

# Increasing Digital Competences within Humanities

-a joint effort between researchers, library experts & infrastructure providers

---

## A Funding Application for Phase 1

### Index

Executive Summary .....	1
Introduction .....	2
The consortium.....	3
Description of work .....	5
Phase 1: Pool digital competences within humanities.....	5
Outline of Phase 2: Initial proposal on course development and integration .....	6
Timeline .....	6
Budget for phase 1.....	7

### Executive Summary

*The objective of this project is to stimulate the uptake of research within Digital Humanities by ensuring easy access to well planned and well documented course material. As Digital Humanities is a very broad area we choose to focus on research based on large text-based collections and on some of the basic skills such as programming.*

*The project has two phases and the present application is for phase one only. Here the present situation for course material in the area of “data analytics within large word corpora” will be analyzed and gaps identified. Phase 1 will result in a catalogue of common Nordic course material as well as a plan for the development of missing courses. In a Phase 2 of the project, some of the identified courses will be adapted and developed based on the sustainability model chosen. The preferred solution will be a corporation with the software carpentry (SC in the following) and extend their portfolio with courses and workshops relevant for Digital Humanities. However other options have opened among these the use of the NEIC code-refinery and of the European DARIAH-Teach platform. Therefore Phase 1 will also produce a report outlining the pro and cons of the different solutions.*

*Part of the work includes identifying the different roles and defining the different responsibilities for non-ECTS and ETCS competence development among different actors such as the researchers, library professionals and technical infrastructure providers.*

## Introduction

A recent white paper<sup>1</sup> from the University of Aarhus, Denmark “A digital strategy for Arts” begins:

*“Digitization and digital media have generated a rapid proliferation of data that is unprecedented in the history of man. This digital surge is transforming knowledge discovery and understanding in every domain of human inquiry. Digital research, computing, data management, and data-intensive methods will therefore become integral parts of internationally-leading research in the humanities and arts.”*

Further on in this white paper one reads:

*“Digitization of the cultural heritage, for instance, and new orientations towards archival studies, distant reading and large-scale quantitative approaches are transforming a number of disciplines, which challenge disciplinary barriers and make interdisciplinary competences and collaboration necessary.  
To be able to benefit from the many new possibilities of having data in digital form, there is a need for support in all research phases in the form of technology, IT-developers, competency development, community building, and fostering of critical thinking about digital technologies and the humanities. “*

Humanities are moving from being a purely qualitative research area to introduce more quantitative methodologies, which naturally require new competences within areas such as statistics, mathematical modelling and computing. The objective of the present application is to build the foundation for the required general competence development. The area is broad, but to ensure success, the focus is in three different areas all being part of what in humanities are denoted “big data”, where different statistical methods are used on large datasets, typically coming from the cultural domain. The objective of the analysis is to get new insight based on the totality rather than on individual pieces of content.

In the project, we will have three focus areas:

1. Corpora building and analysis of text based collections as e.g. 60 mio. pages of digitized newspaper
2. Visualization tools e.g. used to display the result of the analysis of the newspapers or to show the interconnectivity of the Internet based on the harvested web material (e.g. the plus 800 TB of data in the Danish web archive)
3. Computing for humanities as a general discipline

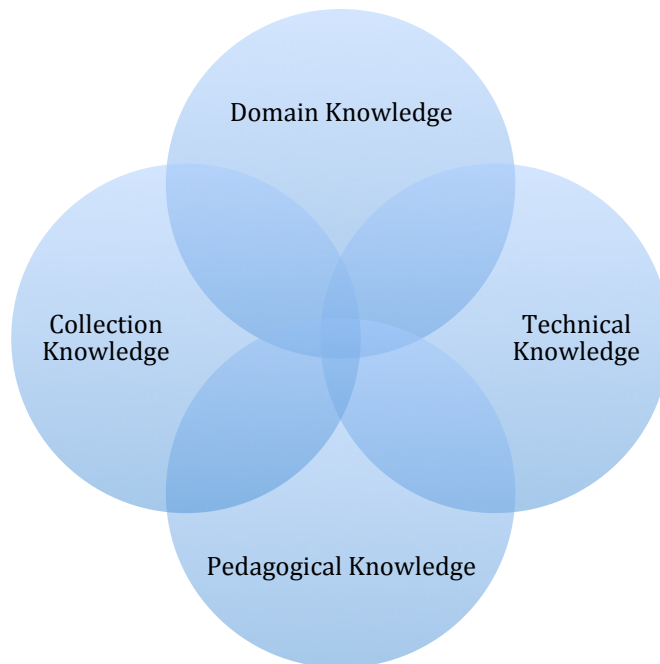
The initial idea behind the initiative was in collaboration with the non-profit organization “Software Carpentry” (SC) to develop courses in the area of Digital Humanities and specifically in the area of quantitative research. The pedagogical basis for this development is a mix of the principles behind Software Carpentry, The Software sustainability Institute in UK and the Nordic tradition for course development. The course portfolio will be developed combining the network behind SC and existing research networks in the Nordic countries consisting of key researchers and groups and already existing work on *Train-the-Trainer* in the European ERICs CLARIN and DARIAH. The initial idea of joint course development with SC is expected to hold; however new actors such as the NEIC code-refinery and the European DARIAH-teach initiative have emerged. Therefore we will produce a report on the possibilities for partnership and thereby implicitly on sustainability, by the end of phase 1.

---

<sup>1</sup> <http://www.forskningsdatabasen.dk/en/catalog/2342565365>

## The consortium

The success of the project depends on the consortium behind. This consortium may change as the project develops and new dimensions are included and some may be left out. The consortium behind this project consists of partners representing four different dimensions, all of which are required in order to develop adequate and relevant courses within Digital Humanities:



**Figure 1: The four dimension of required knowledge**

The **Domain Knowledge** lies within the universities. Some of these are committed to partnerships such as on a European scale; CLARIN and DARIAH, on a Nordic scale CLARIN-Nordic and DARIAH-Nordic or on a national scale such as HELDIG in Finland and DIGHUMLAB in Denmark. To ensure inclusion of the national and international activities, such as DARIAH-Teach, these groups will be involved in the project along with some individual universities holding relevant competences.

The **Pedagogical Knowledge** resides with the researcher but in the context of this project also with external partners such as Software Carpentry and partners within the code refinery and DARIAH-Teach.

The **Technical Knowledge** is represented via those developers, who are based on a broader mandate around super computers, on the Nordic level NeIC and on the national level DeIC (incl. the Kulturarvscluster at KB) in Denmark, SNIC in Sweden, CSC in Finland, RHnet in Iceland and UNINETT Sigma2 in Norway.

The **Collection Knowledge** resides with the national cultural institutions. We here want to highlight, that many cultural institutions work towards integrating tools with their collections (such as the Kulturarvscluster) and work towards establishing technical support for researchers wanting to use their collections.

The boundaries between the four dimensions are not well defined in that some institutions operate on more than one dimension.

Table 2 provides an overview of the involved and committed national, Nordic and international partners. Due to the relatively large amount of engagement a reference group of interested

parties will be established. This reference group will be briefed upon activities and deliverables which the project will provide throughout its dedicated period.

**Table 1: Partner Scheme**

	Infrastructure Provider	Digital Humanity Research Institutions	Cultural Heritage Institutions	Training Institutions
<b>Denmark</b>	DeIC eScience Center <sup>2</sup>	DIGHUMLAB <sup>3</sup> as representative for: University of Aarhus <sup>4</sup> , University of Copenhagen <sup>5</sup> , University of Aalborg and South Denmark University	The Royal Library <sup>6,7</sup>	
<b>Sweden</b>	SNIC <sup>8</sup>	University of Lund, Humanities Lab University of Umeå, Humanities Lab		
<b>Norway</b>	UNINETT Sigma2 <sup>9</sup>	University of Oslo University of Trondheim		
<b>Finland</b>	CSC <sup>10</sup>	HELDIG, University of Helsinki <sup>11</sup>		
<b>Iceland</b>	RHnet <sup>12</sup>	University of Iceland, School of Humanities - Faculty of Icelandic & Comparative Cultural Studies <sup>13</sup>		
<b>UK</b>		University of Sussex <sup>14</sup> Loughborough University <sup>15</sup>	Programming Historian <sup>16</sup> The British Library <sup>17</sup>	University of Edinburgh, The Software Sustainability Institute (SSI) <sup>18</sup>
<b>Nordic</b>	NeIC Training Program Manager <sup>19</sup>	<b>Nordic CLARIN partners:</b> <ul style="list-style-type: none"> <li>• CLARIN.DK (DIGHUMLAB)</li> <li>• Swe-CLARIN</li> <li>• CLARIN-No</li> </ul> <b>Nordic DARIAH Partners:</b> <ul style="list-style-type: none"> <li>• DARIAH-DK (DIGHUMLAB)</li> <li>• Linnæus University, Växjö<sup>20</sup></li> <li>• University of Helsinki<sup>21</sup></li> <li>• University of Oslo</li> </ul>		NeIC Pool Competencies <sup>22</sup>  NeIC Coderefinery <sup>23</sup>
<b>International</b>		The Netherland eScience Center <sup>24</sup>		Software Carpentry <sup>25</sup> and Data Carpentry <sup>26</sup>

<sup>2</sup> Head of DeIC eScience Center, Lene Krøl Andersen

<sup>3</sup> Head of DIGHUMLAB, Birte Christensen-Dalsgaard

<sup>4</sup> Associate Professor, Kristoffer Nielbo & Professor Mads Rosendal Thomsen

<sup>5</sup> Bente Maegaard, University of Copenhagen

<sup>6</sup> Director of IT development, Bjarne Andersen

<sup>7</sup> Senior consultant, Per Møldrup-Dalum

<sup>8</sup> Application expert, Joachim Hein

<sup>9</sup> Senior Advisor, Vigdis Guldseth

<sup>10</sup> Senior Application Scientist, Jussi Enkovaara

<sup>11</sup> Prof. Eero Hyvönen, Director

<sup>12</sup> General Manager Jón Ingi Einarsson

<sup>13</sup> Professor Eiríkur Rögnvaldsson

<sup>14</sup> Dr. James Baker, Creator of Library Carpentry & Lecturer in Digital History

<sup>15</sup> Dr. Melodee Beals, Lecturer in Digital History, Dept. of Politics, History & International Relations School of Social, Political & Geographical Sciences Loughborough University Leicestershire, UK

<sup>16</sup> Managing the Programming Historian, Dr. Adam Crymble (<http://programminghistorian.org/>)

<sup>17</sup> Mahendra Mahey ([mahendra.mahey@bl.uk](mailto:mahendra.mahey@bl.uk)): working to secure funding for DH lesson development

<sup>18</sup> Director, Neil Chue Hong & Training Lead, Dr. Aleksandra Nenadic

<sup>19</sup> To be announced during spring 2017

<sup>20</sup> New member institution from spring 2017

<sup>21</sup> Newspaper group

<sup>22</sup> [https://wiki.neic.no/wiki/Pool\\_Competencies](https://wiki.neic.no/wiki/Pool_Competencies)

<sup>23</sup> <http://coderefinery.org/> Project Lead, Radovan Bast

<sup>24</sup> Mateusz Kuzak ([m.kuzak@esciencecenter.nl](mailto:m.kuzak@esciencecenter.nl)): works with many DH researchers at the Netherland eScience Center

<sup>25</sup> Executive director Jonah Duckles

<sup>26</sup> Executive director Tracy Teal

## Description of work

The work proposed falls in two distinct, but related phases illustrated in Figure 1 and with a detailed description in the following.

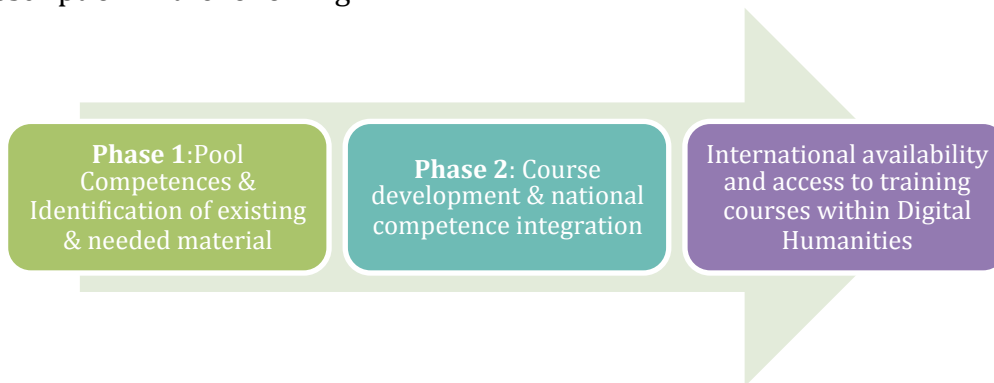


Figure 2: An illustration of phase 1 and 2 of the project (green boxes) and its focus of future accessibility (purple box).

### Phase 1: Pool digital competences within humanities

This initiative possesses a Nordic ambition for basic digital competence integration into humanities research communities. A series of training courses will be identified within the three topics listed in the introduction.

The objective of this phase is to identify research areas and existing competences within Digital Humanities and thus identify relevant areas for competence development. It is important to identify existing activities and existing course development to avoid double work – and it is important to identify both gaps and areas, where existing material might be transformed into DH course material. The objective therefore is to:

- Identify existing courses which can be shared and reused
- Identify needs for new course development

In order to achieve this, we need to:

- Identify categories of research questions based on digital object types as well as research methodologies
- Identify relevant competences in relation to these
- Identify already existing courses to support the competence development
- Identify gaps and describe the content of required new courses
- Identify sustainability models for the course material

The above will result in a catalogue of existing and proposed courses coupled to research groups. The catalogue will include estimates of potential target group and costs for the development of the required course material.

The catalogues will be discussed at a workshop, which will result in a Nordic initiative for shared competence development. After the workshop relevant activities will be identified and a working group will convert these into a detailed proposal for Phase 2 (Appendix).

### Methodology for Phase 1

A set of questions addressing status and needs within the three focus areas, will be developed and raised in each of the Nordic countries, in order to ensure common ground.

To minimize the duration of this initial phase we will kick start the development of the questionnaire with a workshop, where an initial version will be presented and the quality

tested. After the workshop, the questionnaire is ready to send out (a web version) and we have a first mailing list of names. The methodology for collection of respondents for the questionnaire will be snowball sampling.

Based on the responses, a group of researchers in each of the Nordic countries will identify relevant competences and relevant course material for the three topics in focus. All contributions are collected, digested and are grouped in relevant competence area and relevant existing course material. A Nordic workshop for related partners will be held leading to either existing courses or courses to be developed within each of the three topics.

### **Outline of Phase 2: Initial proposal on course development and integration**

Each topic will form a group of experts and these three groups will work in parallel to develop the identified courses. Each group will appoint a responsible person for respectively deliverables and budget.

Sustainability of the course material is important and the activities below assume this is achieved through an agreement with software Carpentry.

Course adjustment and developments will follow the Roadmap for Data Carpentry and Software Carpentry workshop and lesson development with respect to each group's conditions and budgets. Software and/or Data Carpentry will ensure engagement with each of the three groups during both course adjustments of reusable material and development.

All groups will meet during the development phase to ensure consistence and to minimise overlap of content between the courses.

A joint partner summer school will be held and first round of "train the trainers" initiated in order to test and qualify the material and the approach. Hereafter updated versions of course material will be accessible with Software Carpentry and/or Data Carpentry course portfolio(s) and relevant platforms on national, Nordic, European etc. level.

### ***Methodology for Phase 2:***

The proposed course portfolio will be negotiated with SC; based on the available economy for these activities. Courses will be selected, respectively for sharing and further development in close collaboration with Software Carpentry in order to ensure correspondence with used standards and qualification procedures on new materials.

The courses will be based on the pedagogical principles developed by the researchers in the Nordic countries working with pedagogical methodologies and distant learning in dialog with the members of software carpentry.

Project partners behind the different themes will develop the course material and make existing course material shareable through Software Carpentry and also other relevant (national) platforms.

### **Timeline**

Phase 1 will start in August with a first workshop in September involving mainly representatives from the research community. The expected duration of the project is 4 months. It will be finalized with a workshop involving all stakeholders. The results will be

presented and a procedure for writing the proposal for Phase 2 and for selecting relevant course material will be determined based on input from the project coordinator.

Phase 2 is dependent on the outcome of Phase 1 and the success of external funding. However, the project initiators will strive towards a start-up as soon after phase 1 as possible. Phase 2 will hold Train-the-Trainer workshops of selected courses approximately half a year after the Phase 2 initiation. The courses identified to proceed for Phase 2, will be developed and made accessible through relevant platforms and organisations, such as the international non-profit organisation Software Carpentry, Data Carpentry, the NeIC Training Platform, the NeIC Coderefinery platform and also other relevant national eScience platforms.

## Budget for Phase 1

The basis for the budget in Table 2, is that the cost of meetings and training facilitation between partners is applied for through NeIC. Whereas time spend on attending meetings and workshops will be in-kind contributions from participating partners.

We further propose that the project secretariat function will be dedicated to the coming NeIC Training Programme Manager as an in-kind contribution to the project on behalf of the NeIC secretariat and its Nordic member countries. This project will speed up such a new training programme manager's engagement with the Nordic research communities and infrastructure providers significantly and furthermore the project will benefit from having a neutral and Nordic project coordinator and facilitator.

**Table 2: Project activities and deliverables vs. estimated costs (in NOK) for Phase 1.**

Phase	Activities	Deliverables	In kind	NeIC
1	Project lead , DIGHUMLAB	Responsible for deliverables and budget	½ PM <sup>27</sup>	
1	Project secretariat	Responsible for approving project deliverables and releasing funding. Meeting and workshop facilitation. Communication between partners. Coordinating activities and deadlines together with the project lead.		10% NeIC training programme manager <sup>28</sup>
1	National identification of courses, competences & needs	National reports & catalogue development (existing courses coupled to local research groups). Will deliver a catalogue of existing courses in the Nordic Countries.	½ PM <sup>27</sup>	270.000 <sup>29</sup>
1	Two Nordic workshops: 1'st: On the state of art and the formulation of the questionnaire 2'nd: Identifying courses and competences to be shared or developed	Catalogue of existing & proposed courses coupled to research groups. Future recommendations to Phase 2 (appendix). Expected no. of participants: 25 persons . Will after the second workshop deliver a list of recommended courses to be developed as part of this project (big cultural data) along with estimates of the cost and outline of potential sustainability models for each of the recommended courses.	Meeting room facility <sup>30</sup>	45.000 <sup>31</sup>
<b>Total</b>				<b>315.000</b>

<sup>27</sup> DIGHUMLAB; Birte Christensen-Dalsgaard

<sup>28</sup> Anne-Marie Bach, NeIC

<sup>29</sup> 1 PM +10.000 NOK per researcher from each Nordic country (5\*60.000 NOK)

<sup>30</sup> DeIC, if not provided by any other partner

<sup>31</sup> Catalogue development time, per group +travel, accommodations, meeting supplies if needed