

Research data centre services for complex software environments in the humanities

Sven Bingert^{a,*} and Stefan Buddenbohm^b

^a *Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Am Faßberg 11, 37077 Göttingen, Germany*

^b *Max-Planck-Institut zur Erforschung multiethnischer und multireligiöser Gesellschaften, Hermann-Föge-Weg 11, 37073 Göttingen, Germany*

Abstract. Research data is – regardless its disciplinary provenance – very heterogeneous in terms of data formats, applied research methods and content. In contrast to publications, a much broader spectrum of data representations has to be considered by research data infrastructures. Within the Humanities Data Centre project (HDC) an initial service portfolio for research data from the humanities has been developed in its design phase (2014–2016). Not only does this service portfolio has to facilitate the re-use of research data by third parties, a service anybody would expect as conventional for a research data centre, but it also has to provide novel benefits for the providers of research data to promote their contribution of data, documentation and support.

As an example for these intended benefits for research data providers we will introduce two service components: the application preservation and the referencing of complex software environments. Jointly applied they enable researchers to archive and reference to complex representations of research data such as digital editions, virtual research environments, or data visualisations.

Keywords: Digital humanities, humanities, research infrastructure, research data, research data centre, service portfolio, complex software environment, application preservation, Humanities Data Centre, HDC

1. Introduction

Progress and transparency in research largely depends on the capability of researchers to cite and reference certain aggregations of research data, instruments, and publications. Regardless what the subject of referencing may be, it is inevitable to sustain access to a stable object representation and to a transparent and proper documentation. For that purpose infrastructures are necessary that serve the discipline specific procedures of referencing and citing [8]. Whereas these standards and infrastructures are quite established and harmonised for (conventional) research publications – a traditional duty of libraries – the field is only developing with regard to research data [4,11].

The challenges related to archiving and referencing these new types of research data will be described in detail in this paper, thereby focusing on complex software environments (CSE) as representations of research data. The proposed solution of the Humanities Data Centre (HDC) revolves around an application preservation as a means to archive and reference software environments at any point of the research process and an adapted persistent identification approach [3,7]) applying fragment identifiers and template handles.

*Corresponding author. E-mail: sven.bingert@gwdg.de.

The following insights depend in a large part on the design phase of the HDC [4]. In addition to the presented preservation and reference solution for CSEs, we will embed these service components in the larger picture, thus the initial service portfolio of the HDC. Partly the service portfolio covers components someone would expect as conventional, for instance a repository or a bitstream preservation service, but partly the services also feature an experimental character. This reflects not only the HDC as a nascent infrastructure but also the evolving standards in the overarching field of research data management.

All things considered the growing prevalence of complex representations of research data was instructing for the conception of the HDC service portfolio. With regard to these kinds of complex representations it is apparent that ingesting and providing these data not only come with considerable more effort than for research publications but also with the challenge of developing new incentives for the researchers to provide their data. Of course the importance of referencing research data will increase and surely is one of the priority incentives but the service portfolio of a research data centre has to have something more to gain usage and become an attractive research infrastructure (RI).

2. What do researchers reference?

The short answer is: everything. There is no limitation regarding the objects of research and therefore every distinguishable object might be addressed, referenced or cited [7]. The range of objects is almost infinite: files in a file system, database entries, web sites, books, places, events, people or journal articles. The range of referable objects is even more expanding since the establishment of semantic web technologies and standards such as Resource Description Framework (RDF). For a considerable number of object classes, solutions in form of services or tools to create and resolve references are available. The most prominent examples are the International Standard Book Number (ISBN) for books or Persistent Identifiers like the Digital Object Identifier (DOI) or Persistent Uniform Resource Locator (PURL) [9] for digital publications, but also unique stable references to people as realised by the Open Researcher and Contributor ID (ORCID) are prevailing. However the problem remains, that a large share of research data does not fit into these classes. Despite this, the heterogeneity with regard to size, format or structure is not helping matters. Thus far for static objects as representations of research data.

But what stands beyond these conventional, static object classes? As said before researchers want to reference a broad spectrum of research data types that do not fit into the static definition or that are only emerging as referenced objects. The development of new content or data types is closely aligned to the development of the researchers' working environment, methods and instruments and for this reason quite difficult to be foreseen for an RI provider. Also the citation of research data fulfils various purposes ranging from impact and reputation for the researcher to transparency and reproducibility of research results. All together this widens the scope of research data management and subsequently of a research data centre regarding its collection policy.

The preceding point underlines the importance of archiving different aggregations of one and the same data set as it allows to reproduce the whole research process. As a consequence not only conventional formats of data should be taken into account for ingesting in RI but also complex software environments, virtual research environments, complex databases, visualisation frameworks, collections, or processed data. As a golden thread stands out the complex structure of these kinds of data. They can have various consecutive layers, aggregations or components (two examples, a visualisation framework and a digital edition are described in detail below). To some extent software is depending on its environment, but also

on its way of usage. Both is evolving over time, also ontologies, terms or references in databases. The state of complex software environments, like an interactive visualisation tool is also highly fluid but is, as we argue, also a research object in itself.

In which way can these classes of objects be referenced? A reference will likely point to a specific fixed state of the object such as a query term or search string and not to the virtual research environment or the database as such.

A workaround could be a reference pointing to a landing page of a research data centre or a repository providing after that in the next step the query term or search string. This can work out as long as a database remains stable. Beyond question is the inconvenience of this procedure as it requires additional steps in the reception process on side of the reader. Therefore a more convenient solution is needed. This solution must provide a citation which includes descriptors of such a quality that the reference can point to clearly one specific data set, ideally integrated in one element – for instance a PID – that just has to be resolved by the reader.

2.1. Complex research data and implications for re-using

As aforementioned these new types of research data come with new challenges for the RI. We group some characteristic research data types under the label complex software environment but this is not to be seen as an elaborated theoretical concept but merely as an manifestation for the complexity of the research data.

What is expressed by CSE and what represents its new quality? Frankly spoken it is not new at all. Outstanding is the multiple layer character of the data, its nature as being composed of various sub-data types that generate an information content reaching beyond the sum of the individual elements. Below we examine two examples for CSEs. These examples also show very plainly how valuable for a representation of a research project this kind of data can be. The benefit in archiving – or better: keeping it accessible – of these CSEs reaches beyond the re-use or transparency aspects of research data but rather is more about generating impact and reputation for the creator of the CSE. Therefore this kind of research data concerns nearly the same reasons that stand behind research publications like journal articles or monographs.

Notwithstanding archiving and re-using separate elements – like text files, databases, scans, documentation – will likely be also subject to data ingests into the research data centre but in this context we deliberate on the options for archiving and referencing of CSEs as cohesive representations of research data.

2.2. Excursus: Visualisation framework

As an example for the above mentioned new object classes we describe hereinafter a visualisation framework, a CSE, because it not only demonstrates the fluid character of the data but also underlines clearly what layers and dependencies potentially have to be taken into account when a CSE is archived and to be kept available for services and re-use. Connected to the ingest of a CSE may be questions of a technical nature (dependencies, granularity of references, accessibility) or legal questions (license status of content components). All of these questions influence the proposed solution. The visualisation framework described below may be seen in this specific context as synonymous for a CSE. Other manifestations of CSEs can be found in digital editions – one example will be detailed below – or virtual research environments. Our example [2] can be shortly characterised as an attractive presentation of

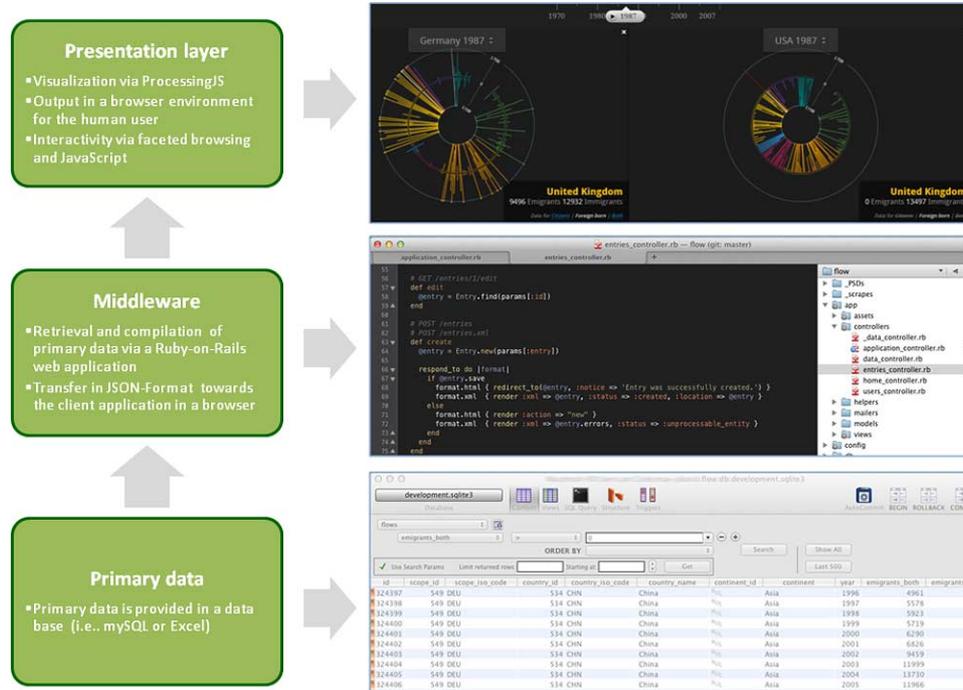


Fig. 1. The multiple layer character of a complex software environment using the example of the visualisation of global migration flows.

research data with interactive components aimed at the human user. Basically it is a database-service visualising the result of a search string.

The illustration in Fig. 1 depicts Global Migration Flows, allowing the user to create individual data sets that visualise migration movements between selected countries and over selected periods of time. The visualisation framework is accessible via a common web browser and is based on data sets from the United Nations Population Division (last accessed on Oct. 20th 2016 <http://www.un.org/en/development/desa/population/theme/international-migration/>) ranging from 1970 to 2011. The data visualisation is closely bound to its presentation environment, therefore an archival solution for this kind of data has to address this multiple layer character to sustain its added value. Usually this kind of data visualisation, based on a browser as access interface, consists at least of three layers (see Fig. 1):

- (1) The normalised and enriched data provided through a database (data layer consisting of primary or processed data), which can be as simple as a spreadsheet or assuming a more complex form of a database.
- (2) A processing layer (middleware) that transfers the normalised and enriched data to the client application of the end user, the web browser. The necessary resources for the development and coding of the processing layer and the following presentation layer cannot be provided in each research project because of spare competences or resources. This advocates the re-use of these kind of application and data for other research projects. In our example the data is retrieved and compiled through a Ruby-on-Rails web application and transferred in a JSON-format towards the client application in a web browser.

- (3) On this rests the user interface, usually a common web browser. The presentation layer is normally out of the scope of action of the researcher but has to be considered if a visualisation application for the end user is aspired. The visualisation framework has no influence on the browsers used by the end user and can only try to cover a range of most common standards in our example applying processing JS and regarding the interactivity with the end user JavaScript.

The above standing remarks allow us to identify two overarching aspects that have to be addressed by any technological solution: the sustainability of the system (cf. Section 3) and the citation of actual states of the system (cf. Section 4). Whereas the latter aspect influences referencing services, the first aspect affects nearly every service as the reproducibility and accessibility of a CSE comes as a basic requirement.

2.3. *Excursus: Digital edition*

As second example for a CSE we describe a digital edition. A digital edition may also be in the reader's eyes closer to the conventional perception of humanities research data. But as we will see later, the challenges for the research data centre coming with the informational and technical structure of the CSE are quite similar to the aforementioned visualisation framework. A digital edition attempts to develop an information science model of the conventional critical edition. Usually it consists of various layers piled up very similar to the example of the visualisation framework. At the bottom the original text serves as data core. All further layers depend on it. After the text is digitised it is usually transformed into an TEI/XML transcript which is employed as main point of reference for the other information layers be they a diplomatic transcript, comments, facsimile or a register.

Figure 2 shows a simplified view of the structure of the digital edition of Theodor Fontanes *Notizbücher* (cf. [5]), a joint research project by the Theodor Fontane Arbeitsstelle of the University of Göttingen and the Göttingen State and University Library. The digital edition can be described along three basic layers:

- (1) The original text and its digital scan serve as base for the TEI/XML transcript.
- (2) The TEI/XML transcript merges not only the visual data from the scan but also includes any other information that haven't been captured by the scan.
- (3) The user gains access to the digital edition through a common web browser. This view presents all data layers of the digital edition, mainly the edited text and the diplomatic transcript but also comments, a register, an apparatus or facsimiles.

Taken all together this allows not only an individual editing of the data by the researcher but also advanced applications, e.g. for a network visualisation or geo browsing. Figure 3 displays the synoptic view on the digital edition as it is presented to the user via his browser. The second illustration (Fig. 4) displays two examples how the data from the digital edition can be used for advanced purposes such as geo browsing or network visualising.

3. Challenges in sustaining complex software environments

Continuous access to static systems in a developing environment will create new problems over time and the systems will be outdated and inaccessible soon [4]. These issues can be divided into technological, legal, financial or social problems. Examples include changes to licenses, the way in which



Fig. 2. Simplified synoptical view of the Theodor Fontane Notizbücher.

Man kann fügen die ganze Stadt befehlt aus der eigentlichen Stadt (alt, unansehnlich) und dem großen Schlosspark in dem nicht nur das Schloß liegt, sondern an dessen Peripherie auch alles andre hübsche gelegen ist: Die ganze Anlage ist groß, in Banlieue-

Fig. 3. Synoptical view of the Theodor Fontane Notizbücher with digital copy, transcript and edited text with apparatus as it is presented in the web browser.

research data are read and interpreted, or a change in the semantics. The non-technological external boundary conditions are not considered here. The technological problems to be expected can be divided into the categories of security, incompatibility and dependencies. Occurring problems regularly result in the inaccessibility of the results, which is not a problem with an accompanying publication. If, however, the system is the research result itself, a long-term availability is necessary. This is in the interest of researchers, funding agencies and research institutions for various reasons such as transparency, reproducibility or re-use of research.

3.1. Approach

There are basically two concepts to tackle hardware compatibility issues: either virtualisation or emulation of components (cf. [12]). While emulation allows the operation of software (such as an OS) on hardware it had not been developed for, it costs extra computing power to emulate the needed hardware environment. A change of the already replaced hardware would cause additional work to re-implement the emulation. Virtualisation comes with the clear benefit that the hardware appears as physical device to the system and to the software modules. It comes with the disadvantage 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 system makes the virtualisation the state of the art approach.

4. Fragment PIDs for citation of unambiguous system states

The necessity to publish data along with a conventional publication or even as a stand alone publication becomes increasingly important. The concept of using Persistent Identifiers (e.g. DOI) for conventional publications is well established, whereas for data publication a different picture is given. This is one of many reasons why PIDs are becoming vital for the data life cycle.

To contribute to the discussion and to fulfil the requirement of referencing and citation of complex software environments we introduce PIDs as our solution [3]. This guarantees that credits and recognition are attributed to the CSE, as a specific research result. Usage of persistent identifiers for references to CSEs additionally allows to create a sustainable archive that, for instance, meets the requirements of the funder. Of course the sustainability depends on the RI that maintains the CSE and the operator of the PID system. We utilise ePIC [10] as PID system, with ePIC PIDs based on Handles [6].

4.1. Persistent identifiers for CSE

Usually PIDs are used to reference an URL. Once a PID is resolved the resolving system can either return the URL or directly forward to it. In our proposed solution the various CSEs can be reached with an URL that consists of the data centre URL with the appending unique connection ID. This comes with the advantage that both PID and connection ID, that must be retained, are under control of the RI.

4.2. Fragment identifiers for CSE

Persistent Identifiers are used in the humanities to identify collections, content or objects, or in our case CSEs. A PID is a alphanumeric number mostly without any semantics. This number can be send to a resolver that returns the referenced resource or object. If the referenced object is e.g. an URL, a Browser would be redirected to the resource. But PIDs not only allow to reference specific objects but may also reference object fragments with the usage of a Fragment PID. This is supported by the Handle infrastructure [6]. By applying this mechanism it is possible to forward information to the resolved object (e.g. URL). These information may be passages of text, illustrations or links to certain sections in digital media by following examples:

<http://hdl.handle.net/21.11101/123456789@page=10>

where <http://hdl.handle.net/> is the resolver, 21.11101/123456789 the PID, @ the delimiter and *page=10* the referenced object, would resolve to

<http://www.domain.org/book1?page=10>

In this example the PID represents a book and the fragment points to a specific page within the book. The system that provides the referenced digital book of course must support this functionality. The naming schema for such a PID like in this example differs between the existing PID systems and is not subject of discussion in this paper. It is important to consider that with this kind of identifiers an unlimited number of fragments (in this example book pages), in an entity can be referenced and provide the level of granularity that is necessary for scientific citations.

5. The humanities data centre as use case for referencing of CSEs

The infrastructure for long-term storage and provision of research data in the humanities is only beginning to emerge and is currently not as progressed as in other domains such as astrophysics (e.g. <http://cdsweb.u-strasbg.fr/>) or climate research (e.g. <https://www.pangaea.de/about/>). Against this background the HDC aims to establish a data centre for humanities research data (cf. Fig. 5).

Therefore one has to keep in mind that the described approach using an application preservation and fragment identifiers is currently only prototypical. Above all hovers the question how to sustain the RI. The layout of the RI not only has to meet the requirements of the researchers but also has to address the conditions and resources available to the RI provider (e.g. in terms of costs, time, knowledge, staff). Accordingly as much as possible re-use suitable and already established solutions and components should be used to compile the initial HDC service portfolio (cf. Fig. 6). The initial service portfolio is designed in order to cover both conventional file based data structures and complex data structures. This is owed to the experience that now a majority of research projects in the humanities create some kind of complex data structures during the course of their research. As detailed above these data structures can be visualisation frameworks, digital editions, database applications or even virtual research environments. Not all of these representations would be in a conventional perspective be seen as research data but nevertheless they are valuable elements to reproduce a research process and are therefore of interest for

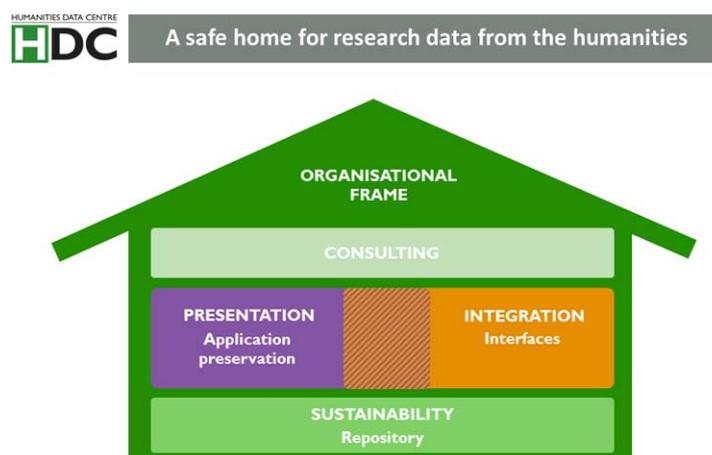


Fig. 5. The HDC seen as stable home for research data.

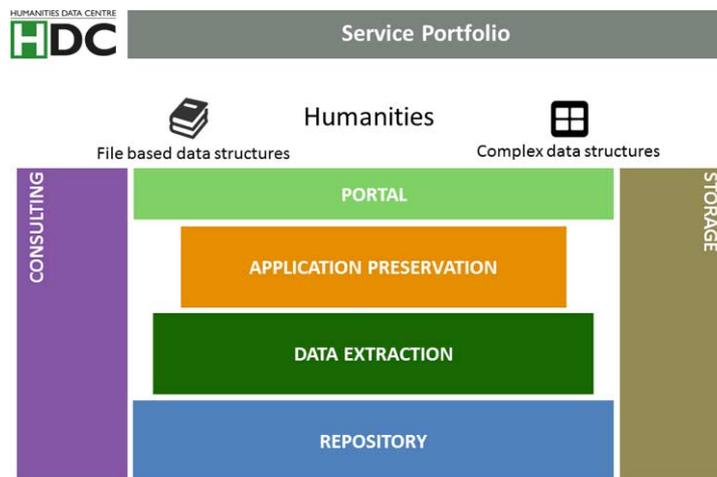


Fig. 6. The HDC service portfolio that is currently under construction.

the research data centre. Another important aspect concerns one of the main incentives for researchers to transfer their data: the opportunity to reference and cite it.

Spoken in service components the HDC's initial portfolio consists of the following services:

Storage The storage layer is a fundamental condition for all other services. Usually it wouldn't appear in a service overview but with look at the long term preservation of research data this component will be developed further to meet the requirements of digital preservation standards. A bottom line service could be to secure research data at bitstream level. The storage layer is not designed to be accessible by users.

Consulting Just as the storage layer, the consulting service functions as enabler for the other services but is also a valuable service in itself. Different than research publications with their limited range of formats, research data comes with an wide array of formats and possible configurations highlighting a higher consulting demand.

Portal Detached from the specific technical implementation of the research data centre, the portal will function as central starting point for users. Whereas application programming interfaces (API) can be used for deploying data to third-party services or converse (e.g. indexing of own content by search engines, harvesting of repository content), the portal is aimed at the human user, permitting the use of the various services.

Application preservation This service enables researchers to archive research data and research results, for instance web applications, in a flexible way for a limited span of time. These applications are stored within a protected environment to allow access despite of security flaws or outdated components. This service component is described below in a more detailed manner to demonstrate exemplary how further services relying on this layer can be implemented, in this case referencing of complex data structures.

Data extraction This is not an independent service but rather describes any curation effort necessary for ingesting research data. This can mean a migration into interoperable formats or the exposure of a "raw data core" from format restrictions to facilitate the re-use.

Repository The repository functions as important component for file-based archiving of research data. The data can be accessed either directly by the user or provided via an API to third party services

like registries for repositories or research data. Different entities for different content types are possible. The DARIAH-DE Repository (cf. <https://de.dariah.eu/repository>) has been chosen as infrastructure partner within the HDC.

For a detailed description of the HDC initial service portfolio consult the project website and the corresponding project report [2]. With regard to referencing of complex software environments (as representations of complex research data) and addressing the above mentioned general challenges of sustainability and referencing we introduce the application preservation as service component below.

5.1. Application preservation

The initial service portfolio of the HDC can on the one hand be described as modular to address complex use cases, and contains on the other hand innovative components such as the application preservation. The application preservation can be used for complex forms of research data that fit into the above exemplified pattern. It guarantees access to a stable version of a component (sustainability) – such as a virtual research environment, a complex database, or a visualisation framework – and enables the user to reference not only the component at large but also individual fixed states (referencing), e.g. to generate a certain query term or a specific visualisation set. Data structures are preserved in their functional handover status and are only technically changed or maintained by the RI to extend its accessibility. Subject of the handover is ideally the complete data structure (e.g. client server structure, dependent libraries and applications). The service focuses on the presentation and reproducibility of research results and methods, not implicitly on the re-usability of the data. It is obvious that preserved applications can only be provided for a limited lapse of time by the RI for reasons of security gaps or outdated components. An ordinary archive case will make use of a combination of the above described services. Nevertheless the application preservation seems as an attractive service as it allows an ample presentation of research compared to archiving of raw data or documentations and therefore demands a solution for referencing which is described in the following section.

5.2. Referencing a CSE employing the application preservation

The architecture of the application preservation prototype as described below is the result of raising requirements by the researchers and of evaluating already available components. Beside its nature as prototype it has also to be seen as a compromise which will improve and be further developed only in practical use. The application preservation offers three ways to access the research result according to the security level. Directly after the project the software is up to date and safe and may be accessed directly (cf. Fig. 7 and option 1). As soon as the software module becomes vulnerable the access will be restricted via an archived browser only (cf. Fig. 7 and option 2 and 3). Basically the application preservation consists of these layers:

- (1) A cloud infrastructure providing the storage and computing capability for the preserved application (its snapshots) and the application itself.
- (2) A set of archived browsers of different brands and versions needed to visualise the application in an optimal way.
- (3) Guacamole [1] serves as bridge between the user environment – typically his browser – and the HDC or RI. It is able to handle RFP/VNC and RDP protocols to enable a remote access from the user to the preserved application without the inconvenience of installing additional software on the user/client site. A simple browser is sufficient to access the research data.

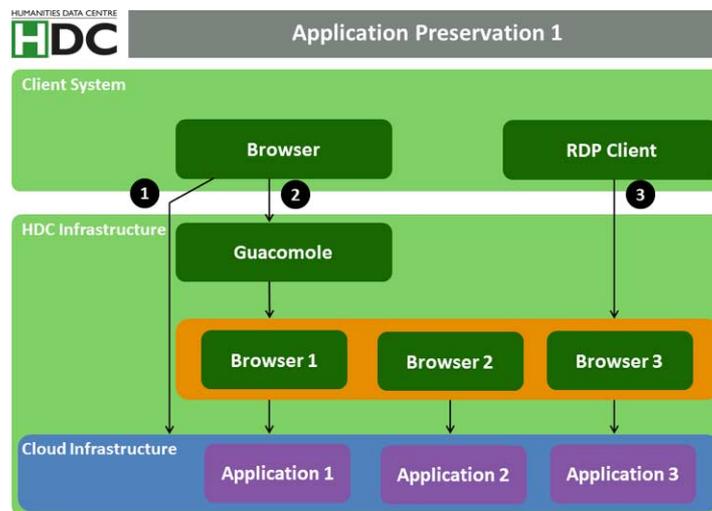


Fig. 7. HDC infrastructure for application preservation. The numbered options depict different ways to access the preserved application.

- (4) PIDs are used to reference the application or a specific system state, e.g. a query term resulting in a specific data visualisation or database result.

The usage of PIDs within the context of the application preservation can be divided into the following access methods:

- (1) A PID pointing directly to the URL of the visualisation (cf. option 1 in Fig. 7). This is the simplest way of referencing but the PID will become unresolvable quickly due to the reason of the application becomes outdated (cf. Section 3 on challenges related to referencing).
- (2) A PID pointing at a specific connection configured in the guacomole client (cf. option 2 in Fig. 7). This allows long term access to the application stored in a secure environment. The downside is an enhanced PID management effort to assure the reference to be stable.
- (3) In case of the third access method (option 3 in Fig. 7) a PID usage is not possible due to a configuration of the RDP client on the user system. It still can be used to create a snapshot of a system state that later can be restored and referenced by a PID.

We also need to distinguish between PIDs generated to reference the software model (the research result) or the HDC/RI infrastructure. In the first case the software offers the service of citable states directly, e.g. by a URL search string. But if this functionality is missing the problem becomes more complex. The user queries or browser interactions have to be monitored and used to recover the desired state. Together with these information the full state of the virtual environment has to be archived (cf. Fig. 8) which will then be referenced by PIDs that may also activate the system via the fragment mechanism. This functionality is only possible for option 2 in Fig. 7.

6. Conclusion

This paper introduced a technological solution for the problem of referencing complex software environments along a presentation of the service portfolio of the recently founded Humanities Data Centre, a RI for humanities research data.

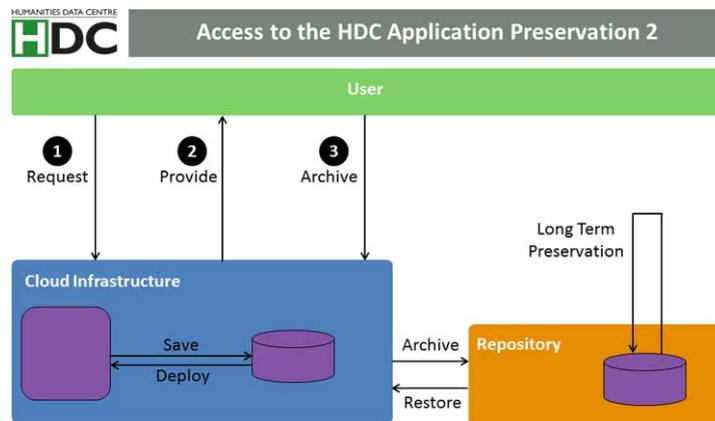


Fig. 8. HDC cloud infrastructure for application preservation.

CSEs in the context of this article serve as exemplary representation for the emerging new forms of complex research data, not only in the humanities. A data visualisation framework and a digital edition have been introduced as individual examples for CSEs. There are other classes of new research data but complex software environments are particularly suitable to demonstrate the challenges related to referencing this kind of research data. There is a growing demand by researchers for archiving and referencing these forms of research data as they allow an ample view on a research project. The referencing of CSEs could become prospectively a substantial element of the scientific impact and reputation of a researcher, therefore RIs have to develop solutions for this.

We continued with a description of the main characteristics of research data in the humanities, focusing on the increasingly complex characteristics of data, which pose new requirements for referencing compared to the long established referencing of conventional types of content, e.g. text and formats, e.g. monographs and journal articles.

Following this, the main challenges in referencing complex forms of research data have been outlined: first ensuring the sustainability of actual software environments and second enabling to reference specific system states, such as a search string or query term in a database to ensure the desired granularity.

At this point it became clear that these new forms of research data pose crucial challenges for archiving and referencing and therefore for RI providers also in other, non-technological fields, e.g. legal issues [8]). For instance the dependencies of CSEs regarding external libraries or the license status of raw data require coverage on side of the RI.

A prototypical solution – the HDC application preservation – was introduced as a technological approach for referencing CSEs. Our approach mainly consists of a cloud infrastructure allowing remote access of users to preserved applications via an easy to use additional access layer. The problem of pointing to a specific system state – such as a database query term or a specific visualisation data set – is in our case solved via Fragment PIDs. Depending on the application Fragment PIDs are used to forward URL extensions, resolve into connections to specific applications or contain pointers that activate the CSE in exactly the state created as a snapshot by the editor of the reference. Although the paper remained on a technological tier, there are challenges to be addressed reaching beyond such as legal, financial, interoperability or organisational tiers. These challenges also have to be addressed by the RI provider but cannot be solved by him alone. Beyond research data centres stand libraries and data centres, in the case of the HDC the Göttingen State and University Library (SUB) and the Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) as additional main providers of RI.

References

- [1] Apache Guacamole (incubating), 2016, [Guacamole.incubator.apache.org](http://guacamole.incubator.apache.org), Retrieved on October 20th 2016 from <http://guacamole.incubator.apache.org/>.
- [2] A. Aschenbrenner, S. Buddenbohm, C. Engelhardt and U. Wuttke, *Humanities Data Centre – Angebote und Abläufe für ein geisteswissenschaftliches Forschungsdatenzentrum*, HDC-Projektbericht Nr. 1, 2015, PURL: [gs-1/13740](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-13740).
- [3] S. Bingert, S. Buddenbohm and D. Kurzawe, *Referencing of Complex Software Environments as Representations of Research Data*, Proceedings of the 20th International Conference on Electronic Publishing, Positioning and Power in Academic Publishing, Players 2016, Agents and Agendas, IOS Press, 2016. doi:[10.3233/978-1-61499-649-1-58](https://doi.org/10.3233/978-1-61499-649-1-58).
- [4] S. Buddenbohm, C. Engelhardt and U. Wuttke, Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum, *Zeitschrift für digitale Geisteswissenschaften* (2016). doi:[10.17175/2016_003](https://doi.org/10.17175/2016_003).
- [5] T. Fontane, Notizbücher, Digitale genetisch-kritische und kommentierte Edition, G. Radecke, ed., Version 0.1 vom 7.12.2015, as of 28.09.2016.
- [6] Handle.Net Registry, 2016, [Handle.net](http://www.handle.net), Retrived on October 20th 2016 from <http://www.handle.net>.
- [7] T. Kalman, European Persistent Identifiers Consortium – PIDs für die Wissenschaft, in: *Langzeitarchivierung Von Forschungsdaten – Standards und Disziplinspezifische Lösungen*, R. Altenhöner and C. Oellers, eds, pp. 151–168.
- [8] H. Neuroth and A. Rapp, Nachhaltigkeit von digitalen Forschungsinfrastrukturen, *BIBLIOTHEK Forschung und Praxis* **40**(2) (2016), 264–270. doi:[10.1515/bfp-2016-0022](https://doi.org/10.1515/bfp-2016-0022).
- [9] N. Paskin, Digital Object Identifier (DOI) system, in: *Encyclopedia of Library and Information Sciences*, 3rd edn, Taylor and Francis, pp. 1586–1592.
- [10] Persistent Identifiers for eResearch Consortium (ePIC), 2016, Retrieved on October, 20th 2016 from <https://www.pidconsortium.eu>.
- [11] P. Sahle and S. Kronenwett, Jenseits der Daten, Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner “Data Center for the Humanities”, *Libreas* **23** (2013), URN:[nbn:de:kobv:11-1002127269](https://nbn-resolving.org/urn:nbn:de:kobv:11-1002127269).
- [12] J. van der Hoeven and H. van Wijngaarden, Modular emulation as a long-term preservation strategy for digital objects, in: *Research and Advanced Technology for Digital Libraries, 9th European Conference, ECDL 2005*, Vienna, Austria, September 18–23, 2005, Proceedings, Lecture Notes in Computer Science, Vol. 3652, 2005.