Strategy for developing digital services:
EHRIS digitization project

Barbara Ruth Cimpa
Abstract:
This master's thesis focuses on the development of a strategy for an information service in the area of digitization of cultural heritage and research, with a web based platform as implementation of the strategy. Via the platform, digital collections and research are presented and shared. Services around the platform are provided. Support activities comprise the whole chain of digitization, from digitization to archival of the digitized objects. The information service aims for a new approach to combine ICT and established research; research and industrial practice were combined and cooperation of participants from various backgrounds was enabled. The strategy and organization of such information service was described using a “quality management system”, standard ISO 9001. The development of the platform followed engineering practices. A guideline for the platform was established as part of the quality management system. On the platform, initial collections were encoded by the humanities' text encoding standard TEI, that bases on XML, and analysis with them were done.

Key words: strategiutveckling, informationstjänster, digitalisering, webbplattform utveckling, ISO 9001, XML/TEI
Acknowledgments

I wish to express my warm and sincere thanks to my supervisors Elena Maceviciute and Mats Dahlström, as well as Mikael Gunnarsson, at the University of Boras for their kind and encouraging support, friendly advice and constructive suggestions that made this master's thesis possible. I owe particular thanks to Elena Maceviciute for her sincerity and encouragement during the entire programme. One simply could not wish for a better or more dedicated supervisor.

In addition, I am very much grateful to the University of Boras' support team, so Ewa Skelte for administrative support and Carina Walden for the quick and friendly technical support.

I wish to express my heartfelt thanks to my friend Isabel Thapa for revising the language of my written thesis. And I would like to thank Nicole Stanbridge for her quick response and review of the policies document.

My professors and lecturers at the University of Boras, colleagues and friends supported me in various aspects of library science and of information science. I owe a lot to them and I am thankful for the many fruitful discussions that we had.

I would like to show my gratitude to my family. My parents, my siblings and my family were extremely encouraging and supportive throughout the time, offering not only support of all types but interesting discussion around the topics of my thesis. I very much appreciate the interest they have taken in my studies.
Table of Contents

1. Introduction..................................................................................................................................1
   1.1 Motivation..................................................................................................................................1
   1.2 Objectives and possible solutions............................................................................................1
   1.3 Previous similar work.................................................................................................................2
   1.4 Main methods and their application............................................................................................4
   1.5 Contribution and material..........................................................................................................5

2 Developing the strategy..................................................................................................................6
   2.1 The institute as context for research..........................................................................................6
       2.1.1 The institute.........................................................................................................................6
       2.1.2 User characteristics related to the strategy..........................................................................6
       2.1.3 Budget considerations..........................................................................................................8
   2.2 Collection development.............................................................................................................8
       2.2.1 Material selection and digitization......................................................................................8
       2.2.2 Preservation and access over time......................................................................................11
       2.2.3 Intellectual access – metadata............................................................................................12
   2.3 Provision of services................................................................................................................15
       2.3.1 Descriptions as QMS and visualization..............................................................................15
       2.3.2 Evaluation and service quality..............................................................................................16

3 The platform as integral part of the strategy..................................................................................17
   3.1 The platform's functions...........................................................................................................17
   3.2 Choice of development methodology and documentation.......................................................18
   3.3 The platform's web design (look and feel) and presentation.......................................................19
   3.4 The platform's user support and guideline..................................................................................20
       3.4.1 Collection digitization........................................................................................................21
       3.4.2 How to encode text.............................................................................................................22
       3.4.3 Collection archival.............................................................................................................23

4 Evaluation of the strategic decisions, via use of the platform (“me”)...........................................24
   4.1 Evaluation during and after development of the platform.........................................................24
   4.2 Evaluation during the digitization of collections.........................................................................26
   4.3 Short evaluation by representatives of target user groups.........................................................28

5 Discussion.......................................................................................................................................29

6 Conclusions.....................................................................................................................................31

7 Summary..........................................................................................................................................33

References...........................................................................................................................................34

Appendix A: selection of technologies and decision for in-house development................................39
   Decision for encoding UTF-8............................................................................................................39
   Decisions concerning TEI................................................................................................................39
   Reason for the use of XML technology XSLT..................................................................................40
   Decision for in-house development and considerations concerning open source software........40
   Decisions on programming language, database, server.................................................................41

Appendix B: platform........................................................................................................................43
   Platform – collection view: metadata.............................................................................................43
   Platform – collection view: transcription and translation...............................................................44
   Platform – collection view: formatted view....................................................................................45
   Platform – collection administration view: metadata.....................................................................46
   Platform – collection administration view: transcription and translation......................................47
   Platform – collection administration view: TEI mark-up sandbox...............................................48
   Platform – collection administration view: TEI file control flow................................................49
   TEI Header, created by the platform.................................................................................................50

Appendix C: documents created for the platform.............................................................................53
List of figures

Table 1: Target groups of EHRIS platform (“x” indicates the importance) .......................................................... 7
Figure 1: The QMS on top of the process network ............................................................................................ 15
Figure 2: Page "legend", the drawing elements of the QMS ............................................................................ 16
Figure 3: Mega menu, with the initially digitized collections ........................................................................... 20
Figure 4: TEI/XML encoding – cross-language search in mark-up ................................................................ 26
Figure 5: Collection view: metadata (default view) .......................................................................................... 43
Figure 6: Collection view: transcription and translation at once .................................................................... 44
Figure 7: TEI/XML encoding - formatted view, transformation of UTF-8 encoded characters ..................... 45
Figure 8: Collection administration view: metadata indication and display file upload ................................... 46
Figure 9: Collection administration view: transcription and translation .......................................................... 47
Figure 10: TEI/XML encoding - format preparation, and transformation of mixed language content .......... 48
Figure 11: TEI file control flow .......................................................................................................................... 49
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>activity based costing</td>
</tr>
<tr>
<td>DELOS</td>
<td>a network of Excellence on Digital Libraries [project]</td>
</tr>
<tr>
<td>DTD</td>
<td>document type definition</td>
</tr>
<tr>
<td>DVD</td>
<td>[optical disc format]</td>
</tr>
<tr>
<td>EHRIS</td>
<td>economics historical research institute South East Asia</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information system</td>
</tr>
<tr>
<td>HTML</td>
<td>hyper text markup language</td>
</tr>
<tr>
<td>ICT</td>
<td>information and communication technology</td>
</tr>
<tr>
<td>ISO</td>
<td>international standardization organization</td>
</tr>
<tr>
<td>KPI</td>
<td>key performance indicator</td>
</tr>
<tr>
<td>NINES</td>
<td>networked infrastructure for nineteenth-century electronic scholarship</td>
</tr>
<tr>
<td>OS</td>
<td>open standards</td>
</tr>
<tr>
<td>OSS</td>
<td>open source software</td>
</tr>
<tr>
<td>PDCA</td>
<td>Plan-Do-Check-Act cycle</td>
</tr>
<tr>
<td>PHP</td>
<td>PHP: Hypertext Preprocessor (originally: Personal Home Page)</td>
</tr>
<tr>
<td>QMS</td>
<td>quality management system</td>
</tr>
<tr>
<td>5S</td>
<td>streams, structures, spaces, scenarios and societies [5S framework for digital libraries]</td>
</tr>
<tr>
<td>SEA</td>
<td>South East Asia</td>
</tr>
<tr>
<td>TEI</td>
<td>text encoding initiative</td>
</tr>
<tr>
<td>USD</td>
<td>United States dollar</td>
</tr>
<tr>
<td>UTF-8</td>
<td>UCS Transformation Format 8-bit [Unicode encoding]</td>
</tr>
<tr>
<td>XAMPP</td>
<td>X (cross-platform); Apache HTTP Server; MySQL; PHP; Perl [development server package]</td>
</tr>
<tr>
<td>XML</td>
<td>extensible Markup Language</td>
</tr>
<tr>
<td>XSLT</td>
<td>extensible style sheet language transformation</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Motivation

In 2011 an idea of the Economics Historical Research Institute South East Asia (EHRIS) started to take shape. The founders of the EHRIS wanted an organization that could create a community of professionals and enthusiasts working on digitization of and access to South East Asia cultural heritage. I saw this as a unique opportunity for my final examination work for the Master’s degree studies in digital libraries as it gave me the chance to develop from scratch a strategy for an institution in a relevant area. I had the advantage of being able to work with a team of software developers and implement the strategy in a digitization platform supporting individual efforts of digitization of cultural artefacts.

The problem I have chosen to solve lies within the area of strategic development of services directed towards digitization. But it can also be seen as a wider problem of modern institutions emerging around the possibilities offered by new technologies. Many of them are small organizations with very limited financial resources that rely on voluntary participation and the enthusiasm of their members. They exploit the possibilities of interactive technologies, such as wikis, to produce a variety of products (e.g. encyclopedias, journals, subject portals, collections of texts or other objects), which are then used freely by the interested community and other users.

The goal of this master's thesis is to develop a strategy for an information service in the area of digitization of cultural heritage and research. Strategy of the information service is a combination of technology, practitioner's expertise and established research. A web platform is developed to implement that strategy. The platform should provide digital services and at the same time, support cooperative work. Research and industrial practice can be combined, where practice offers suitable models, they should be used. An important aspect of the strategy is that such a platform can reach various users groups.

1.2 Objectives and possible solutions

The main goal of the work can be expressed in the following objectives:

1. To develop an institutional strategy of the information service for the first five years. The question answered by this development can be formulated as follows: What are the main elements of the information service's strategy? How to combine the solutions of different know-how and from divers research areas (such as humanities, information science, and information technology) into practical solutions?

2. To create an important part of the strategy using the application of a web based platform that supported digitization, research and cooperation. The question answered by this creation process can be formulated as follows: what could such a web based platform look like to support users to do digitization and text analysis, with satisfying quality? The platform should guide users through the process of digitization and should provide a simple sandbox for text analysis. We see “sandbox” like it was described by Goldberg et al. as a “concept of connecting a[n] [...] application to a restricted environment, within which it has free reign” (1996, p. 3), and that does not impact the application within which the sandbox is embedded.

3. To conduct the initial testing to assess the strategy for the information service and platform through creation of initial digitized collections. The question answered by this development can be formulated as follows: Is text encoding standard TEI (Vanhoutte, 2004) suitable for the needs of the humanities’ text collection creation and combination of the knowledge from different areas?
A strategy will prove its value only after having been followed for a certain period of time. However, the implementation and use of the platform were supposed to give a first impression of the value of the strategic decisions taken. An evaluation of the platform is planned to be carried out during its implementation process. When the platform is up and running, its functionality should be tried out. Three collections are intended to be digitized with the support of the platform. Further on, interviews about the usability and support of the platform for digitization will be carried out, but this is not in the scope of this master's thesis due to time constraints.

1.3 Previous similar work

This master's thesis focuses on strategy development for an information service for digitization work, and its implementation through an internet platform.

Any organization that focuses on offering an internet platform can be seen as similar work, as implementation of strategy development through a platform. YouTube (http://www.youtube.com/), a service to exchange music files, exists solely as an internet platform. Examples in the area of digitization of cultural heritage exist as well. The materials of American Memory (http://memory.loc.gov/ammem/) are available in digital form only. These materials were digitized of historical collections, mainly of the Library of Congress. Another example is Project Gutenberg (http://www.gutenberg.org/) that enables the exchange of works in the public domain, and offers administration services to manage the transcription and assessment of the texts. Project Gutenberg is a volunteer platform with minimal administrative involvement. It had a major influence onto this master's thesis as it is being offered as a pure internet platform and is being maintained by volunteer's work. It runs on a very small budget and offers minimal administrative services.

Another well-known platform run by volunteers is the Wikipedia (http://www.wikipedia.org/). However, this platform may not offer the implementation of “strategy development”. Instead, anybody contribute their knowledge to it. Research is being done onto user's activities, by other initiatives (http://en.wikipedia.org/wiki/Wikipedia:Statistics), but not by the Wikipedia's owners, and less onto the knowledge itself. However, the Wikipedia influenced this master's thesis work by the fact that the control is executed solely by the community. The platform administrators offer the technical maintenance. User participation to that extend could not be implemented in my project for now, please see section “7 Summary” on p. 33.

Platforms that are intended to be run by user interaction, and which are implemented as strategy development, are the so called social networks. At this point in time (2011), Facebook (http://www.facebook.com/) may be one of the most widely used. These platforms enable user interaction in various fields, like blogs, exchange of images or mails. Communities of users with common interests arise and are supported (Facebook: “groups”). However, not all Facebook's users will form a community. The intended platform for this master's thesis may include social network functionality, but for the start and for the purpose of this master's thesis, limits have been set due to time constraints. The intended platform also differs in that it focuses on one narrowly defined purpose. All intended users can share it and thereof, form one community.

For the technical implementation, open source software had to be considered (please see section “2.2.2. Preservation” on p. 11), in the area of digitization and text analysis. Open source software platforms Collex, TextGrid and Omeka are used in the area of digitization. Projects that use these software platforms were scrutinized.

Collex are tools for collaborative work on digitized collections. Collex is used in the project NINES (Networked Infrastructure for Nineteenth-century Electronic Scholarship) (Nowviskie, 2005, p. 2). TextGrid offers access to sources of German universities in the field of humanities (Neuroth et al.,
Both platforms' functionalities accorded to the intended platform. Collex and TextGrid focus on expert users from the field of humanities. The majority of my project's intended users (please see section “2.1.2 User characteristics” on p. 6-7) were presumed not to invest much time into the use of a platform, like humanities experts may do. Therefore, the wide variety of these platforms' functionalities had to be simplified. Technically, applications are strongly connected internally. A small change impacts on the whole application which one had to learn. Hence, that was not possible in the available time.

There is not much to be said about the usage of TextGrid. However, Collex, with the project NINES (http://www.nines.org/about/what-is-nines/), seems to be successful in acquiring and keeping a users' community. Project NINES offers workshops and scholarships. For this master's thesis, it has been considered that making such offerings available strongly contributes to motivating users. A partnership with a university or a research institution may be crucial to make a platform popular. But that is not subject to this master's thesis.

The implementation of projects that use Omeka can be seen as implementation of a strategy. However, projects that use Omeka are of limited, clearly defined content, and most of them are complete (please see examples at http://omeka.org/showcase/). Omeka is mostly intended for publication rather than to support cooperative work. It is designed to upload a collection to be displayed on the web. Omeka is a small and neat platform, but after careful consideration, it was decided that collaboration plays a far too important role to omit.

The creation and outline of the strategy was guided by strategy plans of libraries. The strategy plan of the British library (2005) and a report of the Cornell university library (2003) showed the elements of a strategic plan for information services. The plans are exhaustive in descriptions and references. Another example that focuses rather on the most important issues was the strategy plan of the National library of Scotland (2005). This plan was not followed strictly, but, it provided guidance as to how to write the “policies”, as implementation of the strategy. Documents that arose out of this master's thesis follow a naming convention, please also see section “1.5 Contribution and material” on p. 5-6. The policies' document follows the naming convention and is referred to as document COR.I-020.01.

Similar to our policies, while by far more extensive, were the National library of Australia's web guidelines (n.d.), which comprises the organization and the strategy for the next years. These web guidelines mainly influenced my master's thesis in that all information is open to the public.

A quality management system in line with ISO 9001 requirements is widely used in industrial environments (Beckford, 1998, p. 236). This master's thesis user groups were also practitioners who knew that. A quality management system according to ISO 9001 is applied in various areas and may be flexible enough to be applied to a humanities or research environment. Auer et al. (1996) describe the application of such a quality management system to a research environment.

Guidelines that describe this master's thesis topic area, “digitization”, were mainly the user guide of Minerva (2004). A similar guideline was the digitization guideline of the Deutsche Forschungsgemeinschaft (2009) which provided further details. Both guidelines describe recommendations and lessons learned of digitization projects, and several aspects of both guidelines were taken into account for the digitization work (described in appropriate sections of this master's thesis).
An information services' environment may rather be driven by textual descriptions. But, as flow charts are easy to understand (Wright, 2001, p. 59), we intended to use flow charts to describe guidelines to do work with the platform. Hazen et al. (1998) gives an example of a flow chart in the context of libraries and digitization. However, this flow chart describes only one particular task, and we did not expect to obtain a full description of the entire information service's work flows. Such work flows may be used institution internally and may be only available within the particular institution, but not for us to built the platform's work descriptions on bases of them. Still, the use of a flow chart by Hazen et al. in the context of libraries showed us that drawings were applicable in an information services' environment and we decided to describe our guidelines as flow charts.

Last but not least, in industrial environments quality is usually an integral part of the whole organization. This is reflected by e.g. the use of the plan-do-check-act (PDCA) cycle (Beckford, 1998, p. 67). Activities are planned and assessed, unexpected results may lead to replanning. A similar concept was mentioned for the context of digital libraries; “evaluation” must be planned from the start as an integral part of a task (Maceviciute, 2009a). Within two frameworks to describe digital libraries (“DELOS” and the “5S” framework), quality parameters are collected. In the DELOS framework, “quality” is one of six areas of the “digital library universe” (Agosti et al., p. 9). This view is very close to the PDCA cycle. Hence, we thought that the practitioner's experience to include quality as part of each task, could be integrated into a platform for an information service in the digitization area, too.

However, there was no ready-made package that fulfilled all my wishes. This master's thesis tried to connect them all into a cohesive whole.

1.4 Main methods and their application

EHRIS was chosen as the organization suitable to carry out of the work, based on adapted objectives defined in section 1.2. It was chosen because it was entirely new and the whole process of strategy development and platform creation was in my hands from the start. Its resources very minimal, but it had a suitable goal of cooperation support in the area of digital libraries and digitization, and provided a team of software developers who could work on the software for the platform.

The main method chosen for the strategy development was the strategic planning process for information services according to Bryson (2006).

Bryson describes a strategic planning process for information services (Bryson, 2006, p. 35). Maceviciute (2009b) modified Bryson's model for digital libraries adding a step called “needs assessment”, including the “needs of users”. This attitude of user focus was adopted for the entire planning process. A second aspect concerned the policies and we decided to “develop broad policy statements” (Maceviciute, 2009b), early in the process. “Policies” (COR.I-020.01) were written in parallel to the strategy, and were used as directions for the platform creation.

“A strategy is the pattern or plan that integrates an organization's major goals, policies, and action sequences into a cohesive whole” (Mintzberg, 2003, p. 10). Major goals for the digital service were gathered by talking to founding members. The goals were articulated as “vision”, the first step of Bryson's model, which “outlines the desired future state” and “mission”, and “articulates how the desired future state will be achieved” (Bryson, 2006, p. 35). Vision and mission are listed in the policies (COR.I-020.01).

The third step in Bryson's model is a “situation audit”, which considers external and internal
influences. Strategic objectives arise from the situation audits and are formulated in a forth step. This differed for this master's thesis. Strategic objectives were formulated by the founding members when gathering expectations. Thereby, strategic objectives were known before the strategic planning process began. They were considered “internal influences” and set as strategic objectives “KPIs” (key performance indicators, e.g. Rosam & Peddle, 2003, p. 53-54). The KPIs were used to calculate the budget, where a balance had to be found between the users focus and budget constraints. Here, questions for collection development in respect of digitization were considered. Digitization guidelines like Minerva (2004) and Johnston et al. (2004) showed the direction for the digitization decisions.

After the situation audit, objectives were defined. Strategic objectives span 2 to 3 years (Bryson, 2006, p. 40). These were the final KPIs to reach (INV.I-020.012). Of these, operational objectives were derived, which span shorter terms of 1 to 2 years (Bryson, 2006, p. 40). Operational objectives were defined in the “marketing plan” (INV.I-020.016).

Following the fifth step of Bryson's model, programmes to achieve the operational objects that span durations of 1 to 12 months (Bryson, 2006, p. 40), were designed (INV.I-020.016). The strategic planning process resulted in two programmes built on each other. The basic programme was the development of the strategy and its practical implementation in form of a quality management system (QMS) in line with the ISO 9001 standard.

The second programme was the construction of a platform as an integral part of the strategy of cooperation and presentation. Here the methods of the software/system engineering process according to the ISO/IEC 12207 were applied in a simplified iterative process (Boehm, 1988, p. 64, figure 2). On the platform, XML coding and analysis was done. The TEI guidelines of the TEI Consortium (www.tei-c.org) were guiding for the TEI/XML encoding and analysis. A web based digitization guideline was provided, which was part of the QMS and also followed the ISO 9001 standard.

The sixth step, finally, is the “programme review”. The execution of the programmes is evaluated. Evaluation was intended as described above. Besides, a QMS requires the implementation of regular monitoring and evaluation of strategic objectives. The implementation of the QMS automatically lead to Bryson's sixth step of evaluation. However, an evaluation is planned yearly, and therefore, it is not in the scope of this master's thesis.

1.5 Contribution and material

This master's thesis comprises:

– the creation of the strategy,
– the design and creation of the quality management system and relating documents,
– the design of the platform, as described in the software engineering documentation,
– the project management for the software development,
– the design of the XML encoding and analysis, and
– the evaluation of the platform after completion.

The coding of the platform was executed by a group of software developers. The software developers also contributed by reviewing the software engineering documents. The strategic decisions were made with the contribution of the founding members of the EHRIS institute.

Documents that were created for this master's thesis are work flow drawings and textual instructions
for the QMS, and the software development documentation for the platform. The designed QMS comprises a document control system, which requires a naming convention, shown e. g. by Wang and Tsai (2009, p. 265-266). Besides this master's thesis document itself, the master's thesis documentation material follows the QMS' document control system. Documents are identified by a unique identifier. The identifier is also the file name. The document control system is described in document COR.I-020.19. Documents are listed in the “document list” (INV.I-020.011) and are explained in appendix C.

2 Developing the strategy

2.1 The institute as context for research

2.1.1 The institute

Our framework was the institute “EHRIS”, the Economics Historical Research Institute South East Asia. The institute's vision and mission are described in the policies COR.I-020.01. Awareness for cultural heritage and its value is a major characteristic of the institute. It works on digitization of South East Asian (and other) cultural heritage.

Practically, research on cultural heritage can be done cooperatively, leading to exchanges of ideas and increase of knowledge and expertise. Cultural heritage is rendered usable for analysis by digitization. The institute saw as its main goal to enable people to participate in this. If people do not have specialized skills or expertise, they should still have the opportunity to participate in the institute's activities and be attracted through simple online participation means.

The institute wanted the scope of the collections to cover economic and historical circumstances in South East Asia (SEA). Items of these topics may be common items that are used in daily life. Digital collections are seen as a method to preserve such material that could not be preserved otherwise. Like Smith (2004, selection for preservation) said, “In theory, there is nothing created by the hands of mankind that is not of potential research value for humanities scholars”. In order for the digital collections and work done on it to remain accessible in the future, necessary actions have to be considered.

The development of the platform can be seen as “set-up” of the institute. The platform implements the strategy, and it is the starting point for further activities of the institute.

2.1.2 User characteristics related to the strategy

Through discussions with the founding members, three target user groups for the strategy and consequently for the platform were identified. All user groups had to be open to technologies and to have a basic knowledge of the internet, as well as basic English reading skills. One target group were researchers in the humanities, while we target more likely researchers in the digital humanities. Digitization and research should be supported by ICT, so ICT experts were a second target group. In addition to these two expert target groups, everybody else is encouraged to participate actively (COR.I-020.01, mission 1.). However, with the constraint of the mentioned minimum requirements, in reality a large group of potential users had to be excluded. How to train this large group for participation, and how to reach this group outside the web is an open question.
For now, our three target user groups are seen as follows. “x” indicates the importance:

<table>
<thead>
<tr>
<th>Target user groups ..</th>
<th>. have needs of:</th>
<th>. are early adopters?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICT support</td>
<td>humanities expertise support</td>
</tr>
<tr>
<td>Humanities experts</td>
<td>xxx</td>
<td>x</td>
</tr>
<tr>
<td>ICT experts</td>
<td>-</td>
<td>xxx</td>
</tr>
<tr>
<td>non-experts</td>
<td>xx</td>
<td>xxx</td>
</tr>
<tr>
<td>People in (South East) Asia</td>
<td>..</td>
<td>..</td>
</tr>
</tbody>
</table>

Table 1: Target groups of EHRIS platform (“x” indicates the importance)

This influenced the choice of work processes and tools. Work processes and tools must be considered to be usable for everybody. The transfer of results and procedures to other users, “the community”, is considered highly important; process and language that is understood by experts only must be avoided. Information visualization was considered to allow understanding for non-experts and to quicken understanding for experts of gained information. Therefore, information visualization is not an embellishment but is seen as part of the work on collections. Gained information has to be prepared to be used or at least understood by everybody. Resources use must be kept free of charge and they must be provided usable in a time-efficient manner. Users who are non-experts were considered to expect and accept any level of resource usage provision. Some users may enjoy to invest time into an application to learn it; and there may be people who judged a service after the price to pay for it. But we thought that both, the user groups of ICT experts and humanities experts needed cost-efficient resources, and that ICT experts expected resources to be time-efficient, too. Otherwise, ICT experts were expected to use another solution or create their own solution. While humanities experts were considered to be more tolerant to time-consuming resources use as they may be less used to improve a non-satisfying solution; their budget was considered to be tight and not easily invested into a new resource.

The target group is miscellaneous. Humanities' experts may not be experts in the ICT field and vice versa. Since the use of the internet is wide-spread, more people may know about ICT than have humanities expertise. Regardless of their background, the platform is intended to support users to digitize a collection for preservation and presentation purpose, to generate metadata and to do simple research work. Collaborative work or social network functionality is part of most community pages. These elements will be a part of further development, but not necessarily at the start.

Economic barriers that relate to financial means and to the investment of time (Wilson, 1997, figure 1 p. 552, and p. 559) must be avoided. Access is free of charge for everybody. This golden rule should be kept by all means. Quick access to the collection pages is of high importance. To contribute, a password shall be required. We do not expect that people are going to damage the works of others, but a password may give people the feeling of security for their work and may emphasize the participation in a serious project.
For acceptance and use, the presentation is important. According to Shneiderman: “Well designed [...] systems generate positive feelings of success, competence, [...] in the user community. When an interactive system is well-designed, the interface almost disappears, enabling users to concentrate on [...] exploration, or pleasure.” (Shneiderman, 1998, p. 10). It should be simple and interesting for all user groups to browse the collections and to become aware of their value. At the same time it should be attractive for creators of the collections to see them presented. People in Asia value harmonic arrangements of colours and shapes. This must be kept in mind when designing. It is visualization that makes the platform interesting to visit and use, and, through technologies of visualization, one has the opportunity to present one's collection. By doing so, awareness and interest can be raised. Ideas of visualization should be used everywhere in the platform (COR.I-020.01, mission 3).

2.1.3 Budget considerations

Strategic objectives (Bryson, 2006, p. 40) were set as “KPIs” (key performance indicators, Rosam & Peddle, 2003, p. 53-54), for the first three years (INV.I-020.012). Financial calculations on basis of the KPIs were done (INV.I-020.013).

Costs were calculated as “activity based costing” (“ABC”), Ellis-Newman and Robinson (1998) explain ABC calculations within a library's context. Calculated costs to offer the digital services are approx. 30,000 USD for the first year 2013. From the 2nd year (2014) on, costs are approx. 34,000 USD yearly; given that a “community” gave support. These figures stand against approx. 50,000 USD, 60,000 USD and 90,000 USD costs for the first three years without community support.

The cost difference is considerable. The encouragement of the community is an important factor. The costs with community support (INV.I-020.013, column “reduced”) should not be exceeded. As consequence, the community must be encouraged to support the activity of EHRIS. Realistically, the possibility do to digitization work and research work on the basis of digitized objects may not be a main motivational factor, while, the opportunity for users to promote themselves and their ideas is considered to be one. The support by users is crucial for the implementation of the whole work and specific attention in the devised strategy is given to mobilise users to contribute to the platform.

Therefore, the conferences were planned (INV.I-020.012, COR.I-020.01, mission, 7.). Conferences were considered to make a strong contribution to motivate a community. Experts should be invited to give speeches, making the conference more attractive. However, this is also one of the most costly items; not only of events, but also of the total costs (INV.I-020.013, “ABC conferences, fix”). It would be wise to explore, whether experts could be convinced to contribute for less reimbursement or on a completely different basis.

2.2 Collection development

In this master's thesis, the term “collection” in most cases equals to a digitized book. A collection consists of “objects”. An “object” refers to one page of the digitized book. On the platform, for each collection a TEI encoded file is created, which contains the collection's objects as “<div>” elements in the “<body>” section.

2.2.1 Material selection and digitization

“There should be a knowledge management policy, with a scope statement and criteria for inclusion” (Thomas, 2002, p. 104). Devising material selection criteria, I worked together with the founders of EHRIS and the following resulted from our discussions.
To promote the awareness for cultural heritage of the region of South East Asia (SEA) (COR.I-020.01, vision, and selection (scope statement)), the scope statement of “historical economics SEA” was selected. Minerva (2004, p. 21) gives “pragmatic suggestions” for selection criteria. The topics of “economics” are considered to be of interest for users. Further selection criteria were set, pragmatically, so that as many people as possible can participate. Objects do not have to be unique or of high value, items used on a daily basis or text in common books may be digitized. As for the time frame, one year old objects count as “historical objects”. However, the “historical” time frame may draw users’ attention to older material. Older text material may be in the public domain and “may be re-used without restriction” (Minerva, 2008, p. 21). Hence it can be digitized without further thoughts of copyright.

The selected content must follow ethical rules (COR.I-020.01, code of ethics). A violation of these criteria leads to immediate exclusion of the collection (COR.I-020.01, When do we unfortunately have to exclude participants).

Objects show various aspects that may be rendered in different digital form. An image will capture the aspect of the view. The term “tangialities” refers to “association[s] of [...] information with multimodal sensory experiences”. This information “create[d] a new layer of knowledge [...] more natural and efficient for humans” (Milekic, 2007, p. 371). To achieve this layer special competence may be needed. But we have a requirement that digitization should not require a specialist's skills or special equipment (COR.I-020.01, mission 1.) Therefore, it was decided that the objects will (only) be captured and represented by images. Objects that show text are rendered analysable. This means that text should be transcribed. All text is subject to transcription. Text format attributes should be preserved and therefore captured. The context of an object may convey interesting information that should be preserved within a short but sufficient metadata element set.

The content quantity and quality of a collection was roughly defined, e. g. the “good practices handbook” for digitization (Minerva, 2004) does not mention this type of considerations. They may rather arise in a “platform where all users can publish their own knowledge” as Weel describes “Web 2.0” initiatives (Weel, 2012, chapter 5, The social web). The platform does not provide sufficient user interactions to be classified as Web 2.0 initiative, but, these may be added in the future. Therefore, following decisions were taken:

1. Pure text collections, that is, collections which do not comprise images, are not supported, because the visual impression conveyed by images should not be missed (COR.I-020.01, mission 3.).
2. Three images are the required minimum size of a collection. That should encourage users to participate even if they do not have many objects.
3. No requirements were set for a collection's content quality, only, that the quality was adequate for preservation.
4. According to the digitization recommendations of the library of Congress for image capture, depth 600 dpi and format tiff (Terras, 2008, chapter 3, chapter 4) were recommended. At the same time it was decided that these should be recommendations and not prescriptions. Digitized collections that do not follow them may not be of value, but, the contribution of users is considered of higher priority. Users who may not be able to fulfil these recommendations, should be encouraged to participate nevertheless. That decision was kept as open question. It is to be observed in the future if it will be possible to maintain the balance between the encouragement of user participation and the quality of digitization.

With regards to the composition of objects, the difference between “collection” and “online exhibit” is not made. A book is made of book pages, and a user may register a book as collection in the
platform and consequently add digitized book pages as objects to the collection. However, under appropriate legal conditions (please see the “Creative Commons license” in section “2.2.2 .. access over time” on p. 11), users may use existing digital objects and may compose an exhibit. This is accepted, as it is done in the scholarly world as well. “When electronic sources are brought together for scholarly purposes they become a new, second-generation electronic resource” (Palmer, 2002, scholarly contribution). Solely, duplicate objects within the institute should be avoided (mentioned by Deutsche Forschungsgemeinschaft, 2009, 1.3 “Dublettenprüfung”, while in a worldwide context). That implies that users should not consider to combine digitized objects of other users to prepare own collections within the platform. Recombination in that manner is possible in e. g. “Collex”, a digital humanities' web application. “Collex [...] allows users to collect, annotate, and tag online objects and to repurpose them in illustrated, interlinked essays or exhibits.” (Nowviskie, 2005, p. 1). Collex' users may be mainly humanities researchers and may value highly the work of annotating existing digitized objects. However, EHRIS targets user groups who are not humanities experts and are considered to do less work with annotations. Therefore, they may value higher the digitization itself and may see the use of digitized objects of other users as absence of consideration.

As for avoiding duplicates, it was considered that within the given budget checks cannot be carried out. The support team could not compare each new object image with all existing object images. The responsibility to enforce the requirement to avoid duplicates was transferred to the community: Who notices an infringement, is encouraged to give notice to the support team to take adequate measures.

“Critical digitization” can be seen as carefully performed digitization of unique items (Dahlström, 2010, p. 10). In contrast, “mass digitization” targets rather the quantity (Coyle, 2006), and it is not done to capture unique characteristics. The platform is intended to enable “everyone” to conduct digitization with satisfying quality. This way, a standardization of digitization is supported and such decisions lead to mass digitization. Tanselle (1989) looks at the degree to which a digital representation remains faithful to the original. With the chosen (standardized) practice of image capture, a 1:1 surrogate is produced, but only the aspect of the view is captured. Other unique aspects of objects, which are expressed as “tangialities” (Milekic, 2007), namely the experience of other senses like the handling, material, or size cannot be reached. The context of use of the physical object is lost. Furthermore, only a small set of metadata was meant to be collected. So the level of authenticity may not be too high, it could become “maimed information” (Deegan & Sutherland, 2009, p. 134). Milekic goes further: “apart from cataloguing a museum collection, the value of such repositories for any kind of knowledge transfer is highly questionable” (2007, p. 369).

Apart from the intellectual value of such digitization, the aspect of the work on the physical object needs to be considered. Mass digitization can be of any level of quality. If we consider the target users groups, professional equipment may not be available and trained conservators may rarely be involved. Decisions for suitable measures that a conservator could do (Paris, 2008), will not be taken. Consequently, damages on the object might occur, a problem for “mass digitization” raised by Dahlström: “Perhaps it is vital not to destroy the source document during the digitization process (as mass digitization does)” (2010, p. 10). We also need to bear in mind that the climate in SEA is humid and sensitive material deteriorates quickly. Conservation measures on physical objects are out of the institute's scope (COR.I-020.01, What we unfortunately are not able to provide for now). Nevertheless, users are reminded to pay attention to conservation (please find that suggestion in COR.I-020.04), but in reality, it may not be possible for users to take such measures.

All these consequences are the facts and must be accepted. The objective that everyone should be able to participate is of greater importance. We assume that the digital surrogates are a method to
preserve objects that might not be preserved otherwise (COR.I-020.01, mission 4.). “The purpose of preserving cultural and intellectual resources is to make their use possible at some unknown future time” (Smith, 2004, What is Preservation and Why Does it Matter?). If we see it that way, preservation is accomplished.

The main objective is to collect as much material as possible, digitized by users. This approach to a certain sense may fall under crowdsourcing (Weel, 2012, p. 26: “mobilise the 'wisdom of the crowd'”), since we want to motivate “everyday people using their spare cycles to create content, solve problems, even do [corporate] R & D.” (Howe, 2006, p. 1).

2.2.2 Preservation and access over time

One of the institute's goals is the preservation of objects (COR.I-020.01, vision). All work accomplished by the users should be kept accessible in the future (COR.I-020.01, mission 4.). Provisions must be taken to avoid “problems of media degradation and hardware/software obsolescence” (Smith, 2004, Technical Challenges to Digital Preservation and Why They Matter).

The technical life time of files, “logical preservation” (Smith, 2004, Technical Challenges to Digital Preservation and Why They Matter), must be considered. We decided to use open source software (OSS) and open standards (OS) (Minerva, 2008, p. 32-35). OSS is thought to be more stable than proprietary software, since a community adjusts the software to technological advances. Open standards may support the open source software principles. Open standards, as well as OSS, are thought to remain functional as their community adjusts the standards and reading tools to technological advances. Opting for open source software excludes the use of e.g. functions offered by the “turning the pages” program (British library, n.d.). “Turning the pages” simulates a physical book, but it requires the installation of software that is proprietary. Many OSS and OS can be used without paying royalty fees, and as explained above, no financial burden should be passed on to users and prevent their participation. Therefore, OSS and OS for which no royalty fees apply, must be chosen.

The “code of ethics” (COR.I-020.001, code of ethics) demands that laws must be abided. With the use of royalty free software and standards, users will not have to decide against participation, or infringe a law by using tools illegally. Amongst the open standards that fulfil these considerations, those have to be chosen that are supported by suitable OSS. Conversely, an open source software must be selected that supports suitable open standards. Exceptions are acceptable when other de facto standards exist. Then, a decision leading to long-term accessibility is to be preferred. A collection's “master files” are high resolution images of the collection's objects (Federal Agencies Digitization Guidelines Initiative, n.d.), that are not further treated and serve for preservation purpose. Image format .tiff is the “current archival standard” (Terras, 2008, p. 62, p. 82-87), and, “tiff” is an “open file format”, that is: an open standard (Terras, 2008, p. 86). It meets all requirements and therefore, is recommended to use on the platform.

The physical life time of a file refers to the file on a storage media, the “physical preservation” (Smith, 2004, Technical Challenges to Digital Preservation and Why They Matter). Files may become corrupt for any reason. Therefore, measures have to be taken to prevent the loss of information. Since the service to preserve the master files (COR.I-020.01, mission 4.) is offered, it was decided that a master file collection should be stored in two copies. The second copy can provide access in case the first becomes corrupted. For presentation on the platform, lower resolution files must be derived of the master files. While the master files are created once and then left unchanged, the derived files and other data on the platform may change. We opted for daily back-ups of the data on the platform and of the platform itself. Thus, the changes that users make
are frequently recorded, allowing users to work with a previous status of their data. A daily back-up of the platform for 30 days back is considered sufficient. These daily back-ups and the two master file copies will be preserved in a folder system separate from the presentation folder system. In addition, the back-up activities are assessed on a yearly basis. At that point, all files will be transferred to a new storage media. “Migration” is the preservation strategy mentioned by Minerva: “Having created the digitised material, storage media [...] should be refreshed periodically” (2004, p. 40). However, migration may mean a loss, “ranging from formatting or presentation information to potentially more serious forms of loss” (Smith, 2004, Technical Challenges to Digital Preservation and Why They Matter). We do not plan to migrate data into another format, which surely meant a loss. But we are aware that by migrating digital content to another physical carrier we accidentally may lose data. Based on that strategy, losses may accumulate each year.

Everyone should be allowed to use the digital collections and the work results gained from them in the future. Legal agreements ensuring share whilst protecting creator's rights are targeted (COR.I-020.01, mission 9.). As an example material produced in the European Union falls under “public domain” after 70 years after the creator's death (Minerva, 2008, p. 21). The scope of “historical” material is hoped to draw users' focus to public domain material, but public domain material cannot be considered the norm. Further aspects come to mind, namely that digitized material implies copyright (Minerva, 2008, p. 14) of the creator (Minerva, 2008, p. 16). The copyright owner has the economic rights (Minerva, 2008, p. 17) and decides what to do with the material. By a “licensing” model, e. g. the Creative Commons license (Minerva 2008, p. 30-31), the copyright owner grants the right to use the material. The “Creative Commons Attribution NonCommercial ShareAlike” license provides requested conditions (Creative Commons, 2012). The creator maintains the copyright while the material can be used by everyone for non-commercial purposes. The obtained results can be shared, again. Therefore, it was decided that those who wish to contribute to the platform will have to agree to the license (COR.I-020.01, share and intellectual property position). The creator of a digital collection must have the economic right for the objects to be digitized: the right to digitize the collection and to share the digitized items under the above mentioned Creative Commons license. The same is valid for work done on a digitized collection. A creator of descriptions and analysis data must have the economic right over it and be able to share it under the above mentioned Creative Commons license.

However, according to Neiss, a computer scientist of the National library of Sweden, the most important answer to the question “How does data survive the passing of time...?” is “the organization!”, “routines”, “plan[s]”, “checkpoints” (Neiss, 2011, p. 34-35). This will be described in the section “2.3 Provision of services” on p. 15-16.

2.2.3 Intellectual access – metadata

Metadata, “data about resources that can be used to help support a wide range of operations on those resources” (Johnston et al., 2004, p. 22) is intended to be collected to give information about the object, its creation and use (COR.I-020.01, mission 4.). Descriptive metadata of a text may correspond to its bibliographic data (Johnston et al., 2004, p. 23) and may be taken from union catalogs. However, it was considered that not all texts in the scope of EHRIS collections may be covered by union catalogs and have metadata. Therefore, it remains an open question which metadata users may be able to find; the metadata element set may be adjusted in the future to these data. For objects other than texts, identifying information like product numbers should be added. Administrative metadata are data to manage the digitized object (Johnston et al., 2004, p. 24). To identify a digitized object within the institute, a unique identifier must be set. The unique identifier should “stick” to the digitized object; the file name may encode that information. Also considered as administrative metadata are: source, digital provenance and rights management metadata (Johnston
et al., 2004, p. 24). Source metadata on the object's state and a small set of digital provenance metadata on the digitization process are gathered. Rights management metadata are limited to metadata on the physical object and the digitized object.

Minerva suggests the use of an existing metadata element set (2004, p. 42 respective p. 22 of the pdf file), but when we considered the target user groups, we decided against using such element set. Metadata element sets used by libraries are thought to add high level of understanding and future use to a collection. But for our purposes metadata is produced by the collection's creator. The target group of humanities experts may be used to find catalogue data. The other target groups are expected not to find these data. Nevertheless, the collection's creator should be able to find metadata and register the metadata in the platform. A balance had to be found for conveying enough information, and that information in a sufficiently quality for future research, and at the same time, not to overload the metadata's creator. Therefore, we opted for an own small metadata element set. Nevertheless, this set should provide information in a quality and of a level that Yale university library calls “[f]ull: users can perform precise searches, sorting, etc.” (Yale university library, 2008).

To reach that level, “common questions about the things we collect” (Yale university library, 2008) were answered and a metadata element set was created based on the answers received. We listed this metadata element set in document COR.I-020.18. While we hoped that our metadata element set was self-explanatory, we gave short explanations to the metadata that we expect. However, we will observe how the users understand this metadata element set, and we intend to slowly extend and change it to a subset of one of the metadata element sets that libraries use. The group of developers will be available in the future, and adjustments of the platform to changes of the metadata element set are easily possible.

“Critical digitization” results in a huge amount of intellectual metadata: [it] “maximizes interpretation and metadata” (Dahlström, 2010, p. 10). Dahlström also mentions that “scholarly research may need to be embedded in the objects themselves.” For the start of the platform, we decided to add intellectual metadata to text collections. An open standard enabling us to do this is the “text encoding initiative” (TEI) (Vanhoutte, 2004). Standardized terms that express an interpretation are added to text parts. They are called “tags” or “mark-up”. Using the TEI standard means, to actually ‘do research with text’. Style may convey meaning and the TEI standard provides tags to record style indications (TEI Consortium, 2012b). The visualization of information is supposed to facilitate access to information for users (COR.I-020.01, mission 3.). Hence, besides meaningful tagging, style tagging was used, too.

Tagging may be performed in a simple text editor. The resulting TEI encoded file follows the XML standard, “TEI guidelines [...] provide [...] support for XML” (Vanhoutte, 2004, p. 11). Validators to check the correctness of XML files are available free of charge on the web, idem for TEI encoded files (“TEI by example” http://tbe.kantl.be/TBE/xquery/TBEvalidator.xq). Users need no other software than a standard browser and access to the internet. Hence, the application of TEI is technically not difficult. The application of TEI does not necessarily render a collection “critically digitized” in the meaning of Dahlström (2010, p. 10); however, by the use of the TEI standard users are able to produce tagging results respected by the research community.

Therefore, users should be encouraged to use TEI. If one considered the target user groups, the group of ICT experts should know about the practical work with the XML standard. TEI was invented by “humanities computing scholars” (Vanhoutte, 2004, p. 10). Therefore, the target group of humanities' researchers may comprise of “computing scholars” who know and use TEI, but this cannot be anticipated. And whilst HTML encoding is similar to TEI tagging and many users may know about HTML, it can still not be presumed that all target users have the required skills to apply
TEI. Therefore, in future, detailed instructions and examples should be provided.

TEI is an open standard (Gibson & Ruotolo, 2003, p. 57), hence, TEI encoded files can be considered to remain readable in the future. In theory, users' work is performed for eternity. This should be a motivation for users.

Users may want to work collectively on TEI encoded texts. Technically, this requires a collaborative workflow to manage the TEI files collectively. At the beginning only a text's originator can change his/her file. Collaborative workflow may be implemented later.

Intellectual access to objects is also shaped by the presentation of the objects. Presentation context can be neutral, e.g. a grey and flat background of a web page, or context may be chosen "interpretative" (Dahlström, 2010, p. 3), in which case it will direct the visitors' ideas into a certain direction. As we want to attract attention a non-neutral context for the objects should be chosen. It may look similar to online exhibitions on East Asian topics adding a subtle interpretation. For text collections the interest can be further increased by recommending TEI encoding. Thus, we hope to raise awareness within the targeted groups.

All preparation is meaningless if the result is not used. A study of log analysis gave insights into digital humanities collection usage (Warwick et al., 2007, p. 297). The result was alarming: "roughly a third of the resources remained unused" (Warkwick et al., 2007, p. 298). Dahlström summarizes various factors (size, interoperability, machine interaction and human interaction) as decisive for the success of large-scale projects in the area of humanities (Dahlström, 2011, p. 20). The platform is a small project, but it was designed for a large amount of data. If we use similar criteria for the use of the platforms' collections, then size is hoped to be attained by the platform's support for digitization tasks. Interoperability is ensured legally by the Creative Commons licensing model, technically by use of open source software and open standards, and in the context of exchange with the scholarly world by using TEI. Machine interaction with collections is made possible through the processing of the TEI files, that is, by the use of XML technologies. When the platform goes live, human interaction is enabled only in form of use of TEI encoded files of other users, and as interaction of users with the support team of the platform. For the future direct interactions like in social networks may be intended.

Summing up what has been said, we can note that the level of digitization on the technical side is not high but acceptable. However, we hope that the level of digitization on the intellectual side could be high, under the condition that it is carefully done. This level depends on the person who does the digitization work. The platform shall ensure that a minimum level is guaranteed.
2.3 Provision of services

2.3.1 Descriptions as QMS and visualization

“Quality management system” or “QMS” is a method to describe an organization's activities in the form of processes. “Quality” was seen as “consistency of outputs” (Beckford, 1998, p. 237), whereby outputs were the institute's services and the users' digitization work.

We decided to describe the provision of services, and therefore, the organisation of the institute, as QMS. The QMS was set on top of activities, conceptually, as well as visually (figure 1). This idea was described by e.g. Zhang et al. (2009, p. 701) and it was used in documentation COR.I-020.10.

In 1998, “the ISO 9000 series [...] the [...] most widely used of [QMS] systems” (Beckford, 1998, p. 236). ISO 9000 consists of terms and definitions for QMS, and ISO 9001 describes requirements for QMS that are the bases for formal assessments. Since “there was a trend [...] particularly in South East Asia, to only deal with quality accredited organisations” (Beckford, 1998, p. 237), we expected that now, more than 10 years later, companies in South East Asia have introduced this standard. Therefore, people in South East Asia who are an important target group were expected to be familiar with the ISO 9000 series standard, and thereof, of ISO 9001-based quality management systems. This is one of the main reasons for the use of an ISO 9001-based QMS. In addition, thanks to its popularity, the use of a QMS may also serve to represent the institute in the web; even that a formal assessment and certification (Beckford, 1998, p. 239) was not foreseen for the start up.

The ISO 9001 standard lists elements that a QMS should consist of (e.g. Idris & Ahmad, 2011, Fig. 2). The 7th section “Product Realization” corresponds to core processes. These processes refer to activities to fulfil a client's requirements in an business environment. We refer these activities to activities of the platform's users. The 6th section “Resource Management” corresponds to processes that support the core processes. The 5th section “Management Responsibility” comprises coordinating or decision making processes.

The QMS concept was followed. Core processes are digitization processes. That is what the community does. Support services surround community's activities (COR.I-020.13). These supporting processes are carried out by the support team and comprise of the provision of the archive and user support. Management responsibility processes correspond to processes of the institute's board, the founding members. Activities of the 4th section and of the 8th section “Measurement, Analysis ..” of the ISO standard (Idris & Ahmad, 2011, Fig. 2), are used in all processes. Some are implemented as part of the board's processes.
The writing style we applied was “the one page format”. Activities were described to be visible at one glance, within one DIN A4 page. “Flowcharting is a good method to use with procedures.” Most people intuitively understand flow charts (Wright, 2001, p. 59). The QMS was drawn as flow charts, using standardized drawing elements. The drawing elements, “metaphors” (Marcus, 1994) or icons, are listed in the “legend”. A subset of the ISO 5807:1985 flow chart set was taken and a few additional icons were created.

In the platform, the flow charts are accessible via the main menu. Each flow chart links to the “legend” page (figure 2) that explains how the flow charts are to be understood: Steps of an activity are drawn as squares. Directed edges connect the steps and show the work flow. A rhombus indicates that a decision must be taken during the work flow. Flow charts are connected by linking. A user ‘clicks' onto a “predefined process” step and is forwarded to a more detailed description of the current activity. Similar connections were provided for documentation that needs to be accessed during a work flow. A grey waved icon provides a link to a template or to an explanatory document.

These work flows were created for all involved parties: the community, the support team of the institute, and the “institute” as decision making body. A user starts and performs an activity. As a consequence the support team performs an activity. Zhang et al. show how the output of activities of one party becomes the input for activities of another party (2009, p. 700, Fig. 4).

We decided to add the QMS to the platform. Since the work flows are part of the platform and have been made publicly accessible, they follow an open information policy (COR.I-020.01, Introduction, “framework for […] openness”).

A document control system ensured that all these information flows were recorded. Documents were listed in the “document list” (INV.I-020.11) with a set of metadata and identified by a unique identifier.

According to the ISO-9001 principles, documents can be divided into “documents” and “records”. “Documents” have directive character. They split the policies into executable actions. “Records” are the notes of these actions. To follow naming convention, “documents” start with “COR” or “TMP”. Information from these documents must be transferred to the support team by training (COR.I-020.12). Documentation starting with “INV” means “records”, like the software engineering documentation for this master's thesis.

2.3.2 Evaluation and service quality

There are operational evaluation processes at the EHRIS: the PDCA cycle and user satisfaction assessment.

Service quality shall be ensured by the QMS (COR.I-020.10). The “Plan-Do-Check-Act” cycle (PDCA) is an integral concept (Beckford, 1998, p. 67). The board, namely the “institute” sets key performance indicators and defines the policies (“Plan”). The policies are transferred to the support team and to the community. This transfer is achieved thanks to the provision of the drawings and the documents as well as by training. The various groups of the team provide services (“Do”). Evaluation (“Check”) may uncover issues that relate to users' satisfaction or to the support team's performance. Any issues are subject to corrections (“Act”). Each team is responsible for taking appropriate actions. In order to avoid issues arising again, corrective actions must be reflected by
the policies. This influences the strategy and is discussed in the annual review meeting (COR.I-020.11) by the institute's board.

Data regarding “user satisfaction” (COR.I-020.24) are being collected (TMP.I-020.13). Once per year the community is invited to give feedback through a survey (TMP.I-020.01), and the support team is encouraged to give feedback during the refresh training (TMP.I-020.04). Support team notes on daily actions (TMP.I-020.07) and the result of the yearly archive check (TMP.I-020.10) show the service quality. Figures highlighting the achievement of the strategic objectives can be derived from the database's data.

Attendees of the meeting are the institute's founding members. They themselves belong to the three target user groups of the institute and are experts in various areas. They are invited to give input on changes of law and research and technological advances in their areas (COR.I.-020.11). Attendees are encouraged to suggest at least two ideas to enhance the service. Actions that are feasible within the given budget will be implemented. The revised policies and strategy will be published.

The use of the PDCA cycle within a QMS is shown by Li et al. (2008, p. 1064, fig. 5 “Running model of quality workflow”). Each individual unit follows the PDCA and so does the organisation as a whole. Following the principle of “continuous improvement” ensures that the entire system improves.

3 The platform as integral part of the strategy

3.1 The platform's functions

The platform of the EHRIS is created to provide the following functions according to the institute's strategies:
1. Digitization of the objects
2. Creation of collections
3. Creating context for the objects
4. Displaying collections and objects created by users
5. Managing the collections over time
6. Encoding the text collections by using TEI
7. Managing the descriptions of the quality management system.

The following categories of users have access to the platform with different permission settings:
Everyone has access to the platform on the web and can browse the platform and the collections. Creators of the collections can edit collections and objects on the platform. Creators of the collections can create and edit TEI encoded files for a collection.
Administrators of the platform are allowed to edit all collections and objects on the platform. Administrators need special permissions to support users on the platform and to correct content that is not conform to the institute's policies. In addition, administrators can set user roles. Users who were given a special role for a collection are allowed to edit particular metadata.

The following tools were used to implement the platform:
We decided to develop the platform by using open source software PHP and MySQL. Technology XSLT was applied and TEI version P5 and UTF-8 were used to encode and mark-up text collections. The reasons behind these decisions are explained in appendix A.

We decided to carry out an initial analysis on the TEI encoded files for text collections. As part of the overall strategy, three features were implemented. The platform supports the creation of a TEI
file for a text collection. Another feature was the display of formatted text. The original text may be in any language. If a translation into English exists, it should be displayed with an appearance similar to the original text's format, to give the user an idea of the original in a language that he/she understands. This should be achieved by most simple means. The third feature was an additional search facility. The South East Asian area covers many languages. One uses a language that one knows (English) to do text search, and should receive results of texts written in other languages.

3.2 Choice of development methodology and documentation

The platform was created in a small company. “There is no 'ideal' software process” (Sommerville, 2007, p. 64). The degree of formalism (Erdogmus, 2005, p. 77, figure 1) is high in sequential approaches like the “waterfall” model (Boehm, 1988, p. 62, figure 1), and less in iterative models like the “spiral” model (Boehm, 1988, p. 64, figure 2). Least formalism is found in so called “agile” approaches (Ambler, 2002). Agile methods are incremental and allow fast “responding to change”. Since agile approaches restrain documentation, they rely heavily on communication, - “tacit knowledge” must be shared (Boehm, 2002, p. 65-66).

The company's standard process for software development is a simplified iterative process (INV.020.009, section 4.2.1). Agile approaches have previously not been considered. The company is small and people communicate well with each other. However, knowledge got lost, when developers left the company. We thought the risk was too high to give up documentation entirely. We also decided against an incremental approach, whereby, a developer would only develop a small part of a requirement, and continued to work on it in a next iteration; while during the current iteration, the developer had to switch to other requirements. We realized that our developers did not have this kind of flexibility. The developers prefer to work on one requirement until its completion. We must also say that our developers were experts in their own technical fields who had difficulties switching quickly to requirements of other technical fields. Therefore, we opted for a non-incremental process. Respectively, an increment corresponded to a package of requirements that transformed the whole software into a higher version.

We decided on two deviations of the standard process (INV.020.009, section 4.2.2). Firstly, human resources for testing were not available till 2012. Therefore, rigorous testing was postponed to the end of the development process. Secondly, the intended platform was a web application and developers were considered to picture the architecture. Therefore, only an overview of the architecture was described (INV.020.001). Additional requirement information was conveyed as a “prototype”. “[U]ser-interface prototyping can serve as the baseline for developing the software […] specification.” (Kelly & Neetz, 1988, p. 647). “Screen mock-ups” are images of an application's interfaces that the user will interact with (Hammar Cloyd, 2001, p. 67, figure 3). Thanks to the use of screen mock-ups, requirements were gathered, the web design's “look and feel” was created and functional specifications were defined. This way one could start at a stage as if one had gone through the development process several times and our development took up speed. In addition, the developers' review of the mock-ups served as first usability tests. If the developers did not understand the requirements, we had to consider whether the interface functioned satisfactorily usable.

Development iterations were planned and the requirements were as known and fixed right from the start. However, the future strategy (INV.020.012) states a “switch” to the Java technology. The platform will be re-developed, within another programming environment. Fazit, the development of the platform can be seen as a “throwaway prototyping” approach (Sommerville, 2007, p. 68); that follows the agile development like explained above.
The software production was controlled (Kruchten, 2004) by the collection and analyses of “quality data” (INV.I-020.009, section 4.8, section 5.3). Besides, “configuration management” (INV.I-020.009, section 7) set guidelines to identify working items during the development process (Kruchten, 2004). Working items were the documentation and pieces of software. Meta procedures, “audits”, to control the configuration management procedures (INV.I-020.009, section 7.4) were not considered as the project was small.

3.3 The platform's web design (look and feel) and presentation

Editors, libraries or in general providers of information services, must think about the presentation of a collection. They may take an objective view, “not intervening” (Dahlström, 2010, p. 3), or they may take a subjective position and draw users’ attention to particular aspects of a collection. The design of the presentational background of an object, the web page's “look and feel”, is an interpretation because it will suggest to the user a certain context. We opted for the subjective position to remind visitors of the origin of the objects: South East Asia. Therefore, the context was kept subtle. The object's digitization elements (transcription, translation and metadata) take the main space of a web page. However, by adding any interpretation, collections became context-bound and may become less reusable (Dahlström, 2010, p. 3).

Insights from user behaviour research were applied. As much information as possible is displayed at once, without the need for “scrolling”. However, limits were given by the fact that we required a web page's size to be shown on the (small) screen of a netbook. Still, by showing as much information as possible under these limits, we hoped to increase “serendipity” findings (Thelwall, 2004) or “information encountering[s]” (Erdelez, 1999). The information on the platform for one object or collection may be narrow. “Serendipity” is not supposed to find information that was not on the radar, but it is meant to stimulate imagination for the creation of own collections.

Information concepts may be transferred as icons or “metaphors” (Marcus, 1994). Icons were used where ever possible. Icons replace lengthy text documentation. The information as a whole can be kept short. A big advantage is the universality of an icon. The users' native language may not be the language of the platform, which is English. Icons are not totally language independent, but we expected them to be easier understood and memorable than if we used natural language.

The “Online collections” page is the main view of the platform (appendix B, figure 5) and it shows as much information as possible at one glance. At the top, a row of thumbnail images represent the collection's objects. Users browse the collection by 'clicking' on the thumbnails. The object is shown below the thumbnails. This page arrangement was designed after the Codex Sinaiticus web site, “see the manuscript” page (http://codexsinaiticus.org/en/manuscript.aspx).

On the left-hand side of the platform, the captured image is shown, while metadata are displayed on the right. Icon buttons switch the content of the metadata field. The default setting is the “information” view, icon letter “i”, that shows metadata (appendix B, figure 5). The document icon switches to the transcription. The icon “EN” for “English” allows switching to the translation. Users may like to see transcription and translation at the same time. Therefore, when one button is active and the other is pressed, the metadata field splits vertically (appendix B, figure 6). The upper field will show the transcription and the lower field the translation. In our platform, the icons with letter “i” for the term “information” and “EN” for “English” stand against the language independence of icons. However, we expected the terms “information” and “English” as satisfyingly “universal” in use for our target groups, who were expected to have a least basic English reading skills.
The camera icon is used to switch to the view to do and to show coding and analysis of TEI/XML format instructions (appendix B, figure 7). This idea was designed after the arrangement of the web site of the Deutsches Textarchiv (http://www.deutschestextarchiv.de/oken/biologie/1805/viewer/image/text/40/16/), with options to switch between views of transcription and TEI encoding, while a digitized object stays on the left-hand side of the webpage. In the platform, the captured image is shown on the left-hand side. On the right-hand side, the translation of the transcription is displayed; similarly arranged as the original text on the captured image.

A “mega menu” was used to select collections (figure 3). A vast amount of information is shown at once to increase the above mentioned “serendipity”. Bates (2007) describes users' browsing behavior as follows: users do not browse in a straight-forward way, instead their “browsing consists of a series of glimpses”. Users vaguely scan information that is beyond the information they focus on. For the start of the platform, the mega-menu will not show much information. We hope that later the mega-menu will lead users to collections they have not thought about.

The mega menu has three pages (“tabs”) to help users to preselect the search for a collection; this helps to giving ideas to find interesting collections. A collection's metadata set has metadata “industry”, “type” and “country” (COR.I-020.18). The collections are listed according to these metadata in the mega menu. One tab lists the collections by country of origin. Another tab shows the collections listed by “industry”. “Industry” is intended to describe the context in which the physical objects were used; options are “service”, “trade” and “manufacturing”. A third tab page shows the collections listed by “type”. Text collections are of type “text”. The terms “concreta” and “abstracta” were intended to be used as follows. “Concreta” may be used for tangible items like machines, tools or furniture; while “abstracta” are intangible items like ideas or concepts, an organizational chart or a schedule. It might be interesting to see what creators of collections will find for “industry trade”. We hope that this categorization will lead to innovative ideas for collection topics.

3.4 The platform's user support and guideline

The platform supports the digitization of cultural heritage collections. A set of self-explanatory guidelines were meant to be laid down to provide directions for users, so that all target users groups could quickly understand it. Existing guidelines like the ones used by the National library of Australia (n.d.), Minerva (2004) or Deutsche Forschungsgemeinschaft (2009) are text based and detailed. Therefore, we decided to create our own guideline. We envisioned a guideline based simply on the steps that are considered the most important for digitization. To make it easy to
understand this guideline, we wanted to create it by means of graphical visualization.

As described previously, the organization of the institute was described as QMS. This QMS consists of three sections, the decision making part for the management, the service provision part for the service team, and activities that the user (the “community”) does. The QMS was depicted as flow charts, which were considered to be highly visualized. Therefore, the section of the QMS that comprises of user's activities was adjusted to serve as guideline for users at the same time; explaining how to interact with the platform and how to do other activities in the framework of the institute. The QMS descriptions are linked to and reachable from within the platform. Hence, the guideline is easy to find for present users and will be informative for future users. The guideline is part of the QMS which itself is part of the policies, and therefore, subject to controlled change and review processes.

The next paragraph describes the interaction of users on the platform and how they are being led by the guideline, to do digitization work, and supported by the support team of the institute.

3.4.1 Collection digitization

Guideline
Users are responsible for their collections. Collections must adhere to selection criteria (COR.I-020.04). The selection criteria refer to the institute's “code of ethics” (COR.I-020.01, code of ethics). Both ensure that users behave within a legal frame, particularly for copyright issues, and that users treat each other and others' collections respectfully. Users shall make sure not to use embarrassing content, images or text. South East Asia consists of various cultural areas and languages. The community should be familiar with the various cultural values and languages and needs to get engaged to control themselves. If the community comes forward with a complaint about a collection, the collection is withdrawn from the web (COR.I-020.14).

Digitization of the objects (COR.I-020.05) starts with selection (COR.I-020.03). Selection criteria are listed in detail in form of a check list (COR.I-020.04). A collection must meet the institute's scope statement and the creator must have the right to digitize the objects and bring the digitized objects under the Creative Commons Attribution NonCommercial ShareAlike license (Creative Commons, 2012). It is explained what that meant for the user: that everyone can use one's collection for their own research. The image of a selected object is made accessible and metadata are found and set. A list outlining the metadata element set for a collection and for an object is available for download (COR.I-020.18).

The sub-process to make an image accessible starts with the image capture (COR.I-020.06), and the resulting image is the “master”, which shall not be changed any further. Recommendations to capture a master with archival quality are given in the work flow drawing. The formula to calculate the maximum size of an object was taken from Deutsche Forschungsgemeinschaft (2009, p. 7-8) and was depicted as image. The settings of Terras (2008, chapter 3, chapter 4) are recommended. Explanations are given in document COR.I-020.21. Master images may be sent to the institute for preservation. Then, the support team, “EHRIS”, is responsible for “preservation (archive)” (COR.I-020.15).

The user creates low resolution images of the master file in .jpg format and uploads the images to the platform. Now, the “presentation (web site)” (COR.I-020.14) of the images is the responsibility of the support team.

Platform
A log-in window requires username and password and the user is then transferred to page “admin
collection” (appendix B, figure 8). On this page, an existing collection is chosen or a new collection is created. Small images at the top represent a collection's objects. Below, the collection's metadata are listed. The collection's title and the search criteria industry, type and country are required fields.

Via the button “add object to collection”, one switches to the object administration view. We presume that often the collection's metadata will correspond to its objects' metadata. Therefore, a collection's metadata are transferable to an object (button “take data from collection”). Important metadata cover usage rights. The field “objects_permission” must be set to “Creative Commons Attribution NonCommercial ShareAlike”.

Three web optimized images can be uploaded per object. These images are shown on the platform. Image file names that correspond to the platform's architecture are set automatically by the platform after upload. After saving the object data, a small thumbnail image represents the object. This is located on the top of the collection page.

The user may also want to encode text. By clicking on the “camera” icon, the user switches to the “object capture” view (appendix B, figure 10). This view supports the mark-up of the translation or transcription text. XSLT transformations are performed in a sandbox.

Support service
“Presentation support” (COR.I-020.14) is given as and when a user uploads or changes an object. The support team assesses whether the object is technically accessible, and whether it complies with policies. The selection criteria (COR.I-020.04) are taken as reference for the policies. Objects that do not comply are removed from the web and “quality measures” (COR.I-020.23) are taken.

3.4.2 How to encode text

Guideline
Process COR.I-020.07 is the procedure to make text accessible. It explains the work with the platform for the transcription, translation and XML/TEI coding and analysis. Transcription and translation into English are uploaded to the platform. With one “button-click”, the platform creates a TEI encoded file for a collection. The user may download the file, adjusts it and upload it back onto the platform. Recommendations for text preparation and TEI mark-up are provided (COR.I-020.20). Text with style instruction is displayed as per instructions listed in the provided default XSLT style sheet. Users can adjust the style sheet to use particular formatting instructions. Thereby, if users' changes do not conform to the TEI or in general XML standards, users will not achieve a formatting of their texts; may an “empty” page be shown. However, the XML/TEI transformations were implemented in a manner that they only impact the display section of a web page; not the entire web page and thereof, not the application as a whole. That is, users' actions impact only the display of their own collections. In addition, users who wish to reset their instructions can do so and reset the setting to the default XSLT style sheet of the platform.

Platform
The user must log into the platform and has got access to all information of collections that he/she has created. For these collections, the user has the role of “owner”. In addition, a user has got access to information of collections that he/she was added as “transcribe” or “translator”.

When being the “owner” of a collection, one may use the support of the platform to create a TEI encoded file. The head part of the TEI encoded file, the TEI Header, contains administrative and descriptive metadata about the collection (TEI Consortium, 2012c). These metadata correspond to the data that a user inserted when he/she added the collection to the platform. Therefore, the TEI
Header was created out of the previously inserted metadata. As much metadata as possible were used for the TEI Header (appendix B, TEI Header, created by the platform). The TEI file creation support is optional. Users are welcomed to upload a TEI file. Users decide whether they wish to share their TEI file. If yes, a link to that file will be displayed in the “collection view” page. Per default, the file is not shared.

It was one of our targets to encode and analyse TEI/XML files on the platform. English text should be displayed in a manner as close as possible to the original text. Formatting mark-up may be added to a translation text. This mark-up is analysed by a XSLT style sheet; one such XSLT style sheet is applied to all text sections (“objects”) of a “collection”. Users may utilize the platform's sandbox (appendix B, figure 10) to adjust the mark-up until a desired result has been achieved.

Support services
“Research and share” (COR.I-020.08) provides an overview of user's activities. The “user support” (COR.I-020.16) starts when a user sends a request or proposal to the institute. Requests that comply with our policies are executed. Each year, a survey is being performed with the community (TMP.I-020.01). To record the daily work and to be reminded of perpetual tasks, the support team makes use of the “daily mnemonic” check list (TMP.I-020.07).

3.4.3 Collection archival

Support services
“Preservation” support (COR.I-020.15) is offered. The support team receives a storage media with the master images of a digitized collection and transfers (copies) the data to a long-term storage media (process “archive”). “Daily back-up” is performed each working day for the whole platform. The daily back-up on the 10th of each month is transferred to the long-term storage and the archive consists of two copies of the master file collections and of one daily back-up of the platform each month. This concept for one collection means that the not changing master files are preserved as a mirror, and the “living”/changeable part of the collection, the files on the platform, are preserved monthly.

It was decided ahead of time that the daily back-ups would be kept for 30 days. In case it turns out that this is too much the back-up frequency could be changed from daily to weekly.

The archive’s accessibility is checked on a yearly basis (COR.I-020.15). The platform in the archive and the latest copies of the platform in the short-term storage are checked for accessibility and functionality. These checks are to be carried out by a professional tester (TMP.I-020.11). The overall annual check (TMP.I-020.10) will be discussed in the annual board meeting (COR.I-020.11).

The support team uses the “daily mnemonic” check list (TMP.I-020.07) to record the daily preservation work.

The work flows (COR.I-020.15) show the migration of the archive’s content to another long-term storage once a year. In the beginning, the archive consists of DVDs. Migration is the transfer of data from one set of DVDs to another set of DVDs. A more powerful storage system is envisioned for the future. In which case the question may arise, whether yearly migration makes sense, as a more powerful system is costly. Migration from one powerful costly system to another powerful costly system may not be achievable.
4 Evaluation of the strategic decisions, via use of the platform (“me”)

4.1 Evaluation during and after development of the platform

Technology
The developers prepared the environment. Then I prepared and tried out the XSL transformations (INV.I-020.006 and INV.I-020.007) on the platform. The resulting power of XSLT was astonishing. One could say that 'XSL transformations' are a powerful programming language.

A few issues required other solutions than planned. With the chosen programming language and programming library, it was not possible to do programming work with the TEI attribute “@xml:id” that references unique identifiers. The unique identifier of objects, however, must be referenced. Therefore, the developers used another attribute (@n). However, by TEI standard, this attribute serves the purpose to indicate an order of elements, and hence it is not necessarily unique (TEI Consortium, 2012a). So we decided to narrowed down the definition of this attribute. As a consequence, the TEI collection files became “conformable”, but not “conformant”, mark-up does not adhere to the defined semantic meaning (TEI Consortium, 2012d, 23.3.3.1 Semantic Constraints).

We planned to use pure XSL transformations. The XSLT style sheets should be available on the platform, so that users could apply them as they are. However, when search criteria were set by a user, these parameters had to be transferred to the XSLT style sheet. Meaning that one section of the style sheet is rewritten by programming and this solution would have required more technical background from our users than initially planned. Hence, the question how to enable an easier work flow remains to be solved.

Creating a TEI encoded file with the data in the database was no problem. The resulting TEI file should adhere to the definitions of the TEI Lite standard. These definitions have been established by the TEI Consortium which is the authority for the development of TEI. We decided to use the recent “schema” format (http://www.tei-c.org/release/xml/tei/custom/schema/xsd/teilite.xsd) as reference. However, a validation of the TEI file was not possible. When we changed the reference to the formerly used DTD file (http://www.tei-c.org/release/xml/tei/custom/schema/dtd/teilite.dtd), the TEI file could be validated. This is still an open question. Similar thoughts had influenced the decision to develop our own platform instead of using an existing solution (please see appendix A, “Decision for in-house development and considerations concerning open source software”). TEI is an open standard. The open standard community was neither quested to provide support, nor to work with industrial quality. It is important to bear this in mind to calculate how much effort from the institute's support team may be required.

Since it was possible for the TEI authority to provide a broken reference-file, another lesson may be that people do not reference these web files. It might be interesting to look at the reasons behind this and take this into consideration when preparing the annual survey.

Workflow and control
Once the platform had been used, it became clear that the various options to create TEI encoded files (INV.I-020.006) were not needed. Although, we realized that we did want to have information about the work on the TEI files i.e. the indication of its origin (“creation”, “upload”). The platform's interface was adjusted to that work flow (appendix B, figure 11) (INV.I-020.010).

When a new file was created, the existing file was “caught” by the developers before replacement.
This functionality proved to be very useful. We decided to maintain this function and it is still available on the platform.

Another issue arose with the file control. As soon as a collection was available on the web, the TEI files were downloadable for the community. A link to the file was set and shown. We reconsidered this setting. The files should remain available for the XML search, but the files should no longer be shared automatically, because preparing these files is tremendous work. The control of sharing permissions should be in the hand of the users.

XML/TEI encoding and analysis
An English translation should be provided and displayed in the same format as the original text, to convey a native speaker's experience. We made some attempts with text collections that were in Vietnamese.

XSLT style sheet examples were adjusted following instructions that could be found on the web. Appendix B, figure 7 shows the application of character capitalization transformations. The transformation of the Vietnamese text could be achieved by simply adding Vietnamese characters to the range of existing characters. The transformation also worked with a text of mixed languages (French and Vietnamese), which are encoded in different Unicode spaces (please see appendix A, “Decision for encoding UTF-8”). Appendix B, figure 10 shows the capitalization of the French letter “ç”, surrounded by text that consists mainly of Vietnamese letters.

A default XSLT style sheet is provided on the platform. It is hoped that even users without background knowledge of XSLT may find it equally easy to do simple adjustments, just by adding terms or characters to the default style sheet.

Therefore, technically, the idea could be implemented as planned. However, this is still subject to consideration. We presumed that users would mark-up both the transcriptions and the translations. But as tagging causes considerable work, users may add mark-up only to transcriptions. If, however, users added format mark-up to the English translations and meaningful mark-up to transcriptions; the intellectual metadata would be split. For now, if no translation is provided, the transcription file will be displayed according to format instructions. However, this way the original idea of transferring a native speaker's experience cannot be achieved.

A question remained on whether mark-up could provide a method to do cross-language search. Transcriptions and translation texts should be encoded equally. The translation's sentence structure may be rearranged to allow the order of tags to follow the order in the transcription text. A few transcription texts in Vietnamese, and their English translations were encoded in this manner. The search functioned with XSLT.

The search results list will appear as follows (figure 4): “Original” is the marked-up text, “element” is the mark-up, “meaning” is the translation of the “original”. “Element” is the search criteria, like “role”, “place”, “event”. If an “element” consists of two parts, these are highlighted by different colours.

For example in object number 000005, an encoded term of “element: role” was found. Original text was “Giám Đốc Nguyễn Văn Thế”. The translation (meaning) of “Giám Đôc” is “director” and the transcription (meaning) of “Nguyễn Văn Thế” is “Nguyen Van The” (figure 4).
Whilst in principal the technological implementation was feasible, thanks to the user's effort, the original idea proved to be impossible. The TEI transcription file and the corresponding TEI translation file had to be encoded in the same manner. Similarly to the experiences above, we didn’t take into account how much effort it would take to mark-up two files. That is why the idea was changed. Only the transcriptions file was encoded and the translation was added as mark-up to a term. Since the TEI Lite standard provided no elements to fulfil exactly our purpose, we needed to narrow down definitions here. As a consequence, the TEI files also became “conformable”, but not “conformant” to the TEI definitions (TEI Consortium, 2012d, 23.3.3.1 Semantic Constraints). As we were confronted with this issue twice, one might wish to reconsider the choice of technology as well as the underlying concepts.

4.2 Evaluation during the digitization of collections

Initially three collections were digitized. Two of them were text collections; one was in Vietnamese and the other was in multiple languages Vietnamese, Chinese and French. The third collection was an image collection with some text in Vietnamese, Chinese and French. We decided not to transcribe and analyze the Chinese text, since the expertise was not available. The collections were assessed by the selection criteria (COR.I-020.04) and proved to be within. The collections were in the scope, they had more than three objects (pages) to digitize, and no object of the collections had so far existed on the platform. In addition, we verified that the “code of ethics” (COR.I-020.001, code of ethics) was not violated by the objects. However, questions as to who would be the right-holders were not clear. We shall come back to this later in the section “5 Discussion” on p. 29-30.

Digitization

The chosen collections were bound together as thin books, which were then scanned. The book's pages could not be taken out, since no trained professional would have been available to re-bind them afterwards. Damages did not occur during that procedure, but the books' bindings were loosened. This experience had not been anticipated, and no recommendations for such an event had been given in the guidelines. This proved that the expertise of a trained conservator is urgently needed.

The scan allowed us to capture high resolution master images. These were treated with simple
operations (cut, rotate, simple colour adjustment) and low resolution images were derived and uploaded to the platform. The resulting image quality was satisfying enough for presentation on the platform. Nevertheless, a professional digital image treatment would be preferred and thus, this remains to be discussed. An expert would also be able to supervise the creation of a guideline for digital image treatment. The display image creation for these first digitized collections may be repeated.

Transcriptions were done manually. Optical character recognition (“OCR”) (Tanner, 2004) of Vietnamese text is possible, but errors are numerous. “Extra dashes” on special characters đ, ơ, ư, or pronunciation signs like ạ, ả are taken as noise respectively, the particular character is not recognized. The text encoding was not a problem but time-consuming. Software Unikey (http://www.unikey.org/) was used for the transcriptions. Keyboard key combinations resulted in particular Unicode encoding.

After transcription and translation, the transcriptions were encoded. The support of the platform to create a simple TEI encoded collection file was helpful. “The biggest jump in cost comes in the move from an image front transcription to a carefully edited and tagged transcription” (Rydberg-Cox, 2009, methods). The file then underwent further mark-up. It would be an overstatement to say that the resulting file was “carefully edited”. But still, the fact that the collection-file was “ready” for mark-up, that no further thinking was required for the TEI Header and that each object was “prepared” in its particular <div> element; all this saved a lot of time, not only in terms of the work being actually carried out, but also with respect to the preparatory thinking processes. It was planned that a collection's translations should also be enriched by TEI mark-up. However, TEI encoding took considerable time, even with the platform's support. Based on our initial schedule, enrichment of the translation text by intellectual metadata was not achievable.

Metadata were collected based on the list of COR.I-020.18. One question arose with regards to the “language” indication. Objects may be of multiple languages. This can be highlighted on the platform: language=mul, as it is used in the ISO 639-3 standard. However, this “solution” makes a language detection impossible. Thus, some reconsideration may be required regarding this aspect.

Two collections were “books”. They had a back, a front and even a serial number. But they were not registered within a union catalog and we had to look up bibliographic data. However, this was not difficult. As one might expect, these books mentioned the author, year of publication and the printing agency's address.

The economic rights of these books could not be clearly attributed. They may be orphan works. The books are of the years around 1950 and were produced in Viet Nam. Since then, legal, economic and information infrastructure changes occurred. Minerva mentions that agreements on orphan works exist between countries (2008, p. 23). The country of origin, in this case Viet Nam, may be partner of an agreement or maybe not. Thus, the collection's metadata were set as: “Work still in ©, right-holder could not be traced”. These metadata are visible for the relevant collections on the platform. In addition, people visiting the platform are encouraged to contact the institute if they knew a right-holder. The “code of ethics” demands to “refrain from undertaking[s] which can be considered to be unethical.” (COR.I-020.01, code of ethics). In case that an economic right holder exists and comes forward with a request not to share his property, we hope that we would be able to come to an agreement. Besides, we would hope that if we came into contact with a right holder, more background information of the work could be gathered.
4.3 Short evaluation by representatives of target user groups

Target user groups were described in section “2.1.2 User characteristics” on p. 6-7. Due to time constraints for this master's thesis a profound evaluation could not be carried out. However, we did manage to briefly interview representatives of all three target user groups (ICT and humanities experts as well as non-experts; among them people who live and work in the area of South East Asia (SEA)). This allowed us to gain insights into the value of the platform.

ICT experts reviewed the software engineering documentation and implemented the platform. Their feedback served as first usability tests, which lead to adjustment of the platform already during development phase. The group of developers understood the objective of the platform and its intended use. The developers had some knowledge of XML and understood TEI as particular application of XML. The fact that they enjoyed the design and arrangement of EHRIS web page, that they were motivated to make their own suggestions, and that they were keen to try out XSLT transformations themselves, may suggest that this target user group can be reached.

We got an informal interview with Dr. Andrew Hardy, the representative for research in the humanities, a historian and head of the Hanoi center of the EFEO (Ecole française d'Extrême-Orient). Dr. Hardy's research focuses on the “Long Wall of Quang Ngai”, built in 1819, which became designated “Vietnamese national heritage” in March 2011. The project “History and heritage of central Vietnam: research and training project on the Long Wall of Quang Ngai”, “is a cooperation between the Vietnam Institute of Archaeology (Vietnam Academy of Social Sciences) with the participation of a team of European and Vietnamese researchers and students” (EFEO, n.d.). The EHRIS platform was designed for that kind of project, given its scope and its participants.

Today, only few archives that document the construction of the Long Wall exist. Therefore, Dr. Hardy's research relies on multidisciplinary research in archaeology, anthropology and cartography, involving fieldwork and the gathering of traditions and memories of residents in Quang Ngai. The fieldwork done over the last seven years has resulted in a large amount of data. These include transcriptions of Dr. Hardy's interviews (in Vietnamese) and recordings, additional GIS data, photographs and documents. Consulting with an ICT expert was not helpful, as the ICT expert wanted to know the structure of Dr. Hardy's data ahead of doing the actual research, while, according to Dr. Hardy, such structure arose only over time during the research process. For example, to administer the photographs, Dr. Hardy had started to use a software tool. The tool proved to be useful because it allowed to set keywords for a range of pictures instead of for one photograph only. However, Dr. Hardy's workload as head of the EFEO in Hanoi set a limit to the time he could afford to work with such tools.

When I showed him the TEI mark-up and analysis of EHRIS' lab, Dr. Hardy said that this could be a solution to administer his data. Dr. Hardy knew about mark-up and the use of keywords and therefore, understood the TEI encoding of the Vietnamese text. It was immediately clear to him that such encoding enabled easy search facilities, and would be highly valuable for statistical analysis. With statistical analysis I mean to say questions like: how many people in Quang Ngai had said that the wall was built by a particular community of people. This kind of analysis, enabled by “button clicking”, would speed up the publication of research. But Dr. Hardy also made a remark that the preparation (the encoding work) needed time. Therefore, I do think it would be correct to say that he does see the value of a tool like EHRIS, as long as the project continues for a time period long enough to justify the preparatory work.

Because of Dr. Hardy's time constraints, an extensive demonstration of the use of EHRIS was not possible, but he did reassure us that the Vietnamese population is now very computer-literate. He
said that, nowadays, all of his students used the web and its tools extensively, even if their English skills may not be that good. Dr. Hardy believed that the use of a platform like EHRIS would pose no difficulties.

In conclusion, a platform like EHRIS seems to fulfill the needs of a researcher’s project that incurs a high amount of (fieldwork) data and involves various user groups. The platform is available via the web. It is simple and it focuses predominantly on the most important data and functionality. The encoding and simple analysis (by e.g. XSLT) enabled quick access to results of the fieldwork. In addition, the use of simple ICT functionality to support a researchers’ work might bring added value for scientists. The EHRIS platform supports the transfer of a collection's metadata set to its objects. This functionality is comparable to the tool that Dr. Hardy chose to maintain his photos.

I fully acknowledge that a platform like EHRIS must be set up at an early project stage. This will then allow for data regarding the publication work to be already available in the next stage.

Besides, I asked Dr. Hardy about his experiences with copyright protected material in Viet Nam. According to Dr. Hardy, older materials are mostly under public domain and the cooperation with the National library of Viet Nam was good. Everybody had access to their materials; so the EFEO.

Dr. Hardy explained that nowadays copyright law was followed strictly in Viet Nam. However, as a lot of the research material of the EFEO was under public domain or produced by the EFEO, that rarely lead to issues.

5 Discussion
In this master's thesis, we developed a strategy to support cooperation of ICT and humanities in the area of digitization. The platform was supposed to connect these areas of research and expert experience to form a cohesive whole, as described in section “1.3 Previous similar work” on p. 2-3 and implemented them through the platform. In comparison to existing platforms, our platform should be used for collaborative work of experts of humanities and ICT specialists, as well as of our users (be they experts or laymen). This required a high level of provision of usability and support, but we also had to ensure the delivery of quality work and services. In the humanities this field has already seen considerable progress, for example with the easily understandable interface of Collex or the neat interface of Omeka. However, the idea to allow non-experts to fully participate and collaborate is otherwise rather known from commercial platforms like eBay or Amazon or from social network platforms like Facebook. The other new aspect is the inclusion of quality, just as in industrial environments, where non-experts need to be enabled to participate in the working process and where they must be able to understand instructions quickly.

These considerations have been formulated or implemented in several information services and humanities environments. The different approaches were described in section “1.3 Previous similar work” on p. 2-3. However, in these digital libraries and other research bodies, the focus may rather not have been set on non-experts. One could say that the connection of different areas is one aspect that makes this master's thesis relevant. We conducted a short, informal interview with a humanities researcher, which showed that the approach to combine ICT (text encoding support) and humanities research may be valuable to structure a researcher's data and allow faster research results (please see section “4.3 Short evaluation by representatives of target user groups” on p. 28). Unfortunately, due to time constraints, these aspects could not be evaluated in full depth and are subject to further continuous work.

Besides, we hope that the insights gained from this master's thesis will be of future value. However,
the results of our strategy will be visible only over time. But the development and initial use of the platform gave suggestions for the strategy and its future application. A future outlook on access to works and on user involvement can already be gained from the small scale setting of this master's thesis. We trust that such insights can come from any cooperative digital environment and that this can be seen as the second area of relevance of this master's thesis' work. Description of insights:

The access over time in a cooperative digital environment, as provided by the developed platform, was discussed and decisions were taken. Various questions arose. Access over time pertains to copyright and the users' work on the platform should stay reusable for the future. Therefore, users must agree to set their works under a Creative Commons licence (Minerva 2008, p. 30-31). However, this relates only to the users' work. In case that a digitized object consisted of text and that text was considered to be an orphan work (Minerva, 2008, p. 23), the legal status of the work was not clear. The description “Work still in ©, right-holder could not be traced.”, was noted and was displayed on the web page of the platform. This was considered to be a temporary solution. One has to look at measures to find a right-holder. In case that a right-holder would be found, arguments to convince the right-holder to accept the Creative Commons licence should be ready. A policy how to proceed in such cases is needed.

Access over time also refers to the accessibility of digital material. To ensure future access, open standards (Minerva, 2008, p. 32-35) were used on the platform, with the exception of the .jpg image standard. The legal status of this standard is yet unclear (Terras, 2008, p. 79) and we have to keep an eye on any future developments. If the use of the .jpg image standard required the payment of royalty fees in the future, a repeat creation of display images in an open standard format may be necessary. However, this would require considerable effort and time. May one keep that question open.

Conservation expertise is urgently needed. Especially, if an involvement of the institute in the handling of physical objects after digitization is intended, further questions must be answered. Potential responsibility for the storage of the physical object must be taken into account. This would lead to questions of ownership and of copyright for text objects. SEA countries were affected by changes of infrastructure and legal issues may not be clear. For example, if a right-holder was to be found, whilst the work was still in copyright, the right-holder would have the right to ask his/her property back. This may include the digital object. But the right-holder could also ask for a preservation service of the physical object. This question referred to the vision of the institute and its code of ethics and may require the institute to accord. All members of the institute need to be involved in finding possible answers.

Information sharing is considered to be of great importance for a community. Sharing was implemented on a small scale. This may be further extended to include collaborative work flows for shared objects, as already done by e.g. Collex (Nowviskie, 2005, p. 9-10). In addition, users may want to communicate with each other just the way they do in a social network or at the very least, users may want to forward messages to their preferred social network.

For the time being, this level of interaction between users has not been considered for the platform. The Creative Commons NonCommercial license (Creative Commons, 2012) which is used within the platform, allows long-term access while not commercial usage. A university's social network may come with similar terms of use. For other social networks this may be different. In short, when people upload or forward files to Facebook, Facebook will have rights over these files like the right-holder himself has: “For content that is covered by intellectual property rights, like photos and videos (IP content) […], you grant us a non-exclusive, transferable, sub-licensable, royalty-free,
worldwide license to use any IP content that you post on or in connection with Facebook (IP License). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it” (Facebook, 2011).

Therefore, we decided not to include a social network into the platform. However, the service of a social network could be used and users may be updated via the network. This would not enable communication between users, but push messages from the institute to users. This matter may be reviewed in the future.

Further works on the platform that influenced the strategy include the supported technologies. We wanted to distribute information to the community to raise awareness of cultural heritage. This pertained to the level of digitization that one targeted. “Tangialities” refer to inherent metadata of an object. These metadata, e.g. the texture of an object or the handling of an object within a particular context, reach senses beyond the visual sense (Milekic, 2007). Deegan and Sutherland call information with limited metadata “named information” (2009, p. 134). Our aim was to enable our targeted user groups the participation on the platform. This way a particular technological know-how should not be necessary. Given those circumstances, it is a high aim to include “tangialities” into the platform. However, it may be possible to overcome “named information”. Right-holders may provide additional information about the objects' context. Information on the object and its use may be added visually as background or via interactivity. Small windows may open and provide information. If we provide information in this way, the collections become the character of an exhibit. A balance must be found between the degree of interpretation, the collection of the additional information, and the increased awareness that can possibly be achieved.

The platform could be further extended, if one included more XML technology. So far XSLT has been implemented. It proved to be a simple technology that delivered results just about as planned. The XQuery concept (“query for information”) differs from the XSLT concept (“transform information”). The use of XQuery might bring further findings. Further thought could be given to the creation of an XQuery environment within the platform and this way, an XML database could be applied. The TEI analysis tool of TextGrid (Pempe & Vitt, 2009, p. 10-21) may also be included after a switch to the Java programming language.

At the same time the digitization guideline forms part of the institute's quality management system (QMS). The QMS represents the organisation's work flows and was graphically outlined, in form of flow charts. These flow charts were drawn using a subset of standard ISO 5807:1985 flow charts. As the subset has been extended, it does not conform to the standard, but it could easily be transferred into the standard. Flow charts that adhere to the standard can be converted to a format that permits analysis (e.g. White, 2004, p. 6). This format could be XML (Ko et al., 2009, p. 772), similar to TEI that represents a text. To facilitate internal analysis, costs may be listed per activities or the institute's processes could be compared to processes of other organizations, as described by Srivastava and Mukherjee (2009). It may be interesting to do comparisons with organizations, which try to fulfil similar task purposes. This would give interesting insights into best practices.

6 Conclusions

In the frame of this master's thesis, we developed a strategy for the EHRIS institute's information service as well as a web platform to implement the strategy.

The strategy derived from a proper “strategic planning process”, although, the strategic objectives had not been formulated. Since the strategic objectives were known, they formed the basis for the strategy. In a next step, the strategy was then translated into “policies” which we created in parallel
to the strategy. These policies were the basis for a “quality management system” (QMS) to provide and control the proposed information service. Throughout this whole process this master’s thesis project took shape.

A platform that would permit and support users to do digitization work, should be created to set the policies into practice. None of the existing platforms was considered suitable, that is: royalty-free and satisfyingly usable. Therefore, we decided to develop a platform in-house. Guidelines on how to use the platform and to do activities beyond digitization work were established. A guideline to use the platform, and to do activities beyond the platform for digitization, was created. These guidelines themselves formed part of the QMS. Initial support of the strategy as combination of technology and humanities' concepts was implemented with the XML technology. An evaluation of the platform's performance was carried out during the development process and after its completion. This allowed us to rethink the strategic decisions taken up to now.

The first research question concerned the strategy. The strategy had been developed based on contributions received during discussions with the founding members of the EHRIS institute. The founding members themselves have different backgrounds and belong to all groups of users that we targeted. Emphasis was put on usability and visualization to reach all target groups and to encourage and enable collaboration. These main considerations have always remained on our mind while transferring the strategy into policies and into the QMS. Therefore, we can hope that the resulting strategy will reach the intended community. This question shall remain unanswered and we may come back to it in the future. One important constraint for the information service that influenced the strategy in most aspects was the budget. Budget needs should be kept as low as possible. Consequently, several “tasks” which similar traditional institutions like libraries handle themselves were transferred to the community (the users), for example the “expert support” for the support team or the check of collections for compliance to policies (please find both in COR.I-020.16). This is even more important since the institute’s scope covers an area of different (cultural) values. As a consequence of the budget constraints, the community’s motivation and engagement are crucial for the institute and had to be taken into account everywhere, including in our strategy. Whether these considerations meet the intended community's expectations may be seen. Questions related to these considerations with considerable impact not only on the institute itself, but also on the institute's image within the community, are raised in the section “5 Discussion” on p. 29-30.

The second research question concerned a platform as implementation of the strategy. Since no suitable existing software was found, we developed our own platform in-house with a small team of software engineers. The selected team members had already proven their expertise and they knew and trusted each other. Therefore, the communication within the project team was easily established and project management and development could be carried out as planned. A professional evaluation of the platform, the so-called “testing” could not be done, because testing personnel was not available within this master's thesis' duration. However, this was known from the start and deliberately left outside the scope of this master's thesis.

The platform was developed based on one main consideration, namely to achieve a result within this master's thesis' duration. Therefore, a technology had to be chosen, which would allow quick implementation. However, this decision excluded the use of a more powerful, and maybe more suitable technology. But we did consider a switch to the Java technology within the strategy for the future. Further possibilities are considered in section “5 Discussion” on p. 29-30.

A third research question posed the use of TEI/XML encoding and analysis on the platform. TEI is a
XML standard that is adjusted to the needs of humanities collection research. Simple encoding and analysis with TEI/XML files was implemented as planned. The application of XML technologies on humanities' encoded files gave surprisingly good results. Even the idea of a cross-language search, with the help of TEI encoding was a possibility. The informal interview with a humanities researcher showed that the use of TEI encoding and simple analysis may be of high value. The platform could be used to structure a humanities researcher's (field) data and to do simple analysis of text collection (by e. g. XSLT). This way, research results could be achieved much faster (please see section “4.3 Short evaluation by representatives of target user groups” on p. 28). However, due to time constraints, the interview could not be pursued further; and an adjustment of the platform's data structure architecture, of the TEI/XML encoding support and of the analysis of researchers' needs in practical terms is subject to further work.

The working procedures foreseen, especially for text digitization, proved to be too ambitious. The time and effort that it takes to encode a collection's file by TEI was much more than anticipated. While the platform's support was highly useful, we still need to reconsider the concept of encoding two files (one file with transcriptions, one file with translations) for one collection has to be reconsidered. In light of the above our plans were changed and all encodings were done within one file. Subsequent decisions that had to be taken, lead to the result that the TEI files became “conformable”, but not “conformant” to the TEI definitions (TEI Consortium, 2012d, 23.3.3.1 Semantic Constraints), meaning that they did not conform to the semantic concepts of the TEI Consortium any more. Therefore, some rethinking will be needed, as for the chosen technology and the underlying concepts.

The platform was evaluated after its creation. So far, three collections had been digitized with the support of the platform. The results showed a need for conservation guidelines. This had been considered when we created the strategy, but the urgent need for the expertise of a trained conservator had not been clear at that point in time. The founding members of the institute were informed of this particular need. However, a digitization guideline dealt only with the handling of objects during the digitization process itself. The handling of the physical objects after digitization should not be forgotten. Conditions under which to keep the objects must be evaluated and this led to further thoughts regarding this topic, all described in more detail in section “5 Discussion” on p. 29-30.

7 Summary

This master thesis had two components: the creation of a strategy for an information service in the field of digitization and its application in form of a web platform. Focus was put on the support for collaborative work and on the combination of different areas of knowledge. Thereby, the master's thesis research questions could be answered and, no matter how small the scale of the master thesis had been, it gave insights for strategies of information service organizations. These insights concerned the legal handling of digital objects and the involvement of the community. The involvement and their “place” within such information service provider relates to the organization in general.

For this master's thesis, the information service provider's organization was depicted as QMS. Participation and contribution from the community as well as from the support team was encouraged, but only for their own tasks. A rather hierarchical model was applied. Decisions regarding the organisation as a whole were taken by the board. This was justifiable because the sponsors of the organization are members of the board. However, some rethinking might be needed in the future. A similar network, project NINES which uses the mentioned Collex platform (Nowviskie, 2005, p. 2 and p. 9-10), works based on an equally centralized administration (NINES,
n.d., “Our editorial boards locate reviewers to evaluate [each collection]”). However, giving all decision making power to the executive may not be suitable for this type of organization. Maybe in the future more participation could be granted to the community instead, and if it turned out that the community got absolutely engaged, even all decision making power could be transferred. This master’s thesis clearly provided a lot of food for thought for similar future activities.

References


36


Appendix A: selection of technologies and decision for in-house development

The following section shall explain in more detail our line of thoughts from the selection of technologies down to the creation of the platform itself. The collections’ scope covers the field of cultural heritage of SEA (COR.I-020.01, vision). Any text falling under this scope may be produced in or about the SEA area. Such a text may be written in English, French or Spanish or as a matter of fact in any Asian language, even those based on non-Latin alphabets. In order to be able to cover transcriptions of all possible languages we had to use Unicode. This choice determined partly the application of XML technologies, since XML supports Unicode. Consequently, we needed an environment that supported XML and particularly TEI. The environment we had in mind should offer simple image display functionality and it should allow multi-user access as well as permit user collaboration, at the very least to the level of file exchange.

Off-the-shelf solutions do exist and were indeed available open source and free of charge. However, we considered them to be too sophisticated for our intended user groups as well as for the developers who would have to adjust the underlying technology within the planned schedule. Therefore, we opted for in-house development of a platform based on our own needs. The technologies that we selected had to permit development within the time frame of this master's thesis.

Decision for encoding UTF-8

Unicode allows representation of “characters used in all the major languages written today” (Unicode, 2012c). Unicode is supported by XML (Unicode, 2012a), and therefore, by TEI. As Unicode encoding “UTF-8 is most common on the web” (Unicode, 2012b), we selected this encoding, too.

We started with transcriptions of Vietnamese and French texts. These languages are located in different “blocks” of the Unicode space and Vietnamese characters are even part of “several noncontiguous Unicode ranges” (Vietunicode, 2002): http://vietunicode.sourceforge.net/charset/. The question arose, whether UTF-8 could be used for these languages of different and mixed blocks. We also wondered whether we could use of UTF-8 for transcriptions of mixed content (French and Vietnamese). Fortunately, this latter question could be answered positively, please see section “4.1 Evaluation during and after development of the platform”, p. 25 and appendix B, figure 10.

Decisions concerning TEI

We had to make a decision as to which TEI version to use. The current 2012 version is P5. We thought that the support of version P4 would cease soon. However, the intended user group may still work with P4. One of our target groups, namely the humanities' experts, was thought to be familiar with P4. TEI P4 has been in use since 2001 (Vanhouotte, 2004, p.11), and whereas P5 has only been launched in 2007 (TEI Consortium, n.d.). Today P5 may not yet be widely used, but we chose to start with the latest version. Our important target group of people of South East Asia are known to be early adopters and always interested in the latest versions.

The subset of the TEI standard that one is able to use is listed in a so called “schema” file. The “TEI Lite” schema (http://www.tei-c.org/Guidelines/Customization/Lite/) lists a subset, that for example Vanhouotte believes to fulfil most of a researcher's needs (2004, p.11). We also looked at another subset of the full TEI schema, the “TEI Tite” schema. “What distinguishes Tite from other TEI customizations is that Tite is meant to prescribe exactly one way of encoding a particular feature of a document” (Trolard, 2011). Hence, TEI Tite made it possible to encode text with less own
decision making. We thought that our users might not be familiar with TEI encoding or with the many decisions faced by the humanities. In addition, TEI Tite omits the TEI Header section of a TEI encoded file. Our platform offers a way to create the TEI Header section of a TEI encoded file. In short, TEI Tite may simplify the encoding process by sparing users the decision making. However, we still opted for TEI Lite. While TEI Tite may be easier to use, our target groups should be encouraged to learn and experiment with encoding (COR.I-020.01, mission 6.). Through surveys we are planning to gather feedback on both version and schema. So that in the future, the platform can be adjusted in terms of TEI version and/or TEI definition.

Reason for the use of XML technology XSLT

TEI is an XML standard (Vanhoutte, 2004, p.11). XSLT style sheets are a method to analyze XML files by transformations (Clark, 1999, 1 Introduction). “[Y]ou can simply send an XML file to the browser, with a processing instruction to identify the [XSLT] style sheet to be used, and the browser will automatically invoke a transformation and then display the resulting HTML” (Kay, 2010, Chapter 3: Stylesheet Structure, The <?xml-stylesheet?> Processing Instruction). Our objective was to present text formatted by display instructions, and XSLT was designed to just that. Another programming language to analyze XML files is XQuery. We wanted to analyze XML/TEI encoded files to allow cross-language search. Here, the presentational aspect, played a secondary role and we preferred XQuery, which was designed as a method to do sophisticated data analysis. However, in the end we decided against the implementation of XQuery. To work with XQuery a user would need an XQuery execution environment. “It is possible to […] [use] XQuery operations […] [while] it suffers from being a plugin: it needs to be manually installed […]” (Etter et al., 2011, p.195). We didn’t want to impose the burden of installing an XQuery supporting environment to our users. XSLT transformations also permit data analysis and we decided to use it for both implementations. We offered XSLT files for both implementations ready for use on the platform. Users may then adjust them to their needs. XSLT’s current version is 2.0. Yet, “at the time of writing, [2010,] none of the major browsers includes support for […] XSLT 2.0” (Kay, 2010, Chapter 3: Stylesheet Structure, The <?xml-stylesheet?> Processing Instruction). Therefore, we decided to use XSLT version 1.0.

Decision for in-house development and considerations concerning open source software

Our policies demand the use of open source software (OSS) solutions (COR.I-020.01, mission 9.). Existing platforms, that were in use in the context of humanities research and that provided functionalities similar to our requirements were scrutinized. Both Collex and TextGrid fit the bill and are OSS.

“Collex” are tools for collaborative work on digitized collections. The tools read and compare the TEI mark-up of collections. “Collex allows […] users: […] to collect, tag, and annotate trusted objects (digital texts and images […]” (Nowviskie, 2005, p.2). Users own a pool to which digital objects can be added to. To these digital objects, users can add tags (Nowviskie, 2005, p.9-10); these user activities are analyzed (Nowviskie, 2005, p.9). Collex is open source (Nowviskie, 2005, p.1) and Collex’ functionality corresponds to the institute’s requirements for a platform. Nevertheless, we decided against Collex, because it is written in “ruby on rails” (Nowviskie, 2007, Notes [10]), a programming language, which our developers were not very familiar with.

“TextGrid” is a part of the D-Grid project, which offers access to various German university sources. TextGrid covers sources and collaboration for the humanities (Neuroth et al., 2007, p.62-64). The TextGrid initiative offers a “lab”, a text editor that supports TEI mark-up and tag analysis
Both Collex and TextGrid focus on academic users. As Rydberg-Cox said, “the creation of a scholarly or critical edition [...] is the sort of work that might take many years or even represent a scholar's lifetime achievement.” (2009, methods). The target group of humanities researchers may well do so, but other target groups were thought not to invest that much time into the preparation of the collections and the mark-up. While Collex and TextGrid's lab offered the requested functionality and TEI analysis support, we thought that in most of our cases there may not be that much information available and both Collex and TextGrid were designed for handling considerable amounts of data. From this point of view, Collex and TextGrid were considered to be “over-performing”. Of course this may not really count as an argument against their use. However, the interface had to be simplified to fit to our user groups. Since applications are usually strongly connected internally, this meant that the developers had to learn the whole system. This was not possible in the available time and so we decided against an integration of existing solutions like for example TextGrid's lab.

Another reason why we decided against a ready-made system, and especially against Collex and TextGrid (at least for the moment), was our decision to use OSS. OSS is supported by “teams of volunteer programmers” (Minerva, 2008, p.32). These community members may share experiences. e.g. on how to install, maintain, upgrade and enhance the software, but they are not obliged to do so. There is no maintenance contract or warranty, which means that by using OSS, one is personally responsible for the software's maintenance. Let’s look at the origin of research oriented solutions, like Collex and TextGrid. Collex (“as our Collex development team is incredibly small”, Cohen, 2007, p.70) and TextGrid (TextGrid, n.d., partners) are produced by researchers for researchers. Indeed, these applications may be highly adjusted to the humanities context and collection research. However, the production of the software may not have been carried out by a professional programming team. A small, self-made application may be easier to maintain. It’s true that it may not be of better quality than an existing (university's created) solution, but we would know our system inside out. Therefore, we opted for our own development to launch the platform. If the data storage systematization, the database and the file storage have been well-designed, it would not be difficult to switch to another solution later on.

We were planning to set our in-house developed platform open source. In addition, we also wanted to be able to provide adequate support for users of our code, even though we were not really required to do so. In the beginning, when our platform went live, we did not have sufficient resources to provide this kind of support. Therefore, we decided against the option to set our platform open source. We hope to be able to reconsider this decision in the future.

**Decisions on programming language, database, server**

The two most commonly used OSS programming languages for web applications are PHP and Java (e.g. Hars & Ou, 2001, p.6, table 2, and e. g. SourceForge, 2012). Our developers were familiar with both languages. Java, as object oriented language, supports the representation of thought models (Eckel, 2003, p.16-17) and can, thereof, be seen as more powerful than PHP. To store data relational database systems may be used in web applications. Our platform supports the work with TEI that is XML-based (Vanhoutte, 2004, p.11). Hence, any data could be represented by XML and an XML database instead of a relational database could be chosen. Final component of a web application is a so called “server”. XMAPP is a development software package that offers all three components: server, database and programming language (Abajo et al., 2011, p.151-152). Abajo et al. describe XMAPPs simplicity to install and use as development environment. We opted for a development based on the XAMPP package, because of its simplicity.
and because our developers already knew about it. As a consequence, PHP became our programming language and MySQL our database system.

However, in future, we plan to switch to the more powerful programming language Java (INV.I-020.012). Then, new consideration may be given to the above mentioned TEI lab of TextGrid, which is Java based and provided on an open source license (Göttingen state and university library, 2011).

Given the decision against an XML database system, the XML structure of the TEI files was not used. An XML database system may be considered in the future. Besides all these technical considerations, one may also look again at activities of social networks, as offered by Collex. But first we would have to find out whether our intended user groups would indeed be interested in these activities, and how to integrate them.
Appendix B: platform

Platform – collection view: metadata

Figure 5: Collection view: metadata (default view)
Platform – collection view: transcription and translation

Figure 6: Collection view: transcription and translation at once
Figure 7: TEI/XML encoding - formatted view, transformation of UTF-8 encoded characters
Platform – collection administration view: metadata

Figure 8: Collection administration view: metadata indication and display file upload
Platform – collection administration view: transcription and translation

Figure 9: Collection administration view: transcription and translation
Figure 10: TEI/XML encoding - format preparation, and transformation of mixed language content
A TEI file is created out of transcriptions or translations that are stored in the platform, or the file is uploaded (please see the blue arrows). Files are shared via the platform (please see the yellow arrows). The decision to share may be withdrawn (please see the red arrows).
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title level="m" type="main">Thư từ xã giao</title>
        <title level="m" type="sub">a machine readable transcription</title>
        <author>Thi Đạt Chi Soạn</author>
      </titleStmt>
      <respStmt>
        <resp>TEI-conformant markup</resp>
        <name>Barbara Cimpa</name>
      </respStmt>
      <respStmt>
        <resp>Transcription</resp>
        <name>Barbara Cimpa</name>
      </respStmt>
      <titleStmt>
        <edition n="1">machine readable transcription</edition>
        <date when="2012-03-02">2012-03-02</date>
      </titleStmt>
      <extent>??size of current file?? x KB</extent>
      <publicationStmt>
        <idno type="number">001</idno>
        <publisher>EHRIS institute</publisher>
        <availability status="restricted">
          <p>Creative Commons Attribution-NonCommercial-ShareAlike</p>
        </availability>
        <date when="2011-05-23">2011-05-23</date>
      </publicationStmt>
      <notesStmt><note>??</note></notesStmt>
      <sourceDesc>
        <biblFull xml:lang="vi">
          <titleStmt>
            <title type="main" level="m">Thư từ xã giao</title>
            <author>Thi Đạt Chi Soạn</author>
          </titleStmt>
          <extent>12.7cm * 18 cm , 52 objects</extent>
          <publicationStmt><publisher>Thi Đạt Chi Soạn</publisher><address><name type="country">Viet Nam</name></address><date when="1959-01-01" xml:id="dateSourcePublication">1959</date><availability status="unknown">private</availability></publicationStmt>
        </biblFull>
      </sourceDesc>
      <encodingDesc>
        <projectDesc><p>Digitization to share under EHRIS.</p></projectDesc>
        <samplingDecl><p>No sampling.</p></samplingDecl>
        <editorialDecl>
          <p>Correction was done silently or marked with choice-sic-correct.</p>
          <p>Normalization was done silently. Words in capital letters were transcribed to small letters, and marked with "hi rend=uppercase". Empty spaces were inserted to separate words.</p>
          <p>Quotation marks were kept sometimes, standard. Opening and closing quotation marks are transcribed as ".</p>
        </editorialDecl>
      </encodingDesc>
    </fileDesc>
  </teiHeader>
</TEI>
Hyphenations have been retained in case they do not group syllabus.

Standard values were encoded as follows: Numbers are kept. Values in tags are encoded in standard-iso-format. Unit values are encoded by tags including unit and value.

```
<list>
  <item>Date</item>
  <item>Time</item>
  <item>Currency: "num type=dollar (value, unit)"</item>
  <item>Weight: "num type=kilogram (value, unit)"
  <item>Year: "num type=year (value, unit)"
  <item>Age: "num type=age (value, unit)"
</list>
```

Segmentation was not done.

Interpretation was done as follows: That content was marked:

```
<list>
  <item>Person (relation-to-originator, name): "name type=person (name type=role, name)"
  <item>Product/company name: "name type=brand"
  <item>Places: "name type=place"
  <item>Events: "name type=event"
</list>
```

Ehris classification system
<category xml:id="ecs.industry">industry</category>
<category xml:id="ecs.industry.1">service</category>
<category xml:id="ecs.type">text</category>
<category xml:id="ecs.country">Viet Nam</category>

Transcriptions in vi.
Text besides the transcription in English.

```
<catRef target="#ecs.industry.1 #ecs.type.1 #ecs.country.vn" scheme="#ecs/>
```
<change>
<date when="2012-03-02">2012-03-02</date>
<label>Transcription, mark up</label>
<name>Barbara Cimpa</name>
</change>
</revisionDesc>
</teiHeader>
### Appendix C: documents created for the platform

Please find the documents in this master's thesis attachments 1, 2 and 3.

<table>
<thead>
<tr>
<th>identifier</th>
<th>title</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>COR.I-020.01</td>
<td>Policies</td>
<td>The policies describe the purpose of the institute and the framework for institute's members and users to act within.</td>
</tr>
<tr>
<td>COR.I-020.02</td>
<td>Community overview</td>
<td>The community overview drawing shows the actions that users perform. Users select objects and thereof collections, digitize them, share and do research with them. By sharing and research, a “community” arises. Users participate in conferences, such keeps the community alive.</td>
</tr>
<tr>
<td>COR.I-020.03</td>
<td>(community) selection</td>
<td>The selection drawing shows the areas that users have to consider when selecting material. Objects must be in the scope of the institute's topics (economical historical, of/about the South East Asian region). Objects should be checked whether they are available in an existing collection. Before digitization, objects must be checked for access and use rights. Conservation of the physical object is not done by the institute. Users are reminded to consider actions. Technically, the drawing is a visualization of the selection criteria check list.</td>
</tr>
<tr>
<td>COR.I-020.04</td>
<td>Selection criteria</td>
<td>The selection criteria check list shows questions to answer when selecting material. Users are responsible for that their objects adhere to the list.</td>
</tr>
<tr>
<td>COR.I-020.05</td>
<td>(community) digitization</td>
<td>The digitization process drawing shows the actions that a user performs to prepare material. Digitization starts with selection. Then, the image is made accessible, finally, meta data are found and set. Optionally text is made accessible.</td>
</tr>
<tr>
<td>COR.I-020.06</td>
<td>(community) digitization, make image accessible</td>
<td>The “make image accessible” procedure drawing is a sub process of “digitization”. It shows users' activities to prepare master images for archival and derive images for presentation. The procedure starts with the image capture. Objects are captured by scan or a photo is taken. Resulting image is the “master” in .tiff format. Recommendations are given for settings of scanner or camera and the maximum object size depending on the capture device. The placement of a color reference card next to the object while capturing is suggested. Master images shall be sent to the institute, on a storage media. That is a responsibility change. The community (user) is responsible for the image capture. After sending the master images to the institute, EHRIS becomes responsible for “preservation (archive)”. Action “preservation (archive)” forwards to a sub process within the responsibility of the support team. Users perform digital image enhancement measures onto captured images. The color reference card is used for color adjustments. After that, smaller sized copies in jpg-format are created out of the tiff-images. Three small sized images are used for the representation of one object. When a user uses the platform and adds an object to a collection, he/she will upload these three images. The “presentation (website)” is a sub process within the responsibility of the support team.</td>
</tr>
<tr>
<td>COR.I-020.07</td>
<td>(community) digitization, make text accessible</td>
<td>The “make text accessible” procedure drawing is a sub process of “digitization”. It shows users' interactions with the platform to add transcription and translation, and to encode text for presentation and research. Transcription is done in Unicode. Transcriptions are translated to English. Transcription and translation are uploaded into the platform. The upload can be done by either the creator of a collection or a user with the role “transcribe”</td>
</tr>
</tbody>
</table>
respective “translator” for that collection. The platform supports users to do TEI mark up. Started by button “to XML”, the platform creates a simple XML/TEI tagged text of the given text. The user downloads the tagged text and adjusts it, and uploads it back to the platform.

TEI encoding can be used to add meaning to a text and also to indicate style instructions. The platform offers a simple style sheet (XSL). Users adjust the style sheet with particular formatting instructions. The adjusted style sheet is uploaded. “Presentation (web site)” is a sub process in the institute's responsibility.

The “research & share (upload)” process drawing shows an overview of interaction activities with the platform, to do and share research. Users who want to join the community must register. A user logs into the platform and adds a collection. When a collection is created, objects are digitized and uploaded. TEI encoded files or XSL style sheets are created and uploaded. Users do not have to work on the platform. Users can download the TEI and XSL files and do research on their machines. Users are invited to upload the resulting TEI and XSL files, therefore, they must log in.

The institute overview drawing introduces in short the institute's activities and intention. The digital services and their place in the institution are shown.

The quality management system (QMS) drawing shows how actions of the institute and support team are related to provide the services for the community. The drawing is a detailed view of drawing COR.I-020.13. To ensure the QMS, the flow of information must be ensured. The “Plan do check act” (“PDCA”) principle is a cycle. It starts with “Plan”. Strategy (“KPIs”) and policies are defined. Budget is planned. These definitions are done by the board, the “institute”.

The policy is transferred to the organization by training. Then, service is provided (“Do”). The technical support team maintains the equipment and provides the archive. The awareness team is responsible to raise awareness about the institute and to find partners. The user support team gives direct support and performs the survey. The conference team organizes conferences.

The community is the center. The community selects and contributes collections and does research.

During the service provision, inconsistencies or irregularities may arise. Such is uncovered through control steps (“Check”). Inconsistencies and irregularities relate to users' satisfaction or to quality. User satisfaction is checked by users' feedback. Indirectly, use of (human) resources, contribution and numbers of members give some indication of user satisfaction. Quality is the staff's process performance. It is checked by analyzing process execution data. Irregularities are subject to correction or improvement (“Act”). For that irregularities will not repeat, and for that the system improves itself, corrective and enhancing actions are reflected by the policies and influence the strategy. The board, the “institute”, again, discusses all such information. A new cycle starts with “Plan”.

The annual review meeting process drawing describes the board's annual meeting to build the strategy. Basis for discussions is the annual report about last year's performance, and each attendee's input. Actions to improve the services are discussed and if they are subject to implementation, they lead to a change of the policies.

The training procedure drawing gives a guideline how to prepare, execute and evaluate trainings. Trainings may be regularly performed trainings or decided as additional trainings. Regular trainings are newcomer trainings and refresh
| COR.I-020.13 | Support overview | The support overview drawing shows an overview and shortened view of the quality management system (drawing COR.I-020.10 shows a detailed view). It shows the actions of the institute's support team that surround the users' actions and are driven by the board's policies. |
| COR.I-020.14 | (support) presentation support | The presentation support procedure drawing shows the actions that the support team performs when an user has uploaded or changed an object. The support team checks the object for technical accessibility, and whether it complies with policies. The selection criteria are the reference for the policies. Non-compliance leads to removal of the object of the web. |
| COR.I-020.15 | (support) preservation | The preservation process and procedure drawings show actions that the support team performs to ensure long-term accessibility of the collections. Work flows “archive” and “daily back-up” are sub processes. They are used during the “yearly” check of the digital archive and the “daily” preservation activities. |
| COR.I-020.16 | (support) user support | The user support process drawing shows actions of the direct user support team. User support starts when a user sends a request or proposal to the institute. Requests are executed if they comply to policies. For actions that exceed the support team's expertise and authority, an expert is consulted. The community's feedback is subject to the annual meeting, the feedback is gathered by a survey. |
| COR.I-020.17 | Archive explanation | The “archive explanation” document explains the archive's folder structure. The document is intended for the technical support team. |
| COR.I-020.18 | Meta data explanation | The “meta data explanation” list explains the meta data. Collections and objects apply the same meta data set, a few additional data need to be set. Meta data are grouped into general data about items, data about the physical objects and data about the digitized objects. |
| COR.I-020.20 | Text preparation and mark-up recommendations | The document gives recommendations for the preparation of text. Tools for transcription and translation are mentioned. TEI elements that are used and their extended definitions by the institute are explained by examples. The document is not meant to be complete nor instructive, further experiences will be added. |
| COR.I-020.21 | Digital image treatment recommendation | The document gives recommendations for the preparation of digital objects. Settings to capture the item are described. Recommendations how to treat the captured image are given. Naming of a captured image may encode sorting information related to the collection. The document is not meant to be complete nor instructive, further experiences will be added. |
| COR.I-020.22 | (support) conferences | The conferences procedure drawing shows the actions to prepare a conference and follow-up actions. After the conference, external speakers and community speakers, both, are reimbursed. |
| COR.I-020.23 | Quality measure - PDCA | The quality measure process drawing shows the PDCA ("plan do check act") cycle, from the view of the support team, at the bottom right corner of the drawing: the small quality measure repeats itself in the whole organizations' quality management system (COR.I-020.10). |
| COR.I-020.24 | Annual report | The annual report drawing shows the various sources of the annual report, to summarize a year's performance. Feedback of the community and of the support team is collected to measure user satisfaction. Process data are collected to know about service quality. Data that arise for a planned number of occasions during the year, were marked, so the data of the archive check “archive check (1)”. Strategic figures, the KPIs, are the targeted performance. Actual figures are taken from the database. Additional information is added to the report: “Awareness information” about existing and planned partnerships, financial information about budget use, and the previous meeting's notes. |
| COR.I-020.25 | Internal audit | The internal audit process is executed once every months. It shall uncover “non conformities”, issues that prevent the achievement of the strategic objectives. An “action plan” will be created to overcome these issues. The internal audit is part of the whole organizations' quality management system (COR.I-020.10). |
| INV.I-020.001 | High level design | Scope of this document is the high level design in terms of technology requirements.  
  – An overview of architecture and programming languages is described.  
Technical details are in the responsibility of the programmer, that description is not in the scope of this document. |
| INV.I-020.002 | Requirements non-functional (quality requirements) | Scope of this document are non-functional requirements. The document shall be valid for the whole web application. Additional requirements shall be tested against that document. |
| INV.I-020.003 | Detail level design | Scope of this document is the description of  
  – the database schema  
and further designs, these are:  
  – the page flow  
  – the identifier system and file naming system  
  – the folder structure for images and XML files.  
The designs must be embedded into the overall structure of the web site. The overall website structure is in the responsibility of the programmer, that description is not in the scope of this document. |
| INV.I-020.004 | Functional specification (mock ups) - basic functionality | Scope of this document are requirements in form of mock ups. These are the interfaces that a user will see. Each web page that a users will see is shown. Elements of the pages that are considered obvious (like the exit-cross at top right) are not mentioned further. Elements that are considered not-obvious, or that shall have particular functionality, or that have particular data sources, are explained.  
  – Requirements for the whole application are drawn. |
| INV.I-020.005 | Functional specification (mock ups) v2.0 - multi user environment | Scope of this document are requirements in form of mock ups. These are the interfaces that a user will see. Each web page that a users will see is shown. Elements of the pages that are considered obvious (like the exit-cross at top right) are not mentioned further. Elements that are considered not-obvious, or that shall have particular functionality, or that have particular data sources, are explained.  
  – Requirements for the extension of the existing application to a multi-user environment are described. Users shall a/ be able to administer their collections and b/ shall have a place explaining the framework of the community.  
  – Database changes are described. |
<table>
<thead>
<tr>
<th>INV.I-020.006</th>
<th>Functional specification (mock ups) v2.0 - “lab” preparation, formatted view</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of this document are requirements in form of mock ups. These are the interfaces that a user will see. Each web page that a users will see is shown. Elements of the pages that are considered obvious (like the exit-cross at top right) are not mentioned further. Elements that are considered not-obvious, or that shall have particular functionality, or that have particular data sources, are explained.</td>
<td></td>
</tr>
<tr>
<td>- Requirements for the extension of the existing application to include “lab preparation” are described. The lab preparation refers to the collection administration pages. String processing and to the use of XML functionality shall be added. The user is supported to encode a text. The encoded text is transformed by an XSL style sheet.</td>
<td></td>
</tr>
<tr>
<td>- Folder structure adjustments are described.</td>
<td></td>
</tr>
<tr>
<td>- Please refer to the page flow for v2.0 in INV.I-020.008.</td>
<td></td>
</tr>
<tr>
<td>Programming libraries for XML processing might have to be included. That is subject to the programmer's responsibility, that description is not in the scope of this document.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV.I-020.007</th>
<th>Functional specification (mock ups) v2.0 - “lab”, search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of this document are requirements in form of mock ups. These are the interfaces that a user will see. Each web page that a users will see is shown. Elements of the pages that are considered obvious (like the exit-cross at top right) are not mentioned further. Elements that are considered not-obvious, or that shall have particular functionality, or that have particular data sources, are explained.</td>
<td></td>
</tr>
<tr>
<td>- Requirements for the extension of the existing application to include the “lab” are described. The lab search refers to the use of a sequence of XSL style sheets. Input is an XML file, that is transformed into a further XML file. Settings for the first transformation can be set by the user. The final output shows a list of particularly encoded terms of text collections.</td>
<td></td>
</tr>
<tr>
<td>- The search in XML node trees is described.</td>
<td></td>
</tr>
<tr>
<td>- Please refer to the page flow for v2.0 in INV.I-020.008.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV.I-020.008</th>
<th>Detail level design v2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of this document are detail level design changes for version 2.0, the multi user environment and the inclusion of the lab (XML processing). That document describes</td>
<td></td>
</tr>
<tr>
<td>- the new page flow, and roles to access restricted pages. Pages are referred to by names used in version 1.0. New pages to create shall be named as indicated. Former requirements stay valid. The web site shall use the former database model and former technologies. Adjustments of the program and the database model, as far not described in the FCS, shall be obvious by reading the new page flow, that description is not in the scope of this document.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV.I-020.009</th>
<th>Project plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of this document are decisions taken for the software development.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV.I-020.010</th>
<th>Functional specification v2.0, detail design TEI file creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of this document are functional specifications (mock ups) and detail level design changes for version 2.0, the creation of the TEI file and interface, in the collection administration view. That document describes</td>
<td></td>
</tr>
<tr>
<td>- changes of the database</td>
<td></td>
</tr>
<tr>
<td>- changes of the folder structure</td>
<td></td>
</tr>
<tr>
<td>- changes of the interface and functionality.</td>
<td></td>
</tr>
<tr>
<td>Former requirements stay valid. The TEI creation is simplified.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV.I-020.011</th>
<th>Document list</th>
</tr>
</thead>
<tbody>
<tr>
<td>All documents are listed in the “document list”. The list shows meta data</td>
<td></td>
</tr>
<tr>
<td>Identifier</td>
<td>Title</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>INI.I-020.012</td>
<td>Strategy (&quot;KPIs&quot;) (2011-2014)</td>
</tr>
<tr>
<td>INI.I-020.013</td>
<td>Financial plan (2011-2014)</td>
</tr>
<tr>
<td>INI.I-020.014</td>
<td>Human resources plan (2011-2014)</td>
</tr>
<tr>
<td>INI.I-020.015</td>
<td>Yearly schedule of events (2011-2014)</td>
</tr>
<tr>
<td>INI.I-020.016</td>
<td>Marketing plan (2011-2014)</td>
</tr>
<tr>
<td>TMP.I-020.01</td>
<td>Survey</td>
</tr>
<tr>
<td>TMP.I-020.02</td>
<td>Training list - template</td>
</tr>
<tr>
<td>TMP.I-020.03</td>
<td>Training newcomer</td>
</tr>
<tr>
<td>TMP.I-020.04</td>
<td>Yearly refresh &amp; team feedback</td>
</tr>
</tbody>
</table>
| TMP.I-020.05 | Meeting evaluation | Attendees and speaker of a meeting evaluate the meeting with the
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMP.I-020.06</td>
<td>Annual review meeting attendee’s input</td>
<td>The attendee's input document supports attendees of the board's yearly meeting to prepare their input. Input might be given about changes of law, and about research and technological advances. Additionally, attendees are pleased to consider at least two ideas to enhance the service.</td>
</tr>
<tr>
<td>TMP.I-020.07</td>
<td>User/technical support - daily mnemonic</td>
<td>The support team's daily activities are recorded in the daily mnemonic checklist. Activities are a check of the platform's accessibility, a count of the collections inserted into the archive, the back-up of the platform and further tasks. The daily mnemonic list is used by the user support team and by the technical support team.</td>
</tr>
<tr>
<td>TMP.I-020.08</td>
<td>User support – Expert support list</td>
<td>The support team receives questions of the community. An expert's expertise might be needed to find a solution. Questions and experts' solutions are recorded in the expert support list. The list is subject to the annual meeting.</td>
</tr>
<tr>
<td>TMP.I-020.09</td>
<td>Technical support – yearly archive check</td>
<td>Master image collections are checked yearly for accessibility. The assessment's result is recorded in this yearly archive check document.</td>
</tr>
<tr>
<td>TMP.I-020.10</td>
<td>Technical support – yearly arc. check summary</td>
<td>The result of the yearly check, of master image collections and of the platform, is summarized in the yearly archive check summary document. The document shows the archive's status and information of the assessment like the time to perform the check. The summary is subject to the annual meeting.</td>
</tr>
<tr>
<td>TMP.I-020.11</td>
<td>Tester – yearly daily-back-up check</td>
<td>The platform's back-up files are checked yearly for accessibility. The check is performed by a professional software tester. The assessment's result is recorded in that yearly daily-back-up check document.</td>
</tr>
<tr>
<td>TMP.I-020.12</td>
<td>Conference sign-in form</td>
<td>The sign-in form is used by start of a conference. It records participants' expectations. The information is compared to participants' evaluation at the end of a conference.</td>
</tr>
<tr>
<td>TMP.I-020.13</td>
<td>Annual report template</td>
<td>The annual report summarizes a year's performance of the institute.</td>
</tr>
<tr>
<td>TMP.I-020.14</td>
<td>Annual notes template</td>
<td>The annual notes record the annual meeting's decisions.</td>
</tr>
<tr>
<td>TMP.I-020.15</td>
<td>Bi-monthly quality report (internal audit)</td>
<td>The report controls strategic objectives and other quality influencing figures. That “internal audit” is performed once in two months.</td>
</tr>
</tbody>
</table>