
Humanities Data Centre (HDC) – Developing Services for Heterogenous Humanities Research Data

Stephan Buddenbohm

stefan.buddenbohm@sub.uni-goettingen.de
Göttingen State and University Library, Germany

Claudia Engelhardt

claudia.engelhardt@sub.uni-goettingen.de
Göttingen State and University Library, Germany

Sven Bingert

sven.bingert@gwdg.de
Gesellschaft für wissenschaftliche Datenverarbeitung
mbH Göttingen, Germany

Research data play a fundamental role in science and the arts and humanities, a large part of them being either digitised or of digital provenance nowadays. Digital research data are fragile, which means that their thorough management and preservation is crucial to prevent the loss of data and information, avoid redundant data collection, enable more efficient research, ensure the verifiability and reproducibility of research results as well as the citability and re-usability of research data.

The management and long-term preservation of digital research data requires an infrastructure that takes into account the digital nature and heterogeneity of the data and the diverse requirements of the research community. Research data centres, such as the Humanities Data Centre (HDC), are key players in this context. They provide a safe and trustworthy place for researchers to deposit their data as well as to search for and get access to previously deposited data. They also act as centres of expertise, offering training and support to their community. The importance of case-specific consultation must not be underestimated as the heterogeneity of digital research data and the wide range of data management tasks across the data life cycle call for extensive support by data experts.

Comprising of a large number of disciplines and employing a variety of different research methods, the arts and humanities produce very heterogeneous research data. A great deal of them are file-based data,

such as text, audio or video files. But there is an increasing number of more complex research data, such as digital editions or visualisation frameworks, that consist of various interwoven layers of different types of data. In these instances, it is often hard or impossible to distinguish between the “primary data” and the software application that is used to process and display the data. While there are more or less established solutions for file-based data (e.g. repositories), the curation and long-term preservation of complex data types still poses a challenge to infrastructure providers.

In view of this background, the HDC developed an initial service portfolio in its design phase that responds to the heterogeneity and complexity of arts and humanities research data by offering consultation and training, repository services for file-based data and the application preservation. The latter is designed to sustain representations of complex research data such as digital editions or visualisation frameworks.

What are the challenges for the research data centre in this? Continuous access to static systems in a developing environment will create problems sooner or later - the systems will eventually become outdated and inaccessible. On the technological level, there are basically two concepts to preserve applications: either virtualisation or emulation of components. While emulation allows the operation of software (such as an operating system) on hardware and software environments it had not been developed for, it costs extra computing power to emulate the required environment, to name only one disadvantage. Virtualisation comes with the clear benefit that the hardware appears as a physical device to the system and to the software modules. The disadvantage is that the virtualised environment must be sufficiently performant directly on the hardware. But for most of the use cases considered here, only standard hardware is being used. Hence the advantage of running more than one virtualised system and the relatively simple management of these systems makes the virtualisation the state of the art approach. The poster illustrates how the application preservation of the HDC uses virtualisation to preserve complex research data applications in a protected environment and ensure their representation, citation and use for a finite period of time.