Processing of Information: Linguistics, HKI and Digital Humanities. Developments in Cologne

Øyvind Eide / Jürgen Hermes / Andreas Witt

Résumé

L'article intitulé «Traitement de l'information: Spinfo, HKI et humanités numériques - l'expérience de Cologne» présente l'histoire du développement des humanités numériques au sein de l'Université de Cologne. L'institutionnalisation des humanités numériques a commencé encore à l'époque où dans le monde germanophone le périmètre de la discipline était en train d'être défini par les travaux de quelques pionniers. Parmi eux, il convient de souligner le rôle d'Elisabeth Burr, active notamment à Tubingue, Duisbourg, Brême et Leipzig.

L'article retrace le développement des humanités numériques à Cologne à partir de leurs débuts dans les années soixante du 20ème siècle, en passant par leur consolidation dans les années quatre-vingt-dix, jusqu'aux deux dernières décennies, quand Cologne est devenu un centre important de cette discipline. Le processus illustre comment une nouvelle discipline scientifique peut s'institutionnaliser au sein d'une université allemande. L'article décrit la perspective de deux domaines fondateurs: le traitement linguistique de l'information (en allemand: Sprachliche Informationsverarbeitung, Spinfo) et le traitement historico-culturel de l'information (en allemand: Historisch-Kulturwissenschaftliche Informationsverarbeitung, HKI) et leur synthèse, qui a abouti en 2017 à la création de l'Institut des Humanités Numériques (Digital Humanities), qui aujourd'hui est - du point de vue interne - une composante de la Faculté de Philosophie de l'Université de Cologne et - du point de vue externe - une partie intégrante de la communauté internationale des humanités numériques.

Traduction: Paweł Kamocki.

Abstract

Artikkelen „Informasjonsbearbeiding: Spinfo, HKI og digital humaniora -- en historie fra Köln" følger de historiske linjene i digital humaniora (DH)-utdanningen ved Universitetet i Köln, hvor institusjonaliseringen av DH begynte på et tidspunkt da fagretningen i det tyskspråklige området fortsatt var preget av noen få pionerer. I første rekke blant disse pionerene var Elisabeth Burr, som framfor alt var aktiv i Tübingen, Duisburg, Bremen og Leipzig.

Artikkelen viser at utviklingen i Köln startet allerede på 1960-tallet. På 1990-tallet kom den institusjonelle forankringen som etablerte Köln som det viktig DH-sentret det har vært de siste 20 årene. Dette viser også hvordan et nytt fag kan bli institusjonalisert på et tysk universitet. Artikkelen klargjør de ulike perspektivene vi finner i de to retningene språklig informasjonsbearbeiding (Spinfo) og historisk-kulturvitenskaplig informasjonsbearbeiding (HKI). Vi ser også

215

hvordan en syntese mellom de to, som til slutt resulterte i etableringen av Institutt for Digital Humaniora (IDH) i 2017, ble til. Fagretningen er således en fast bestanddel av det humanistiske fakultetet internt i Köln, men er samtidig, sett fra utsiden, en fast del av det internasjonale DH-samfunnet.

1 Introduction

In 2017, a new department was created at the University of Cologne: the Department for Digital Humanities (German: "Institut für Digital Humanities, IDH"). 103 new students were enrolled in the two BA study programmes organized by the department for the winter semester 2020/21, alongside 17 MA students. This chapter is dedicated to the history behind this department, the reasons for establishing it and how it is currently thriving in the context of the so-called DH continuum. The DH continuum consists of different units at the University which are all, in different ways and to varying extent, focusing on DH. While the main parts of this chapter are related to teaching, some remarks will also be made on research.

Histories have starting points and turning points. This is also the case with the history of DH (which is additionally complicated due to the fact that the area has been developing since long before the term 'DH' was coined). The term DH was introduced to a larger audience in the meaning intended here only around 2004–2005. There is no earlier single term covering what is today known as DH. Computing in the humanities (CHum) covers parts of it, meaning parts which are close to what was the focus in Cologne since the 1990s. Given the central role of computational linguistics in Cologne, the CHum term applied to the Cologne situation is understood to include computational linguistics as well.

2 Origins: the invention of the Cologne Model in the 20th century

As is the case also in many other universities, CHum first gained a foothold at the University of Cologne in the form of computational linguistics. Already since the early 1960s, Paul Otto Samuelsdorff, one of the pioneers in this area of research in Germany, taught the programming languages *Fortran* and *Lisp* (and later also *Pascal* and *Prolog*) and the benefits of using a computer for the analysis of natural languages and its applications. One of his students in the 1970s was Jürgen Rolshoven, who later – in 1989

– was appointed to a professorship at the University of Cologne, where he founded the Chair of Linguistic Information Processing (in German "Sprachliche Informationsverarbeitung", Spinfo). Like Elisabeth Burr, who also started her career mainly focusing on linguistics of the Romance languages and then dived into what later developed into DH, Jürgen Rolshoven had a background in Romance studies, and did his PhD and postdoc in this field. Thus, his professorship was primarily oriented towards computational linguistics and computational philology rather than computer science.

From the beginning, however, his vision was to offer specialized IT facilities also for students and staff not primarily oriented towards linguistics, i.e., for the entire Faculty of Arts and Humanities in Cologne, which is one of the largest of its kind in Germany. The Internet age was just emerging and the Spinfo staff were pioneers in that period. Both the Ethernet cabling of the main faculty building (Philosophikum) and the establishment of the first computer pool in this "headquarter" of the faculty are due to the Spinfo department.[1] The pool, which can be used for digital teaching as well as for self-study, is still maintained by the IDH.

Another decisive, and in retrospect perhaps the most important, step towards establishing CHum in Cologne was the development of a study programme called Information Processing. It was designed by Rolshoven and his staff and started in 1997 as a Magister programme, which was a pre-Bologna structure approximately at the level of a complete BA+MA study, but without the separation along the lines of today's BA and MA levels. In this Cologne Magister programme, three fields of study were combined. These could be chosen from the entire range of the Faculty of Humanities as well as beyond, also including, for instance, Geography, Sociology, or Education. The Information Processing programme accommodated this by offering not only a language orientation but also another, complementary orientation. For the latter, a professorship for Historical-Cultural Information Processing was established in 1997 and held for the first two years by Guido Mensching, whose postdoctoral qualification (Habilitation) was supervised by Rolshoven. He left Cologne in 1999, after being appointed professor of Romance linguistics at the FU Berlin.

---

[1] For the history of the early years of CHum in Cologne see Mensching et al. (2018: 9ff).

3 Antagonisms: the first 15 years of the new Millennium

In 2000, Manfred Thaller, who came from a position as the inaugural director of the Humanities Information Technology Centre (The HIT Centre) at the University of Bergen in Norway (established in 1997),[2] was appointed to this professorship. The denomination of his professorship was "Historisch-Kulturwissenschaftiche Informationsverarbeitung"[3], short HKI, making Thaller the first full CHum professor outside of linguistics in Germany.

Thus, at the beginning of the millennium, Cologne happened to have two CHum professors, but there was still a long way to go before a department would be founded. In fact, the two professorships were not even assigned to any institute at first, they were solitary units within the faculty. Later, the Spinfo was included in the Department of Linguistics and the HKI ended up at the Institute of Archaeology. It further complicated the situation that these two departments, while wandering through different subject groups (the faculty is divided into eight of them), never actually found themselves in the same one. Thus, they always belonged to different organizational structures within the faculty.

The programs were also organized separately: In their first year, Magister students of Information Processing had only two courses in common – Introduction to Information Technology and a programming course. Later they had to specialize in courses offered by the Spinfo that more or less covered Computational Linguistics, or in courses offered by HKI, which more or less covered the other areas ("Professor for the rest", cf. Thaller 2017: 54). Information Processing could be studied as a full-fledged subject at the Faculty of Arts and Humanities and was thus – already at that time – equal to the traditional subjects. At most other universities at the time, CHum could only be studied as an auxiliary minor or even in the form of short methodological courses, often offered by administrative IT units. Although both professorships strongly supported Information Processing as a full-fledged subject, their very different methodologies prevented stronger integration in joint courses (cf. Thaller 2017: 55f).

---

[2] The pre-history of the HIT Centre goes back to 1972, when "NAVFs edb-senter for humanistisk forskning" was established.

[3] An English translation of this term would be "Information Processing with a focus on historical and cultural disciplines".

It was only with the Bologna reform that the need to more firmly define the content or the competences to be taught forced the resources of the two professorships to be more closely connected, resulting in a BA programme in Information Processing introduced in 2007, in which all students now took courses in both the previously separated fields. In the two-subject MA designed at the same time, students could again specialize in Spinfo or HKI. HKI also participated in the BA-MA joint degree program in Media Informatics, which was introduced in the same year. On behalf of the faculty, HKI also established the IT-Certificate of the Faculty of Arts and Humanities as an offer for the Studium Integrale (SI), available to all students at the faculty as well as beyond. The BA programmes were more than well received, virtually from the beginning there were always more interested students than places. However, the majority of MA students wanted a more integrated programme, which was implemented as part of the 2015 reaccreditation as a single-subject MA in Information Processing. This MA programme consists of a combination of HKI and Spinfo courses.

In parallel to the establishment of the study programmes, both chairs devoted themselves to research in different areas. Initially, Rolshoven was involved in the development of computational approaches to the Generative Grammar, with a strong connection to compiler generators (cf. Rolshoven 1991: 133ff). He also specialized in digitization of literary and lexical resources for small language communities, especially those of Romansch (cf. Rolshoven 2012: 119ff). In later years, Rolshoven's research focused primarily on the conceptual pair "pattern and meaning", which he interpreted in an interdisciplinary way – especially in cooperation with bioinformatics (cf. Wiehe 2018: 327ff).

Thaller's broad research interests, with history as one of their starting points, were focused on the development of computer tools for researchers and students in the humanities since the 1970s.[4] During his 15 years in Cologne, he also initiated and took part in a number of large research and infrastructure projects covering broad areas spanning the humanities. In addition to that he also organized a number of academic events leading to important publications such as "Controversies around the Digital Humanities" (Thaller 2012).

Around 2008, the Cologne Faculty of Arts and Humanities decided to actively promote the establishments of research centres in order to enable

---

[4]   See Thaller (2017) for further details.

research through focus-based organizational units. In 2009, this initiative ultimately gave rise to the Cologne Center for eHumanities (CCeH), in which both chairs of information processing participated as members of the Board. Thaller became the spokesperson for this centre for the first four years. Patrick Sahle was appointed the first managing director, funded by overhead from Thaller's research projects. The CCeH grew rapidly, especially after Thaller's successor as CCeH spokesperson, Andreas Speer, in 2015 succeeded in anchoring the *coordination office DH* of the *North Rhine-Westphalian Academy of Sciences, Humanities and the Arts*[5] to the CCeH, and still plays an important role in the Cologne DH continuum. The third player in this relatively unique construct became the Data Centre for the Humanities (DCH), which owes its establishment largely to the *KA3*[6] project. This project, also started in 2015, was led by Nikolaus Himmelmann, and goes back to a call for proposals by the BMBF to promote the establishment of DH centres. The development shows that Digital Humanities indeed had influential advocates, who helped to build a sustainable structure also at the faculty beyond the two professorships mentioned earlier.

4 Synthesis: the foundation of the IDH

As a final step of merging the two branches of CHum in Cologne, which in fact was already intended in the original layout developed by Rolshoven in the 1990s, the IDH was founded in the summer of 2017, shortly after the two professorships were filled by the successors of Rolshoven and Thaller, namely, two of the authors of this chapter (Eide and Witt). The third author (Hermes) has been central to the history of CHum in Cologne since 1998. He was a staff member of Spinfo since 2002, and became the managing director of the institute when it was established.

While the links with DH in the individual disciplines were important, the cooperation with the other DH institutions (see *DH continuum* below) had to be maintained and strengthened. The University of Cologne supported this endeavour by making available an entire floor (13 offices) of the building at Universitätsstraße 22, directly on the university's central campus. In addition to the IDH, the management of the CCeH and DCH moved to this

---

[5]   See <https://www.awk.nrw.de/about-us.html>.

[6]   See   <https://dch.phil-fak.uni-koeln.de/forschung/ka3-koelner-zentrum-analyse-und-archivierung-von-av-daten>.

floor, ensuring short distances and a lively exchange between these institutions. Other rooms (including rooms for project work, for student meetings, a VR Lab, a library, and several meeting rooms of different sizes) are located in the immediate vicinity. The COVID-19 pandemic which started in spring 2020 disrupted this spatial proximity for a time, but it will continue to be a basis for the DH integration in Cologne in the years to come.

5 Teaching CHUM in Cologne

The first CHum study programme in Cologne, established in 1997 and still going strong, is called *Informationsverarbeitung* (hereinafter: IV). The term has no direct English translation, but the term "Information Processing" conveys the meaning fairly well and is normally used. It started as a Magister programme and was converted into a consecutive Bachelor-Master programme in the course of the Bologna reform in 2007. At the same time, the Media Informatics (further MI) degree programme was launched in conjunction with Media Cultural Studies. Both subjects were reaccredited in 2015 with slight adjustments and will be continued into the future. IV can be combined with any other subject from the Faculty of Arts and the Humanities in the BA, whereas the IV MA is exclusively dedicated to this field of study. The MI programme is combined in the BA and MA with content from Media Culture and Theatre. The focus in MI is more on the visual and performative aspects of the development of computer systems for the humanities, whereas IV is more oriented towards language and linguistics.

Both study programmes are very popular. More than four hundred students are enrolled in the two BAs, distributed roughly equally between Information Processing and Media Informatics. About 50 students follow the Master's programme IV, and about 25 the Master's programme MI. In addition, IDH teaches modules in other Master's programmes. These include the Linguistics (specialization in Computational Linguistics), the Digital and Computational Archaeology, and the Medieval Studies MA programmes.

The University of Cologne is thus a major centre of education for the next generation in CHum, for well-trained project staff and research software developers as well as for more traditional academic careers. However, by no means all graduates end up in academia; the majority are employed

by companies, public administration, and other types of jobs outside academia, often already after graduating from the BA programme.

5.1 BA

There are currently approx. 450 students enrolled in these two BA programmes. Both fields of study have a similar progression: in the first year, basic computer and application systems, basic technology and methods are taught, with some project work. In the second year, the students are taught object-oriented programming (currently Java), and – depending on which of the study programmes the student is following – attend seminars in computational linguistics or in visual programming. The third year is for advanced development seminars and exercises, mostly organized as group work.

5.2 MA

The MA study programmes in IV and MI build on the corresponding BA programmes and include advanced project work and in-depth theoretical studies and discussions. Candidates from outside our BA programmes must document university level competence in an object-oriented programming language as well as in a humanities discipline in order to be enrolled. There are currently about 75 students enrolled in our MA programmes.

5.3 IT-Certificate

IDH also offers an IT certificate which does not educate programmers, but offers all students at the university four sequential modules giving basic and advanced education in computer-based methodology for the humanities and advanced use of relevant tools. The Certificate consists of four courses which can be included in any BA or MA level study at the university. Almost 300 students completed between one and four modules in 2019. Specialized IT certificates for employees, teaching students, and other groups are under development and partly realized based on project funding.

5.4 Modules for other study programmes

The modules of the IT certificate are accessible to all students. The IDH also offers some selected modules from our study programmes to other

study programmes at the university. The MA programme in medieval studies includes two optional Digital Humanities modules offered by us. There has also been a long cooperation with archaeology which will be described further below; it is still operative, with exchange of modules and co-teaching of courses. The Department of Archaeology, specifically the professor in computational archaeology, is currently offering a specialization in Computational Archaeology within the MA in Archaeology. A new international MA study programme in Computational Archaeology will enroll the first students in the winter semester 2021–22. IDH supported the development of this new study programme and offers modules for it, also through co-teaching.

5.5 Research education

There are currently two PhD programmes connected to our department, with ca. 10 students enrolled. We also offer infrastructure for post-doctoral studies. Furthermore, students at all levels, from BA onwards, are actively integrated into research projects, both through their studies and through working as student assistants.

6 The Cologne DH continuum, and beyond

Taking into account the number of study programmes and students IDH is responsible for, the department is rather small. This is partly due to the larger DH continuum in Cologne. In addition to the archaeological department with a dedicated professorship in computational archaeology, IDH also works with a number of university departments and other units, including the regional computing centre, the university library, and the Centre for Data and Simulation Science. The Institute also cooperates with academic departments outside the Faculty of Arts and Humanities: most noteworthy with the Department for Computer Science, but also with the Departments for Numerical Mathematics and Genetics.

IDH's two closest cooperation partners, however, are the Data Centre for the Humanities (DCH) and the Cologne Centre for eHumanities (CCeH). Employees from both of these institutions participate in activities of the Department in a number of ways. They teach courses based on their specific expertise; for instance, research data management or XML technologies. Several of the employees of DCH and CCeH are students in the study programmes at IDH at BA, MA, and PhD levels. This often adds

relevant exposure of research projects to the students' competence, while also providing research projects with engaged participants. Particularly important is the availability of programmers through the study programmes. IDH has an extended scholarly and social cooperation with other departments, with co-location and shared seminar series and workshops. Similar cooperation takes place with the digital art history archive Prometheus, connected to system development, visual programming and image analysis, and research data management.

The Centre for Data and Simulation Science (CDS) was founded in January 2018 as a research centre integrating researchers and research activities across all the faculties of the university. The Centre is organized around five scientific domains, of which one is Digital Humanities. Furthermore, there is a cross-cutting scholarly platform "Reflecting and Shaping the Digital Age," which provides a forum for the investigation of the societal implications of research in data science and scientific computing, and the media upheavals associated with the digital age.[7] The IDH has a permanent seat at the Board of Directors and has already worked through the centre in establishing lecture series, cooperative research activities, and project proposals.

This continuum forms the narrow context around the study programmes. The wider context includes the specific media-heavy business sector in Cologne, DH as a topic at universities in Germany, as well as digitization more generally, and international DH. The University of Cologne has always been active in national and international cooperation and Eide was for many years engaged in the management of international DH organizations, an interest he shared with Elisabeth Burr, with whom he cooperated closely. Cologne also has an active role in infrastructure projects, including NFDIs[8]. These contexts form the basis for the job opportunities for the students. For the majority of them, the opportunities are outside academia. It also enables relevant institutions to hire young professionals, not only research software engineers, with a good understanding of the humanities and cultural disciplines. This forms the basis for several projects at IDH, CCeH, and DCH, but also at the faculty of arts in Cologne more generally, and even in partner institutions outside the university, such as the German

---

[7] See <https://cds.uni-koeln.de/en/>.

[8] Nationale Forschungsdateninfrastruktur (National Research Data Infrastructure), see <https://www.nfdi.de/en-gb>.

Archaeological Institute (DAI), an institution with a long history of cooperation with Cologne in general and HKI, and later IDH, in particular.

7 Conclusion

The use of computers in the humanities has been part of the research agenda at several universities since the 1970s, and in some cases even before. Already in the last century, study programmes and other forms for teaching were fairly common. What has often been the case, however, is an inclusion of DH mainly through third-party-funded and/or short-term centres and other units. The establishment of two full professorships already in the 1990s, and the ongoing support from the University of Cologne since then, has established the use of computers in research in the humanities as a topic all professors in Cologne are used to, for most of them ever since they took up their positions as professors. The same can be said about the teaching of DH - and not just within the university. The study programmes offered by IDH are well known outside the university, not only in Cologne, but also in other parts of Germany. This has led to a normalization of the study programmes which offer other universities some suggestions as to where they might go, and also, what they might not want to offer.

DH teaching in Cologne is integrated in the local context partly based on the specific nature of the job market in Cologne. It is also part of an international development, as we see it, for example, in the annual summer school "Deep Learning for Language Analysis" hosted by the IDH since 2018. This Summer school series does not have the long tradition of Elisabeth Burr's "Culture & Technology – European Summer University in Digital Humanities", which was established as early as 2009. However, in September 2021 the Cologne summer school took place for the 4th time.

The ARTEST project also exemplifies the international orientation of IDG. It is coordinated by the department (Philosophische Fakultät) and will support the development of MA study programmes in DH and digital cultural heritage at four Russian and three Mongolian universities. The cooperation between Russia and DH Cologne goes back more than 20 years, whereas for Mongolia it is completely new. Together with the DH Summer School in Lagos, Nigeria, which IDH co-organized twice, this is an indicator of DH in the 21st century as a truly global movement with significant variation as to how it is implemented in different countries and

even in different cities within a country. This works as a continuation of Elisabeth Burr's long involvement in multicultural and multilingual research and teaching. She inspired these activities in Cologne both directly, at a person-to-person level, and indirectly, through her long-term involvement with international organizations such as EADH and ADHO.

Bibliography

Mensching, Guido / Lalande, Jean-Yves /Hermes, Jürgen / Neuefeind, Claes (eds.) (2018): Sprache—Mensch—Maschine. Beiträge zu Sprache und Sprachwissenschaft, Computerlinguistik und Informationstechnologie. Köln: Universität zu Köln <https://kups.ub.uni-koeln.de/9849/7/Festschrift_Rolsthoven.pdf> [20.03.2023].

Rolshoven, Jürgen (2012): "Die Digitale Rätoromanische Chrestomathie", in: Ladinia 36: 119-151.

Rolshoven, Jürgen (1991): "GB und sprachliche Informationsverarbeitung mit LPS", in: Rolvshoven, Jürgen / Seelbach, Dieter (eds.): Romanistische Computerlinguistik: Theorien und Implementationen. Tübingen: Max Niemeyer Verlag 133-158.

Thaller, Manfred (2012): "Controversies around the digital humanities: An agenda", in: Historical Social Research 37, 3: 7–23.

Thaller, Manfred (2017): "Between the Chairs: An Interdisciplinary Career", in: Historical Social Research, Supplement 29: 7–109.

Wiehe, Thomas (2018): "Strings and Structures – eine Begegnung von Computerlinguistik und Bioinformatik", in: Mensching, Guido / Lalande, Jean-Yves / Hermes, Jürgen / Neuefeind, Claes (eds.) (2018): Sprache—Mensch—Maschine. Beiträge zu Sprache und Sprachissenschaft, Computerlinguistik und Informationstechnologie. Universität zu Köln 327-331.