

Full Description

The present state of knowledge about the longevity of many materials used to store media (optical disks, hard drives, magnetic tape) remains limited, and usually such technologies are not designed for safeguarding data for a long-term. One should keep in mind that although certain methods such as error-correcting codes can add some redundancy, the problem of digital data preservation is not simply reducible to the preservation of the corresponding binary code. One equally needs to preserve the corresponding software (especially for multimedia data) and sometimes specialized hardware. One example that led to the loss of many digital creations was the sudden drop of flash support in 2011 leading to a rapid loss of income of many hosting websites whose works were not all archived in the ensuing debacle. Many modern systems cannot run the old files as the common software required are nowadays considered security risks. Moreover, the expertise on how to run the outdated software and hardware also needs to be preserved. Any digital archive that is not continuously maintained is at risk of becoming incompatible with the present-day technologies in sometimes just a few years.

The task of digital archivist also involves a specific difficulty related to the fact that the digital data *prima facie* are freely reproducible and do not involve any notion of original. This fact is at odds with the established archival practice of being in charge of preserving particular original items such as documents and other material artifacts. The absence of digital "originals" among other things affects the accountability of archivists. As a part of the proposed project we will research whether recent blockchain and NFT technologies aiming at reconstructing the concept of original ownership in a digital environment could lead to new forms of accountability. A closely related problem concerns the possibility of discrete modifications of digital data, whether intended or not. Existing public online archives often give an impression that they are static and durable when it might not be the case (especially when they are owned and maintained by private actors with financial interests). There is a strong temptation for certain actors to censor or otherwise modify older documents (including texts, audio and video recordings) to comply with the present-day tastes or regulations. Even though digital signatures and hashes (potentially stored on a blockchain) can guarantee authenticity in the medium term, the evolution of cryptography questions the ability of such signatures to last more than a few decades.

It should be stressed that on the top of the various technical issues related to the durability of digital data the problem of durability of digital archives has important social, cognitive and other human-related dimensions. Indeed, the archived data are always supposed to be in some sense *meaningful* to those who wants to preserve them. The endurance of digital data (along with the necessary soft- and hardware) does not, by itself, guarantee that the preserved data remain meaningful over time. A perfectly preserved clay tablet with a cuneiform text is meaningless unless historians and linguists manage to translate it to modern languages and explain its original meaning in modern terms. The relevant concept of meaning is broader than the standard formal semantic notions because it assumes a multi-linguistic environment and a broad context of interpretations. The talk of invariant meaning is somewhat naive in such contexts since our understanding of older artifacts typically changes through the historical time. What allows old texts and other old artifacts to endure through time is the continuity of traditions of interpreting them rather than a timeless meaning attached to these items once and for all. Historians and archeologists have today a significant expertise in deciphering ancient texts and providing reliable material evidences, including material textual evidences, supporting their claims. Multi-lingual secondary sources and translations of the same artifacts also have a significant role to play, not just historically but also for contemporary practice, as with the digitalization and translation of the Annals of the Joseon Dynasty). We plan to research how established methodologies of historical research can be applied in historical studies that involve using digital data sources, possibly with some significant revisions.

The above primarily concerns storing and preservation of artifactual data such as humanly-generated texts, images and the like. A similar argument can be made with respect to empirical data of any sort obtained via scientific measurements, observations and experiments. It goes without saying that storing of such data needs to be complemented with safeguarding exact technical descriptions of all minute details of how the given data set has been first obtained: this includes the relevant experimental design, experimental methodologies, technical specifications of experimental and observational equipment, the exact dates and locations of experiments. A long-term archiving of such data, however, requires more than that. Namely, it requires to preserve in some appropriate form for the scientific ideas and the reasons for collecting data of this particular type, for making such-and-such particular experiment or observation. Unless such a broader context is made explicit and taken into consideration, a given set of empirical data can be hardly seen as scientifically meaningful and used properly. Reconstruction of such a context after a span of historical time may once again require the expertise of historians able to reconstruct older and possibly outdated scientific theories and ideas related to a given piece of older scientific data . In order to make today's scientific data useful in the future we need to facilitate the work of future historians of science by archiving these data in a transparent form. Thus the expertise of today's historians is just as relevant to the purpose of long-term storage of empirical data as it is relevant to the storage of textual or other artifactual data.

The DDS project will offer a state of the art of the issues pertaining to very long term digital conservation. It will also make an inventory of contemporary answers and potential leads. Moreover, the DDS project will gather a community of people from a large variety of fields interested by these questions. Building such a community with strong interdisciplinary discussions is one central goal of the DDS project.