Classification Research

Knowledge Organization Systems and Digital Humanities: 10th ISKO-France Colloquium, 2015

Michèle Hudon

EBSI, Université de Montréal, C.P. 6128, Succursale centre-ville, Montréal, QC, Canada, H3C 3J7, <michele.hudon@umontreal.ca>

On November 5 and 6, 2015, 69 registrants took part in the 10th biennial ISKO-France Colloquium (http://www.isko-france.asso.fr/), to discuss knowledge organization (KO) and its relation with digital humanities (DH). The event was organized by the Laboratoire interdisciplinaire des sciences de l'éducation et de la communication (LISEC), Université de Strasbourg (http://www.lisec-recherche.eu), and chaired by Emmanuelle Chevry Pebayle, Associate Professor and a member of the LISEC research team. Most of the fifty-six participants were French, with others hailing from Italy, Spain, Great Britain, Algeria, Nigeria, Canada and Brazil.

Twenty-seven papers, selected through a standard blind review process, were presented; nine were given in English. All presentations have been written up as full papers, and proceedings will be published in 2016. This report, prepared using pre-print versions, summarizes the twenty-four papers, which linked most directly to the theme of the meeting. To facilitate reading, the titles of papers presented in French have been translated into English; translated titles are identified by asterisks (*).

As a domain, DH is characterized by its fuzzy borders and its vast power of attraction of researchers and practitioners working in different disciplines. It was not surprising then to hear about a wide variety of projects, having little in common with many others, but each seen as being a part of or an application of DH. A few theoretical papers provided useful elements of definition of what the field of DH really is. Most papers, however, focused on specific disciplines, software applications, methodologies, etc.

In *Digital Humanities and Research Infrastructures: Their Impact on Knowledge Organization*, the first keynote speaker, Laurence Favier (Université Lille 3, France) emphasized the distinction between DH and the digital library as a general concept. She claimed, however, that the growing importance of the research infrastructure, defined as ecology of complex adaptive systems needed to facilitate collaborative work, data sharing and preservation, could eventually obliterate this distinction. Favier suggested that KO processes and tools must be adapted to the new infrastructure, taking into account not the evolution of science itself, but rather the transformation of the conditions of production of this science.

The second keynote speaker, Julianne Nyhan (University College London), in "Knowledge Organization Systems and Digital Humanities: Looking Back to Look Forward," took the participants back in time. Nyhan is interested in the evaluation and recognition of collaboration, a topic of much interest nowadays; the increasingly larger size of research teams means rising complications when comes the time to distinguish, assess and credit individual contributions. The speaker adopted a historical point of view to relate Roberto Busa's *Index Thomisticus*, collaborative work, and recognition of women's roles and contributions. Using this case study, Nyhan explored the broader issue of gender and proposed a definition of what seems to "count" as a worthwhile contribution in Busa's project and possibly in more current ones.

Mostly theoretical papers included "The Epistemological Construction of 'analyse documentaire' and Influence of Jean-Claude Gardin in Brazil." José A. Guimaraes and Johanna Smit (Sao Paulo State University, Brazil) described the role played by Gardin in the expansion and epistemology of subject access to information (SAI) defined as: "the analysis, description and representation of the content of the documents, as well as its inevitable interfaces with information theories and storage and retrieval systems." Gardin was interested in the search for methodologies and process of analysis, and he

https://doi.org/10.5771/0943-7444-2016-4-296 - am 13.01.2026, 10:28:19. https://www.inlibra.com/de/agb - Open Access -

was a pioneer in the use of technology in linguistics. It was thought-provoking to hear Guimaraes expose, to a mostly French audience, Gardin's influence on the Brazilian TEMMA Research Group established in the early 80s; TEMMA's researchers linked Gardin's theoretical bases to terminology, theories of language, logic, diplomatics, archival science and communication studies, something that was never done in the same way in France. This presentation led many in the audience to the realization that French pioneers were little known and would certainly deserve more attention on the national stage.

In *The Disciplinary and Topical Structuring of Digital Humanities*, Eglantine Schmitt (Université de Technologie de Compiègne, France) also adopted a historical approach, combined with an epistemological point of view, to discuss the relation between digital technologies and disciplines traditionally seen as part of the humanities. A significant contribution of her presentation was a demonstration that what the field of DH pretends to include, in terms of disciplines, does not correspond to what is found in the scientific literature, especially the literature published in French. A visual representation of the results of Schmitt's bibliometric analysis of the literature published in English showed numerous clusters, with an important node around publishing and archiving, and smaller nodes around geography, pedagogy, cognitive sciences, computer science and data mining. In the French literature, however, the representation was not as dense, the clusters fewer, the nodes smaller and focused on publishing and on digital libraries and repositories. Furthermore, in the French literature, the integration of social and humanities disciplines with computer technologies was not as explicit as could be expected. Schmitt suggested that, in English as in French, the field of DH may have been created as a way to legitimize a program around the growing ambitions of researchers in traditional disciplines, and to establish a so-called transdisciplinary domain by simply naming it. Without risking a definition, Schmitt nevertheless acknowledged that the field does exist, but that it represents multiple realities, some of which may even contradict each other. The concept evokes epistemological transformations, relating to the very nature of knowledge produced by innovative tools and methods, as well as sociological transformations in the ways of conducting research.

In their theoretical paper, Grant Campbell, Alexandre Fortier and Jacquelyn Burkell (all of Western University, Ontario, Canada) used J.K. Feibleman's theory of integrative levels to reveal the potential consequences on the life of an individual of his leaving multiple traces on social networks and elsewhere on the web. The title of this presentation, *The Illusion of Anonymity in the Crowd* set the stage for the demonstration of how refined methods of big data analysis have come to represent a growing threat to private life and individual autonomy. Feibleman's theory shows that data aggregation at a higher level does not render lower level data invisible or non-pertinent; without realizing it, the individual can still be identified, and his behaviour analyzed and exploited by commercial or political interests.

General content organization, an ongoing research area in DH from the very beginning, was addressed by Angel Castellanos, Ana Garcia-Serrano and Juan Cigarran (ESTI Informatica Madrid, Spain) in "Concept-Based Organization for Semi-Automatic Knowledge Inference in Digital Humanities: Modeling and Visualization." In their paper, the authors first described the problems engendered by probabilistic techniques that have become the standard for automated content organization in DH. These problems include the need to fix the number of topics to be detected and the difficulty for disciplinary specialists to visualize and interpret the results. The researchers then proposed a novel semantic approach based on Formal Concept Analysis (FCA) for the modelling and organization of contents; they believe that, since the approach relies on formal concepts close to concepts in the real world, they may be more easily inspected and evaluated by human experts and humanists than probabilistic models. An experimental study carried out on an Archivo General de Simancas (AGS) dataset proved the viability of FCA, providing valuable results for the historian experts, and facilitating inference of new knowledge.

Gabriela Previdello and Marilda Lopez Ginez de Lara's (University of Sao Paulo, Brazil) paper, "Foundational Ontologies and Knowledge Organization in Digital Arts," opened on a characterization of contemporary arts and particularly digital arts. What distinguishes contemporary art is the fact that the "contemporary condition" profoundly modifies structures and does not support analyses used with other art systems. Digital arts, born from communication technologies, are characterized by their ephemerality and manifestations through the networks. Digital works, meant to be deconstructed and reconstructed, are therefore resistant to the usual categorizations. It seems easy to understand why static KO structures, even those conceived by and for art domains, are not appropriate. Previdello and Ginez de Lara explored the possibility of using foundational ontologies, defined as representations of reality in axiomatic models described in first-order logic or descriptive logic, to organize and provide access to contemporary and digital arts. The researchers also analyzed existing dynamic organization models and introduced some examples of digital arts knowledge platforms, namely Highlike.org, Netzspannung.org and the Software studies project (http://lab.softwarestudies.com/); these initiatives all dismiss logical organization at a certain point, and rely on random or free immersion approaches to in-

https://doi.org/10.5771/0943-7444-2016-4-296 - am 13.01.2026, 10:28:19. https://www.inlibra.com/de/agb - Open Access -

formation. Despite the fact that their use was not apparent in these initiatives, the authors maintained their hypothesis that foundational ontologies, such as DOLCE (http://www.loa.istc.cnr.it/old/DOLCE.html), could help identify more flexible ways to account for the specificities of newer art forms.

Most presenters also focused on specific fields of knowledge and presented projects in various stages of progress. Corinne Manchio's paper, *A Study of Machiavelli's Language from a DH Perspective*, connected the methods of philology and computer applications. Manchio illustrated the possibilities offered by technology for more varied, flexible and efficient processing of traditional sources, a corpus of Machievelli's letters for example. An important contribution of this paper was the author's description of communication difficulties arising when disciplinary specialists, who do not share the same culture and language of specialty, work in collaboration. Manchio also showed that there is no magic in the application of digital technologies and that a considerable implication on the part of the researchers is needed to consolidate the epistemic bases of new forms of KO. Marianne Vergez-Couret and colleagues (Université de Toulouse, France), in *Digitization and Optical Character Recognition in Text Written in Regional Languages*, compared two OCR tools, namely Jochre (https://github.com/urieli/jochre) and Tesseract (https://github.com/tesseract-ocr), used to digitize documents written in Occitan and in Alsacian. One of the hypotheses being tested was that OCR tools developed for one regional language in France could be used effectively with another French regional language, especially one which has similar characteristics; results showed that this was indeed the case under certain conditions.

Lorna Balkan (UK Data Archive) described an experiment that used KEA (Keyword Extraction Algorithm) to index a corpus of question texts/variables in "Automatic indexing of a variable/question bank collection using KEA." She presented the results of a comparison of machine-generated versus human-generated indexes as well as observations based on qualitative evaluation and error analysis. This work confirmed yet again that automated indexing did not match the performance of human indexers, but could be useful as support to human indexing. Camille Rondot and Emeline Seignobos (CELSA Paris Sorbonne) showed how a straight digitization project was likely to modify existing KO and research practices in archival collections. In *Archives and Digitization … the Digitization of Press Releases Emanating from the Council of Ministers*, the authors explained that, once digitized and offered on a web-based platform, the newly created corpus could free itself from traditional archival organization practices, based on provenance, to allow for true subject access. At Archives nationales de France, the many steps leading to the implementation of a virtual collection of press releases covering the period 1945 to 1981 were the occasion of a dialogue between archival theory and practices and information and communication sciences. The project demonstrated the feasibility of providing various types of access in archival collections, thus facilitating research.

Assunta Caruso and Antonietta Folino (University of Calabre, Italy) walked the participants through a mapping project involving two thesauri in the domain of tourism. In "Mapping Tourism for Semantic Interoperability," the authors explained how they applied Part 2 of ISO25964 *Thesauri and Interoperability with Other Vocabularies* (2013) to map the Italian INMOTO Tourism Thesaurus to the international World Tourism Organization's *Thésaurus du tourisme et des loisirs* (http://www.idestur.org.br/download/20120219151302.pdf). The project's objectives were to measure the semantic coverage of two multilingual terminology sources in the same domain, and to qualify semantic correspondences. The methodology, based on the direct-linked model of inter-vocabulary equivalence, involved lexical comparison and conceptual correspondence identification. This paper was very interesting as it is one of the few integral applications of ISO25964-2 principles reported to this day in the literature.

In *Opening Data and GeoHumanities*, Gabriella Salzano aand Joumana Boustani (Université de Paris-Est, Marne-la-Vallée) illustrated the concept of transdisciplinarity through their description of the steps leading to the creation of web portals for the dissemination of public data relating to professional guidance and career counselling. The case study addressed two questions: 1)How and why would DH researchers be involved in the implementation of such a portal?; and 2) How could these portals assist a wide variety of DH researchers? Salzano and Boustani explored the GeoHumanities approach, formalized in 2013 by the Alliance of Digital Humanities Organizations, to answer their questions; the approach utilizes geo-temporal analysis as well as interactive mapping and data visualization tools. Michèle Hudon and Michelle Cumyn (respectively Université de Montréal and Université Laval, Canada) presented *Organization and Access to Legal Information: The Potential Usefulness of Faceted Analysis*. The paper described a five-year (2015-2020) interdisciplinary project conceived with the objective of improving access to the vast quantity of legal information available in digital form and widely circulated to various categories of users (general public, students, law students, legal practitioners, law teachers and researchers). Beyond the strictly functional, classification of legal materials is a domain that remains largely unexplored; the researchers plan to use facets and faceted

https://doi.org/10.5771/0943-7444-2016-4-298 - am 13.01.2026, 10:28:19. https://www.inlibra.com/de/agb - Open Access -

analysis to do so. Marina Caputo and Marion Lamé (University of Pérouse, Italy), used the case of antiquities to demonstrate, in *DH, Knowledge Organization and Disciplinary Convergence*, the ability of digital processing to group various types of knowledge relating to a single object, whatever the sources and original locations. The authors suggested that phenomena offered a relevant basis to structure KOS, claiming that in such systems it would be easier for domain specialists to define their contribution when describing a phenomenon (or object) from different angles.

The communication and sharing of information deemed "sensitive" for various reasons is an issue that, so far, has rarely been discussed in the DH literature. Aminata Kane (Université Lille 3, France), in *Externalisation of "Sensitive Heritage": Between Reticence and Right of Access*, evoked the potential impacts of broadening access to sources related to Franco-African colonialism. Kane, on the basis of a survey, suggested that DH practices should take into account the sacred character of certain cultural sources and their significance for specific categories of users. She also implied that policies and guidelines relating to the communication of archives should consider this issue.

Participants heard about a specific institution, the École nationale supérieure vétérinaire d'Alger (National School of Veterinary Sciences, Algiers), from Radia Bernaoui (ENSV, Alger). In her paper, Bernaoui presented the results of a survey of users' needs and expectations, distributed in the preliminary planning stages of a digital library at the school. Not surprisingly, the users expressed their interest, but it was not clear that such a library would eventually be made available.

Collaborative work is a defining characteristic of DH. Most presenters evoked the potential difficulties of interdisciplinary collaboration, and several of them addressed this issue specifically, looking at practices and at facilitating software. The collaborative project of Nicolas Chauveau and his colleagues (Université Paul-Valéry, Montpellier, France) focuses on the processing of ancient Egyptian. In the paper presented by Chauveau, *VEgA, a Digital Platform to Increase Knowledge and Communication Around Ancient Egyptian*, the researchers explained how they had developed and were using an online dictionary, VÉgA, in what they considered to be a perfect example of DH application. Indeed, the tool is accessible to a wide variety of users, and it stimulates collaboration between egyptologists, who are not aware of all that data processing can do, and digital design specialists who do not know what is most relevant for egyptologists. The main interest of the paper lies in its description of actual ways of supporting collaborative work through design. Amel Fraisse and Patrick Paroubek (respectively Université Lille 3 and LIMSI-CNRS, France), proposed a new management paradigm, collaborative and incremental, applicable to KO. To develop their proposal, they used two types of experiment: one based on crowdsourcing, the other on a game proposing a specific goal, in the domain of climate changes. The results showed that the collaborative methods led to results, in terms of content organization, equivalent in quality to results obtained by KO experts. Mohammed Ourabah Soualah and Mohammed Hassoun (ENSSIB and Université Lumière, Lyon, France) presented a project of *Dynamic Cataloguing of Ancient Arabic Manuscripts* supported by automated analysis of annotations provided by domain specialists, cataloguers and various categories of technicians. The paper highlighted the fact that, with certain types of documents, bibliographic description may never be complete and should be a dynamic process. Contrary to traditional cataloguing, dynamic cataloguing implies that the description remains open to changes and transformations, evolving with new transcriptions and with new information uncovered by the scientific community. The technological solution is based on natural language processing techniques, applied to various types of metadata provided by the variety of users with access to the collaborative platform.

Collaborative work necessarily involves communication of information through networks, with potential problems linked to data security. Stephen Ojo Aderibigbe (University Ile-Ife, Nigeria) was the only speaker who clearly addressed this issue in "A Friend-to-Friend Approach for Secured Knowledge Sharing." In his paper, Aderibigbe reminded the audience of the importance of social networks for data and knowledge exchange even among domain specialists, who may or may not be aware of the various security and privacy problems to which they expose themselves and their work. The author and his colleagues have used F2F collaborative theoretics and the distributed harsh table and predecessor replication techniques based on the underlying philosophy of social trust to construct a "trust-aware model." Aderibigbe recognized that the objective is a very ambitious one: to propose a model allowing for the computation of trust.

Another issue of importance that is not, as yet, widely represented in the literature, is that of evaluation. Bernard Ijesunor Akhigbe and his colleagues (University Ile-Ife, Nigeria), believe that DH systems, as any other systems, must be evaluated for quality, pertinence and usability. In their paper, "Towards a Conceptual Praxis-Oriented Evaluative Framework: A Web Analytics Approach," they introduced a user-centric evaluation framework based on web analytics methodology and information processing theories. The framework, described with much technical details, was presented as a potential base-

https://doi.org/10.5771/0943-7444-2016-4-299 - am 13.01.2026, 10:28:19. https://www.inlibra.com/de/agb - Open Access -

line contribution to the evaluation of the plethora of digital tools that are available to DH.

Even if the presentation did not relate directly to these problematic issues, we can establish a link between the paper authored by Cécile Delay-Artous (Laboratoire dynamiques économiques et sociales des transports, France) and the preceding ones. Delay-Artous talked about the difficulty of identifying and locating data due to the multiplicity and diversity of DH projects in France and in Europe. The open access and open data movements add to the complexity of the situation. This leads to the difficulty of controlling the authenticity and preserving the integrity of data, as well as to security problems. In *Where are the Data?*, Delay-Artous presented her project of mapping DH initiatives and actors based in France. The laborious exercise allowed her to highlight the difficulty of negotiating the labyrinth and to circumscribe the actual role of each individual and the limits of their projects. Interdisciplinary projects, of course, represent an even greater challenge!

Not directly related to the theme but of special interest to the KO community, a paper by Edouard Laroche-Joubert (Université Lille 3, France) focused on our old friend the library catalogue. Laroche-Joubert reminded the audience of the important role still played by the catalogue as interface between the information seeker and information in the digital environment. Laroche-Joubert is trying to describe how, through library catalogues, users can conflate virtual and physical document spaces. The main contribution of the author of *Semio-Praxiologic Contribution to Spatial Knowledge Organization Through Online Communication in Libraries* is a detailed analysis, from a semiotic-graphical perspective, of several library catalogues, with a view to illustrating some of their deficiencies. Laroche-Joubert concluded that, with the possible exception of the VisualCatalog (for a description see http://bbf.enssib.fr/consulter/bbf-2005-04-0005-001), the graphical power offered by current software and systems to represent the spatial organization of knowledge in libraries is most definitely underused.

The variety of papers, projects and perspectives presented during the 2015 ISKO-France Colloquium stayed well within the limits of the most general definition of the field of DH: collaborative research and teaching activities that combine computing and information technologies with scholarly practices in the field of humanities. Elements common to most papers can be highlighted: a focus on disciplines traditionally considered as forming part of the humanities (such as history, philology, antiquities, law), a demonstration of the predominance of collaborative work, detailed descriptions of how technology (such as automated language processing) can be applied with benefit, and numerous references to visualization.

Within DH, the production of knowledge is profoundly transformed and the contours of traditional disciplines are regularly redesigned. Researchers appear to have focused until now on the efficiency of practices and tools, without enough consideration given to their potential impact on knowledge management and organization. The KO community has become acutely aware of the unavoidability of significant consequences on its own practices, systems and effectiveness. The cases and examples provided during the Colloquium stress the necessity to design adaptable and dynamic processes and systems; only then will they have a chance to become valuable components of the new knowledge creation, management, distribution and preservation infrastructure described by Professor Favier in her keynote paper.

https://doi.org/10.5771/0943-7444-2016-4-296 - am 13.01.2026, 10:28:19. https://www.inlibra.com/de/agb - Open Access -