

Conference at GHI Washington DC, USA
Digital Hermeneutics in History: Theory and Practice
10. – 12. October 2019

Pascal Föhr PhD, State Archive of Solothurn, Switzerland

Slides in Prezi

Historical Source Criticism in the Digital Age

Good morning, ladies and gentlemen.

The organizers asked me to present the results of my thesis. Given the limited time available, I decided to focus on the following aspects:

Table of content

- Properties of a digital object (as a precondition)
- Selected findings with short explanations (just to give you ideas for further discussions)
- Selected findings with longer explanations
- Conclusion

characteristics

First, I defined the characteristics of a digital object:

Digital sources are genuinely digital, multimodal and multidimensional, not traceable, manipulable, volatile, independent of data carriers, lossless duplicable and (at the same time) reusable and dependent on a specific digital information system. The characteristics from which digital objects differ from previously known objects are marked.

The further research is based on this definition.

Source types

Three new source types could be defined:

1. The integrating source type: a digital object can integrate other objects in a way that is not visibly traceable. A website can integrate external sources without having them to be stored locally. The user can't trace which elements of the website originate from where.
2. The reducing source type: If several clones of a digital object exist, they can be reduced to a single object without loss.
3. The source type of invisible coded information: Digital sources exist at different layers at the same time and every layer contains different information. The invisible code can also contain actions that are only triggered by certain actions of the user (mouseover).

Functions

So far, primary, secondary and tertiary sources have been classified. Digital sources can, however, take over two additional functions, the quaternary and the quintary sources. Both are generated by an algorithm in an information technology system. The quaternary source is the product of a process, the quintary source is an algorithm created by an algorithm.

My most important finding is source safeguarding:

[Source safeguarding]

Most physical sources can be retrieved, read and reused at a later date without additional procedures if stored correctly. This is not possible with digital objects because they require a digital information system to be reused. This consists of hardware and software components that are subject to a rapid aging process and may no longer function within a relatively short time. Therefore, it is not sufficient to obtain only the binary code in order to be able to guarantee future retrieval, reading and interpretation. In addition, the characteristics of volatility and incomprehensible manipulability of a digital object fundamentally requires source safeguarding in order to guarantee the availability, integrity and authenticity of the digital object for as long as possible. Above all, research resources on the Web can be changed or deleted within a very short period of time and therefore may not exist later or not as they existed at the time of use by the user. Source safeguarding is therefore of great importance not only for the user's own critical examination, but also for source criticism carried out by a third party.

How a digital object has to be secured is not as trivial as it seems, because it is necessary to decide on form, format, additional hardware and software components, code levels, dependencies, storage location and supplementary metadata. Safeguarding often creates an additional, reborn-digital object that has not existed before and that in many cases can no longer be used in the same way as the primary object. Nevertheless, the greatest possible consistency must be maintained. This is understood to mean that all information is available in order to be able to present it in its context, in the intended manner and with complete meaningfulness.

Due to its importance for the traceability and authenticity of digital sources, safeguarding is of outstanding importance and must therefore be defined as an independent process step in the historical-critical method, so that access to a digital research resource and its availability can be guaranteed permanently. I refer to the repeatedly heard statement by Jeff Rothenberg: "Digital objects last forever – or five years, whichever comes first".

Creating metadata is part of the safeguarding process:

[metadata]

The evaluation of the authenticity of an object depends to a large extent on the available metadata. Various tasks can be performed by metadata: the unique identification of an object, an object description within an collection, embedding in a context, support for retrieval and identification of the object, structuring in collections within a repository, proof of authenticity with history documentation, support for ensuring object integrity, registering access, access and other rights, simplifying the exchange of information and cataloguing technical parameters describing how the object should be presented.

Metadata are particularly useful for scientific publications if they can be read automatically. Therefore, the object itself and the publication platform used should always provide all relevant metadata in a standardized form.

[infrastructures and tools]

In order to secure digital sources, appropriate infrastructures must be set up, including those for individual researchers, which is known as *Researcher Driven Archiving*. The server infrastructure and the backup process must be trustworthy and the objects must be retrievable with a persistent identifier. The infrastructure must not be obtained from private institutions, because it could be switched off tomorrow and all data would be lost. Only state archives or similar institutions offer the necessary security, because they guarantee their permanent existence by law. They would fulfil the requirements for the trustworthy and permanent archiving of digital objects. Unfortunately, only a few archives are equipped with digital long-term archives and even less take up assets of individual researchers.

It would therefore be advantageous if a software, a *Personal Trusted Archive*, would be available in which digital objects could be secured with a *Trusted Saving Procedure* in a local, trustworthy container.

Stored objects must be retrievable. Although DOI and URN are widely used, they have one weakness: they depend on resolver services, which must always be available and updated when changes are made. A possible solution would be a unique address of the resource: a combination of the ISIL-number of the first responsible archive, the person, the project and the object. For example: CH-101111-1.00120.A0Y4.02I8. The object is accessed under the URL of the archive <https://tdr.staso.ch> – of course the URL of the archive must be blocked for trade. The object has a unique address and is therefore identifiable, it's independent of the storage location and can be accessed directly without a resolver service.

Like all sources, digital sources must also be critically examined.

[Source criticism]

The examination of a digital object is based on the questions of the traditional historic method, but the examination is only possible with information technology methods.

Authenticity and integrity also play an important role in digital sources:

[authenticity]

Most analogue sources can be assumed to be authentic, which is why they can be falsified. Digital sources should generally be mistrusted, which is why they should be verified. One way of proving authenticity in a comprehensible way is to approximate authenticity. On the **basis of 7 criteria** a source is comprehensibly assessed and evaluated. The result gives the researcher an approximate value of the authenticity of the source. This process, I call it the *Authenticity Approximation* is nothing new in itself, and every researcher works intuitively with it. Moreover, it cannot be applied to all sources, the effort would be too high. But in the case of critical and key sources, it would make sense to write down one's own considerations, especially for third parties.

[integrities]

A digital object does not only have a computational or technical integrity (Yes/No). From an intellectual point of view, one can speak of various integrities: the integrity of the content, of the representation and of the usage as well as the integrity of the information systems involved are of interest. For example, the computational and representation integrity of an object may be violated, but if the content integrity persists, the source could be used. In art history, on the other hand, representation integrity is of utmost importance and the same object would probably be unusable.

Another important finding is known in our circle, so I will discuss it only briefly:

[competences / skills]

The demands on a historian or scientist in dealing with digital sources are high, which unfortunately is still underestimated by the responsible authorities from a large number of educational institutions. Today, new key competences are required:

- **Information competence** as the ability to determine the need for information on a question, to develop a search strategy, to know and be able to use the appropriate sources of information, to evaluate and select the data obtained and to process the information in a way that the question can be answered effectively.
- **(Digital) media competence** as the ability to use all kinds of analogue and digital media for the communication and behaviour of people in an active and appropriating way.
- **Communication competence** as the ability to use communication channels and media for interacting with colleagues and the public and to filter and (re-)organise all available information, suggestions and hints.

- **Competence in computer science or information technology** as the ability to critically examine and assess digital information systems and their products, the digital objects.
This requires not only basic knowledge of how hardware and software work, but also programming skills.
- **Data competence** as the ability to adequately assess and visualise digital databases in terms of their origin, quality and intended use, using appropriate means.
- **Legal competence** (in particular copyright law, data protection and privacy)
- **Archival competence**, because with the safeguarding of digital objects, the scientist also pursues a kind of archival appraisal of records.
In many cases objects that are safeguarded, did not exist in this way until then.

In the following examples, I ask you to consider what skills are required for safeguarding, creating metadata, and critical examination:

1. [Online article: Website NZZ]
2. [Maps of Cashmere in Google Maps Switzerland and India]
3. [DB Black Box / Dynamic Visualization]
4. [Games]
5. [Venice Time Machine]

A wonderful project, but from a methodological point of view problematic: 80 kilometres of documents were digitized and processed to create a digital visualisation of the evolution of Venice over centuries. But I do not believe that every source was critically examined thoroughly. But you still have to do it, especially because you have sources like this! In digital history, too, the 'analogue' competencies are still needed.

These examples show that different and new competences and skills are necessary for different research questions and that these must somehow be acquired or trained – in addition to the skills already being trained

[publication and discussion]

Digital objects require digital forms of publication and visualization, because they cannot be represented analogously at all. Also, new possibilities of interaction between author and reader have emerged and are emerging, even while working on the publication. These new forms should be recognized as part of scientific work, which is only to a small extent the case today.

[conclusion]

To summarize this summary:

1. digital sources differ significantly from previously known sources
2. digital sources must be safeguarded in specific infrastructures
3. historical science urgently needs to develop new competences
4. historians, archivists and computer scientists should exchange information constantly

Thank you very much

Pascal Föhr, pfhist@foehr.ch
hsc.hypotheses.org, PFHist (Twitter), LinkedIn
Dissertation: <https://edoc.unibas.ch/64111/> - ISBN: 978-3864881534