## **Digital Humanities Strategies in Transcultural Studies**

Matthias Arnold,<sup>1</sup> Eric Decker,<sup>1</sup> Armin Volkmann<sup>2</sup>

#### Abstract

In this article the authors discuss the Digital Humanities strategies of Heidelberg University's interdisciplinary Cluster of Excellence "Asia and Europe". It describes the structures, methods, workflows and tools that have been established in Transcultural Studies research to profit from Digital Humanities research. It introduces a schematic workflow of typical humanities research within the Cluster's digital research environment developed by the Heidelberg Research Architecture (HRA). The authors discuss challenges and solutions on the way towards sustainable digital research and education in the field of Transcultural Studies.

#### Introduction

The Cluster of Excellence "Asia and Europe in a Global Context: The Dynamics of Transculturality" at Heidelberg University bundles the rich expertise in Asian studies developed within Heidelberg University with the broad disciplinary scope of humanities and social sciences. It institutionalizes interdisciplinary research and strengthens the transcultural focus in interdisciplinary collaborations. It established ways to enhance productivity of collaborative scholarship, make collaboration easier at a larger scale, and to provide easy access to expert knowledge of other fields in all forms of scholarly exchange. A central challenge was to enable this kind of collaboration and to prove that it leads to sustainable results. To be able to collaborate and to make research more efficient researchers need common workflows, and an environment to share material, for example a digital repository.

From the outset, the Cluster established a technical core unit to support information technology in the humanities and to foster and enhance digital scholarship on an institutional level: the Heidelberg Research Architecture (HRA). It was later reorganised into two separate but closely collaborating units, IT-support and Digital Humanities (DH). Since then, HRA refers to the DH unit at the Cluster. The HRA team, consisting of members with humanities and computer science backgrounds, often serves as the "digital partner" for the Cluster's digital humanities projects. The HRA provides counselling, training and a digital

<sup>&</sup>lt;sup>1</sup> Heidelberg University, Heidelberg Research Architecture

<sup>&</sup>lt;sup>2</sup> Junior Research Group Digital Humanities and Cultural Heritage

research infrastructure.<sup>3</sup> In terms of infrastructure the HRA adopts and implements existing software solutions, tailors these according to the need of research projects, or —if necessary— develops new open source solutions based on international standards. Where possible, the HRA invests in the enhancement of already existing systems.

Since the early 2000s, Digital Humanities has established itself as an independent discipline.<sup>4</sup> Therefore, the cluster established a Junior Research Group Digital Humanities<sup>5</sup> to bridge the gap between the Humanities and Computer Science. The range of involved Humanity disciplines within DH includes studies of almost all disciplines, such as Language, Literature and Linguistic Studies, as well as cultural orientated disciplines (e.g. like History, Archaeology, Ethnology, contemporary cultural studies and art disciplines), but also extending to the Transcultural Studies.

A key task of DH is to analyse the issues concerning disciplines of the humanities with new digital procedures and methods, and thereby generate new research approaches for innovative and interdisciplinary theorization. By confronting humanistic research issues with technically orientated solutions from Computer Science, the innovative potential for research increases in accordance to the focus and advancement of digital investment within the humanities. On the other hand, in the field of Computer Science digital methods can be applied and reviewed based on humanistic research questions. Ideally this leads to questioning and reviewing methodical and theoretical foundations of the involved disciplines. In the same way, the Transcultural Studies aim to transfer from one discipline to another in a close and reciprocal dialogue.

A common trait of Digital Humanities and Transcultural Studies have in common is the interdisciplinary approach to research questions, where different disciplines with different methodological approaches collaborate asynchronously with internationally dispersed researchers. However, in Transcultural Studies, research topics are more concerned with cultural borders, or "contact zones", where classic definitions of culture, language, terminology are less stable, get mixed or develop different hues. This makes it harder to

<sup>&</sup>lt;sup>3</sup> Ministerium für Wissenschaft, Forschung und Kunst, "E-Science: Wissenschaft unter neuen Rahmenbedingungen. Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg," May 2014, fig. 8, https://idw-online.de/de/attachmentdata37340.pdf.

<sup>&</sup>lt;sup>4</sup> E.g. journal "Digital Scholarship in the Humanities," accessed September 6, 2016, http://dsh.oxfordjournals.org/; or Patrick Sahle, "DH Studieren! Auf dem Weg zu einem Kern- und Referenzcurriculum der Digital Humanities," *DARIAH-DE Working Papers*, DARIAH-DE Working Papers, no. 1 (2013): 39.

<sup>&</sup>lt;sup>5</sup> "Cluster Asia and Europe: JRG Digital Humanities with Focus on 'Archaeological Information Systems' and 'Digital Cultural Heritage,'" accessed September 13, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/jrg/jrg-digital-humanities.html.

precisely describe them in digital environments or analyse them by computational methods. Multilingual data adds just one more facet of complexity. And overall, in the involved disciplines we still have many differences, focused on different research areas, research approaches, research methodologies. The most subjects don't have any guidelines for standards in the application or routines in the procedure of digitally-based studies. To overcome this, the HRA is developing re-usable solutions for Transcultural Studies, which are close to the DH.

In recent years many efforts where done in order to establish the DH at Heidelberg University.<sup>6</sup> The HRA is working on a strategy for digital research infrastructure in the context of Transcultural Studies and other international partner disciplines. In order to do so, HRA conceptualised so called life cycles for the specific needs of heterogeneous projects. This process made very clear that different individual solutions in the field of Transcultural Studies based on the generic schema are needed.

It also became clear that new forms of research-based and collaborative teaching are needed to educate a new generation of students. It is important to guide and train them in using digital sources, both as part of a modern research-oriented education, and as part of a sustainability stragtegy. A number of courses were already held at the Cluster to provide insights in digital key qualification with DH-theories, methods, and tool application.<sup>7</sup>

#### DH challenges specific to Transcultural Studies

At the Cluster of Excellence "Asia and Europe" scholars from various humanities disciplines are collaborating in about 60 research projects. While most of them are using computers to work on their material, some are applying more advanced computational methods and apply methods from the Digital Humanities. These DH projects collect, digitize, and organize different types of research material throughout the whole lifetime of the project. Interdisciplinary groups of scholars and students collaboratively annotate and analyse images, texts, video and audio material. Collaboration<sup>8</sup> often includes scholars from partner institutions on a national and international level.

In the following section, we will take a closer look on the challenges met by DH projects within Transcultural Studies at Heidelberg's Cluster "Asia and Europe". The illustration of a

<sup>&</sup>lt;sup>6</sup> See Armin Volkmann, "White Paper der Digital Humanities an der Universität Heidelberg" (Heidelberg, February 28, 2015), 4–6, urn:nbn:de:bsz:16-heidok-188407.

<sup>&</sup>lt;sup>7</sup> For example the lecture and seminar "Introduction to Digital Humanities", Armin Volkmann, summer term 2015, cf. http://lsf.uni-

heidelberg.de/qisserver/rds?state=verpublish&status=init&vmfile=no&publishid=207651&moduleCall= webInfo&publishConfFile=webInfo&publishSubDir=veranstaltung.

<sup>&</sup>lt;sup>8</sup> For an excellent recent discussion of different kinds of collaboration in the Digital Humanities see Jennifer Edmond, "Collaboration and Infrastructure," in *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture (Chichester, West Sussex; Malden, MA; Oxford: Wiley Blackwell, 2016), 54–65.

prototypical research workflow will help to identify and discuss the typical challenges of DH projects.

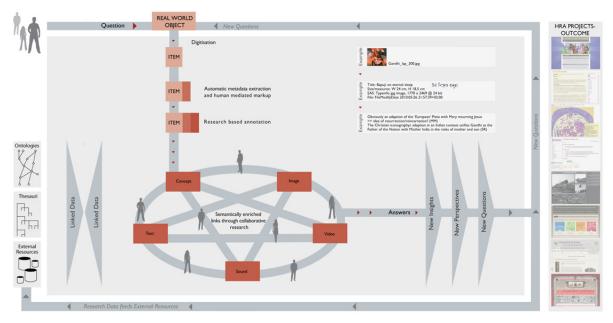


Figure 1: Schematic workflow: From research question through digitization, metadata enrichment and collaborative research to new insights, new perspectives, and new questions, the DH accompanies all stages of the research process.

The schema starts with a set of questions with which a group of scholars approach a collection of research material. During the process of data collection within the digital research environment, they organize the resources and digitize analogue material. In the next phase they analyse and discuss the material, and finally compile the result in a form of publication that stimulates new research questions. At the Cluster, the HRA accompanies the process shown above during all project stages.

Although only a simplified description, it already indicates several challenges that projects face when they apply DH methods in the context of an interdisciplinary transcultural research institution.

#### The conceptual phase - research question and transcultural approach

The first challenge is the formulation of research questions, which are relevant to the researchers involved and help define the project design. In a typical humanities project, these questions evolve while the project is carried out. Often, the core research issues are reformulated, expanded, and refined several times. Answers, and knowledge gained through research, are often multifaceted, context dependent, and explained in rich vocabulary. The "vagueness" of complex natural language used in humanities research results and their interpretation (Hermeneutics) is often conflicting with the demand for precise, structured and machine-readable datasets in digital applications. As a result, research questions often need

to be further refined and split into several smaller aspects that can be handled more precisely and open up the possibility of additional quantitative and qualitative analysis. The constant dialogue between HRA as experts for DH projects and the project members as "domain experts" proved to be most fruitful. This collaborative effort helps to develop realistic expectations of what can be achieved within the project's budget and time frame.

#### The collection phase - data processing and transcultural survey

The main challenge during the collection phase of any research project in Transcultural Studies lies in the broad spectrum of disciplines involved, the heterogeneous character of the data and the resulting diversity of media types and languages. Nowadays it is common practice for humanities scholars to have some sort of personal digital data and resource collections. Typically, folder names are used for expressing personal organization criteria — sometimes with a disciplinary flavour—, or full bibliographic entries constitute a file name. These are, however, unfavourable preconditions for collaborative interdisciplinary research, especially in DH projects. To prevent interdisciplinary misunderstandings and to set a common basis for research on shared collections the HRA fosters the development of a common terminology for basic concepts, the use of controlled authoritative vocabularies, and international open metadata standards, especially for encoding multilingual metadata (language, script, transliteration schema) and analyzing complex data.<sup>9</sup>

For the creation and transformation of standardized metadata the HRA provides training, a number of research tools, and the digital infrastructure *Tamboti*.<sup>10</sup> *Tamboti* is the online metadata framework the HRA developed specifically to manage the data collections of Cluster projects. Its modular plugin architecture is based on eXist-DB,<sup>11</sup> an open-source database for data encoded in the Extensible Markup Language (XML).<sup>12</sup> During the collection phase projects can make use of the MODS-Editor for bibliographic data,<sup>13</sup> the *Ziziphus* VRA Core 4 Editor for the "description of works of visual culture as well as the images that document them,"<sup>14</sup> and the *CSV2XML* app for the transformation of flat tabular

<sup>&</sup>lt;sup>9</sup> For a detailed explanation of national metadata standards see "DFG Practical Guidelines on Digitisation" (Deutsche Forschungsgemeinschaft (DFG), February 2013), http://www.dfg.de/formulare/12 151/index.jsp.

<sup>&</sup>lt;sup>10</sup> For the Heidelberg Instance see "Tamboti Metadata Framework," accessed September 7, 2016, http://tamboti.uni-hd.de/ The open source software is published under an Apache-2-License and available on GitHub: https://github.com/exc-asia-and-europe/tamboti, accessed September 7, 2016. <sup>11</sup> "eXistdb - The Open Source Native XML Database," accessed September 7, 2016, http://existdb.org.

 <sup>&</sup>lt;sup>12</sup> "Extensible Markup Language (XML)," accessed September 7, 2016, https://www.w3.org/XML/.
 <sup>13</sup> The Metadata Object Description Schema (MODS) standard is an XML "schema for a bibliographic element set that may be used (...) particularly for library applications." It is maintained by the Library of Congress "Metadata Object Description Schema: MODS (Library of Congress)," accessed September 7, 2016, http://www.loc.gov/standards/mods/; Source code of the MODS editor is published under an Apache-2-License and available on GitHub "Exc-Asia-and-Europe/mods-Editor," accessed September 7, 2016, https://github.com/exc-asia-and-europe/mods-editor.
 <sup>14</sup> The VRA Core data standard was first developed by the Visual Resources Association in 1996 and

<sup>&</sup>lt;sup>14</sup> The VRA Core data standard was first developed by the Visual Resources Association in 1996 and its current XML version VRA Core 4 was released in 2007. It is hosted at the Library of Congress "VRA CORE - a Data Standard for the Description of Works of Visual Culture: Official Web Site

data into structured XML<sup>15</sup> to create standardized metadata records. All records are organized and shared in Tamboti. It is also possible to synchronize bibliographic records created with Zotero<sup>16</sup> with *Tamboti* via an API.<sup>17</sup>

Simple Search				
Simple Search	Advanced Search Query History My List	Tamboti Collections	About Tamboti Sea	arch Help
Search for record	Is in MODS or TEI or VRA or Wiki 🔹 format, using the or 💌 search operator, searching	for Image red	ords (VRA	core
and • krishr	in the Name Field (MODS, TEI, VRA)	Bibliographic		
inside /reso				
	• descending • Video annotatio			
+ 4	Add Search Field Clear Search Search	lext	annotation	(IEI
	ecords 145 to 154  >   >  of 155 records found in /resources			
	Krishna (Depicted): Archibald Orr Ewing & Co. (Client): Textile label of Archd. Orr Ewing & C		Title Words	
145 🖂 📟	Reside a contraction of the start of the sta	0.	Names	
146 🗐 🔲	Krishna (Depicted); Leela (Depicted): Krishna Leela: Krishna plays flute while girls enjoy			
		$ \longrightarrow $	<ul> <li>Dates</li> </ul>	
147 🛱 📘	Meera Shyam, contributor Manmohan Krishna N/A		<ul> <li>Subjects</li> </ul>	
148 🛱 🛄	The Hare Krishna Movement. Lifting of Govardhana Hill. The Hare Krishna Movement2011-10	0-27	<ul> <li>Languages</li> </ul>	
149 🖺 📕	Zimbo, contributor Homi Wadia, J. B. H. Wadia, Chand Pandit, Chitra, Azaad Irani, Krishna I N/A	Kumari, and Achala Sachdev	Genres	
150 🖪 📕	Nidimoru, Raj and D.K. Krishna, Flavors, performed by Reef Karim, Anjan Srivastav, Bharati /	Achrekar, and Leslie Lewis.		
C A	USA: Net Effect, 2004.			
	Vamsi, Krishna. Shakti: the power (शक्ती), produced by Boney Kapoor and Sridevi, performed	d by Aishwariya Rai, Shah		
151 🖪 📕	Rukh Khan, and Karisma Kapoor. India: Sridevi Productions, 2002.			
151 🖪 🗎				
151 🖪 🗎	Ray, Satyajit. Satyaji ta Rāye'ra Prati dvandvī, produced by Priya Films, performed by Indirā D	Debī Caudhurāņī, Bhaskar		
152 🖹 🗎		Debī Caudhurāņī, Bhaskar		
- 4	Ray, Satyajit. Satyaji <i>ta Ráye'ra Prati dvandví</i> , produced by Priya Films, performed by Indirá D Roy Chowdhury, and Krishna Bose. India: Shradha Home Video, 2005. The Encyclopædia Britannica. 11th edition. Cambridge: University Press New York, 1910-11.	$\longrightarrow$		
152 🖹 📕	Ray, Satyajit. Satyaji <i>ta Rāye'ra Prati dvandvī</i> , produced by Priya Films, performed by Indirā D Roy Chowdhury, and Krishna Bose. India: Shradha Home Video, 2005.	dern period has been so		

Figure 2: Screenshot of a search for "Krishna" in Tamboti showing image, bibliographic, video annotation and text records on a single result page.

(Library of Congress)," accessed September 7, 2016, https://www.loc.gov/standards/vracore/; For more information see VRA Core Oversight Committee, "VRA Core 4.0 Introduction" (Visual Resources Association, October 28, 2014), https://www.loc.gov/standards/vracore/VRA\_Core4\_Intro.pdf; Ziziphus is an open-source VRA Core 4 metadata editor. It is web-based and applies forms for editing and storing metadata directly in XML. Ziziphus is built on eXist-db and operates with the Tamboti metadata framework. Source code is published under an Apache-2-License and available on GitHub "Exc-Asia-and-Europe/ziziphus," accessed September 7, 2016, https://github.com/exc-asia-and-europe/ziziphus; It is implemented in the Heidelberg Tamboti instance "Tamboti Metadata Framework"; For more information see "Ziziphus Help," accessed September 7, 2016, http://ziziphus-help.uni-hd.de/.

<sup>15</sup> CSV2XML is a tool to generate XML files of any XML schema by using mappings and templates. Its first use case is the "VRA Core 4 XML Transform Tool," which converts descriptive work and image metadata from flat tables (.csv) to structured VRA Core 4 XML. Source code is available on GitHub "Exc-Asia-and-europe/csv2xml," accessed September 7, 2016, https://github.com/exc-asia-and-europe/csv2xml; For more information see Matthias Arnold, "Transform Tool: From Flat Data to Structured XML and RDF," Blog, *Yet Another DH Blog*, (April 5, 2016), http://hg7.blog.uni-heidelberg.de/2016/04/transform-tool-from-flat-data-to-structured-xml-and-rdf/.

<sup>16</sup> Zotero is a project of the Roy Rosenzweig Center for History and New Media, cf. "Zotero," accessed September 7, 2016, https://www.zotero.org/.

<sup>17</sup> For more information about application programming interfaces see, for example, "What Are APIs and How Do They Work?," *ProgrammableWeb*, accessed September 13, 2016, http://www.programmablow.ob.com/api.university/what are apis and how do they work

http://www.programmableweb.com/api-university/what-are-apis-and-how-do-they-work.

In addition, the HRA runs a small medialab providing an infrastructure for digitization and guidance for skills that are not yet part of the Transcultural Studies curriculum.<sup>18</sup>

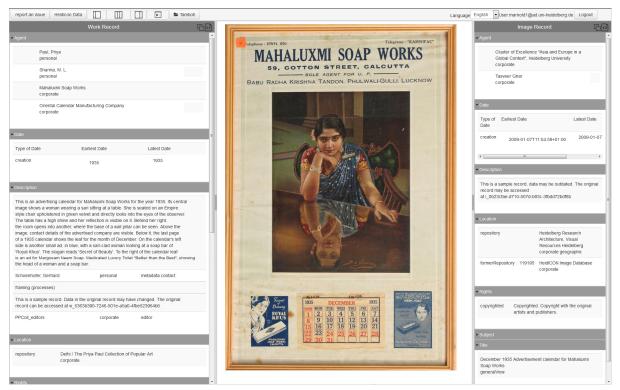


Figure 3: Interface of the Ziziphus image metadata editor, showing an Advertisement calendar for Mahaluxmi Soap Works, 1935. The Priya Paul Collection of Popular Art. <u>Stable link</u>.

In Transcultural Studies the collection phase is often an essential part of a research project. In contrast to the recent trend of funding cuts for projects mainly concerned with digitization and data collection there is still a need for the creation of digital collections in this field. This is mainly because researchers are often working with material that was either little studied, unpublished, or hitherto unknown.

Besides establishing collections and systematically enriching items with standardized metadata—for example, large controlled vocabularies like the Getty Thesauri are implemented into the user interface of the *Ziziphus* editor—the HRA is also actively involved in the enhancement and expansion of authoritative vocabularies. For example, new records were added to the German Integrated Authority File ("Gemeinsame Normdatei", GND)<sup>19</sup> based on research and enriching metadata records in the Priya Paul Collection<sup>20</sup> via

<sup>&</sup>lt;sup>18</sup> See note 5 above.

<sup>&</sup>lt;sup>19</sup> "DNB, Katalog der Deutschen Nationalbibliothek," accessed September 7, 2016, https://portal.dnb.de.

<sup>&</sup>lt;sup>20</sup> For more information see "The Priya Paul Collection of Popular Art," accessed September 7, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/heidelberg-research-

architecture/databases/transcultural-image-database/collections/priya-paul-collection.html; The collection is freely available at "Tamboti Metadata Framework - Priya Paul Collection," accessed September 7, 2016, http://kjc-sv016.kjc.uni-

workflows established in collaboration with Heidelberg's University Library.<sup>21</sup> One aspect of the "Early Chinese Periodicals Online (ECPO)" project<sup>22</sup> is a collaboration with various institutes at the Academia Sinica in Taiwan.<sup>23</sup> Together with the Academia Sinica Digital Centre, ECPO is mapping its keywords to the Chinese translation of Getty's Art and Architecture Thesaurus (AAT).<sup>24</sup> At the same time, ECPO contributes specific terms related to Chinese history, periodicals studies, and women studies.

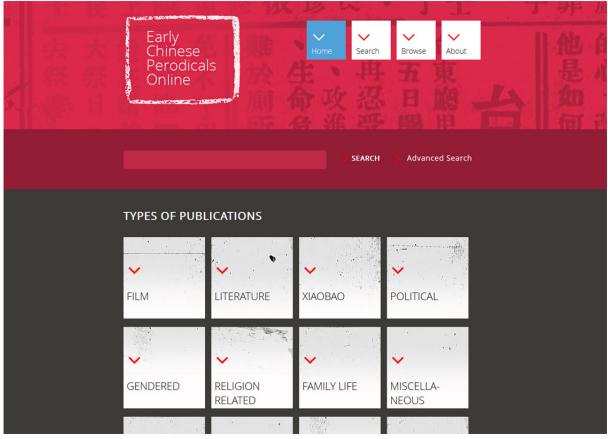


Figure 4: Home page of the Early Chinese Periodicals Online (ECPO) database, http://ecpo.uni-hd.de.

heidelberg.de:8080/exist/apps/tamboti/modules/search/index.html?collection=/resources/commons/Pr iya\_Paul\_Collection.

<sup>21</sup> "Universitätsbibliothek Heidelberg," accessed September 7, 2016, http://www.ub.uni-heidelberg.de/.
<sup>22</sup> The project website was launched in Summer 2016: "Early Chinese Periodicals Online (ECPO)," accessed September 7, 2016, http://ecpo.uni-hd.de/; ECPO is building up on and significantly enhancing two earlier initiatives, namely "Chinese Entertainment Newspapers (Xiaobao)," accessed September 7, 2016, http://xiaobao.uni-hd.de; and "Chinese Women's Magazines in the Late Qing and Early Republican Period," accessed September 7, 2016, http://womag.uni-hd.de; For the latter see Doris Sung, Liying Sun, and Matthias Arnold, "The Birth of a Database of Historical Periodicals: Chinese Women's Magazines in the Late Qing and Early Republican Period," *Tulsa Studies in Women's Literature* 33, no. 2 (2014): 227–37, doi:10.1353/tsw.2014.0004.

<sup>23</sup> "Academia Sinica," accessed September 7, 2016, https://www.sinica.edu.tw/main\_e.shtml.
 <sup>24</sup> The Art and Architecture Thesaurus (AAT) is one of the three published vocabularies edited and maintained by the Getty Research Institute "Getty Research Institute (GRI)," accessed September 7, 2016, http://www.getty.edu/research/; The English version is available at "Art & Architecture Thesaurus (Getty Research Institute)," accessed September 7, 2016,

https://www.getty.edu/research/tools/vocabularies/aat/; its Chinese translation can be accessed at

"AAT-Taiwan 藝術與建築索引典," accessed September 7, 2016, http://aat.teldap.tw/index.php.

The discussion phase - information aggregation, contextualization, and transcultural dialog

With the use of defined vocabularies inter-disciplinary projects can establish a common working language which enables collaborative analysis and curation of research material even with individual team members working in various locations. Digital annotation tools offer an effective way of working with research material and discussing it in detail.<sup>25</sup> Systems for image and video annotation like HyperImage<sup>26</sup> and Pandora<sup>27</sup> allow groups of researchers to collaboratively annotate shared resources online. These platforms feature asynchronous discussions and provide means to reference parts of objects, like details of images or sequences in films.

The challenge of new tools lies in establishing adequate workflows and designing them in accordance with the researchers' needs. For most members of Cluster projects the Digital Humanities aspect was new, especially working collaboratively in an online research environment.<sup>28</sup> Therefore, the HRA engages in the conceptual design of the DH projects, helps developing project specific workflows and provides training and documentation. These outcomes are continuously enhanced to make sure subsequent projects can profit from the lessons learned.

The two main tools for annotation and discussion of image and video material currently in use at the Cluster are HyperImage and Pandora.

HyperImage is a system that allows registered users to annotate and interlink specific regions of images and to combine multiple image and text elements on so-called "lighttables". Annotations are stored in a standardized XML format. It allows the creation of linear and non-linear narrations with network-like argumentation structures. The result can be published on the internet. Several HyperImage projects have been realized at the Cluster in research and teaching. The latest is "Hachiman Digital Handscrolls" where seven different

<sup>25</sup> The development and implementation of tools to support the practical use of research anotations is an important aspect of HRA work. For a recent discussion of the different aspects of annotations see Domenico Fiormonte, Teresa Numerico, and Francesca Tomasi, *The Digital Humanist: A Critical Inquiry*, trans. Desmond Schmidt and Christopher Ferguson, First published, Digital Humanities / Global Computing (Brooklyn, NY: punctum books, 2015), chap. 4.3 "Markup standards and languages" and 4.8 "Text analysis and text mining"; The team is closely connected with the "DARIAH-DE Cluster 6: Annotieren, analysieren, visualisieren," accessed September 7, 2016, https://wiki.de.dariah.eu/display/publicde/Cluster+6%3A+Annotieren%2C+analysieren%2C+visualisier en; and actively involved in discussion groups, cf. Michael Bender et al., "Wissenschaftliche Annotationen: Formen – Funktionen – Anforderungen," Blog, *DHd Blog*, (July 17, 2015), http://dhdblog.org/?p=5388; and the working group "Wissenschaftliche Annotationstechniken," accessed September 7, 2016, http://wissann.de/ of the Society for Informatics; The main standard in use is "Web Annotations", cf. "W3C Web Annotation Working Group," accessed September 7, 2016, https://www.w3.org/annotation/; and "Web Annotation Data Model," accessed September 7, 2016, https://www.w3.org/TR/annotation-model/.

<sup>&</sup>lt;sup>26</sup> The open-source HyperImage Virtual Research Environment supports the linking of (audio)-visual objects, texts and mixed-media documents, cf. "HyperImage," accessed September 7, 2016, http://hyperimage.ws/en/.

 <sup>&</sup>lt;sup>27</sup> Pandora is an open-source media archive platform to collaboratively create metadata and time-based annotations to video material, cf. "Pan.do/ra," accessed September 7, 2016, http://pan.do/ra.
 <sup>28</sup> On Collaboration see note 6 above.

medieval Japanese handscrolls all showing the Hachiman legend were densely annotated.<sup>29</sup> The approach has later been adopted for use in class where students collaboratively annotated an eighth Hachiman handscroll over the course of a semester. Based on the requirements formulated by projects using the software in Transcultural Studies, the HRA, together with bitGilde,<sup>30</sup> who maintain HyperImage, develops and implements new features and functionalities. These become part of the open-source software and can be used by subsequent projects.

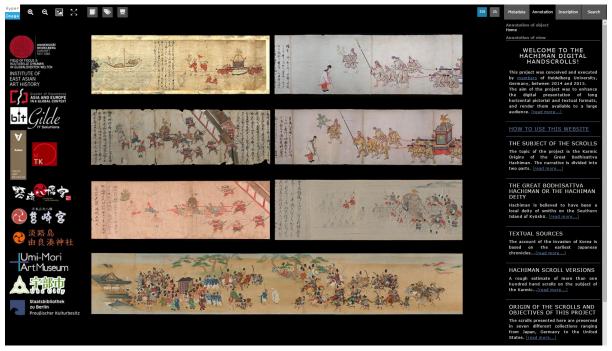


Figure 5: Front-page of the HyperImage publication of the Hachiman Digital Handscrolls project, <u>http://www.zo.uni-heidelberg.de/iko/hdh/</u>.

With the video annotation database Pandora the HRA introduced another digital tool to Transcultural Studies research and teaching in Heidelberg. Pandora allows registered users to add time-based annotations to a video on multiple annotation layers, like place, keyword, description, transcription, subtitle etc. Annotations can relate to the whole film or just a segment, from individual frames to full scenes, and all statements get an author attribution and a timestamp. Since some HTML5 Tags<sup>31</sup> are allowed in annotations, videos cannot only

<sup>&</sup>lt;sup>29</sup> For the project see "Hachiman Digital Handscrolls," accessed September 7, 2016,

http://www.zo.uni-heidelberg.de/iko/hdh/ The online publication is available at http://hachiman.unihd.de; See also Jens-Martin Loebel et al., "Hachiman Digital Handscrolls: Semantische Anreicherung mit Hyperimage und Yenda," in *EVA 2014 Berlin. Proceedings: Elektronische Medien & Kunst, Kultur, Historie* (EVA Berlin 2014, Berlin: Gesellschaft zur Förderung angewandter Informatik e.V. (GFal); Staatliche Museen zu Berlin; Fraunhofer-Institut für Graphische Datenverarbeitung (IGD), 2014), 262– 67, http://hyperimage.ws/wp-content/uploads/Loebel\_JM\_HDH-HyperImage-Yenda 2014 PREPRINT.pdf (in German).

<sup>&</sup>lt;sup>30</sup> "bitGilde IT Solutions," accessed September 7, 2016, http://bitgilde.de/en/.

<sup>&</sup>lt;sup>31</sup> For the definition of "tags" in HTML5 see "HTML5 - A Vocabulary and Associated APIs for HTML and XHTML," chap. "8.2.1 Elements," accessed September 7, 2016, https://www.w3.org/TR/html5/.

be annotated with text but also with images, and in addition annotations may also contain hyperlinks. The system has been used in different teaching and research contexts including musicology, islamic studies, anthropology, and history for subjects like analysis of Japanese Visual Kei music videos, islamic sermon broadcasts, nepalese marriage videos and a Japanese propaganda film.<sup>32</sup>

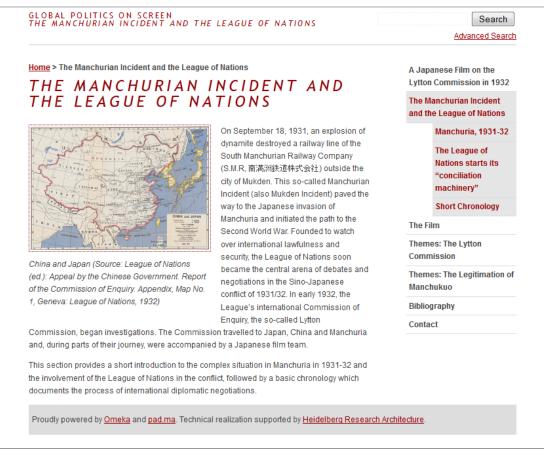


Figure 6: Subpage of the film analysis project "Global Politics on Screen: A Japanese Film on the Lytton Commission in 1932, <u>http://lytton-project.uni-hd.de/</u>.

Through a close cooperation with the chair of Visual and Media Anthropology (VMA)<sup>33</sup> the HRA established a tutor program for DH-tools. The chair's strong commitment to research based teaching was awarded in the state funded program "Welcome to Research".<sup>34</sup> Initially

 <sup>&</sup>lt;sup>32</sup> "Global Politics on Screen - A Japanese Film on the Lytton Commission in 1932," accessed
 September 7, 2016, http://lytton-project.uni-hd.de.
 <sup>33</sup> The Chair of Visual and Media Anthropology is one of the new professorships established by the

<sup>&</sup>lt;sup>33</sup> The Chair of Visual and Media Anthropology is one of the new professorships established by the Cluster to foster academic exchange and deepen the already existing expertise in the participating disciplines, cf. "Cluster Asia and Europe - Uni Heidelberg: Visual and Media Anthropology," accessed September 7, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/hcts-professorships/visualand-media-anthropology.html; For more informationa about the Chair's Research based teaching see "Research Based Teaching at the Chair of Visual and Media Anthropology," accessed September 7, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/hcts-professorships/visualanthropology/teaching/research-based-teaching-new.html.

<sup>&</sup>lt;sup>34</sup> "Welcome to Research' at the Chair of Visual and Media Anthropology," accessed September 7, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/hcts-professorships/visual-and-media-

a VMA initiative, the tutor program is now an interdisciplinary undertaking which includes classes from Chinese Studies as well as East Asian Art History.

Another challenging aspect of collaborative projects to mention is, as Jennifer Edmond phrases it, that "project objectives propose interesting research questions" for all disciplines or specialties involved in a project, and "that team members maintain a clear sense of their own roles and respect for those of their fellow team members." Teamwork should not "degenerate into the negotiations of uncomfortable tacit hierarchies, where some contributors (regardless of their expertise or seniority) feel like service providers working in the shadow of otherwise autonomous project leaders."<sup>35</sup> While this is true for all phases of DH projects, the discussion phase can be best used to prevent irritations between project members.

### The publication phase - transcultural analysis and dissemination

A successful traditional Humanities project usually ends with a paper publication: research results are presented in articles, reports, or books, which are handed over to a publisher who produces paper copies. These copies are then made available through book vendors and finally taken care of by institutional libraries or private persons. The infrastructure, workflows and legal frameworks for publishing and sustaining paper publications are well established and have evolved over hundreds of years.

When it comes to digital publications the situation is fundamentally different. It already starts with the question of what an academic digital publication is. There is a broad spectrum of what could be a digital publication. On the one side there are the mostly static publications that "mimic" traditional paper publications and are basically digitized versions of printed publications, e.g. in pdf format. These are sometimes available for books, but more often for articles or reports. Current journals systems extend the paper horizon by linking authors, reviews, citations or subjects, but these can usually be found in STEM-fields.<sup>36</sup>

On the other hand, there is a rich variety of published information that go beyond the limitations of a paper publication.<sup>37</sup> This includes collections of data sets, all kinds of multi-

<sup>36</sup> For a first study on digital publishing within the Digital Humanities see Ann Katrin Lorenz, "Interdisziplinäre E-Publikationen – interdisziplinäre Evaluation? Ein Blick auf die Bewertung fächerübergreifender Forschungsleistungen am Beispiel der Digital Humanities," *DARIAH-DE Working Papers*, no. 9 (2015): 11; STEM refers to Science, Technology, Engineering and Mathematics, cf. Heather B. Gonzalez and Jeffrey J. Kuenzi, "Science, Technology, Engineering, and Mathematics (STEM) Education A Primer," CRS Report for Congress (FAS, 2012), http://archive.org/details/R42642ScienceTechnologyEngineeringandMathematicsSTEMEducationAPrimer-crs.

anthropology/teaching/research-based-teaching-new/welcome-to-research.html; For the programme see "Heidelberg University: Welcome to Research - Research-Oriented Teaching Concepts," accessed September 13, 2016, https://www.uni-

heidelberg.de/research/young\_researchers/research\_oriented\_teaching/welcome\_research/. <sup>35</sup> Edmond, "Collaboration and Infrastructure," 56–57.

<sup>&</sup>lt;sup>37</sup> For a stimulating discussion of challenges in the current academic publishing world see Kathleen Fitzpatrick, *Planned Obsolescence: Publishing, Technology, and the Future of the Academy* (New

media files, source code, desktop software, online platforms, etc. It has become much easier to not only publish research results, but also research data.<sup>38</sup> The interesting thing about digital publications is that the target audience does not necessarily have to be human. Huge scientific datasets can be published and made accessible through application programming interfaces (API). This is not limited to measurement data from the sciences, but also applies to humanities datasets that can be published as structured linked (open) data.<sup>39</sup>

One of the problems, however, with digital publications in the humanities is still their acceptance within academia. In an environment where a large part of a publication's prestige is derived from the name of the print publishing house, young academics might think twice before investing their time in experimental new technologies. The risk one's research is dismissed as non-academic might be just too high.

Changing this environment is a slow, but continuously ongoing process and has many facets beyond publishing culture. It affects aspects of teaching, grading, and even administration.

The whole evaluation and grading process for the Humanities is optimized for print publications. That already starts in undergraduate classes. There are established criteria and procedures on how to grade written essays. But what happens when the test achievement ("Prüfungsleistung") has no longer the form of a linear print essay but comes e.g. in the form of a completely non-linear HyperImage publication? When the medium is so different from well-known print formats that criteria for "what is a good publication" no longer fit, new criteria have to be developed, and tested, and adjusted. At the Cluster of Excellence "Asia and Europe" we have already gone several times through this process of establishing and refining these criteria in a series of classes held by the Chair of Visual and Media Anthropology in cooperation with the HRA.

Having established an evaluation framework, the next challenge awaits. Your university's regulations might require that a physical paper copy of the graded work has to be stored for at least 10 years. This of course makes a whole lot of sense for traditional print works, but becomes tricky when a work comes in the form of a rather unprintable medium like a HyperImage reader. If you need to guarantee a ten-year accessibility to a digital test

York University Press, 2009),

http://mediacommons.futureofthebook.org/mcpress/plannedobsolescence/.

<sup>&</sup>lt;sup>38</sup> Within Germany, the heiDATA Dataverse Network, established in 2015 at the University of Heidelberg and managed by the Competence Centre for Research Data is still a pioneering initiative. It offers a research data repository to provide open access to and persistent identifiers of data sets in a long-term preservation environment. Cf. "Competence Centre for Research Data - Heidelberg University," accessed September 8, 2016, http://www.data.uni-heidelberg.de/index.en.html; and "heiDATA Dataverse Network," accessed September 8, 2016, https://heidata.uni-heidelberg.de/dvn/. <sup>39</sup> For more information about linked data from a library perspective see, for example, Carol Jean Godby, Shenghui Wang, and Mixter, Jeffrey K., *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*, Synthesis Lectures on the Semantic Web, Theory and Technology (Morgan & Claypool Publishers, 2015); For an application of linked data in Japanese Art History see Biligsaikhan Batjargal et al., "Linked Data Driven Multilingual Access to Diverse Japanese Ukiyo-E Databases by Generating Links Dynamically," *Literary and Linguistic Computing* 28, no. 4 (December 1, 2013): 522–30, doi:10.1093/Ilc/fqt058; and John Resig, "Japanese Woodblock Print Search - Ukiyo-E Search," Portal, *Ukiyo-E Search*, (2012), http://ukiyo-e.org/.

achievement ("Prüfungsleistung") an infrastructure that goes beyond a folder and a bookshelf is required, which brings us to the discussion of sustainability.

For something to be accepted as a scientific publication it has to be citable, the data and the interpretation on which the findings are based have to be reproducible. That means they have to be published in a sustainable way.<sup>40</sup> With all the new digital media types that are now available for publication sustainability is an issue that still requires some thought and investment. Having a digital result available (accessible and working as intended) by the end of a project does not mean that this is the case five years later. Usually, dependencies on a variety of different and fast changing technology result in ongoing maintenance costs after the project has ended. Keeping digital outcome alive not only requires a budget and technical expertise but often includes decision-making from the domain experts who created the original outcome.<sup>41</sup> Due to the nature of academic careers, however, those people have moved on and are often no longer available.

Sustainable digital outcome is therefore not only a technical challenge but has also institutional and legal dimensions. One of the current challenges in DH is setting up institutions that have the technical expertise to run, preserve or at least archive digital project outcome and bring legal mechanisms in place that settle the use of the material.<sup>42</sup> Nevertheless, the variety of technical solutions, software dependencies, and fast

Sustaining research data is also an important issue at the Cluster. The HRA collaborated with a number of projects that produced quite different kinds of digital publication. The HyperImage Reader, that is used for the web-presentation has evolved from the proprietary

technological change make it very expensive to sustain non-standardized project outcomes.

Adobe Flash to the open W3C standard HTML5. The separation of research data and presentation technology made it possible, that older projects previously published in the Flash format can easily be re-published with the HTML5 version of the Reader.<sup>43</sup> Other teams have collected and digitized material that was either dispersed, hardly accessible, or even yet completely unknown, and created public online resources with additional metadata.<sup>44</sup> The visual essays produced in classes of Anthropology and Visual and Media

<sup>&</sup>lt;sup>40</sup> "Long-term availability" of digital content is one of the eight core points on the DFG's "Checklist for Applicants and Reviewers", cf. "DFG Practical Guidelines on Digitisation," chap. 7.

<sup>&</sup>lt;sup>41</sup> Changes in the underlying database technologies or data migration to other platforms usually require a revision and refinement of the original data-structure. This is especially true with non-standard based data structures.

<sup>&</sup>lt;sup>42</sup> Cf. Puhl et al., "Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften," chap. 5.4 and the Heidelberg heiDATA initiative (note 34, above).
<sup>43</sup> For example "Hachiman Digital Handscrolls" (see note 26 above). "Cluster Asia and Europe - Uni Heidelberg: A New Mirror of Kāšī," accessed September 8, 2016, http://www.asia-europe.uni-heidelberg.de/en/research/heidelberg-research-architecture/detail/m/a-new-mirror-of-kasi.html is one of the older HyperImage projects that has already been republished as HTML5 version.
<sup>44</sup> "The Above Narders Collection" is an Onen Access unbeide where the complete isourcels of the

<sup>&</sup>lt;sup>44</sup> "The Abou Naddara Collection" is an Open Access website where the complete journals of the Egyptian satirical writer James Sanua are published, cf. "The Abou Naddara Collection," accessed

Anthropology<sup>45</sup> using Atomic Wiki<sup>46</sup> resulted in a kind of publication that is un-printable, because they contain audio and video, or are created in a non-linear structure.



Figure 7: Home page of "Ethnographic Photography in Asia and Europe - A transcultural approach", realized in the HRA instance of Atomic Wiki, http://ethno-foto.uni-hd.de.

In the "Global Politics on Screen" project mentioned above<sup>47</sup> a combination of two platforms, namely Omeka<sup>48</sup> and Pandaora were used. While Omeka is the platform to host the processed project findings, Pandaora provides the reader with the fully annotated video source material. Yet another new form of publication are the various kinds of research

September 8, 2016, http://abou-naddara.uni-hd.de; The "Early Chinese Periodicals Online (ECPO)" project provides free online access to over 200 publications, and for selected periodicals even annotated each individual item, like article, image, or advertisement with Chinese and English metadata: "Early Chinese Periodicals Online (ECPO)" (see note 19 above).

<sup>&</sup>lt;sup>45</sup> See note 30 above.

<sup>&</sup>lt;sup>46</sup> Atomic Wiki is the wiki engine integrated into eXist-db, cf. "Atomic Wiki," accessed September 8, 2016, http://atomic.exist-db.org/blogs/eXist/; and "AtomicWiki: A Wiki Based on eXist-Db Which Stores Data as Atom Feeds," accessed September 8, 2016,

https://github.com/wolfgangmm/AtomicWiki.

<sup>&</sup>lt;sup>47</sup> See note 29 above.

<sup>&</sup>lt;sup>48</sup> "Omeka," accessed September 8, 2016, https://omeka.org/.

annotations that are created with DH tools. As "nano-publications" or "micro-publications" they themselves become citable independent research statements.<sup>49</sup>

On the technical side, one way to address the challenge of rapidly developing technologies and the effects on sustainability is the strict separation of the data layer from a management layer and a presentation layer. User-interfaces, as important as they are in guiding users to the information and attracting them to a resource,<sup>50</sup> are ephemeral by nature; they can easily change and follow trends in design. The data itself, however, does not change. It can be further expanded, or annotated, or reviewed, but it does not follow fashions or trends. We therefore believe, that if data is stored in standardized formats following international established (meta-)data standards it can be successfully sustained.

However, international standards are also not static; new versions are developed and published. But since data is encoded according to a standard, it can be migrated and transformed with the standard. Therefore, an important aspect of sustainability is the creation of solutions that allow uncomplicated migration and transformation of data. From a technical point of view, most transformations are achievable, but on a pragmatic level the cost factor soon becomes an issue and this is where non-standardized solutions often become prohibitively expensive.

The implementation of open standards and the use of Open Source Software is part of the HRA strategy to sustain what has already been accomplished. By publishing our own development under Open Source licenses on platforms like GitHub we want to encourage re-use and improvement of the software.<sup>51</sup>

The HRA has developed a CSV2XML tool that helps transforming data from flat tables (spreadsheets) into structured XML. It is generic and can be used with any XML Schema, and its first implementation uses the VRA Core 4 XML standard.<sup>52</sup>

<sup>&</sup>lt;sup>49</sup> On the discussion of micro- and nanopublications see Tim Clark, Paolo N. Ciccarese, and Carole A. Goble, "Micropublications: A Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications," *Journal of Biomedical Semantics* 5, no. 28 (May 15, 2013), doi:10.1186/2041-1480-5-28; and Nora Schmidt, "Tackling Complexity in an Interdisciplinary Scholarly Network: Requirements for Semantic Publishing," *First Monday* 21, no. 5 (April 24, 2016), http://firstmonday.org/ojs/index.php/fm/article/view/6102.

<sup>&</sup>lt;sup>50</sup> Researchers like Johanna Drucker studied this aspect for a long time, cf. Johanna Drucker, "Humanities Approaches to Interface Theory," *Culture Machine* 12 (2011): 20; For a recent academic conference see "Digital Scholarly Editions as Interfaces - Zentrum Für Informationsmodellierung - Austrian Centre for Digital Humanities," accessed September 8, 2016, https://informationsmodellierung.uni-graz.at/de/aktuelles/digital-scholarly-editions-as-interfaces/.

<sup>&</sup>lt;sup>51</sup> Source code of all software developed by the HRA is available under the "Apache License, Version 2.0" on GitHub, cf. "Exc-Asia-and-Europe," accessed September 8, 2016, https://github.com/exc-asia-and-europe; For the License see "Apache License, Version 2.0," accessed September 8, 2016, http://www.apache.org/licenses/LICENSE-2.0.

<sup>&</sup>lt;sup>52</sup> See note 14 above.

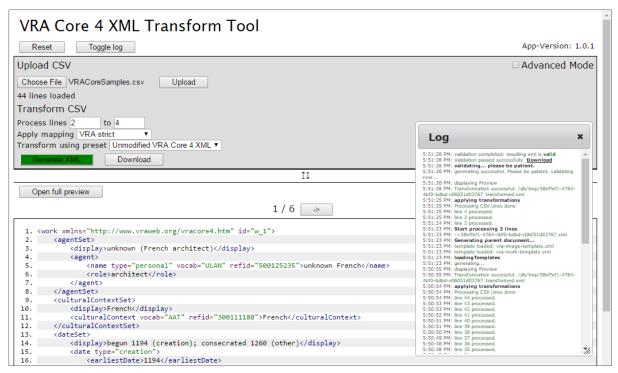


Figure 8: The VRA Core 4 XML Transform Tool is the first productive implementation of the CSV2XML tool, http://kjc-ws2.kjc.uni-heidelberg.de:8081/exist/apps/csv2xml/index.xq.

# Conclusion

In this article we have described the structures, methods, workflows and tools that have been established in Transcultural Studies digital research. With the introduction of a schematic workflow of typical humanities research within the Cluster's digital research environment we demonstrated how Transcultural Studies projects profit from Digital Humanities methods. The Heidelberg Research Architecture is a competent partner during all phases of this workflow. We have shown a number of digital outcomes of projects successfully implementing Digital Humanities methods in co-operation with the HRA.

During that process, not all approaches yielded the expected outcome. There are many lessons we've learned, for example we much more emphasize the importance of generic and re-usable workflows, or focus on the implementation of existing open standards and actively engage in further developing them. We are also actively initiating strategic co-operations with external partners.

However, some of the experiences are even more important to us:

Firstly, finding personnel is difficult because the expected DH competences are not yet part of the curricula at Heidelberg. Therefore, a longer-term investment in teaching and training students is essential.

Secondly, in terms of DH Strategy, dialog within projects is central. A core feature of the HRA is to offer experienced staff that engages as "intermediaries" in projects, to "talk across disciplinary cultures and encourage open-mindedness," as Edmond phrases it. She

continues: "and though these individuals cannot create the esteem and trust required to ensure that interdisciplinary collaborations remain enriching and productive, they can ensure that flashpoints are recognized and managed, and they can capitalize upon opportunities to develop a common language."<sup>53</sup> This "bridging the different sides" is exercised together with providing a central infrastructure that handles data, data management, and data presentation separately. Specific solutions developed as part of actual collaborations in research projects are used as basis for more generic ones that are available for re-use by other research projects. The joint dialog between Transcultural Studies and Digital Humanities is essential and enables project partners to learn from each other.

The growth in the complexity of research projects requires a higher degree of specialization and the strengthening of research disciplines. But at the same time, this also means that projects more and more require expertise from different research areas. Interdisciplinary therefore does not mean a blurring of the distinctive features of research areas, or merging them. In fact, projects in interdisciplinary settings like Digital Humanities and Transcultural studies need strong disciplines. (Only) with strong disciplines can strong interdisciplinary projects be realized.<sup>54</sup> Collaboratively working in interdisciplinary teams significantly increases their potential, but it also requires competencies in teamwork as crucial skill.<sup>55</sup> Intermediaries can foster the communication within research teams, but eventually, the ability needs to be learned by the researchers.<sup>56</sup>

And thirdly, the use of different technological solutions for individual projects are prohibitive expensive to sustain. Therefore, we invest in in-house development capacity for a manageable set of core technologies and apply internationally established data standards. The last eight years were a constant process of review and improvement. The HRA evolved from a department with combined IT service and database development into a specialized Digital Humanities unit and a separate IT department. This reorganisation was an important step and a prerequisite for sustaining the achievements of research introduced in this paper on an institutional level. Nevertheless, sustainability remains THE central challenge in all digital research. It is a multi-layered problem which requires solutions on technical, institutional, and educational levels.

<sup>&</sup>lt;sup>53</sup> Edmond, "Collaboration and Infrastructure," 57; "This variety of perspectives that digital humanities collaboration brings together requires the interweaving of very different intellectual positions and working cultures (...). The key differentiator (...) is (...) that they ensure from the outset that the project objectives propose interesting research questions or otherwise substantive contributions for each discipline or specialty involved, and that team members maintain a clear sense of their own roles and respect for those of their fellow team members. (...) digital humanities teams (...) require trust and harmonization between individual and group goals." Ibid., 56.

 <sup>&</sup>lt;sup>54</sup> Quoting Prof. Barbara Mittler, from a discussion at the Cluster of Excellence, June 23, 2016.
 <sup>55</sup> Lynne Siemens, "It's a Team If You Use 'reply All' ': An Exploration of Research Teams in Digital Humanities Environments," *Lit Linguist Computing* 24, no. 2 (June 1, 2009): 225–33, doi:10.1093/llc/fqp009.

<sup>&</sup>lt;sup>56</sup> Johanna Drucker, "Blind Spots: Humanists Must Plan Their Digital Future," *The Chronicle of Higher Education*, The Chronicle Review, 55, no. 30 (April 3, 2009): B6, http://humanitiesblast.com/wp-content/uploads/2011/10/Blind-Spots.pdf.