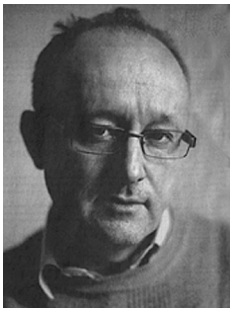
Web 2.0 and the Semantic Web in Research from a Historical Perspective: The Designs of Paul Otlet (1868-1944) for Telecommunication and Machine Readable Documentation to Organize Research and Society

Charles van den Heuvel

Royal Netherlands Academy of Arts and Sciences (KNAW),
Virtual Knowledge Studio for Humanities and Social Sciences, Cruquiusweg 31,
1019 AT Amsterdam <charles.vandenheuvel@vks.knaw.nl>



Charles van den Heuvel is senior researcher for the Virtual Knowledge Studio for the Humanities and Social Sciences of the Royal Netherlands Academy of Arts and Sciences, where he is involved in various projects on visualization in history, annotation of digital maps and images, web archiving for research, Web 2.0 in digital humanities and in cultural heritage. He studied Art History and Archeology at Groningen University, The Netherlands (PhD 1991). He publishes regularly on history of architecture, fortification and town planning, history of cartography, history of science and history of information science, in particular on the visualizations of Paul Otlet.

Van den Heuvel, Charles. **Web 2.0 and the Semantic Web in Research from a Historical Perspective: The Designs of Paul Otlet (1868-1944) for Telecommunication and Machine Readable Documentation to Organize Research and Society.** *Knowledge Organization, 36(4)*, 214-226. 29 references.

ABSTRACT: Tim Berners-Lee described in *Weaving the Web* his future vision of the World Wide Web in two parts. In the first one, nowadays called Web 2.0, people collaborate and enrich data together in a shared information space. In the second part, exchanges extend to computers, resulting in a "Semantic Web" (Berners-Lee 2000a, 157). Most historical studies of World Wide Web begin with the American roots of the Internet in ARPANET or follow a historiographical line of post war information revolutionaries, from Vannevar Bush to Tim Berners-Lee. This paper follows an alternative line. At the end of the nineteenth and in the first decades of the twentieth century various European scholars, like Patrick Geddes, Paul Otlet, Otto Neurath, and Wilhelm Ostwald explored the organisation, enrichment and dissemination of knowledge on a global level to come to a peaceful, universal society. We focus on Paul Otlet (1868-1944) who developed a knowledge infrastructure to update information mechanically and manually in collaboratories of scholars. First the *Understanding Infrastructure* (2007) report, that Paul N. Edwards et al. wrote on behalf of NSF, will be used to position Otlet's knowledge organization in their sketched development from information systems to information internetworks or webs. Secondly, the relevance of Otlet's knowledge infrastructure will be assessed for Web 2.0 and Semantic Web applications for research. The hypothesis will be put forward that the instruments and protocols envisioned by Otlet to enhance collaborative knowledge production, can still be relevant for current conceptualizations of "scientific authority" in data sharing and annotation in Web 2.0 applications and the modeling of the Semantic Web.

ACKNOWLEDGEMENTS: This paper builds partly upon the publication Heuvel (2008) and the chapter by the author: "Web Archiving in Research and Historical Global Collaboratories," in N. Brügger ed. *Web Histories*, Peter Lang Publishing (forthcoming). I would like to my colleagues at the Virtual Knowledge Studio Sally Wyatt and Nick Jankowski for reading earlier versions of the manuscript. Finally, I am grateful to Boyd Rayward for his useful comments and for bringing figure 7 to my attention. I am indebted to the Mundaneum in Mons (Belgium) for the use of Otlet's images that illustrate this text.

1.0 Cyberinfrastructures for research in the humanities and social sciences

Tim Berners-Lee describes in *Weaving the Web* his future vision of the World Wide Web as a dream in two parts. In the first one, the Web becomes a more powerful means for collaboration between people. He imagines information space as something to which everyone has immediate access, not just to browse, but also to create. In the second part of the dream, collaborations extend to computers. Machines become capable of analyzing all the data on the web—the content, links and transactions between people and computers"—resulting in a "Semantic Web" (Berners-Lee 2000a, 157).

The fulfillment of the latter part of Berners-Lee's dream still seems far away. However, the first part of the dream is becoming reality. Given the success of Wikipedia, YouTube, Flickr, Facebook and other Web 2.0 applications, it is fair to say that people are augmenting the World Wide Web with data and information and that they use it more and more as an instrument to collaborate. However its potential for research, especially for research in the humanities and social sciences, is not fully exploited yet.

The report of the American Council of Learned Societies' Commission on Cyber infrastructure for Humanities and Social Sciences: *Our Cultural Commonwealth* mentioned another argument: "the conservative culture of scholarship" in the humanities and social sciences compared to the natural sciences (ACLS 2006, 21) as one of the reasons to explain why researchers in these knowledge domains might be hesitant in using the web and other digital resources for their research. Paul Wouters in a critique on this report countered this argument by pointing to the cyber infrastructures themselves instead of to the researchers. In his view their reluctance is the result of a mismatch between these scholars and the provided tools and services. Tools and services will only be picked up if they serve researchers in their daily work (Wouters 2007) and can be trusted. Researchers in the humanities and social sciences for the greater part use small, heterogeneous datasets that are often highly ambiguous in meaning. Especially humanities and social sciences scholars are often concerned with how meaning is created, communicated, manipulated and perceived. Therefore the cyberinfrastructures around such datasets require both sufficient information to generalize findings and tools to put these into context, for example by using annotation. This requires an infrastructure that allows both for critical mass and standardization and for heterogeneity and contextualization.

2.0 Historical lessons for new scientific cyberinfrastructures

In September 2006, a three-day NSF-funded workshop took place at the University of Michigan with the title: "History and Theory of Infrastructure. Lessons for New Scientific Cyberinfrastructures," in which historians and sociologists of infrastructures, information scientists and domain experts tried to understand the dynamics and tensions within existing infrastructures and to develop a research agenda to support the NSF in their mission to create a robust scientific cyberinfrastructure. The workshop resulted in 2007 in a report: *Understanding Infrastructure: Dynamics, Tensions and Design* that challenges the notion that infrastructures can just be designed mechanically. Instead it describes infrastructural development in various phases from an historical perspective from Large Technological Systems (LTS) towards "internetworks" or webs. The authors claim that historical infrastructures "become ubiquitous, accessible, reliable and transparent as they mature" and allow various (dissimilar) systems to be linked into networks, and networks into webs or internetworks (Edwards et al. 2007, i-ii and 11-13). The Internet and Web are seen as genuine infrastructures resulting from these historical processes of transformation of systems. The report gives an alternative explanation of the development of the Internet and the WWW, different to the beaten historiographical track following protagonists, considered to be relevant for its "design," running from Vannevar Bush to Tim Burners-Lee. Moreover, it gives examples of: "analog information internetworks" (a term coined by Downey in 2001) since the nineteenth century that preceded virtual infrastructures.

First, the *Understanding Infrastructure* report will be used as a theoretical framework to analyze and situate Otlet's designs for telecommunication and machine readable documentation within the described development from systems to analog information internetworks or webs. Secondly, the historical model used in the report to describe this development towards the Internet and World Wide Web as "a key transition from homogenous, centrally controlled to heterogeneous, widely distributed network in which central control may be partially or wholly replaced by

coordination" (Edwards et al. 2007, 11-12) will be questioned. Despite claims of developers, web applications such as Web 2.0 and the Semantic Web, the organization of the underlying cyberinfrastructures Internet and World Wide Web is becoming less transparent and more controlled in an implicit way. The hypothesis will be put forward that Web 2.0 and Semantic Web applications for research might benefit from Otlet's designs to enhance scientific authority via transparent typologies of (the provenance) of annotations and protocols to regulate data enrichment.

3.0 Otlet's designs for analog information internetworks

The authors of the report *Understanding Infrastructure* claimed that: "Robust cyberinfrastructure will develop only when social, organizational and cultural issues are resolved in tandem with the creation of technological services" (Edwards et al. 2007, i). When Paul Otlet set out to find new ways to organize and disseminate knowledge on a global level and to involve scholars in the creation of a more civilized, universal society, he was aware that the creation of his infrastructure was not just a technical matter. He developed a set of strategies (financial, organizational, technical, epistemic and the creation of political will) to make his knowledge infrastructure as robust and sustainable as possible.

By the end of the nineteenth century, Otlet took various actions to embody the knowledge of the world. First there was the creation of a bibliographical database, the Universal Bibliographical Repertory that soon was augmented by a Universal Iconographic Repertory and an Encyclopedic Repertory of Dossiers (Rayward 1975). To order these databases, Otlet developed the Universal Decimal Classification (UDC) on the basis of Melvil Dewey's Decimal Classification System. It was typical for Otlet to organize and materialize such activities both in institutions and buildings. In 1895, the year of the start of the bibliographical repertory, Otlet founded the International Institute of Bibliography (IBB). Fifteen years later, Otlet, together with the Belgian Senator and Nobel Prize Winner for Peace Henri La Fontaine (1854-1943), conceived the project of the "Palais Mondial" that would bring all their initiatives for knowledge organization on a global level together. Later Otlet dubbed the Palais Mondial, "the Mundaneum" which had a turbulent history of closures and re-openings until it more or less died with the outbreak of World War II and Otlet's death in 1944 (Rayward 1975). In the last decade of the twentieth century it was reopened in Mons (Belgium), where its archives are housed today.

On the one hand the Mundaneum had been a project for a real building that would combine a World Library, World Museum, World Archive, World University and a World Headquarters for International Organizations. On the other hand the word Mundaneum was used as an architectural metaphor of knowledge organization and dissemination on a global level. In *Monde: Essai d'universalisme* of 1935, Otlet wrote: The Mundaneum is 'an idea, an institution, a method, a material body of work, a building and a network' (Otlet 1935, 448–52; Rayward 2003).

The notion of network is important. The Mundaneum was not just conceived as a separate building, but as a conglomeration of buildings. These buildings with their international collections and international organizations to be globally effective had to be linked to create 'le reseau' (the network). This network would link the world's citizens in a hierarchical structure from a personal office, via many different forms of Mundaneum (from small to large) to the World City, which would be an architectural reality (Figure 1).

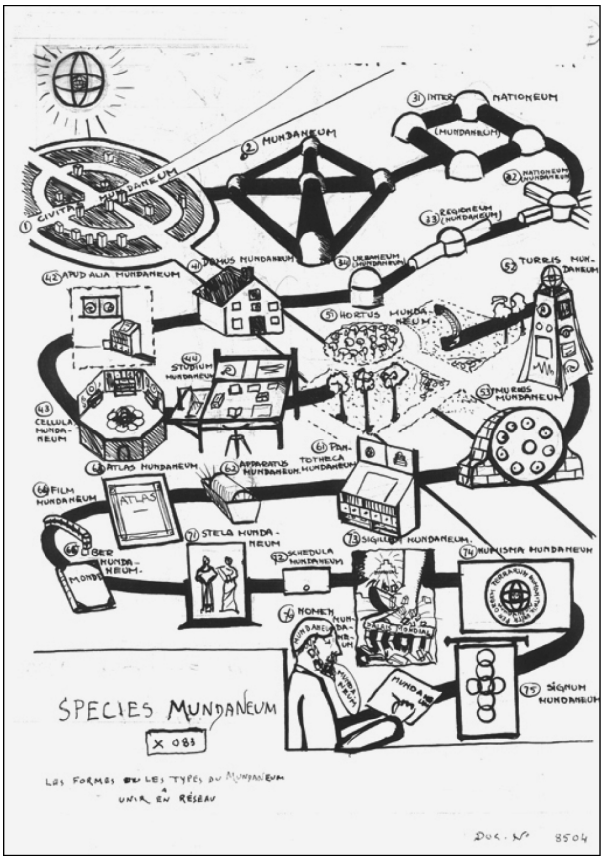


Figure 1. *Species of Mundaneum in Network EUM- 14- 121 (O nr. 8504) – Mons Mundaneum* ©

However, the Mundaneum was more than a conglomeration of buildings. It was primarily considered to be an infrastructure for a networked knowledge-based global society, consisting of material and virtual components. This material and virtual Mundaneum was the driving force for reorganizing the documentation of all knowledge in a planned manner, integrating technical and epistemic strategies.

How fundamental this re-organization would be, becomes clear from Otlet's description of the ultimate problem of documentation (Otlet 1935, 390- 91, transl. Rayward 1990, 1):

> Man would no longer need documentation if he were to become an omniscient being like God himself. A less ultimate degree would create instrumentation acting across distance which would combine at the same time radio, x-rays, cinema and microscopic photography. All the things of the universe and all those of man would be registered from afar as they were produced. Thus the moving image of the world would be established—its memory, its true duplicate. From afar anyone would be able to read the passage, expanded or limited to the desired subject that could be projected on his individual screen. Thus, in his armchair, anyone would be able to contemplate the whole of creation or particular parts of it.

This integration was visualized in its ultimate form in one of the last sketches that Otlet made of the Mundaneum in the year 1943 (Figure 2). The Mundaneum is depicted as a transmitter of knowledge by sound (radio-telephone) and by image (radio-television)—Otlet uses the term Thinking Machine—that allows people all over the world to participate actively.

However, the thinking of what the inclusion of all these various media implied on a technical and epistemic level goes much further back. As early as the end of the nineteenth century, Otlet started to think about a fundamental change of bibliography by involving researchers all over the world to describe and purify information. Otlet envisioned different technical means and media for transferring knowledge leading to new forms of documentation.

4.0 New forms of documentation and mechanical operations

Fundamental for an understanding of Otlet's knowledge architecture is the notion that the book is nothing more than a container of ideas that might be conveyed in a more efficient way. This efficiency was based on what Otlet called the Monographic Principle. It means that texts, but also other forms of information such as formulas, charts, images, schemes etc. should be dissected into their basic elements and recorded on standardized cards or sheets of paper. These chunks of information could then be reassembled over and over again in new combinations of publication formats, comparable (not similar) to the use of hypertext on the World Wide Web (Rayward 1994). Otlet was thinking of mechanical ways to dis-

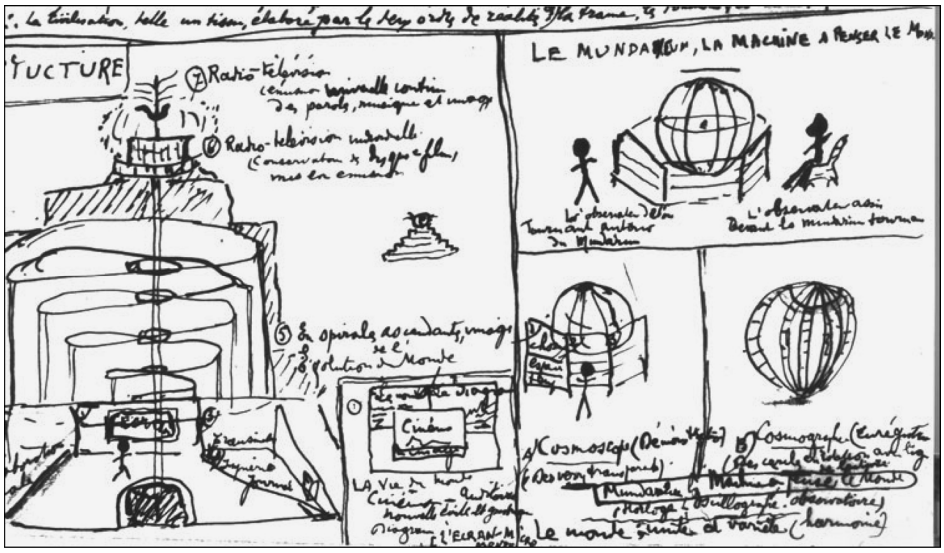


Figure 2. *Otlet—Mundaneum transmitter of knowledge (1943)*
EUM – 14-120, nr 112 – Mons Mundaneum ©

sect and reassemble information dynamically in standardized formats, comparable to the way the browser builds up a web page after a search.

Although Otlet could never have foreseen the World Wide Web, he was convinced that the latest media of his time would transform radically, probably even replace, the book in its role of disseminating knowledge. Otlet's ideas about substitutes for the book were worked out further in the lecture *La function et les transformations du Livre* on 14 November 1908 in the Maison du Livre, in Brussels. Otlet stated that the most important transformations of the book would not take place in its conventional form but in the appearance of substitutes for it (Otlet 1909, 28). The content could easily be expressed by other means than the book. Otlet acknowledged that the telephone was an important means to transmit sound, and the gramophone an important means for storing and reproducing sound, including words, but the technological development that he expected to effect the most radical transformation of the book was wireless. Wireless (radio) had the potential to become "a universal network that would permit the dissemination of knowledge without limitation" (Otlet 1909, 29; Heuvel 2008, 140).

Otlet realized that these alternative book forms and substitutes for the book required new and more efficient ways of organizing knowledge (Otlet 1934, 216-46). This implied the integration of the latest technology for processing and publishing information with new concepts of documentation.The idea of the world-wide dissemination of knowledge in the form of sound and image by wireless that Otlet had expressed in 1908 would return in his discussion of the mechanical, collective brain, 'Le Cerveau mécanique,' that he envisioned in his *Traité de documentation* (1934), and which can be compared to the modern computer. He proposed seven functions for such a device, which would consist of a collection of machines that simultaneously or sequentially could perform the following operations: (1) transform sound into writing; (2) reproduce this writing into as many copies as was needed; (3) create documents in such a way that each item of information had its own identity as part of a collection and could be retrieved as necessary; (4) assign classification numbers to each item of information, with perforations of the documents corresponding to these index numbers; (5) automatically classify and file documents; (6) automatically retrieve documents for consultation and present them either direct to the enquirer or via a machine enabling written additions to be made to them; and (7) mechanically manipulate at will all the listed items of information in order to obtain new combinations of facts, relationships of ideas and new operations to be carried out with the help of numbers (Otlet 1934, 391; Rayward 1990, 1). It is important to note that Otlet envisioned this mechanical brain, as a collective brain of machines and people working together. As such it not only anticipated the computer, but even certain characteristics of the Semantic Web and Web 2.0 can be recognized.

5.0 Towards an analog Semantic Web

The World Wide Web Consortium defines the Semantic Web as a "Web of data" and its main goal as follows (W3C, 2001):

> The vision of the Semantic Web is to extend principles of the Web from documents to data. Data should be accessed using the general Web architecture using, e.g., URI-s; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data.

It is possible to recognize in this definition some interesting parallels with the analog infrastructure of Paul Otlet. Agreed, Paul Otlet speaks of documents, not data. His most comprehensive publication is even called the *Traité de Documentation* (1934). However, Otlet does not see the document as a singular object, even when he says that every object can be in principle be a document (Buckland 1991).

Due to the aforementioned Monographic Principle, documents could be dissected into different parts according to their basic elements (which could be described as data) and recombined again with parts of other documents (Figures 3 and 4). In the definition of the W3C this requires a common framework to share, reuse and to produce data to be processed automatically as well as manually. In Otlet's description of the mechanical brain, machines collectively could classify and file documents that could be manipulated mechanically in order to obtain new combinations of facts and ideas. Moreover, the user in Otlet's view was also able to add information. Whereas Otlet's knowledge construction is

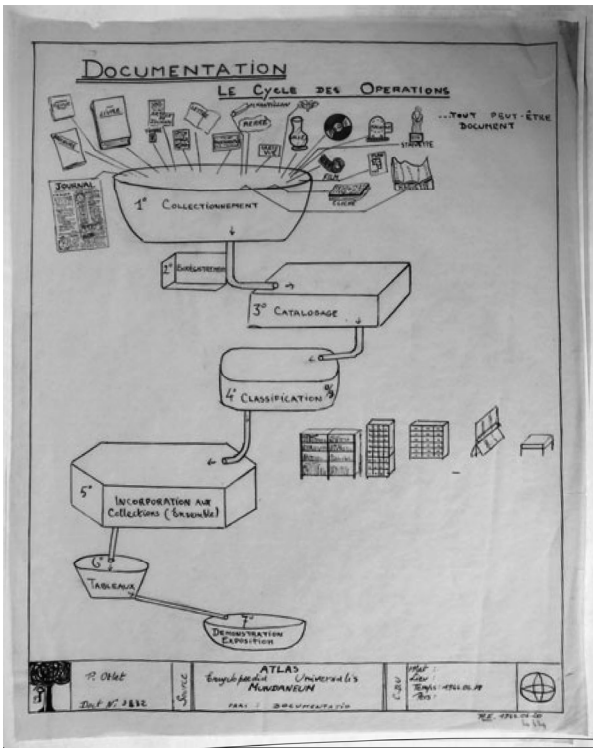


Figure 3. *Otlet – Documentation – Cycle of Operations*
Everything can be a document
EUM 3- 414 (O. number 3832) – Mons Mundaneum ©

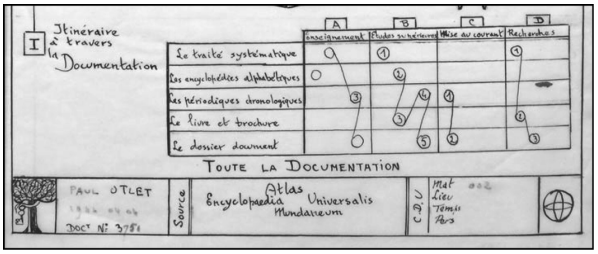


Figure 4. *Otlet – Itinerary of links through documentation*
EUM 3- 52 (O. number 3751) – Mons Mundaneum ©

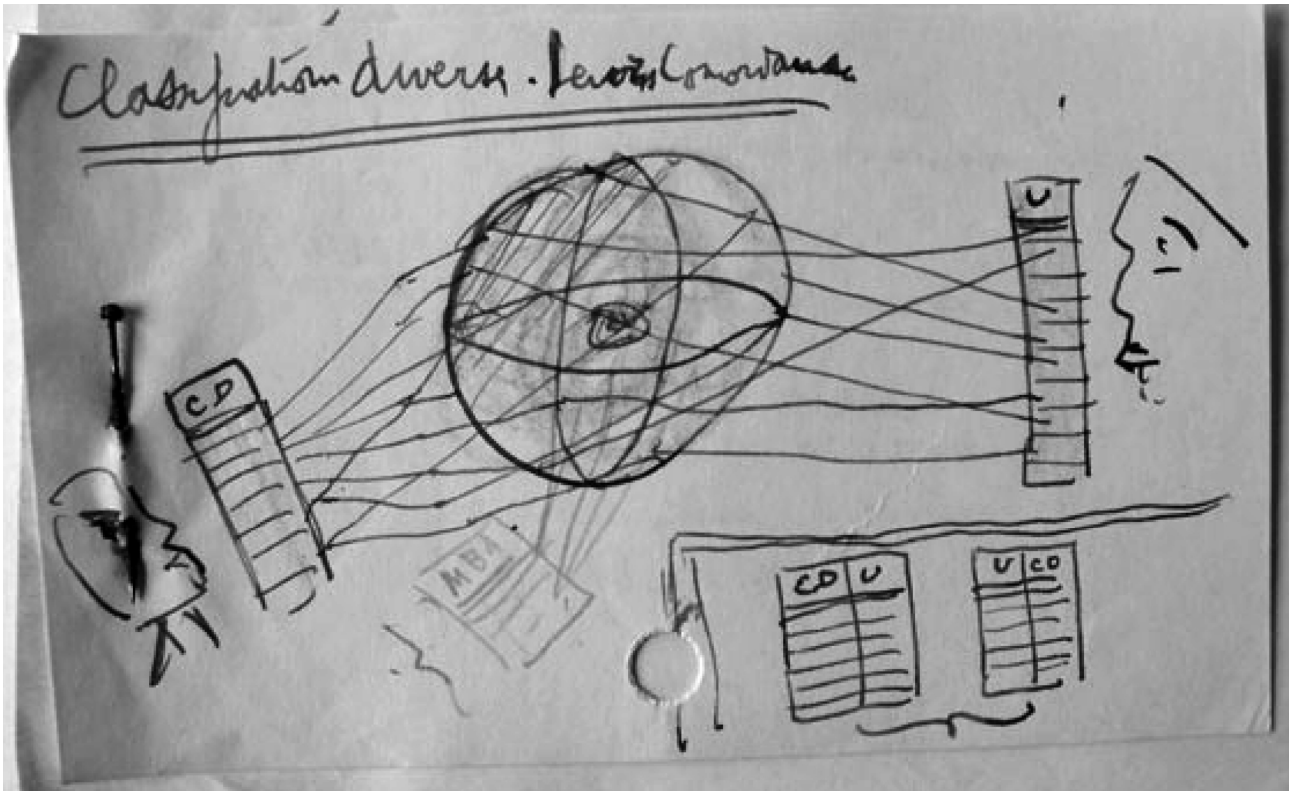


Figure 5. *Otlet linking various classification systems*
EUM- 1-4-164 – Mons Mundaneum ©

built up by decimal classification numbers, the Semantic Web uses the Resource Description Framework (RDF) to present and exchange data in a standardized way. Similar to the RDF data model, the Universal Decimal Classification makes statements about the resources. The UDC does not just order subjects or topics in classes by numeric codes, it also allows for linking to additional facets, such as place, language, physical characteristics via its auxiliary tables of connector signs (Rayward 1994). This resembles the linkage in so-called RDF triples. By indicating and visualizing relationships between classification number (and parts thereof) Otlet defines two ends of the link expressing a semantic relationship; while with the help of auxiliary tables the character of the relationship itself, such as format, can be described. This linkage system relating parts by numeric codes and connectors, such as "+, / and :," was foreseen to update the UDC and metadata descriptions further. This could be done automatically, connecting various classification systems (Figure 5) but also and manually by collaborative working groups.

This can be compared to the way the Linking Open Data Project group is creating RDF Data on the Web. Otlet had set up a similar collaboratory around his "collective, mechanical brain" that was approved during the World Document Conference in 1937 (Rayward 1983). In the Universal Network of Documentation public and private documentation centers and intellectuals would work together.

6.0 Otlet's research collaboratory and multimedia: Documentation and telecommunication

For Otlet, technological and epistemic strategies to process information, such as the Monographic Principle, did not just aim at improving efficiency but also at enhancing the quality of knowledge. Scholars played a crucial role in this and should in Otlet's view collaborate in search of "objective truth." This requires, according to him the organization of the research by coordinating methods and standardizing instruments. The book plays a role in the codification of knowledge, but Otlet uses it in a broader sense as "all sorts of registration of thought," including personal notes for research. He suggests replacing the term book with "document intellectuel" (Otlet 1913, 383). Moreover, the intellectual document is not limited to text, but also includes registrations of sound and images on gramophone disks, photograph and film for scientific communication. To bring all these views and formats of intellectual documents together everyone needs to work together on the creation of *Le Livre universal de la Science* (The Universal Book of Science), "an unlimited work, always up-to-date, constantly growing, concentrating, absorbing, synthesizing, systematizing every intellectual product from the moment it is born"(Otlet 1913, 385). Otlet kept on speculating about immaterial ways that knowledge might be produced collectively and distributed globally for the rest of his life.

In his *Traité de documentation* Otlet suggested the use of the telephone, telephotography and radiotelephotography for collective use in networks (Otlet 1934, 236–7). Two related figures with the title Documentation and Telecommunication of around 1937 intended for his unpublished *Encyclopedia Universalis Mundaneum* show an early variant of multimedia in which telephone, radio, gramophone, film and television are combined and transmitted as part of courses and teleconferencing. While the first image focused on the telecommunication of knowledge in ephemeral, non-book formats; the second one visualized the linkage of users to multimedia repositories in the "Universal Network of Documentation" (Figure 6)

In this network, documentation (collections and catalogues) is composed in a universal format in the Mundaneum and connected with other bibliographic repertories. Then machines assist in complementary operations of analysis and synthesis of the encyclopedic whole and extract desired elements mechanically. Finally transmitters send the desired knowledge elements through the universal network, where they are received by stations. However, Otlet envisioned this network of universal documentation not just as enabling to use, but also to contribute collectively valuable information to enhance the knowledge of the world. In order to enable data-enrichment from smaller centers of expertise or from home, Otlet designed the "Mondothèque" (Figure 7). It is a work station with spaces for a small library containing reference works and essential books, for the documentary encyclopedia, for the large volumes of the Atlas Mundaneum, for small (museum) objects and for drawers with bibliographical cards and microfilms all ordered according to the catalogue rules of the UDC. On its side shelves sit instruments, such a radio, telephone, television and film equipment to enable other forms of documentation . Sketches on the top of the image show connections of the multimedia device with the Mundaneum (compare Figure 1). Moreover, Otlet ex-

plains explicitly that the drawing makes part of the universal network and that it is related to visualization of documentation and telecommunication with the same document number (compare Figure 6) . Texts on the panels on top of the Mundotheca, remind its user of his bigger mission to contribute to global knowledge.

7.0 Personal classifications and Web 2.0

For Otlet the process of documentation did not only involve the creation of a knowledge system; it was a social system aimed at creating a better society: "One can imagine a social state that makes progress in its whole by an instrumentation based on very high lev-

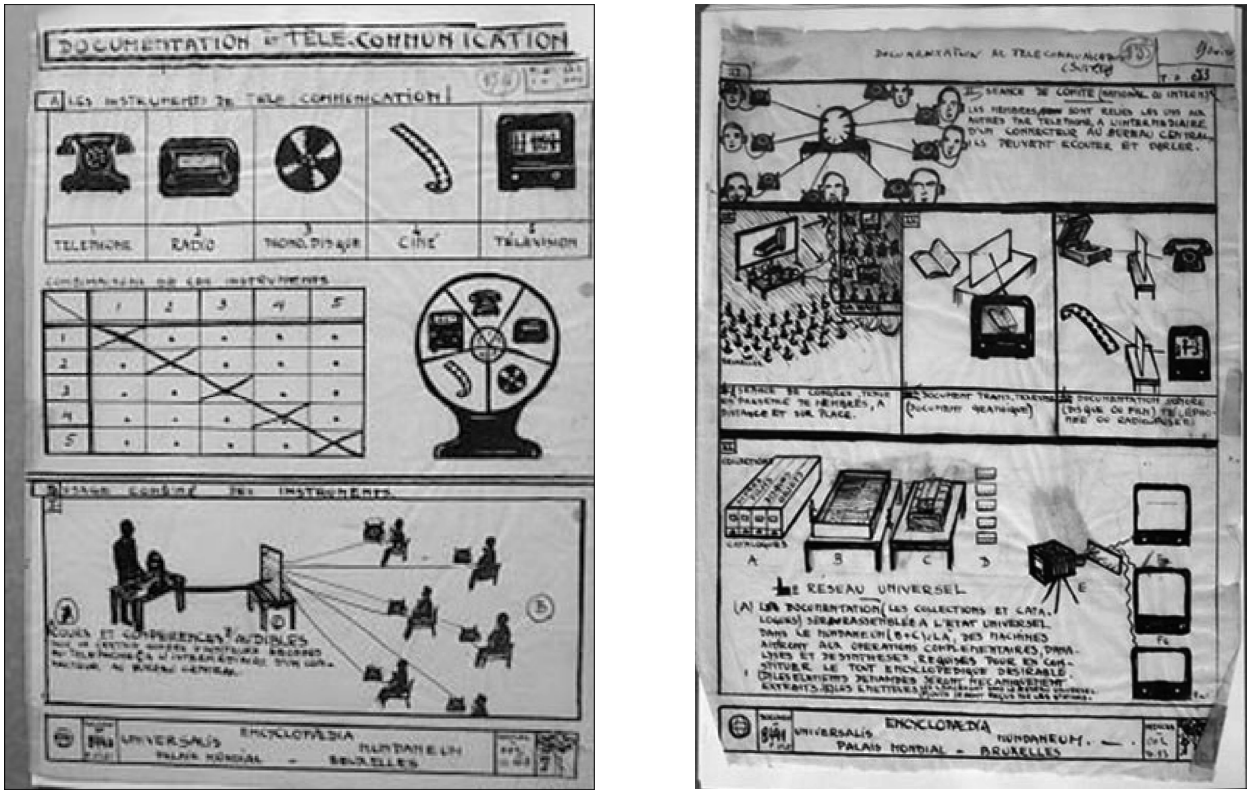


Figure 6. *Documentation and telecommunication 1 and 2*
EUM 3 –14- 132 and 133 (O. nrs 8440 and 8841) Mons Mundaneum ©

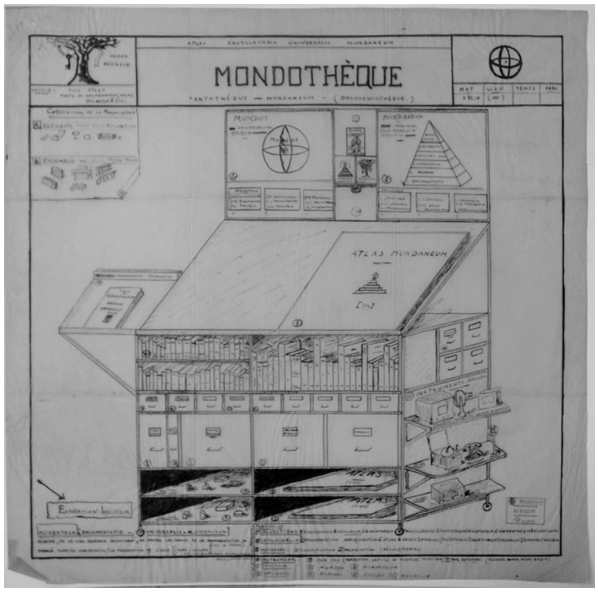


Figure 7. *Mundotheca: A private multimedia work station connected to the Mundaneum*
Atlas EUM (part of old number 8841)
Mons Mundaneum ©

els of abstraction that would be made available to everyone" (Otlet 1935, 238-39). Otlet claimed that this "machine intellectuelle" should not just serve the work of collaborating scholars, but in principal everyone (Otlet 1935, 238): "Like the technical machine allows not qualified workers to make perfect products, the intellectual machine does not require a specific education of the one that uses it." Otlet recognized the value of extracting personal notes from documents: "Preserved, classified, revised, continuously enriched with other notes derived from other sources, they could become a real book: a particular book for each person of which one could say (Otlet 1934, 319): 'My Book', 'My Encyclopedia' […] an artificial memory of everything one desires to recall." Otlet also believed in the value of preserving these annotations for collective memory. By classifying and storing the notes together with bibliographical descriptions of documents, "One could avoid new transcriptions often subject to errors, keep up with facts and ideas annotated at various moments. The confirmations by others that may also express different aspects of the same thing" (Otlet 1934, 319). Individuals could revise documents in the form of annotations, ranging from additions to various points of view (Figure 8).

The process seems at first sight similar to Wikipedia in which the involvement of more people adding and editing certain lemmas in the end to the improvement of those lemmas in particular and the digital encyclopedia in general. Alex Wright even states that Otlet's vision allows marrying top-down classification systems such as the UDC with socially-constructed information spaces, such as MySpace.com, Flickr.com and del.icio.us with their own folksonomies and tags (Wright 2007, 192). In a technical sense, Wright is correct. Otlet's vision indeed allows linking classification systems such as the UDC with socially-constructed information spaces. However, it would be wrong to read Otlet's Universal Network of Documentation simply as a wiki and personal classifications as folksonomies.

Otlet's concept of personal knowledge organization is strongly related to, is actually a microcosm of his Universal Decimal Classification system. For Otlet the producer of the knowledge is foremost an outsider of the system whose contribution would only be recognized after a long process of editing by, what

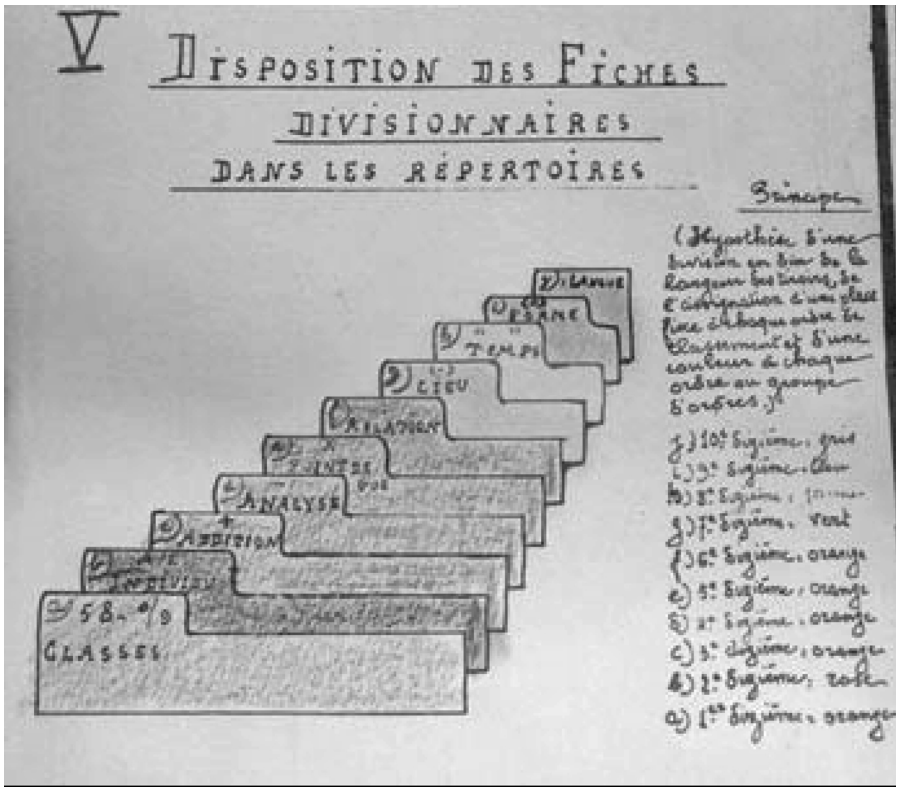


Figure 8. *Otlet – Card box with tabs indicating provenance and intentions of annotations in relation to knowledge class*
EUM – Affiche Table de classification – 58 botanique (detail) Mons Mundaneum ©

nowadays would be called, domain experts. Compared to Wikipedia, Otlet's knowledge system and collaboratory is more top down, but at the same time also more transparent. Edits and annotations do not merge directly with the information, but remain visible in an ordered way, describing the provenance and intention of the proposed data enrichment of the Universal Documentation Network. This aspect is important, with regard to the statement above that the concern of scholars in humanities and social sciences scholars with the creation, manipulation, communication and perception of meaning might explain their reluctance to adopting digital infrastructures. By studying historical knowledge infrastructures in a critical way we might make what Downey called the "hidden workers" in physical technologies and virtual webs, visual again (Downey 2007). The scholarly practices of researchers, and especially of potential researchers need to be brought to the foreground. Transparency in how knowledge is created, communicated and evaluated might enhance trust in the Semantic Web and Web 2.0 for research. In the last part the hypothesis will be put forward, that despite claims of developers of such web applications, the cyberinfrastructures for research are becoming less transparent and might benefit from design features around Otlet's analog information infrastructure.

8.0 Otlet's analog information infrastructure, the Semantic Web and Web 2.0 in research

Otlet described the basic principles of the Universal Network of Documentation, in which public and private documentation centers and individual scholars would join forces as follows (Otlet 1937, 14):

> a – cooperation and exchange, b – centralization and decentralization, c – private and official activities, d – freedom and discipline to meet practical actions and e- open access and paid usage.

In that respect you might, in terms of the *Understanding Infrastructure* report, position Otlet's Universal Network of Documentation as an infrastructure in between a common network and an internetwork or web. There is certainly control and moderation, but the network does not only allow for integration of systems, such as the various classification systems and the technical components of the aforementioned mechanical brain, but also of other dissimilar networks. In addition to that, it can be argued that the Internet and the World Wide Web also do not always meet the characteristics described in the *Understanding Infrastructure* report (Edwards et al. 2007, 12) to distinguish them as internetworks or webs. The historical description from systems to internetworks or webs is an interesting alternative to the historio- (hagio-) graphical line of protagonists of the Internet and World Wide Web. However, the authors of *Understanding Infrastructure*, despite their important notion that infrastructure should be sensitive to discontinuities, describe this historical process itself in a linear way, with the Internet and World Wide Web as logical outcomes. These cyberinfrastructures are used implicitly as yardsticks to measure the development from systems into "genuine" or "matured" internetworks or webs in retrospect. However, the Internet and World Wide Web do not develop in a progressive way, following the criteria for consolidation stipulated in *Understanding Infrastructure* (Edwards et al. 2007, 12). Ted Nelson, whose hypertext project *Xanadu* inspired Tim Berners-Lee, described the structure of the World Wide Web, as it had become, as hierarchical "decorated directories." (Nelson 2003; quoted in Wright 2007, 227).

Lawrence Lessig, who analyzed issues of openness and control in the development of the Internet and the World Wide Web from a legal perspective, observed "that the invisible hand of cyberspace is building an architecture that is quite the opposite of what it was at cyberspace's birth" (Lessig 1999, 6). In a later publication, Lessig (2002, 25) sketched a rather pessimistic development of the Internet and World Wide Web in which free and controlled layers are mixed more and more with such serious consequences for society, that innovation and the future of ideas are at stake. It means that these cyberinfrastructures are less "reconfigurable," one of the criteria according to the authors of *Understanding Infrastructures* of network infrastructures, and diminishes the potentiality of Web 2.0 in general and for research in particular.

The same applies to the Semantic Web, the knowledge structure for which also does not meet all the criteria of a developed internetwork or web. In 1998, Tim Berners-Lee wrote: "The Semantic Web is what we will get if we perform the same globalization process to Knowledge Representation that the Web initially did to Hypertext. We remove the centralized concepts of absolute truth, total knowledge, and total provability, and see what we can do with limited knowledge" (Berners-Lee, 1998).

Ted Nelson described the semantic web as a product of "tekkie committees that will decide the world's

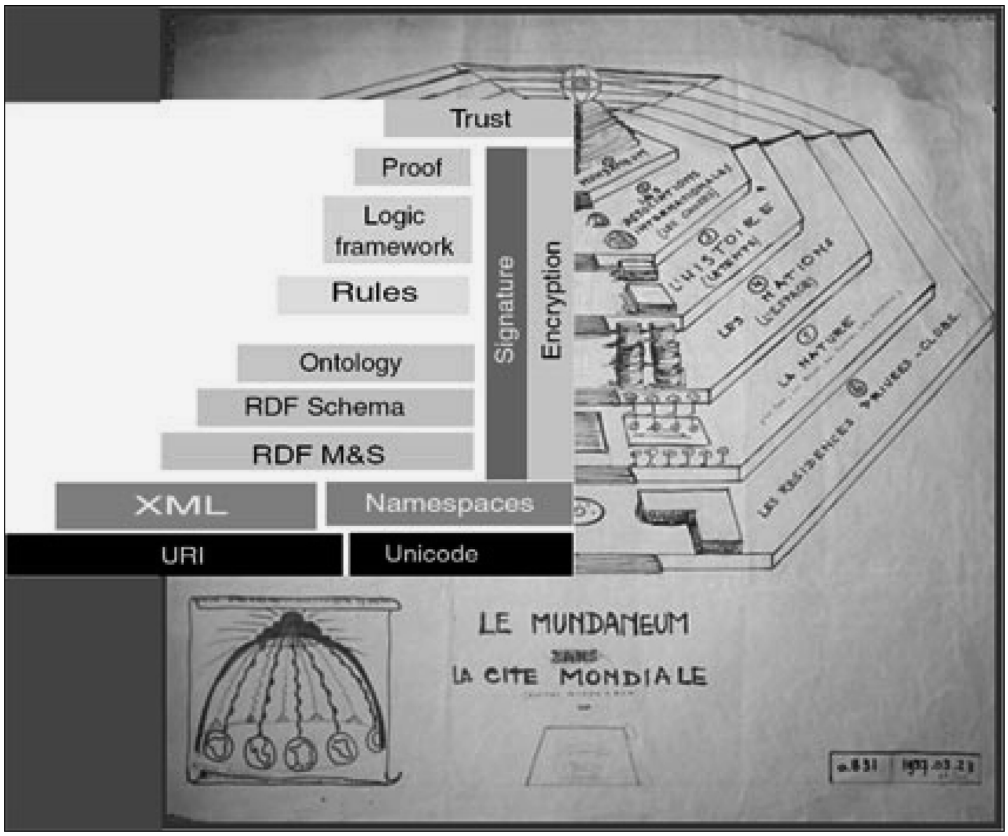


Figure 9. *RDF Stack compared to Otlet's hierarchical knowledge construction of the Mundaneum*
EUM – Affiches (O nr. 831) – Mons Mundaneum ©

true concepts once and for all" (Nelson, 2003). Indeed, by looking at the visualization of the "stack pattern" of the Semantic Web Architecture adopted by the W3C, a pyramidal structure of components, similar to that of Otlet can be recognized (Berners-Lee 2000 b) (Figure 9). Despite Berners-Lee's recommendation to remove centralized concepts of absolute "truth," from the design schemes for the Semantic Web, it might be argued that this "stack pattern " contains positivist elements comparable to the knowledge representations of Otlet. Berners-Lee's image shows a hierarchy of concepts with "trust" and "proof " at top (Heuvel 2008, 149).

The attempts of Tim Berners-Lee and other builders of the Semantic Web to avoid any form of hierarchy and centralization in its knowledge structure are undoubtedly genuine. However, if hierarchy and centralization cannot be completely avoided to validate the building components of the Semantic Web, this should be made explicit. Trust and proof are implicit and it is not clear how the quality of RDF stacks as basis components can be assessed, which compromises the accountability of the Semantic Web for research as a whole.

This is not an a-historical attempt to sneak in through the backdoor Otlet's hierarchical and centralized model that is deeply rooted in nineteenth century Positivism and twentieth century Modernism. However, similar to Otlet's designs for protocols to maintain authority in updates of his knowledge infrastructure, we might perhaps focus more on software developments that make mechanical or manual data-enrichment of the Semantic Web and Web 2.0 more transparent and choices more explicit. There are some interesting projects that might be taken as a point of departure for creating trust and authority in data-enrichment in a distributed context. The HarVANA project for instance allows visualizing the provenance of links, simply by grouping them in colours (Hunter et al. 2008). In the GenTech Datamodel, to give another example, group members can give certain values to links based on expected expertise (GENTECH).

By integrating such software developments in mechanical and manual data-enrichment of larger cyber-infrastructures for research, trust and quality might be enhanced and the necessary critical mass might be obtained that the small, heterogeneous datasets of

the humanities often lack. Katy Börner envisions an infrastructure, called semantic association network, in which the enriching node by scholars becomes the heart of scholarly activity (Börner 2006, 198):

> Somewhere in the not too distant future, reporting a scholarly result might not involve writing a paper. Instead, scholars may add a <<knowledge nugget node>> or an <<association link>> to a complex semantic association network of humanity's knowledge [...] The nodes in this network will describe tangible objects (e.g., a pottery piece found at a certain place by an archaeologist together with information about its origin and intermediate positions/usages up to today) or intangible objects (e.g., a formula). [...] Each node and each link would have information on who added, modified or deleted it. A scholar's reputation would depend on the number of nodes and/or links s/he contributed and their usefulness for humanity.

This vision stands in a longer tradition and is not unlike what Otlet had in mind with his knowledge network.

9.0 Epilogue: Web 2.0 and the Semantic Web in research from a historical perspective

The *Understanding Infrastructure* report provided a useful touch stone for understanding Otlet's Universal Network of Documentation. Following the criteria outlines in this report it was possible to situate Otlet's network as an analog infrastructure in between a common network and an internetwork or web. However, we also challenged the implicit linearity of the historical model presented in *Understanding Infrastructure* and its underlying claims about openness, distributed character and the (re-)configurability of the Internet and the World Wide Web. We showed that these cyberinfrastructures do not meet all the criteria of an internetwork or web described in the report and in that respect are not different from Otlet's network. Finally, we concluded that instruments and protocols envisioned by Otlet to enhance collaborative knowledge production can still be relevant for current conceptualizations of transparency and "scientific authority" in data sharing and annotation in Web 2.0 applications and in the modeling of the Semantic Web for research.

The comparison of texts and visualizations of Otlet's centralized and hierarchical knowledge constructions with design features of the World Wide Web and the Semantic Web makes clear that claims from developers about the non-hierarchical, distributed characteristics of these cyberinfrastructures can be studied in a critical way from an historical perspective. It can even be argued, on the basis of Otlet's designs and protocols for his knowledge infrastructure, that some historical solutions for data-enrichment for research, at least conceptually, were more advanced than "new invented" e-research tools and applications. History is not linear. Or as T.S. Eliot put it eloquently: "Time present and time past are both perhaps present in time future. And time future contained in time past"DDD (Eliot 1943, 3). Therefore recent scientific infrastructures, such as Web 2.0 and Semantic Web applications for research, can only be fully understood by studying them from a historical perspective.

References

ACLS. 2006. *Our cultural commonwealth: the report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences*

Berners-Lee, Tim. 1998. *Design issues* 'What the semantic web can represent,' paragraph: 'Knowledge Representation goes global," http://www.w3.org/DesignIssues/RDFnot.html [last modified 17 Sept. 1998].

Berners-Lee, Tim. 2000a. *Weaving the Web. the original design and ultimate destiny of the World Wide Web by its inventor.* London: Harper Business. [first paperback edition]

Berners-Lee, Tim. 2000b., www.dajobe.org/talks/sw-vienna/slide10.html.

Börner, K. 2006. Semantic association networks: using Semantic Web technology to improve scholarly knowledge and expertise management, in Geroimenko, Vladimir and Chen, Chaomei eds.,. *Visualizing the Semantic Web*, 2nd ed. Berlin: Springer Verlag, pp. 183-198.

Buckland, Michael. 1991. Information as thing. *Journal of the American Society for Information Science* 42: 351-60.

Downey, Greg. 2001. Virtual webs, physical technologies, and hidden workers. the spaces of labor in information internetworks. *Technology and culture* 42: 209-35.

Edwards, Paul.N, Jackson, S. J., Bowker, G., and Knobel, C.P. 2007. *Understanding infrastructure. dynamics, tensions and design*. Report of a work-

shop on History&Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. January – NSF grant 0630263 Consulted April 2009) http://www.educause.edu/Resources/UnderstandingInfrastructureDyn/154606.

Eliot, Thomas Stearns. 1943. Four quartets I, in *Burnt Norton.* New York: Harcourt, Brace and Co.

GENTECH, URL (consulted April 2009) http://www.ngsgenealogy.org/ngsgentech/projects/Gdm/Gdm.cfm.

Heuvel, Charles van den. 2008. Building society, constructing knowledge, weaving the web: Otlet's visualizations of a global information society and his concept of a universal civilization. In Rayward, W. Boyd, ed., *European modernism and the information society*. London: Ashgate Publishers, pp. 127-153.

Hunter, J., Khan, I., and Gerber, A.. 2008. HarvANA—harvesting community tags to enrich collection metadata, paper presented at ACM IEEE Joint Conference on Digital Libraries, JCDL Pittsburgh, 6-10 June.

Lessig, Lawrence. 1999. *Code and other laws of cyberspace*. New York: Basic Books.

Lessig, Lawrence. 2002. *The future of ideas: the fate of the commons in a connected world*. New York: Vintage Books.

Linking Open Data Project, URL (Consulted April 2009) http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

Nelson, Ted .H. 2003. *I Don't Buy In*, URL (Consulted April 2009) http://ted.hyperland.com/buyin.txt.

Otlet, Paul. 1909. *La function et les transformations du Livre. Résumé de la conférence faite à la Maison du Livre 14 November 1908. Musée du Livre 11.* Bruxelles: Musée du Livre.

Otlet, Paul. 1913. *Le livre dans les sciences, Conférence faite à la maison du Livre par M. Paul Otlet, Musée du Livre 25-26*. Bruxelles: Musée du Livre.

Otlet, Paul. 1934. *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelles: Editiones Mundaneum, Palais Mondial.

Otlet, Paul. 1935. *Monde, essai d'universalisme: connaissance du monde, sentiment du monde, action organisée et plan du monde.* Bruxelles: Editiones Mundaneum/D.van Keerberghen&Fils.

Otlet, Paul. 1937. Réseau universel de documentation: documentatio universalis (Rete), *IID communicationes* 4 no. 1: 13-16.

Rayward, W. Boyd. 1975. *The universe of information. the work of Paul Otlet for documentation and international organization*. Moscow: FID.

Rayward, W. Boyd. 1983. The International Exposition and The World Documentation Congress, Paris, 1937. *Library quarterly* 53: 254-68.

Rayward, W. Boyd. 1990. *International organisation and dissemination of knowledge, selected essays of Paul Otlet, translated and edited with an introduction by W. Boyd Rayward.* FID 684, Amsterdam, New York, Oxford, Tokyo: Elsevier.

Rayward, W. Boyd. 1994. Visions of Xanadu: Paul Otlet (1868-1944) and hypertext. *Journal of the American Society for Information Science* 45: 235-50.

Rayward, W. Boyd. 2003. Knowledge organization and a new world polity: the rise and the fall and the rise of the ideas of Paul Otlet. *Transnational associations/Associations transnationales : la revue de l'Union des associations internationales, numéro spécial 1-2/2003 (L' oeuvre de Paul Otlet):4-15.*

W3C. 2001.Semantic Web Activity. URL (Consulted April 2009) http://www.w3.org/2001/sw/.

Wouters, Paul. 2007. Forging new academic identities: key dilemmas and challenges in digital scholarship, paper presented at Simpson Center, University of Washington, Seattle, May 1.

Wright, Alex. 2007. *Glut: mastering information through the ages*. Washington D.C.: Joseph Henry Press.