Scholarly Information Management: A Proposal[*]

Paolo D'Iorio

External Image
Please see:
http://rup.rice.edu/image/cnx_images/shapes-buybutton.jpg

This paper will present some ideas about scholarly information management and outline the conceptual model of a digital research infrastructure for the humanities. An infrastructure is usually defined as a well-coordinated system of buildings, equipment, services, procedures, etc., that facilitates a certain activity. It includes physical and organizational structures and it refers above all to public works such as highways, bridges, airports, etc. It can be conceived as an underlying support as well as something that establishes a horizontal network of connections between different elements.[1] So what is the traditional infrastructure for the humanities and how it is made? In the traditional non-digital world a scholar consults primary sources inarchives or libraries; in the libraries he also reads secondary sources like journals, monographs, and different published editions, which he can also find in bookstores; in the University he transmits this knowledge to students; while conferences give him the chance to share knowledge with colleagues; publishers sell his works on the book market in accordance with copyright law: all of these material and organisational elements—archives, libraries, bookstores, universities, courses, conferences, publishing houses, intellectual property law and others—constitute the traditional infrastructure for research in the humanities, and it has been developing over the course of two thousand years.

Our question is: is it possible to transpose scholarship into an electronic environment? That is, to reproduce the traditional infrastructure of the humanities in a digital medium? Can we switch to virtual to solve some of the problems with the traditional infrastructure without losing any of its virtues? By switching to virtual I mean not only accessing sources, as we scholars usually do in archives and libraries, but also publishing new work in ways that will stand the test of time and win prestige (as we're always trying to do when we submit a manuscript to a publisher), and educating younger generations, as is the mission of our universities. In 2000, I posed this question in my book *HyperNietzsche*,[2] and my conclusion was "yes," so I immediately began to develop it. Of course I was not the only one: you know better than I do that numerous different models and initiatives are under development in this field, and of course I couldn't achieve everything that I wanted to; but from this experience I gained some valuable ideas, which seem to form a coherent model and which could be useful for future development.

---

[*]Version 1.2: May 14, 2010 5:43 am +0000

[†]http://creativecommons.org/licenses/by/3.0/

[1]In Italian the prefix *infra* expresses this duality very well because it contains both the sense of "under," coming from Latin, and the sense of "between," as used in the time of Dante.

[2]*HyperNietzsche. Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques*, edited by Paolo D'Iorio, Paris, PUF, 2000, 200 p. (free digital version available at the address: http://www.hypernietzsche.org/doc/puf/ (<http://www.hypernietzsche.org/doc/puf/>)).

The model I propose is called Scholarsource and it is divided into three parts, which correspond to three subquestions: how is it possible? who can do it? and how can the information be organized and managed? The three corresponding parts are:

1. The Conditions of Possibility of Scholarship
2. Scholarly Communities on the Web
3. Scholarly Information Management

The three parts of the model are each imbued with a different status. The first point—the conditions of possibility of scholarship—is a necessary requirement for any environment that aims to support humanities scholarship. The second and third points, on the contrary, indicate only a possible way to realize scholarship in the digital era and can be replaced by different strategies. In some ways the first point is more philosophical, while the second is a sociological one and the third is more technical; but, as usual, these disciplinary distinctions are not very precise.

1 Conditions of Possibility

Borrowing the phrase from Immanuel Kant, but using it in a non-Kantian sense, "conditions of possibility of scholarship" is used here to mean the principles that undergird scholarship. It is those general rules without which either our infrastructure will not function or it will produce something different from scholarship. By way of explanation, let us have a look at a successful example of transposition of a traditional activity into a digital environment: eBay. eBay is a digital infrastructure for selling and buying. It was made possible because its inventors could identify and reproduce in a digital environment all of the key requirements for a successful business relationship, that is trust—trust in payment and in merchandise delivery. Once successful in ensuring trust, additional features could be added, such as price comparisons, email reminders and advanced searching. Without the trust rating system, though, all of the additional features would have been useless because people would probably not have used eBay at all. Research infrastructures in the humanities have in many cases been driven more by capacity than by exigency, with each advance in technology inspiring a new set of aspirations and plans and producing new sophisticated features. Before adding additional features, though, we have to ensure that the key requirements for scholarship are fulfilled. In other words, we have to start by identifying the conditions necessary for conducting scholarship, and only then will we have the basis upon which to develop and evaluate digital infrastructures for the humanities. Three of these requirements—Quoting, Consensus and Dissemination/Preservation—will be discussed here.

Quoting is the first requirement for the activity we call scholarship. Scholarship is a conversation based on hypothesis, arguments and facts. We should remember that facts, in the humanities, are often contained in documents, in texts. Emma Bovary drinks arsenic and dies. That is a fact. What exactly this fact means might be a matter of interpretation and dispute, but that she "drinks arsenic" is indeed a fact. But to be sure of this, you must be able to consult and quote the first edition of Flaubert's *Madame Bovary*. Quoting requires stability of bibliographic references and, most of all, stability of texts. Printed texts can normally ensure both of these. But what about on the Web? On the Web, this type of quoting can be quite difficult. Web pages change every day, appearing, disappearing, reappearing under other names and addresses. Nevertheless, it is certainly not impossible to create special systems, like little islands in the Web, to ensure the stability of electronic documents and web addresses. Technical solutions exist. From a technological point of view, the URL/DNS technologies are perfectly sufficient to ensure the stability of web addresses[3] and a simple checksum system is able to verify that documents were not changed over the

---

[3]There are plenty of initiatives to ensure the stability of Web addresses, such as DOIs (Digital Object Identifiers) and many more. This is a hot topic in the librarian community. My personal opinion is that trying to tackle this problem by inventing a new naming system is fundamentally useless because the URL (Uniform Resource Locator) or more precisely, the URI (Uniform Resource Identifier) can already identify documents in a stable and univocal manner. What's more, DOIs are managed by a commercial organization that has the same if not more chances of disappearing as each single repository that manages its own URIs; anyway, it doesn't give any more guarantee of stability than IANA (Internet Assigned Numbers Authority) and it is a contractor ICANN (Internet Corporation for Assigned Names and Numbers) who manages IP addresses and domain names. In

time. But the existence of technical solutions alone is not sufficient. To create an island in the Web where documents and their addresses are stable requires both the decision to install these technologies and a strong commitment not to alter them over the time. We already have the technical solutions, but we need coherent scientific policy decisions.[4]

Consensus is the second important requirement for scholarship. As with other social activities, to receive support and to be included in the common research enterprises, scholars must produce works recognized as interesting by their colleagues. But more than in other social activities, in scholarship this recognition is understood to be based on evidence and must be as fair and transparent as possible. In reality, of course, this is very difficult to realize. Peer review and other systems of fair evaluation are continuously contested. The Web holds a real possibility for change, a chance to organise consensus in a better way: easier, more transparent, more efficient. In this case the Web is an improvement over the traditional system, allowing for new possibilities: article ranking according to quotation, impact factor based on semantic tagging, number of citations or downloads of an article. And some journals, for example *Nature*, are already experimenting with new types of digital peer review. It is important to note that in this case as well, policy decisions, more than technology, will determine the success of these new systems.

The third condition for the possibility of scholarship is *Preservation*. Scholarship is essentially an historical activity and we must be certain that our electronic documents survive us. Stanford's LOCKSS ("Lots of Copies Keep Stuff Safe") project supposes that the best way to preserve an electronic document is not to keep a unique copy in a very safe place, but to let people make thousands of copies and to spread them all over the world. They are right; the thirty centuries of our cultural heritage confirm the premise of LOCKSS. We have lost only what we did not copy. Why did we lose almost all the work of Heraclitus, while we preserved almost all of Aristotle's work? Because Heraclitus had a DRMS (Digital Rights Management System) strategy: he stored the unique copy of his work in a safe place (the temple at Ephesus) and people could not copy it. As a result, his work was lost and only the quotations made by other authors still survive. Aristotle, on the contrary, was a copyleft guy: allowing his students to copy his works whenever they wanted. Copy after copy of Aristotle's works have been passed down to us. The worst enemies of preservation are copyright, DRMS, and all of the technical means that now prevent copying. The true friends of preservation are free sharing and the copyleft movement.

I am wondering if we should add a fourth condition of possibility: *Dissemination*. Dissemination seems to be a fourth requirement because it is hard to imagine modern science or scholarship without public diffusion of its results. Modern science is a public conversation based on evidence in which both primary sources and research results must be easily accessible to all. On the one hand, it is impossible to provide arguments or proofs based on documents which are not accessible; on the other, to be taken in account, research results have to be published. Yet as we agreed that dissemination through copy is the best—and ultimately the only—way to achieve preservation, then dissemination is not a fourth condition of possibility, but a different name for the third one. Preservation and dissemination are indeed the same thing—two faces of the same coin—and the third condition of possibility is twofold: Dissemination/Preservation. In this case, we have to conclude that the electronic medium—the Web in particular—is the best medium for both dissemination *and* for preservation, if we agree that digital preservation is best achieved through copying. This means that to maximize preservation and dissemination, we should remove all legal obstacles that so often obstruct access to sources and the diffusion of research results. In the long run, open access is not an option for digital scholarship: it is a requirement. And this is yet another problem of policy, not of technology.

---

the end, DOIs are "identifiers of identifiers" that just shift the problem to a new layer, and their supporters seem to ignore the story of the anthropologist and the Indian: "What does the world rest on?" the anthropologist asks the Indian. "The Great World Tortoise." "And what does the Great World Tortoise stand on?" "Another tortoise..."

[4]Digital libraries should simply have the same policy as that of prestigious traditional libraries, which are not used to lose or alter the content of their books, or to change their signatures (at least not without writing a table of concordance with the old one).

2 Scholarly Communities on the Web

Now the question is: who will resolve all these policy problems? In order to fulfill the conditions of digital scholarship, scholars will undoubtedly have to come to agreements with their libraries, publishers and other stakeholders. First of all, though, we scholars must come to an agreement amongst ourselves: we need to form Open Scholarly Communities on the Web to lead the change into digital scholarship. These are free international associations of specialists who work on a specific author or area of research. They collaborate with libraries, universities, and publishers, but they themselves fix their own priorities, preserve the stability of texts and authorship, guarantee scholarly standards, and ensure open dissemination and thereby the long term preservation of content. Scholarly communities on the Web do not yet exist and they will be difficult to create: levels of digital literacy and awareness vary greatly between scholars, public institutions are not always open to allowing online access to their holdings, and there are also few publishers who seem willing to accompany scholars into the digital era. (This is perhaps understandable, as they don't want to change their old business model; but less understandable is that scholars are often bound hand and foot to the publishers and seem happy to stay that way.) The model for the creation of these open scholarly communities on the Web can be found in the tradition of the academic societies of the seventeenth century: the *Accademia dei Lincei*, the *Académie Française*, the *Royal Society*. These were the social networks when modern science was born. After the time of the Academies, a new model emerged and it seems no longer suitable for the transition into the digital era.

3 Scholarly Information Management

Up to this point, we have discussed some of the basic and necessary requirements for scholarship. We have seen that these requirements necessitate a certain number of policy decisions and that the principal actors who should make these decisions are the scholarly communities on the Web. They are the real stakeholders—those who really know what is at stake, who care about it and who are willing to act (or at least they should know, and should care and should be willing to). Now the question is, how can we conceive, from a technological point of view, the realization of our island of selected scholarly knowledge? For a moment let us imagine that Nietzsche Source will be the island in which Nietzsche specialists consult and publish reliable editions and scholarly articles, the same for Wittgenstein Source.[5] Together, the Scholarsource Federation will be the archipelago containing documents that can be quoted in a stable way, which receive the consensus of a scholarly community and which will be disseminated *and* preserved. What other functions could it feature? How should the content be organized? My initial thought was that all the scholarly communities should rely on the same software, like Facebook, MySpace, Wikipedia, etc.[6] But scholarship is more complex and scholars are too different to adopt a "one size fits all" strategy. Each Island should therefore chose or develop its own software. So what will be suggested here is only a conceptual model of scholarly information management containing some general structures and features that are not necessary elements like the conditions of possibilities, but only possible forms of organizing scholarly content and which can be realized with different technologies. They can be divided in three categories: 1) Ontologies; 2) Capacities; 3) Interfaces.

3.1 Ontologies

Each of the sites of the Scholarsource Federation should use a very general ontology—The Scholarship Ontology—that expresses the distinction between research objects (primary sources), research results (secondary sources), and the authors of both (scholars). It also describes the kinds of relationships between these sources and their authors, such as "related to," "describe," "criticize," "comment," etc. The primary sources are what we want to speak about and the secondary sources are the product of the different ways

---

[5]A first version is available at http://www.nietzschesource.org (<http://www.nietzschesource.org>) and http://www.wittgensteinsource. org (<http://www.wittgensteinsource. org>); see also the other website published within the Discovery project: http://www.discovery-project.eu (<http://www.discovery-project.eu>).

[6]Like Scholarsource, they are islands in the Web, each one of them with its own features, rules, values.

in which scholars can speak about the primary sources. Along with this general ontology each node of the Federation will use narrower domain source ontologies. These more specific ontologies can be bibliographic, specifying the different types of sources used by the community (commentary, articles, critical editions, etc.) or theoretical, expressing the concepts used by the concerned authors and their relationship (philosophical, historical, linguistic, and so on.)

Let's focus on the Scholarship Ontology. If a scholarly community intends to conduct research on a certain topic, it first needs to define which documents or objects to consider as its primary sources. When a research line is about to be developed and consolidated, a catalogue of primary sources is compiled, usually by archivists or librarians.The catalogue of primary sources lists the relevant classes of objects and often includes the complete list of their instances. For example, in the case of the work of Wittgenstein, scholars interested in studying his philosophy have created a detailed catalogue of his writings, divided according to the different types of documents (books, manuscripts, typescripts...) and including a complete list and description of each manuscript. Catalogues of secondary sourcescome later, and are written by scholars or librarians, generally in the form of a bibliography listing the most relevant scholarly contributions written on Wittgenstein (editions, monographs, articles, reviews...) existing at a given moment. The distinction between primary and secondary has a fundamental epistemic value. According to Karl Popper, what distinguishes science from other human conversation is the capacity to indicate the conditions of its own falsification. In scholarship, the conditions of falsification normally include the verification of hypotheses on the basis of a collection of documents recognized by a scholarly community as relevant primary sources. Thus we can affirm that the distinction between primary and secondary sources exhibits the conditions for falsifying a theory in the humanities.

The idea of collecting in one place all of the primary and secondary sources needed for conducting research on a given subject is intrinsic to the history of the organisation of knowledge, because scholars and librarians know that it is a very effective means of producing new knowledge. Now, to what extent are the traditional research environments, that is the libraries, able to represent the fundamental distinction between primary and secondary sources and to help researchers orient themselves in the information? Before answering this question, we should mention that even if manuscripts, artifacts and paintings are considered, almost without exception, only as primary sources, most printed documents have no fixed status and can be considered as primary or secondary sources according to different research topics and scholarly communities. For example, an article written by Nietzsche on Plato is a primary source to Nietzsche scholars, but it is a secondary source to Plato scholars. Traditional physical libraries are generally unable to reconfigure the disposition of their books according to the needs of the scholars. Nevertheless, they put in place a certain number of strategies to permit scholars to find their way amongst the mass of collected documents:

1. *Research Libraries.* The most successful strategy is to dedicate some libraries to a single research topic. While a general library allows users to consult numerous collections dealing with a wide variety of subjects, the purpose of a research library is to focus on a single subject and to provide scholars with access to all the primary documents and reference works they need to conduct research therein.

2. *PhysicalArrangement.* Open-shelf libraries often arrange their books in a way that puts the primary sources next to the relevant secondary sources. In the Dewey classification, for example, the critical essays on an author usually follow his collected works.

3. *Cataloguing.* Independently from the physical arrangement of the books, catalogues of primary and secondary sources or subject catalogues help scholars retrieve a relevant publication and relocate the information according to their research needs.

Digital libraries can do even better. Not only can they unite the collections of different libraries, but they can also easily reconfigure their holdings according to any scheme, taking into account the different status of a given source within different research contexts. If a scholar enters our network of semantic digital research libraries through the door of, for example, Plato scholars (using the Plato Source ontology), the information and resources would appear to him in a certain configuration. For example, Plato's *Dialogs* would appear as primary sources and articles by Nietzsche on Plato would be listed under the secondary sources and be accompanied by other critical essays on Plato. But if the scholar enters through the Nietzsche door (using

the Nietzsche Source ontology), the same material would be presented in a different way—with Nietzsche's articles on Plato appearing as primary sources (within the class "published works") and related to all critical essays and other secondary sources *on* Nietzsche (not on Plato), while Plato's *Dialogs* would be included in the class Nietzsche's "personal library." So in this way we could transpose the structure of traditional scholarship onto the Web, preserving the different epistemic values and relationships which scholars attribute to their sources, and improving the way in which the documents can be dynamically rearranged according to these relationships. Furthermore, all digital objects would appear as generic resources having the same epistemic status and the user could search them using a minimum set of shared, standard metadata, such as title, author, date of publication, etc. In this way our infrastructure can be very specialized and targeted to the needs of specialist scholarly communities, and at the same time be fully interoperable with general digital libraries and aggregators. So the general library will serve all kinds of readers and ensure interoperability while the specialized research libraries (concerning Plato, Nietzsche, Wittgenstein, etc.) can permit scholars to find their way in an electronic environment structured according to the standard classification used in their communities. The same thing would, of course, presumably be possible concerning theoretical ontologies.[7]

3.2 Capacities

Now we stand in front of the shelves of our digital library and we see how ontologies can help dynamically arrange the books according to the glasses we use to perceive them. What about opening the books? What happens when we start to navigate not only in the library but in the documents contained in the library? The first of the core features listed under Capacities provides an answer to this question.

3.2.1 Scholarly Navigation

The traditional scholarly infrastructure has been useful because with a simple bibliographical reference at the bottom of a page, an author was able to refer in a very precise manner to a specific passage contained in another article or in a book. A scholar in pre-digital times did not navigate the library by following a list of "hits" like the kind produced by Google. Scholarly knowledge is not structured like a list or a tree, but rather like a graph. In mathematics, a graph indicates a set of objects connected by links, where the links can be labelled. These links not only indicate the connection between two objects, but explain the type of—or reason for—the connection. The structure of a set of documents connected by references in footnotes, which indicate both a link and the reason for the link, can be formally described as a graph. Understanding this helps to dispel a common misunderstanding—that the difference between printed books and hypertext is that a book ensures a sequential reading whereas hypertext introduces non-sequential reading. Nothing could be more false in the realm of scholarly research, because a key characteristic of scholarly reading is precisely that it is non-sequential. A classicist at work in the library is likely to have a dozen or more books open on the table and to jump from one to the other: he verifies, he looks for connections, he follows links made explicit through the venerable tradition of scholarly citation chaining.

Now that we have a clear picture of how the scholar works, the question becomes: how can we transpose this good old system of scholarly citation into a digital infrastructure, producing a new referencing system that employs all of the powers of the Internet? I proposed a feature called dynamic contextualisation at the level of database programming and scholarly navigation at the level of user interface. Thanks to this feature, when a user selects a critical essay he will be automatically presented with a list of all the primary sources cited in the essay, a list of all the articles cited by the selected essay, and, more importantly, a list of all the essays in which other authors cite the essay currently being viewed. When a user selects a manuscript page, the system will immediately present all the transcriptions, editions and translations available for that page, as well as all critical essays commenting the selected page.

Often research infrastructures for the humanities are completely based on search engines; to the point that they are actually more search infrastructures than research infrastructures. Scholarly navigation attempts

---

[7]A first example of the Scholarship Ontology has been produced and formalized within the HyperNietzsche and the Discovery project. Within the Discovery project were also written and tested some philosophical ontologies devoted to Nietzsche, Wittgenstein, ancient and early modern philosophy.

to provide a complementary model, in which you don't need to search words to find that fundamental piece of information that allows the production of new interpretations, that is: who has previously commented on this passage and how?

This system of dynamic contextualization can also be combined with the domain scholarship ontologies mentioned above. For example, if Nietzsche is cited in an essay published in the Wittgenstein research library, the reader could mouse click to the Nietzsche research library and go right to the original source in Nietzsche. There he will have translations of the passage in different languages and commentaries from Nietzsche experts. Scholarship, indeed, is the capacity to analyze the same object with different criteria, and different objects with the same criteria, and this is important not only from a methodological but from an epistemic and cognitive point of view. The objects of the hard and human sciences always result from a process whereby meaning is constructed within a research community. The increase in the number of contributions concerning a certain object actually represents a progressive transformation of this object, insofar as each essay discovers unknown properties. To know that an aphorism is genetically or thematically related to other texts and manuscripts can radically change our comprehension of this object of study:it is as if one had identified a gene on the basis of a certain number of characteristics and then ten scientific articles illustrated hitherto unknown properties and unsuspected relations with other genes, thus appreciably transforming its very definition. This is the epistemological value of the Scholarly Navigation, which permits one to follow very concretely and very closely the epistemological process of object construction.

Dynamic contextualisation can also be seen as a new form of scholarly citation in the digital era, more powerful than the old citation system because it is bi-directional and dynamic. Bi-directional means that the system can not only point towards a textual passage but also go backwards to the origin of all the references that quote it. Dynamic means that the list of articles referring to a certain passage is updated automatically without the need to peruse all journals and monographs manually, as in the case of the *Science Citation Index*. With this system you can develop automatic bibliometric surveys without using core journals arbitrarily chosen and manually browsed, and it would be the actual give-and-take of real academic discourse registered automatically on the network through citations that would determine the reputation of scholars— and not a tiny number of core journals chosen by the editors of the *Science Citation Index*. I am against the use of impact factor for the evaluation of scholarship, for a number of reasons I will not mention today, but if we are going to use impact factor, the dynamic contextualisation could offer a fairer way to realize it.

3.2.2 Semantic Knowledge Management

A scholarly system of information management should be capable of managing semantically structured information. This function can either be linked to the previous one, the scholarly navigation, or can be implemented in the form of a traditional search engine, but it is important to use Semantic Web technologies, because, as we mentioned, scholarly knowledge comes naturally in the form of graphs with labelled arcs.

There are different kinds of links to express the range of relations between primary and secondary sources. For example, we could distinguish between positive and negative citations of an article; or between philolog- ical, rhetorical, or philosophical analyses of a text passage; or between archaeological, historical, or stylistic analyses of a painting or an artefact. We need to have software agents to exploit these relations, like a bib- liometric application that takes into account not just of the number of citations of an article, but also their quality—positive/negative, agree/disagree, etc.—to calculate a weighted impact factor; or an application to manage index of concepts according to the philosophical domain ontologies. If we codify all this information using a standard language (like RDF), all the computers connected to the Internet could refer to and analyze it, and everybody could program applications to use it in ways that we can't even imagine.

This was the original idea of Tim Berners Lee, and, as you may have noticed, the title of my text is an homage to the thirtieth anniversary of the paper in which he described the project of the World Wide Web. That paper—judged "vague, but exciting" by his boss, Mike Sendall—already contained the fundamental idea of what Berners-Lee later developed under the name of Semantic Web. Tim Berners Lee was fully aware that structuring knowledge in form of a tree would "not allow the system to model the real world."[8]

---

[8]Information Management: A Proposal, Tim Berners-Lee, CERN, March 1989, May 1990,

3.3 Interfaces

Interface is a very important element for any information system in general and particularly for a website aiming to represent the complex knowledge relationships used in science or scholarship. In this case, a single unique interface cannot fit the needs of different scholarly communities and therefore I don't have in mind to present you the ideal interface. On the contrary, I would like only to report my experience, my own trial and error process, in designing an interface capable of representing the concept of dynamic contextualization, hoping that it could be useful for designing similar websites. Dynamic contextualization is a coherent and rigorous concept, but, as it turned out, quite difficult to transpose in an intuitive and easy navigable interface. In the HyperNietzsche website, designed in 2003, contextual information was displayed using a vertical bar on the left of the screen: while navigating the website, the contextualization sidebar presented the user with all of the contributions related to the document in the form of a list of hyperlinks.[9] It seemed a simple, reasonable and standard solution (if standard means the fact that a lot of websites were designed using a left sidebar and users were more and more familiar with it). Nevertheless, users experienced difficulties with the navigation and they were not even able to visualise the facsimile or the transcription of a Nietzsche manuscript. Starting from the version 0.4 of HyperNietzsche, we therefore introduced a series of new Web pages, called "views," which didn't contain contextual information and made navigation easier and more perspicuous. Finally, at the end of 2007, we decided to radically modify the interface and the conception of the website to mark this turning point, we changed the name of the project from HyperNietzsche to Nietzsche Source.[10]

To understand why the HyperNietzsche interface was not satisfying, let us try to consider the principles on which it was built: we will see that the difficulty here was probably not the design of the sidebar, but the organization of the content—that is, the general structure of knowledge that this design was expected to express. In the print culture, scholarly knowledge came under the form of well-defined genres shaped by the physical structure of the book: treatises, critical editions, journals, collected papers, catalogues, etc. The problem is that these genres stored in the same container heterogeneous kinds of information. For example, a critical edition contains in a single book several types of scholarly contributions: manuscript transcriptions, texts editions, philological commentaries, critical commentaries, cross references, bibliographical references, introductive or critical essays, and so on. From a logical point of view—and even more from an information technology perspective—this way of collecting and mixing different types of scholarly contributions is not satisfying because then it is difficult to query, assembly and redeploy them according to different purposes. In theory, digital technologies will allow users to collect and compare different editions or translations of the same texts, or to read all the philological commentaries concerning a certain text but excluding the philosophical ones, or to create a diagram showing all the cross references concerning a certain text, etc.; but for this to be accomplished the different kinds of scholarly contributions and their parts need to have been clearly distinguished previously. Otherwise, as happens in digitization projects like Google Books and many others, digital technologies cannot deploy all their possibilities and the user is only allowed to search words, getting endless lists of occurrences without being able to retrieve and compare the information he needs. In order to allow advanced scholarly information retrieval, in HyperNietzsche I established a scholarly ontology containing a catalogue of all the different types of primary and secondary sources used by Nietzsche specialists (see above, "Ontologies"), and built the database which powered the HyperNietzsche website

http://www.w3.org/History/1989/proposal.html (<http://www.w3.org/History/1989/proposal.html>); Tim Berners-Lee and Mark Fischetti, *Weaving the Web*, London/New York, Texere, 2000, pp. 229-251.

[9]See Paolo D'Iorio, "Nietzsche on New Paths: The HyperNietzsche Project and Open Scholarship on the Web," in Maria Cristina Fornari (ed.), *Friedrich Nietzsche. Edizioni e interpretazioni*, Pisa, ETS, 2006, pp. 475-496, also available at the address: http://www.hypernietzsche.org/doc/files/new-paths.pdf (<http://www.hypernietzsche.org/doc/files/new-paths.pdf>).

[10]Besides, "hyper," "hypertext" has always been rather vague and foggy concept and the attempts to make it more precise haven't been particularly successful; and now it sounds quite retro. "Source," on the contrary, is an old idea in the humanities but one that is just as relevant as it has always been. More vintage than retro. It also has a technological meaning (code source) and a political one (open source), but it is true that the principal meaning refers to knowledge in general and to the philological sources in particular. It suggests the concrete and documented nature of research and it also indicates that in the websites bearing this name we could find the essential primary and secondary sources for anyone who wants to study the life and work of an author.

on this ontology. In this way it was possible to perform all the kinds of queries needed for the dynamic contextualization, e.g., to retrieve all philosophical commentaries concerning a certain page of Nietzsche, or all reviews concerning a certain article, etc. As we explained, these queries would have not been possible without such a disassembling of the machine of scholarship into its constitutive parts that were hidden in the form of a book. But what we didn't understand at that time was that this way of structuring information, which was completely appropriate to construct the database, could not be suitable for interface design. We reassembled the machine of scholarship in a fully hypertextextual way, transposing the logical structure of dynamic contextualization directly in the interface, abstracting from old forms of knowledge organization like editions or journals. This was not a good idea. The use, the manipulation, the construction of knowledge objects don't depend on logic, but on history. Scholars cannot work well if their materials are organized in conceptual structures which are too innovative, too different from the long-term scholarships practices of work with objects showing a certain layout and presenting a certain affordability.[11]

We finally came to the idea that, without renouncing the novelty of the system we were designing at the database level, the interface should support as much as possible the habits and expectations of the scholars. The solution was to separate navigation from contextualisation. The interface of a suitable scholarly information management system should thus be divided into two communicating parts: a part A to browse and navigate easily in the documents and a part B to contextualise and compare them. In the part A, the electronic medium should try to recreate the traditional formats of scholarly communication: improving them, if possible, but without altering their form and usability. When browsing documents, the interface should be designed using common templates which make the navigation intuitive for those who have a normal practice on the Web. Functions are reduced to a minimum and contextualization is absent. This part is divided into different subparts corresponding to the traditional formats of scholarly communication. The most common of these are:

1. The *Facsimile Edition*, which usually contains a catalogue, a material description and a digital reproduction of all the primary sources, be they documents, artifacts, movies, etc.
2. The *Critical Edition*, which publishes a textual version of the primary sources including a critical apparatus, commentary and often a critical introduction.
3. The *Genetic Edition*, which reconstructs and represents the genesis of the work.
4. *Translations* will render the meaning of primary sources or of an edition in other languages.
5. If the primary sources contain the *personal library* of an author, the catalogue of the library along with digital reproduction of the books, transcription of the annotations, commentaries and a general introduction can from a separate format.
6. A *Journal* will publish essays, reviews and commentaries.
7. *Bibliographies* will contain lists of secondary sources compiled according different subjects.

From each page of part A, a link allows the user to switch to the corresponding page of part B (and vice-versa). In part B, all the documents which in part A appeared organized in different formats are completely atomized. It is now possible to use a set of tools to retrieve them according to different criteria and above all to contextualize and compare them. As an interface for scholarly navigation, this time we will use a synoptic mask divided into several columns. The synoptic representation is widespread in erudite tradition at least since the time of synoptic gospels, and scholars should therefore not be lost. With this mask they will be able to compare not only different versions of a text, but any kind of contributions. If in the first column of the synoptic mask we select, for example, a passage of an article published in the journal (format 6 of part A) containing the reference to a Nietzsche aphorism, the second column will automatically display the related aphorism extracting it from the critical edition (from format 2), while the third column will reproduce the genetic path (extracting from the genetic edition, format 3) containing all the preparatory jottings Nietzsche used to write it. If it happens that the first step in the genesis of the aphorism was the page

---

[11]Even when new media permit a different and more logical organization of content, at the beginning new media mimic the old ones: it is well known that the first printed books imitated manuscripts books, and the first CD-ROMs tried to reproduce the look and feel of printed books.

from an other author's work contained in Nietzsche's personal library, a column could display the facsimile of such a page containing, if present, the annotation Nietzsche wrote on it (from format 5). On the side of secondary sources, the user can choose to display in a column the text of other articles criticizing, praising or complementing that precise passage of the selected article (from format 6) and, finally, an additional column could list a bibliography of other articles written by the same author of by different authors of the same subject.

From a technical point of view, each format of part A can be a subpart of a unique website or an autonomous website hosted by a different server and created and managed by a different scholar or research team. And the synoptic view of the part B can collect contextual information coming from different websites for comparisons, e.g., different transcriptions of the same manuscript published by different critical editions produced by different teams.[12]

---

[12]To be able to communicate, part A and B should simply use a compatible scholarly ontology and a common communication protocol which can be a reduced and customized version of the Open Archives Initiative Protocol for Metadata Harvesting.