EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Exploratory Workshop Scheme

Standing Committee for the Humanities (SCH)

ESF Exploratory Workshop on

Digital Palaeography

Würzburg (Germany), 20–22 July 2011

Convened by:

Dr. Malte Rehbein

SCIENTIFIC REPORT

1 Executive summary

1.1 Background: Palaeography and Digital Palaeography

Palaeography is the study of old handwriting. Palaeographers usually answer questions of high importance for understanding historical manuscripts: what was written, when and where was this written and by whom. Palaeography is hence a basic prerequisite for any kind of work with primary sources for which it is roughly estimated that the libraries and archives of the medieval Western world alone hold between 500,000 and 1 million manuscripts, charters, rolls and registers. Without the endless opportunities for discovery and renewal contained in such sources, scholarship could only end up going round in circles, accumulating glosses of glosses or gratuitous extrapolations, as it sometimes will.

Constructing readings of ancient documents and exploring their scribal practices and contexts is a difficult, complex, and time-consuming task, often involving reference to a variety of linguistic and archaeological data sets, and the invocation of previous knowledge of similar documentary material. Due to the involved reading process, it is difficult to record how the final interpretation of the document was reached, and which competing hypotheses were presented, adopted, or discarded in the process. It is also difficult to acknowledge and present the probabilities and uncertainties which were called on to resolve a final reading of a text.

Hence, scholars world-wide have begun to develop and employ new or methods using advanced digital technology for palaeographic research. This “digital palaeography” is about to improve and enhance the traditional methodology, and, as of today, there are several projects concerned with it. These encompass a wide range of scientific, interdisciplinary approaches, many of which are represented in this Exploratory Workshop, but despite such single projects, a comprehensive approach is still missing.

1.2 The Exploratory Workshop

This Exploratory Workshop was convened by Dr Malte Rehbein (Julius-Maximilians-Universität Würzburg, Germany) and hosted at the University Library in Würzburg from 20th to 22nd of July, 2011. The surroundings permitted a fruitful, constructive, and in some cases controversial discussion and led, also in conjunction with a supporting programme, to many new contacts, sharing of knowledge and first ideas towards a research agenda to be developed at an international level.

The Workshop brought together 23 researchers from ten countries. The participants represented the fields of manuscript studies in general, palaeography in particular, computer sciences and Digital Humanities as well as different disciplines of the Humanities: history, art history and literature.

The Workshop was opened by the keynote presentation “Palaeography today. Old questions and new technologies” which gave input and an anchor point for the recurring discussions about the role and limitations of computational means in palaeography. Following this, the remainder of the Workshop was structured in five sessions:

- Three *thematic sessions* (“Enhancing Palaeography”, “Crossing the disciplines”, and “The Digital World”) in which presentations (nine overall) prepared beforehand by the participants gave input for discussion, both immediately following and concluding each session;

- One *practical session* in which the participants were provided with case-studies to deepen the issues raised and to validate the theories developed before;
- A final session called *Perspectives* which served as a summary of the whole Workshop and which led to the outline of a research agenda on Digital Palaeography.

The thematic sessions ranged from theoretical considerations through presentations of new and recently finished projects to ideas and prolegomena for future research. Topics covered encompass web-based resources and databases for palaeographic and manuscripts studies, physical and genetic examination of manuscripts, optical character recognition of script, computational methods for joining fragments, spatial exploration tools and automatic layout and character analysis.

1.3 Key Outcomes

The Workshop met its primary objectives, namely to bring together scholars from various disciplines and interests in order to explore the potential of Digital Palaeography, to establish a European community of practice and to identify opportunities and needs for future international and interdisciplinary collaborations.

The outcomes of the Workshop can be summarized as follows

- Outline of a research agenda encompassing the fields: international community building, conducting projects and experiments, and (re-) thinking the theoretical foundation of (digital) palaeography;
- A joint proposal for an international follow-up workshop on “Computation and Palaeography” with a stronger focus on computer sciences;
- The creation of a web-based platform for information interchange;
- The establishment of a working group on “describing handwriting”, one of the desiderata identified during the Workshop;
- Publication of the proceedings in 2012.

In addition, the workshop has already earned international response even outside countries with invited participants. See the conference website for an overview: http://www.zde.uni-wuerzburg.de/veranstaltungen/digital_palaeography/

2 Scientific content of the event

The aim of this report is to provide an overview of the scientific content of the Exploratory Workshop on "Digital Palaeography", held at Würzburg University in July 2011, to present the individual contributions in that context and to give a synopsis of the subsequent discussions that followed each presentation and thematic session.

2.1 Introduction

The workshop was opened by an introductory presentation by the workshop convenor, Malte Rehbein (Würzburg), who gave a brief overview of the state-of-the-art in Digital Palaeography, groundwork achieved so far and the history of this workshop. He claimed that research in Digital Palaeography is making progress but that in order to develop it as a research field, a community effort is still missing. Preparatory work on the latter has been undertaken, however, for instance by the "Codicology and Palaeography in the Digital Age" initiative. Hence the main objectives of this workshop: to bring together a variety of people, ideas, experiences and interests to promote innovation, creative outcome and synergies.

In his stimulating and thought-provoking keynote address, Eef Overgaauw (Berlin) spoke about "Palaeography Today. Old questions and new technologies". The use of new technologies in palaeography, he argued, will help answer unsolved problems of the past but will not necessarily lead to new research questions. Whether the areas of computers in palaeography he named (statistical methods, digital reproduction of manuscripts and databases) are sufficient or not was disputed in the following discussion and needed to be revised and extended in the course of the workshop (e.g. towards social media and artificial intelligence). His distinction between palaeography as a scholarship and palaeography as a connoisseurship (which are in his point of view not exclusive but complementary) turned out to be a useful anchor point in the recurring discussions about the role and limitations of computational means in palaeography.

2.2 Enhancing Palaeography

The first thematic session of the workshop was devoted to the question how the digital medium – especially by allowing complex operations on large sets of data – can support the fundamental questions of palaeographers such as what is written, when and where was it written, by whom was it written, was script A written by the same scribe as script B, how did scripts develop etc. Approaches highlighted in this session included the creation and provision of digital resources, corpora-based analyses and codicology.

Stewart Brookes (London) presented "Digital Resource for Palaeography, Manuscripts and Diplomatic", a relatively new research programme, funded for four years by the European Research Council and directed by Peter A. Stokes (London). Its aim is the creation of a web-based framework for the study of scripts with 11th century English vernacular minuscule as main focus, including digital images and formalised data of about 1,200 scribal hands. Although being only a side product, one of the major outcomes of the project might be, as was commonly agreed in the discussion, a standard terminology for the description of scripts. This is not only a pre-requisite for the resource developed in this project but for quantitative palaeography on a larger scale in general. Especially if shared and interoperable knowledge between individuals, projects and data-sets became reality, it may constitute a qualitative leap in palaeographical research and scholarship.

Picking up this thread, Wendy Scase (Birmingham) stated that many resources, especially manuscript corpora, exist but are not yet targeted and could not be exploited for palaeography. In her talk “New Methodologies for Effective Exploitation of Digital Manuscript Corpora”, she argued that although only a minority of existing images and metadata were created specifically to support palaeography teaching and research, much of these materials can be of immense value to palaeography. Here, the discussion focused again on questions of interoperability of data which is apparently still an issue as standards (besides terminology for palaeography, image and manuscript metadata and cataloguing information were mentioned) are either non-existing or unsatisfying or – which might even be a bigger problem – not accepted. As another obstacle for comparative research (which is essential for palaeography), copyright and monetary issues (raised by repository owners) were named and put onto the agenda for further discussion.

After the presentation by Paola Errani (Cesena): “Parchment and Scribes in the Malatestian Scriptorium”, the open discussion incorporated codicological data, i.e. research on the physical structure of manuscripts as. Errani presented a project that aimed at identifying one of the codicological features of the parchment used in the scriptorium of Malatesta: the thickness of the leaves composing the quires of each manuscript. She suggested that the sources of supply and methods of treatment of parchment were unique or at least standard and relatively invariant throughout the life of the Malatestian scriptorium. It was, however, questioned whether such data provides recognizable patterns to be used for identification or dating.

One of the chief overall outcomes of the presentations and discussions of this first thematic session was that the palaeographic method could be enhanced by 1) bridging the different levels of investigation outline here, e.g. by combining palaeographic data with codicological information and manuscript meta data and by 2) moving from singular approaches to investigations based on large-scale sets of data. The computer in this context, however, is regarded only as an auxiliary means or tool, with the interpretation of the collated and visualised data still being the task of the trained palaeographer.

2.3 Crossing the Disciplines

The second thematic session shifted the viewpoint away from the underlying research questions of palaeography and how these can be supported towards methods developed in other disciplines (here: biology, computer sciences, professional character recognition software) and their application to the field of palaeography. While the first session was driven by the viewpoint of the Humanities, here the focus was technology driven.

Timothy Stinson (Raleigh) presented recent observations based on preliminary tests in “DNA Analysis and the Study of Medieval Parchment Books”, and offered suggestions for defining and implementing future genetic studies of parchment. Potential applications of the genetic analysis of parchment, he proposed, include not only codicological studies, but also the mapping of trade routes and the study of medieval animals and animal husbandry as well as parchment's nonpareil value as archaeological evidence. The vision of dateable and localizable manuscripts was clearly recognized, but more research is required as the experiments are at a very early stage only and more data is needed.

By digitization, on the other hand, a lot of data has been produced already and even manuscript digitization is turning into mass production. Moving from mere image digitization towards fulltext transcriptions, however, is a tedious enterprise that is still far away from being industrialized. Torsten Schaßan (Wolfenbüttel) gave an overview of "OCR [optical character recognition] for manuscript and early prints" based on recent evaluations at German libraries. It was concluded that systems for OCR, even for early prints, let alone for manuscripts are still far away from being productive and that for the moment, double key is more accurate, especially for smaller projects. Again, improvement of such OCR-systems in the future might rely on huge databases, but in this case the previously mentioned questions of data standards and interoperability of databases would need to be addressed first.

The project presented by computer scientist Lior Wolf (Tel-Aviv) aims at "Identifying Join Candidates in the Cairo Geniza", a collection containing approximately 350,000 fragments of mainly Jewish texts discovered in the late 19th century. The proposed method for this enterprise is based on a combination of local descriptors, different algorithms and learning techniques. Here, the role of computers is different from the previous presentations as the software decides itself what data to use and automatically suggests solutions. Following up Overgaauw's categorization from the beginning, a new level of employing computers for palaeography hence emerges, that of artificial intelligence (AI). Although, the application of AI for palaeography is certainly only at its very beginning, its potential is clearly visible as was proved by the Genizah project.

The overall discussion was followed by a "practical session" in which thoughts and ideas developed so far were validated against some case-studies compiled and provided by Peter A. Stokes (London) from his research. The third session was titled "The Digital World" and provided a thematic continuation of the previous one. Matthieu Exbrayat (Orléans) presented "Spatial Exploration Tools in the Graphem Project", a project only recently finished. The aims of Graphem included the study of various pattern recognition techniques applied to digital paleography both for feature extraction, clustering and categorization of historical scripts, and visual exploration of their results. Analogue to the Genizah project, Graphem brought together various teams of computer scientist with humanities scholars. The discussion highlighted especially this cross-domain experience as being most valuable to produce innovative methods and tools. It was noticed, however, that results from the Graphem project have mostly been published in computer science journals and – like much other research in Digital Humanities – have not yet made their way into the "traditional" Humanities publications.

Ségolène Tarte (Oxford) added another aspect to this discussion. "Interpreting Ancient Documents: Of Avatars, Uncertainty, and Knowledge Creation" presented the cognitive processes papyrologists tap into when deciphering, transcribing and interpreting ancient documentary artefacts and how a new ontology developed from there influences scholarly practice. Here, her agreeable statement that knowledge creation needs to be repeatable, verifiable and understandable moved the discussion towards objectivity versus subjective interpretation. It was pointed out that even image digitization must be regarded as an interpretation on its own. Hence, guidelines, standard requirements and transparency in manuscript digitization are an essential prerequisite for bringing manuscript studies to the next level. The current situation, however, shows that in many digitization enterprises even the motivation behind the acquisition of digital images remains unclear.

The final presentation by Melanie Gau and Robert Sablatnig (Vienna) showed another example of cross-disciplinary work. It picked up the considerations about artificial intelligence but went a step further still by suggesting a new paradigm in which the computer does not simply use calculation or visual means to provide aids for decision making, but in which the computer is trained to think like a palaeographer. This certainly is a research topic for the future, but it has its (obviously successful) analogy in computer chess: programs are shifting, or have been already, from brute-force calculation to human thinking. Two aspects which may be essential human, namely intuition and curiosity (or interest) might, however, never been taught to a computer.

2.4 Perspectives

It was astonishing throughout the whole discussions on this topic of digital palaeography that questions that were originally meant quite practically (e.g. what is the best technology to solve a concrete palaeographic problem) often led to fundamental theoretical considerations. The traditional "palaeographic method" as such was particularly questioned and – in its historical dimensions – regarded as not being always transparent to others. Here, new technology with its requirement of formalisation and unambiguous terminology might not only lead to new knowledge but might also help to sharpen a critical assessment of "old" knowledge.

In the concluding session of the workshop, moderated by Malte Rehbein (Würzburg), results have been compiled and summarized. One key outcome of the workshop as a whole is the need to develop a "research agenda" for Digital Palaeography as outlined in Figure 1, including:

1. International community building;
2. Conducting projects and experiments;
3. (Re-) Thinking the theoretical foundation of (digital) palaeography.

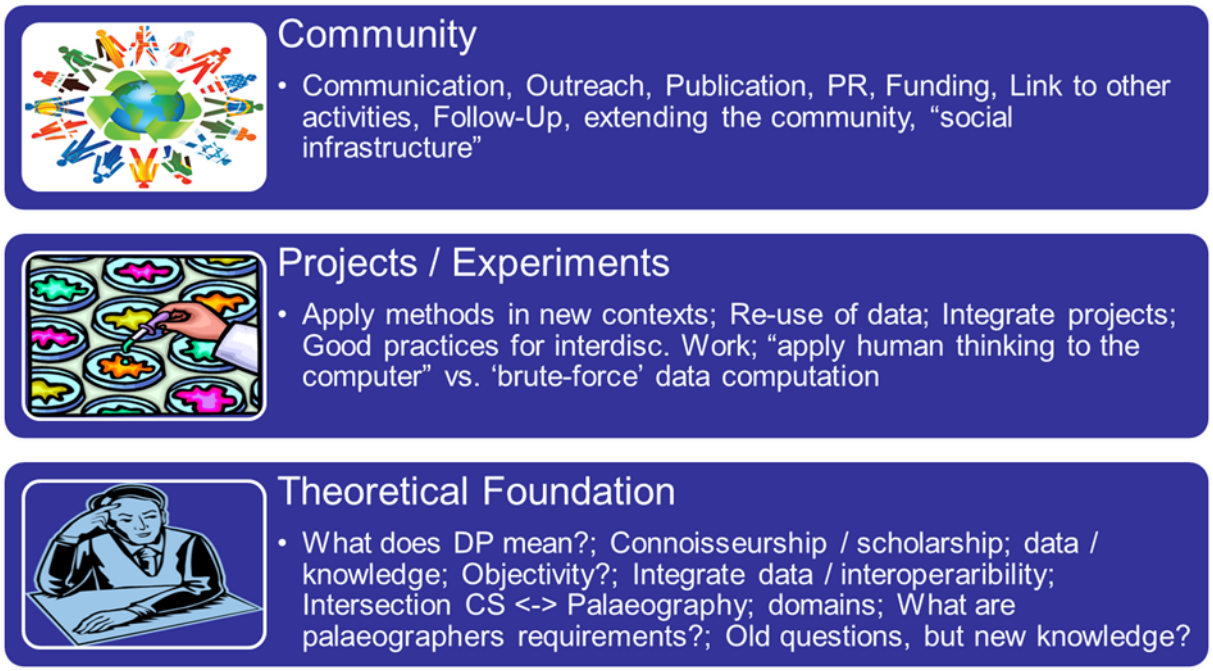

Figure 1: Future directions of Digital Palaeography as developed during the concluding discussion.

3 Assessment of the results, contribution to the future direction of the field, outcome

The workshop supported the departure from individual and singular approaches towards the development of an international community of researchers and research in the field of Digital Palaeography. It also promises to have impact on the traditional palaeographic method.

Already during the Würzburg workshop, many new contacts have been established and ideas for "bilateral" projects developed. This encompasses the application of methods developed in one project for the use, *mutatis mutandis*, in different contexts.

Some actions towards the proposed research agenda have already been specified or even realised. These include:

- An application for a Dagstuhl Perspectives workshop on Computation and Palaeography has been submitted jointly by Tal Hassner and Lior Wolf from Tel-Aviv University, Peter Stokes from King's College London and the convenor of this Exploratory Workshop, Malte Rehbein, Würzburg University. This joint effort also indicates the international dimension of this research as well as its interdisciplinary dimension. The next workshop shall have a stronger focus on Computer Sciences and might especially address the potential of Artificial Intelligence towards script and scribe identification in palaeography and forensics to which the presentations by Wolf and Gau/Sablatnig have demonstrated first ideas;
- A website has been published which is dedicated to the workshop and which provides summaries of the presentations along with additional material and responses to the workshop published in different media. It is available at http://www.zde.uni-wuerzburg.de/veranstaltungen/digital_palaeography/;
- Liaison with related research initiatives on international dimension, such as DigiPal (Digital Resource and Database of Palaeography, Manuscripts and Diplomatic), funded by the European Research Council (realised), DARIAH (Digital Research Infrastructure for the Arts and Humanities), funded under FP7 (specified), and especially NeDiMAH (Network for Digital Methods in the Arts and Humanities) a recently established ESF Research Networking Programme (specified);
- A working group on "describing handwriting" is to be convened by Peter Stokes (initial material for which has already been published at http://digipal.eu/blogs/blog/describing-handwriting-part-i/);
- Publication of selected presentations the proceedings of this workshop is planned and discussions with potential publishers are already well underway.

4 Final programme

	Wednesday, 20 July 2011
Afternoon	*Arrival*
17.00-18.00	Malte Rehbein, Welcome by Convenor; participant's introduction; "Briefing" on workshop programme and possible outcome; discussion
18:00-18.45	Eef Overgaauw, "Palaeography today. Old questions and new technologies"
18.45-19.00	Discussion
	Thursday, 21 July 2011
09.30-10.00	Claudine Moulin, Presentation of the European Science Foundation
	Thematic Session 1: "Enhancing Palaeography"
10.00-10.30	Stewart Brookes, "Digital Resource for Palaeography, Manuscripts and Diplomatic"
10.30-11.00	Wendy Scase, "New Methodologies for effective exploitation of Digital Manuscript Corpora"
11.00-11.15	*Coffee / Tea Break*
11.15-11:45	Paola Errani, "Parchment and Scribes in the Malatestian Scriptorium"
11:45-12.15	Discussion & Summary of Session 1
12.15-13.30	*Lunch*
	Thematic Session 2: Crossing the Disciplines
13.30-14.00	Timothy Stinson, "DNA Analysis and the Study of Medieval Parchment Books"
14.00-14.30	Torsten Schaßan, "OCR for manuscripts and early prints"
14.30-15.00	Lior Wolf, "Identifying Join Candidates in the Cairo Genizah"
15.00-15.15	*Coffee / Tea Break*
15.15-15.45	Discussion & Summary of Session 2
	Natasa Golob, "A few art historical reflections"
15.45-17.45	Peter A. Stokes, Practical Session "Case study"
19.00	*Dinner*
	Friday, 22 July 2011
	Thematic Session 3: "The Digital World"
09.30-10.00	Matthieu Exbrayat, "Spatial exploration tools in the GRAPHEM Project"
10.00-10.30	Ségolène Tarte, "Interpreting Ancient Documents: Of Avatars, Uncertainty, and Knowledge Creation"
10.30-10.45	*Coffee / Tea Break*
10.45-11.15	Melanie Gau, Robert Sablatnig, "Investigation of Historic Documents with Focus on Automatic Layout and Character Analysis"
11.15-11.45	Discussion & Summary of Session 3
11.45-12.30	*Lunch*

	Thematic Session 4: Conclusion
12.30-14.30	Malte Rehbein, "Perspectives" and overall discussion
14.30-15.00	Wrap-up
15.00	*End of Workshop and departure*

5 Final list of participants

1. Dáibhí Ó Cróinín, National University of Ireland, Galway
2. Dominique Stutzmann, Centre National de la Recherche Scientifique
3. Holger Essler, Universität Würzburg
4. Lior Wolf, Tel Aviv University
5. Marc Smith, Ecole National des Chartes
6. Matthieu Exbrayat, Université d'Orléans
7. Melanie Gau, Vienna University of Technology
8. Paoloa Errani, Biblioteca Malatestiana, Cesena
9. Robert Sablatnig, Vienna University of Technology
10. Timothy Stinson, North Carolina State University
11. Torsten Schassan, Herzog August Bibliothek Wolfenbüttel
12. Wendy Scase, University of Birmingham
13. Peter Stokes, King's College London
14. Stewart Brookes, King's College London
15. Malte Rehbein, Universität Würzburg (Covenor)
16. Nataša Golob, Ljubljana University
17. Eef Overgaauw, Staatsbibliothek Berlin
18. Ingo Kottsieper, Akademie der Wissenschaften zu Göttingen
19. Marjorie Burghart, EHESS (pôle de Lyon)
20. Ségolène Tarte, Oxford e-Research Centre
21. Britta Mischke, Universität Bonn
22. Jinna Smit, University of Amsterdam
23. Claudine Moulin, Universität Trier (ESF Rapporteur)

6 Statistical information on participants

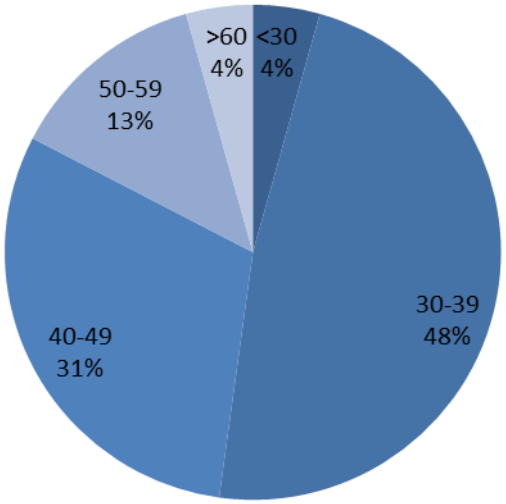

Figure 2: Age Distribution (in years).

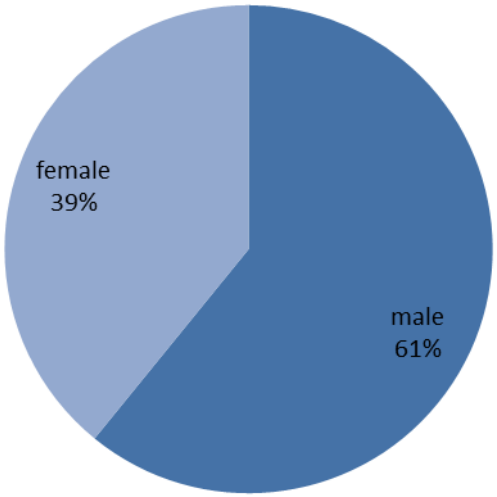

Figure 3: Gender.

Figure 4: Countries of Origin (without participant from US). Map created with communitywalk.com.

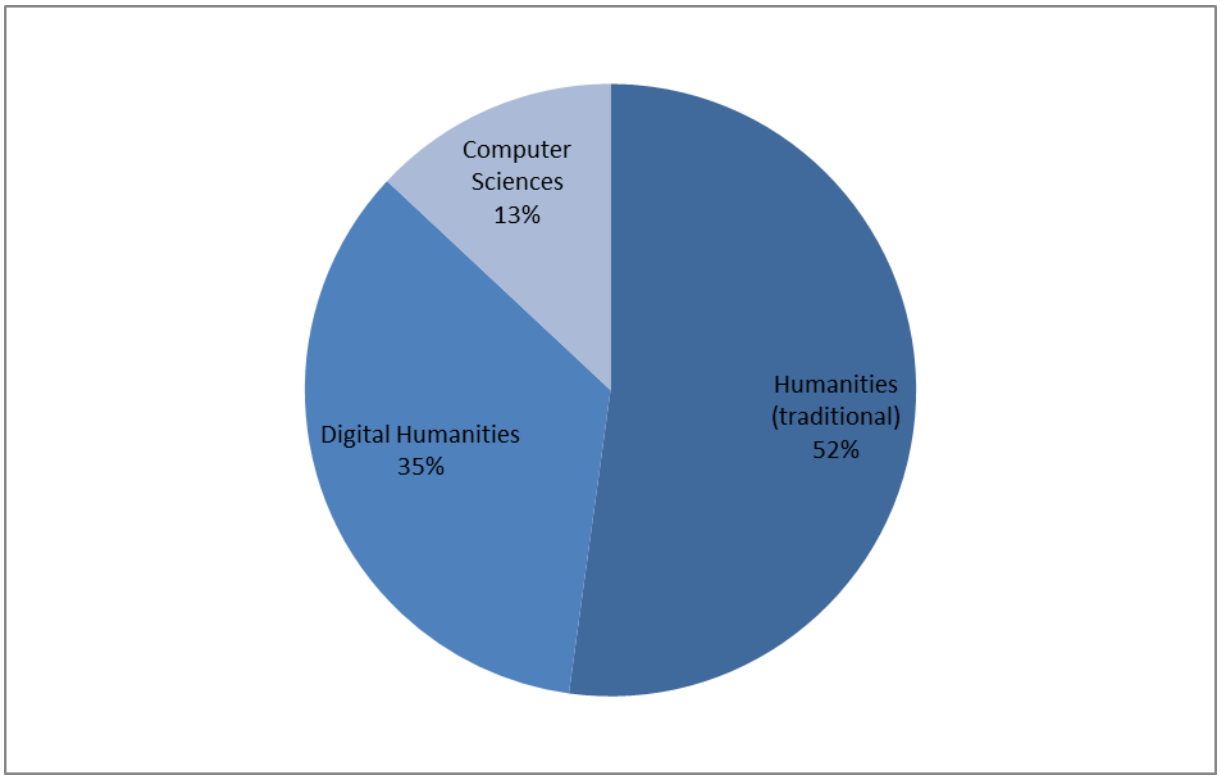

Figure 5: Participants by Scientific Speciality.