, WHAT AND HOW?

Former "Society of American Archivists" (SAA) president David B. Gracy II, when commenting on the use and necessity of archives in 1987, pointed out, that

> *"archives, like deposits that archaeologists dig up, are documentary remains of the past pulled together and preserved for use in the future. (…) The greatest service of archivists is contributing to the continuity of culture by stimulating connections between the useful information from the past and the challenging needs of the present. (…) Archivists serve not only contemporary (often transient) needs but also the possible needs of the researchers in the far-distant future"*

[Gracy II 1987]

Judging by the intent that is documented in laws in countries around the globe (e.g. laws regulating national libraries or limiting copyright), this obligation of preserving for the public good is still a duty today.

While this idea is well established in digital scientific and cultural endeavors, digital industries and – to a certain extent – private users, too – do not always see preservation as an important goal. James Newman, in one of the few early works on digital game preservation, argues that this lack of focus is due to the fast-paced nature of the creative market and a medium that has strong characteristics of obsolescence and supersession [Newman 2012]. This adds an additional layer of complexity to preservation efforts.

All digital artefacts are part of the digital ecosystem. They need bits and pieces of the system to properly function. Using the digital ecosystem as a basis to define the scope of the effort therefore offers insights into the individual parts that have to be archived to preserve a "digital first" artefact:

First and foremost the *hard- and software* needs to be preserved and made long-term accessible in working order, including digital and analog *artefacts associated with it* (e.g. manuals, which are "infrastructure" needed for the operation of the software), the *information exchanged* over the digital lifespan (e.g. databases) and finally *meta-information* from users and their interactions with the preserved artefact (e.g. reviews, forums).

In principle these aspects of preservation of "digital first" ("born digital") objects are well established in the literature (e.g. [Webb 2003]) and work well for static information like a digital picture. Interactive digital objects like games – and especially MMOs – offer a glimpse of what is to come even for "simple" digital objects in a digital ecosystem that is dominated more and more by social platforms with dynamic content and less by individual webpages with static objects. These should be preserved using the initial environment or, if necessary, have to be migrated in a standardized way into new rendering environments (e.g. [Ioannides et al. 2010b]). While there were a variety of EU projects dealing with interactive virtual representations of non-digital objects, especially text (from reUSE 2003[9] up to READ 2016[10]; the author's own elib.at digitization repository profited from these initiatives), even recent EU projects focusing on the digital infrastructure, like e-Infrastructures [University of Vienna 2016]; the author was part of the Open Access focus in Austria), did not provide solutions for the preservation of large scale social web objects in Austria [University of Vienna 2016], like social media platforms or MMOs (perhaps this will change with the follow-up e-Infrastructures Austria Plus or other initiatives like the Designing Digital Heritage Network [Ioannides et al. 2010a]).

---

[1] https://books.google.at
[2] https://archive.org
[3] runeberg.org
[4] https://www.gutenberg.org
[5] Elib.at: https://www.univie.ac.at/elib
[6] https://www.computerspielemuseum.de
[7] http://www.literature.at
[8] https://phaidra.univie.ac.at/
[9] https://www.uibk.ac.at/reuse/
[10] https://read.transkribus.eu/

As complex archiving situations will become more common, MMOs are the ideal testing environment on how such repositories should be constituted and how the individual parts should be joined together. All of these parts involve a variety of challenges that can broadly be divided in four main categories (for an overview of the basic challenges see e.g. [Lessig 2006; Mazziotti 2008; Langley 2019]):

- *Legal Barriers e.g.*
    - o   laws prohibit digital archiving longer than the digital artefacts sometimes exist
    - o   so called "abandonware" without clear ownership
    - o   general trademark and copyright provisions in conflict with archiving efforts
- *Software Barriers e.g.*
    - o   "Digital Rights Management" (DRM) & other Copyright Protection Measures make archiving and emulation difficult
    - o   time-consuming reverse engineering, as source code no longer available
    - o   ports, patches or game modifications ("mods") no longer available or only available for one version of the game
- *Hardware Barriers e.g.*
    - o   physical access to storage media is no longer possible
    - o   lack of parts for continuous repairs
    - o   problems with storage
- *Costs over Time e.g.*
    - o   initial costs
    - o   monthly costs and
    - o   costs of porting and emulating
    - o   personnel costs

The *preservation of the meta-information*, like video reviews or forum discussions is not included as those four categories have to be applied in full for each case individually (e.g. preserving a forum discussion makes it necessary to preserve the forum, too).

Preservation has to take into account the short lifecycle of digital information and measures guarding against the aforementioned obsolescence and supersession.

To get a better understanding of the challenges, the preservation of computer games is used as an example.

CASE STUDY: ARCHIVING COMPUTER GAMES (MMO)

Using computer games as an example is instructional, as they are at the forefront of the technical development cycle (e.g. VR-games), are on the one hand locked into a mostly commercial system with fast and short development cycles but also have a thriving open source modding and programming community. Computer games have no established and standardized archiving strategies and a lot of games have been lost due to the above mentioned challenges – especially the unclear legal situations. In some cases, games are simply abandoned and are either lost or curated by enthusiasts in abandonware archives.

There are generally 3 large groups of questions to answer, when preserving games, that deal with different aspects, games need in order to function (see fig. 1): Hardware, Software and Player Interaction.

arcade machines, old motherboards, spare parts, joysticks and gamepads, cables and connectors

hardware & periphery

source code, emulators, game art, copy protection, game rules and instructions, boxed sets

software (code, art, rules, instructions)

forum posts, playthrough videos, game modifications, fan art and stories, fan pages, faqs

player experiences, lore, metagaming, modifications („mods")

Fig. 1: Ingredients for Game Revival and Examples. (Own work. Hausar 2018)

There are a variety of game types to compare but especially important for archiving and revival purposes is the difference between those that can be played without an active online connection to the internet (ecosystem 1) and those, that need this connection to work (ecosystem 2).

The first type of game-ecosystem, as diverse as the individual challenges in archiving games may be, have mainly challenges in the field of hardware, from early arcade games (e.g. preserving the arcade machines) to e.g. Commodore 64 and Amiga 500 games, which need the hardware and the input devices to be preserved alongside the code. The code also often has to be adapted, emulators created and old input devices reengineered in order to work. The game content is often scripted and the main differences are how players react to it. Here different out-of-game player content should be archived to offer insights into the range of reactions. While ludologists have tried to capture authentic player (e.g. with "Electroencephalography" (EEGs) [Balducci et al. 2017]) interaction in controlled environments for some time it remains a distraction and a daunting task. The technically and financially low threshold for making digital videos, the popularity of streaming and online sharing of everyday experiences offers a new option as the player does not see the observation as unnatural. Most of the observation data is captured and shared by the observed themselves (e.g. Twitch Gaming Channels) and can be downloaded for archiving under open licenses.

The second type of game-ecosystem offers its own unique challenges. Here the difference is mainly between scripted games, where content is mainly scripted and players cooperate and compete against the scripted content (e.g. "World of Warcraft" (WoW)) and sandbox games (e.g. EVE Online), where the content is mainly created through the interaction of players amongst each other's on centralized servers. Another difference is scale, meaning the number of players that are able to play together over the net, ranging from one on one games up to Massive Multiplayer Online Games with thousands of players. These games generally have centralized hard- and software (main problems in archiving are legal in this case) but the interactions in gaming sessions with multiple players can differ widely even with scripted content and are therefore difficult to archive (WoW has around 9600000 subscribers).

Massive Multiplayer Online Games as server-based games are not reliant on a specific client infrastructure. This is an advantage from the perspective of the archivist. As mentioned above, depending on the type of MMO, content is either provided and scripted mainly by the game company (e.g. WoW) or in the case of a sandbox-type game through the interaction of players using in-game tools (e.g. EVE Online). Massive multiplayer online games focus less on a gripping storyline to motivate players than on player cooperation or adversity.

There are a lot of different game design approaches: Some games let players cooperate in scripted quests against the game's "Artificial Intelligence" (AI), other games focus more on an open sandbox, where players are free to use the game mechanics to build their own stories. Some games divide their players into regional groups, others let all

players join one server. Games are played in a persistent online world hosted by the respective game company servers.

As the second type of game ecosystem is a much harder challenge, it will be used as an example for the complexities of digital archiving and reviving:

Disregarding the legal complexities of getting the rights to the server data and game source code, a server cluster has to be provided to run the game. For a test environment with only a few players the initial costs for hardware are around EUR 1500 and costs for keeping the cluster up is around EUR 250 per server per month, if you use commercial hardware. This setup allows a few players to replay an MMO from scratch but does not re-create player stories and out of game content.

Player experiences can only be archived as recreations based upon playthroughs, player podcasts, streams and videos as well as community sites like forums, websites and fan-art. This content is hosted on a variety of external platforms that have to be recreated as well. One example is the World of Warcraft Wiki, which is a fan run site collecting stories, personal accounts, video and chat interactions and game advice. The community content is provided under open licenses and can be archived in bulk over the API (e.g. [Littmann et al. 2018]).

But how can you get people besides scientists to engage with the archived content?

ACTORS AND APPROACHES

Then and now it is necessary to raise awareness for the value of archives and the content that can be found, both analog and digital, as it is not always immediately clear how each citizen benefits (concerning the continuity of the challenges see e.g. [Roe 2016]). Due to the overall hesitant way in which traditional archiving institutions have participated in the preservation efforts so far, a variety of different and complementary strategies have emerged to finance these efforts, some of which are listed below:

- *Legal framework for digital preservation* and financing of long-term archiving infrastructures: This approach is for example used in Germany, where designated institutions are granted powers to archive content (e.g. legal right to hack copy protection granted to the German National Library for archiving purposes). This approach balances the rights of the individual author against the rights of all citizens to knowledge, culture and information, a principle that is strongly rooted in the continental European law traditions and present in most constitutions. It is a difficult deliberation that has to be made on a case-by-case basis (e.g. article 2 of the UNESCO Guidelines [Webb 2003]), but it is rewarding: If the rights of all citizens would not have been deemed more important than one author's rights, Kafka's works would have been lost to us as they would have been burned according to the express wishes of the author.

- *Offering grants* aimed at preservation efforts: Mostly used in concert with scientific research, grants specifically for archiving are established e.g. in Austria through the "Austrian Research Promotion Agency" (FFG).

- Requirement for *open access and archiving strategy* when receiving public grants: Another way to promote archiving in the scientific context is to mandate that proposals have to contain an archiving strategy.

- *Rebuild and Open Source*: Enthusiasts and citizen-scientists sometimes reverse engineer games and open source the code, so others can build upon and expand on the original game (e.g. FreeCiv as an open source fan-project recreating the iconic original Civilisation game).

- *Emulate it online* so it is possible to preserve at least the emulation, if constraints do not allow for any other solution and make it easily accessible: Out-of-copyright games can legally be recreated online so a broad new user base can experience them online (e.g. Internet Archive).

- *Archive as physical objects* e.g. in a museum or in the form of retro-games: A good example are old arcade machines, that can be experienced best in a real-world environment (e.g. in the Computerspielemuseum).

- *Update and sell old games* as a business model: Preserving old games offers businesses the chance for a second commercial cycle as retro-games thus making it also in the interest of companies to preserve their old games (e.g. "Good Old Games"[11]). While this in itself is not a solution to long term archiving, companies might not

---

[11] https://www.gog.com

simply delete their old games during their copyright term, thus making any legal preservation effort extremely difficult and expensive.

● *Semi-legal or illegal preservation* through abandonware or piracy: Black market economies offer illegal alternatives if accompanying factors are excluding certain groups (e.g. prices or availability).

Most of these approaches are sadly driven by private actors and not by official institutions. This is a problem – both from a legal and a long-term preservationist perspective – that has to be addressed on a country and international level. Initiatives like the Internet Archive, with mirrors in Europe and Cairo, are de facto filling the void until official institutions can catch up. Even then, a comprehensive legal framework is a necessity to legalize preservation of "digital first" content.

CONCLUSIONS

The status quo of digital preservation is severely hampered by the general approach of lawmakers towards the digital ecosystem. Recent developments like the efforts for a European Copyright Directive are not encouraging. There are a lot of undecided questions concerning the nature of the digital domain that have to be clarified to allow institutions like the National Libraries to officially include "digital first" artefacts under their existing mission statements:

As National Libraries archive on a nation-based approach, clarification is needed, which online content is seen as being in that nation (e.g. game company headquarter; player base; domain-based like *.at* or *.ac.at*).

Clarification (and legal security) is needed on how the different national and international actors should work together in the preservation effort

Finally, further clarification is needed on the general trade-off between copyright-exceptions and the right of society to knowledge, culture and art to determine, for which society of the future the content will be archived.

Looking back to 2003 and with full knowledge of all the effort put into digital preservation since then, the warnings from article 3 of the UNESCO preservation guidelines remain still relevant:

> *"The world's digital heritage is at risk of being lost to posterity. Contributing factors include the rapid obsolescence of the hardware and software which brings it to life, uncertainties about resources, responsibility and methods for maintenance and preservation, and the lack of supportive legislation. Attitudinal change has fallen behind technological change. Digital evolution has been too rapid and costly for governments and institutions to develop timely and informed preservation strategies. The threat to the economic, social, intellectual and cultural potential of the heritage – the building blocks of the future – has not been fully grasped."*

[Webb 2003]

ACKNOWLEDGEMENTS

REFERENCES

Austrian National Library. 2019. *The legal foundation of the Austrian National Library - Österreichische Nationalbibliothek*. https://www.onb.ac.at/en/about-us/the-legal-foundation-of-the-austrian-national-library (17.07.2019).
Fabrizio Balducci et al. 2017. Affective level design for a role-playing videogame evaluated by a brain–computer interface and machine learning methods, in: The Visual Computer 33/4, 413–427.
Domo. 2017. *Data Never Sleeps 6*. https://www.domo.com/learn/data-never-sleeps-6 (17.07.2019).
Gracy II, D.B. 1987. Is There a Future in the Use of Archives? *Archivaria Summer 1987*, 24, 3–9.

https://archivaria.ca/archivar/index.php/archivaria/article/view/11411 (17.07.2019).

G. Hausar. 2016. *Weitere Überlegungen zu digitalen Geschichts- und Geisteswissenschaften. digigeist Gastbeitrag Gernot Hausar (Wien)*. https://digigeist.hypotheses.org/308 (17.07.2019).

M. Ioannides et al. (Eds.). 2010a. Same But Different – Comparing Rendering Environments for Interactive Digital Objects. Digital Heritage.

M. Ioannides et al. (Eds.). 2010b. Designing Postdigital Curators: Establishing an Interdisciplinary Games and Mixed Reality Cultural Heritage Network. Advances in Digital Cultural Heritage.

S. Langley. 2019. Planning for the End from the Start: An Argument for Digital Stewardship, Long-Term Thinking and Alternative Capture Approaches for Digital Content, in: Horst Kremers (Ed.). Digital Cultural Heritage, Cham, 209–237.

L. Lessig. 2006. Code. v2. 2. Auflage, New York.

J. Littman et al. 2018. API-based social media collecting as a form of web archiving, in: International Journal on Digital Libraries 19/1, 21–38.

G. Mazziotti. 2008. Freedom of use vs. DRM Technology, in: Mazziotti, G. (Ed.). EU Digital Copyright Law and the End-User, Berlin, Heidelberg, 179–229.

C. Webb. 2003. Guidelines for the preservation of digital heritage. Information Society Division United Nations Educational, Scientific and Cultural Organization (UNESCO) – National Library of Australia. http://unesdoc.unesco.org/images/0013/001300/130071e.pdf (17.07.2019).

J. Newman. 2012. *Best before. Videogames, supersession and obsolescence.* Routledge, Milton Park, Abingdon, Oxon, New York.

K.D. Roe. 2016. Why Archives? *The American Archivist 79*, 1, 6–13. DOI: 10.17723/0360-9081.79.1.6

University of Vienna. 2016. e-Infrastructures - Deliverables, https://e-infrastructures.univie.ac.at/en/the-project/deliverables/ (17.07.2019).