

Kim H.Veltman

## **Four Ways that Digital Communications are Transforming Scholarship Sources, Names, Claims and Scope**

Paper submitted to INET 2004, Barcelona

Written: Arlington and Maastricht, 2002.

---

- 
0. Introduction
  1. Sources
    - Authenticity
    - Criticism, Satire and Caricature
  2. Variant Names and Categories
    - a. Personal Knowledge
    - b. Future of the Book
    - c. Collaborative Knowledge
    - d. Enduring Knowledge
  3. Differing Claims
  4. Scope
  5. Why
  6. Conclusions
- 

### **0. Introduction**

The implications of Information and Communications Technology (ICT e.g. computers linked by Internet) are a matter of great debate. At one end of the spectrum are persons in cultural studies who argue that culture and society entail only human beings. Technology, they claim, is insignificant and to study its effects is only for technophiles. They deny McLuhan's insights that the medium is the message<sup>1</sup> and argue that "digital" is merely a fashion, a passing mode with no enduring significance. Then there are techno-optimists such as de Kerckhove,<sup>2</sup> who see only the happy consequences.

On the other end of the spectrum are techno-determinists, who claim that technology is changing everything. Even within this group there is a wide range of views with thinkers such as Negroponte,<sup>3</sup> Dertouzos<sup>4</sup> and Castells.<sup>5</sup> Networked computers, they tell us are bringing new modes of access, collaboration, and are transforming our notions of work, play, entertainment and life-long learning. Many of these claims are very broad with no precise explanation of precisely how ICT is changing methods.

This essay argues that there are four very specific and fundamental ways in which ICT is transforming the nature of scholarship, learning and education. First, they are changing the way we deal with sources, i.e. the way we deal with evidence in making claims. Second, they offer a means of transcending the enormous hurdles posed by variant names in traditional linear systems. Third, they provide new ways of representing differing and conflicting claims, one of the central problems of all scholarly discourse. Fourth, ICT is changing the scope of scholarship by providing access to an immense range of materials.

Text	Claims and Examples
Notes	References to Evidence Elsewhere
Bibliography	Further References to Evidence Elsewhere

Figure 1. Basic elements of an analog publication.

### 1) Sources

Learning in general and scholarship in particular might reasonably be defined as a quest for truth through making claims, which are carefully supported by evidence. In an analog publication (in an article or book), there is typically a main body of the text, which makes claims with examples that are then supported by notes and bibliography. Traditionally these notes do not cite the full evidence. Typically they are references to evidence in other publications or sources outside the body of the article or book in question. Similarly, bibliographies refer to further evidence outside the document in question. Thus paradoxically the nature of scholarship in printed publications is to refer to evidence, which is absent from the text itself (figure 1).

In the case of written materials, for instance, there is often a short quote from a document in the text of a book and the notes or references then refer to the full title and details of the document cited. In the case of visual materials, texts typically contained verbal descriptions thereof until the sixteenth century; engravings thereof from the seventeenth through nineteenth centuries and black white photographs from the last quarter of the nineteenth and through most of the twentieth centuries. Only in the last few years are colour photographs attaining a level of quality that they are considered as serious evidence by professional art historians.

Precisely how scholarly evidence is used depends, of course, on the kind of publication in question. And while we are not concerned with a full classification of the genre it is quite simple to identify some of the major classes (figure 2). In each of these cases, the same basic principle holds: the analog text makes claims for which the sources are referred to in the notes and bibliography, but are not themselves present.

Kind of Publication	Function
Article	Reports results of study, makes claims
Refereed Article	Reports results of study, makes claims certified by experts
Review Article	Reviews, surveys trends in a field
Book Article	Reports or reviews trend, theme via a number of authors
Textbook	Surveys main schools, claims in a field
Thesis, Dissertation	Makes, defends new claim(s)
Monograph	Reports on substantial research in a field
Fundamental Study	Establishes or redefines a field/discipline

Figure 2. Some basic kinds of scholarly publications and their function.

The earliest electronic publications typically used the concept of hyperlinks to link texts with notes and bibliographies: i.e. they merely adapted the characteristics of the earlier medium into the new medium just as Gutenberg and the pioneers of the printing press in Europe adopted the characteristics and limitations of manuscript layout into their new printed medium.

Potentially, however, electronic publications and digital versions transform the relation of claims to evidence by providing a direct link with sources or at least their facsimiles and digital surrogates. When a claim is made in a digital text one can hyperlink a reference to the original document or source. Equally well one can go back to a facsimile of the entire document or source or even to a live video image of the original. Hence, whereas analog scholarship traditionally refers to evidence and sources, which are absent, digital versions of scholarship refer to sources, which may be absent physically and yet are available in full.

This applies not only to texts but equally to pictures, and potentially to sound and other media. Hence, where a book version of a slide lecture typically provided only a small sample of the images it discussed, an electronic version of the same can include all the original images of the slide lecture and in addition might offer access to many other related images.

A concrete example helps to make this point more vividly. Giesecke,<sup>6</sup> a standard book on the history of printing, discusses in some detail Gutenberg's efforts and procedures in printing the first edition of the *Old and New Testaments* which history remembers as the *Gutenberg Bible*. He cites and illustrates specific examples of how Gutenberg adopted characteristics from the late mediaeval manuscript tradition in creating his publication. Reference is made to the illuminated (i.e. hand painted) initials that decorate the book but no coloured images of the *Gutenberg Bible* are actually provided.

If Giesecke were writing his fundamental study in digital form today he could go directly to at least four full text (and full image) versions of the *Gutenberg Bible* conserved at Cambridge, Göttingen, Keio University and London respectively.<sup>7</sup> A fifth copy is being scanned at the Library of Congress at 767 Megabytes per page. This is such a high resolution that a scholar can have access to details to which they are unlikely to be given access in the case of the original.

As a serious scholar Giesecke would have liked to include many other concrete examples from the *Gutenberg Bible* and from the many early books printed before 1501 (incunabula), which he considered in his standard work. But here the economic limitations of printing in analog form entered just as they did at the time of Gutenberg who, it is sobering to recall, went bankrupt in introducing what some have described as the most important advance in Western technology to this day.

In the case of digital versions there is no reason why these limitations should hold. The institutions which have scanned in the *Gutenberg Bible* have done so using public money and thus, appropriately these materials are being made available without cost. Hence,

whereas analog publications have economic limitations that determine the amount of material they can include, digital publications need not have such limitations. We noted that analog editions try to overcome these limitations by referring to evidence outside their covers. But when this evidence is in publications, which exist in a different location than the analog edition, then we are faced with an analog equivalent of broken links, references to things, which we cannot check up and certify for ourselves.

By contrast, digital editions potentially allow every claim to be linked with the evidence on which it is based. Indeed, in the vision of Tim Berners Lee, every link on the World Wide Web should lead not merely to a Uniform Resource Locator (URL) but to a Uniform or Universal Resource Indicator (URI). Each statement and claim should lead back to its source.

To continue with the earlier example this means that a digital version of Giesecke's book on the early history of printing can potentially be linked to full text copies of every book he cites. Moreover, the *Gutenberg Bible* can in turn be linked to facsimile images of the manuscript versions on which it is based and these in turn to the whole Biblical tradition of the Church Fathers. If this sounds unrealistic it should be recalled that the complete *Patrologia Graecae* and *Patrologia Latinae* were among the first major achievements of digital editions in the 1980s.<sup>8</sup>

To be sure there are many things that have not yet been translated into digital form, just as there were once many manuscripts not available in printed form. But that is changing. In 1630, the Herzog August Bibliothek, then one of the largest libraries of the world only had 130,000 books. In the spring of 2002 the Joint Systems Information Committee (JISC) in the United Kingdom made a licencing agreement for 120,000 full text books in digital form. In China there is a project to scan in 500,000 full text-books. Gradually all materials now available in analog form can be available in digital form.

When that happens the criteria for good scholarship will inevitably be transformed. In traditional analog publications, scholarship is judged on the basis of notes and bibliography, i.e. on the basis of how serious, careful and precise are the references to the evidence on which the claims in the analog publication are based. Herein lies an unstated assumption that if one does not agree with a scholar's interpretation or is uncertain about the accuracy thereof one is provided with the means to check the source of these claims.

In the case of digital publications one can foresee a day where a direct link to all the sources and evidence on which one's claims are based will be seen as an essential feature of scholarly activity and discourse. Only those who link us to their sources will be seen as real scholars. Referring to evidence, which is not accessible, will no longer be sufficient.

### **Authenticity**

Sceptics will object that linking to digital surrogates is more problematic than it seems because of the problem of authenticity: How can I know that the full text of a source which appears on my screen is in fact an accurate replica of the original? This is a serious

problem.<sup>9</sup> We should recall, however, that the problem applies equally to other media such as manuscript or print as Lorenzo Valla revealed during the Renaissance when he demonstrated that the so-called *Donation of Constantine* was in fact a forgery from a later century.<sup>10</sup> Not just the veracity of the image and text is important but also the veracity of its date.<sup>11</sup>

There is much discussion about difficulties in determining the authority and reliability of sources. If one simply makes a direct link to the original holder of the object or content as a criterion of scholarship, many aspects of this problem are resolved. For instance, if one wishes to discuss seriously Leonardo da Vinci's *Mona Lisa*, then it is expected that one links directly with an image thereof furnished by the Louvre where the painting is housed. To protect the integrity of its image the Louvre can add watermarking techniques to its digital surrogates to determine whether persons have tampered with an original image. In addition to image watermarking there is now also text watermarking.<sup>12</sup> Hereby, all other surrogates, which cannot reveal their exact lineage from the untampered original image made by the Louvre implicitly disqualify themselves. To achieve this requires accurate metadata.<sup>13</sup>

Thus while authenticity will always remain a concern, the process of linking claims with the evidence on which they are based helps to diminish the problem. Indeed in exceptional cases such as the *Mona Lisa* one can imagine taking this linking of claims and evidence yet one stage further such that there is a web-cam which continuously links viewers via the Internet with the original object and thus allows one to confront digital surrogates in the forms of colour images directly with live camera feeds of the image represented.

Scholarship in analog articles and books referred to evidence beyond itself which was often not accessible to a reader. Scholarship in digital publications can potentially link readers with images of every object concerning which claims are being made. This is a first fundamental change in scholarship brought about by ICT. It brings a new coupling of claims and evidence, a new framework for demonstrating and testing the truth of any claim.

### **Criticism, Satire, and Caricature**

Related to the problem of authenticity in general are problems of criticism, satire and caricature in particular. Today a photographer A not infrequently discovers that a person B has made a copy of their work and then added critical comments, satirical additions, caricatures or even complete distortions. Unless the two sites are carefully distinguished an outsider C could readily conclude that the distortions are part of A's work rather than B's commentary. This brings into play the question of moral rights.<sup>14</sup>

Our scholarly rule of direct links with sources would largely resolve such problems. If person A publishes a photograph, person B should be free to transform that photograph into whatever they wish on condition that they reveal: 1) that the source is A; that the changes are the responsibility of B. This prevents an outsider attributing B's changes to

A. While B is free to produce as much criticism, satire and caricature as they wish, if they go beyond the limits of decorum and propriety it reflects badly on B.

The same holds true in the case of politicians, governments, companies and multinationals. Those in power inevitably provoke envy, satire and criticism from those with less power. One response is censorship and oppression. However, in a free society anyone can criticize as long as they give their sources. Others can then check their claims (in this case, criticisms) against the original evidence (e.g. the home page of the politician or company etc.) and draw their own conclusions. If the critic goes too far, it is they in turn who invite criticism of themselves. The free system thus has an inherent equilibrating function.<sup>15</sup> As long as the distinction between source and critic is clear, then the risk of confusing the two is minimal. Given proper treatment of sources the grounds for many recent legal squabbles about criticism disappear.

## 2) Variant Names and Categories

With learning in general and with scholarship in particular, there is a basic distinction between information and knowledge. Information is about isolated words: knowledge is about organized words both in terms of a coherent structure (syntax) and a meaningful structure (semantics). The moment we speak of meanings of sentences and stories we enter into the realm of interpretation, which varies from one person to another and varies at the local, regional, national and international levels. This is a vast topic beyond the scope of this paper where we focus on a more basic step in the process of knowledge organisation.<sup>16</sup>

To organise any knowledge (irrespective of the size of the unit), we need a) to agree on basic names for persons and places and b) to find simple categories under which to organise the material. Unless we are carefully trained, we tend to be flexible (terminologists and librarians might say undisciplined) in our choice of these terms. This poses fundamental problems at the level of personal knowledge, collaborative knowledge and even regarding the enduring knowledge of memory institutions.

### Personal Knowledge

At the simplest level, variant names arise from different spellings. For instance, the mediaeval author of a standard textbook on optics, who was also the Archbishop of Canterbury is variously written as: 1) John Peckham; 2) John Pecham; 3) John Peccam; 4) Ioannes Peccamus; 5) Petsan; 6) Pisanus; 7) Giovanni Pisano Vescovo, 8) Ioannes Archiepiscopus Cantuariensis and 9) Monsignor Giovanni, Archivescovo Cantauriese.

Unless I am extremely well trained the natural tendency is to write the name one way today and another way the next time, especially since I can easily recognize most of the variants as being the same person. If, however, I ask a computer to find me exact matches of a name then it will not find near variants and if I ask it to find variants it is very unlikely to find variants which are based on translations into other languages, e.g. from

Latin into French or English. This general problem is made worse by the use of professional and personal nicknames.

In the past, electronic versions of text were mainly word files. This made even the finding of standard spellings difficult. The rise of Standardized General Markup Language (SGML) and more recently eXtensible Markup Language (XML) means that it is now much easier to create the equivalents of databases for most digital materials. These developments introduce two significant possibilities. First, they allow the creation of dictionaries, which provide both a generally accepted standard name and then list all the other variant names. This list can continue to grow and thus accommodate further equivalent names. Second, when I am writing a new text, this dictionary of names can act as a kind of pre-emptive spell checker, such that if I type in a variant, it offers a standardized version of the name.

The enormous difference between such a tool and typical spell checkers today is that it maintains an active record of the serious variants, which can then be used when I wish to search earlier materials prior to such a standardized procedure. In short, we gain in regularity without losing access to the richness of the past.

In future, this technology could be taken even further. In addition to typical spelling differences between British and American English (e.g. flavour vs. flavor; transport vs. transportation), there are many differences in language at the regional and local level. If these variants are captured as dictionaries in database form, a future version of today's spell checker would allow me to adapt any local term to a standard term at the regional, national and international level or conversely: i.e. I could even write my text in dialect and have it appear in the official form of the language or I could translate a formal text into any given dialect.<sup>17</sup>

Ultimately, the technology can go even further and allow me to write names with spelling variants, prompting me only when it is uncertain about the standard name intended. Essentially it would do the reverse of today's spell check software. Instead of reducing all variants to one standard, it would collect all variants and link them with one standard. The rationale for such effort only becomes apparent when we begin searching across large text corpora at different levels of knowledge (cf. enduring knowledge below). In the past, variations in my personal expression meant that only my standardized examples were likely to be found. Personal expression was opposed to standardisation. In the past, I would type in a standard spelling and receive only the direct matches for that spelling, and/or minor variants.

With the new approach personal expression and standardisation are no longer opposed. I can type in a term at one level (local, regional, national or international) and retrieve materials at all levels. I can also type in a national term and then limit retrieval to a particular region or locality and further limit this with chronological constraints: e.g. all cases between 1850 -1900. In future, this approach can be extended to multilingual contexts. Instead of having to search through dozens and potentially thousands of specialized etymological dictionaries at national, regional and local levels, one could do

so in a single operation. Needless to say, the advent of such search and retrieve techniques will again transform scholarship because they introduce a dimension of dynamic etymology into our studies.

The full potentials of such an approach become even more evident in attempting to organize knowledge. If I am organising material concerning automobiles, I might list it under *automobiles* one day; list it elsewhere another day under *autos* and list it a third time under *cars*. If that is all I do in life then it will be fairly simple to remember those three categories and where I put material on those categories.<sup>18</sup> If, however, I am studying new media and try to make a list of the basic categories for developments in the field, then that list soon expands to a hundred categories and if I continue to have three loose words for each of these categories, then I will have three hundred places under which to list things and it becomes a very difficult indeed to remember where I placed any given item. As the number of these categories and the size of these lists expands, then I inevitably reach a point where my attempts to organize knowledge fail. The power that comes from naming is cancelled by the confusion that arises from not having a standard term for any category.

As these lists become more complex they require hierarchical classes which are dealt with in classification systems. Successful use of such systems requires authority files (i.e. lists where one term serves as a preferred standard and the rest as variants). All this has traditionally been the domain of the librarian. One scenario would be to expect that every person should become a librarian, but this would hardly be realistic.

A more practical solution is to use the method outlined above of linking standard names with their variant names.<sup>19</sup> Every time I type in a variant name, the system classes it using standard names, which are subject headings. If the system makes an inappropriate suggestion then I can correct it manually using terms from the authority file. Adaptive learning algorithms can then be used to factor such corrections into future iterations of the system. As a result of this approach the natural human tendency to list a given idea under a series of different terms is no longer a hindrance in knowledge organization. Indeed, because the variants are now built into the system, the more variants I use, the greater are the number of paths by which I can retrieve my own knowledge and that of others. In the past, I had many scattered attempts to organise knowledge under different headings which were non-cumulative. In future such attempts will be cumulative.

The new technologies also promise to transform the way I organise materials within these categories. In the past, I would simply have a list in which I jotted down newspaper articles, books, journal articles and television documentaries indiscriminately. Trying to find back something specific in lists of hundreds or thousands of items then became a considerable challenge.

If such materials are now translated into an XML format, whereby their components can be placed in databases, i.e. indexed systematically, then there are new possibilities. The different materials can be listed in terms of media types, levels of knowledge and indexed geographically and chronologically. As a result I could effectively ask to see only the



(refereed) journal articles of the past decade. Or I could even have a search engine retrieve all books published in Germany these past five years.

Thus the confused fixed lists of the past are transformed into ordered virtual lists which are generated as subsets on demand. This kind of overview will also be a first step towards distinguishing between different levels of authority in knowledge materials: e.g. refereed journals are more serious than journals which in turn are usually more serious than magazine or newspaper articles. Such hierarchies can also help to distinguish between profession discussion groups, moderated lists and regular chat groups.

### **Future of the Book**

These developments will also transform the book as we now know it.<sup>20</sup> As noted earlier a traditional printed book is essentially three things: 1) claims and examples in the text itself; 2) references to sources of evidence in the notes and 3) further references in the bibliography (figure 1).

In our day, one of the basic problems of writing books on rapidly evolving topics such as computers, the genome or bio-informatics is that any examples are rapidly outdated. In future the text part of the printed book may become reduced to a series of claims. As they become available examples could then be updated dynamically in a digital appendage to the physical book. In the case of notes, direct quotes in the text could continue to appear in the notes section. Meanwhile, references to newspaper articles, journals and books would be accessible by medium, with both geographical and temporal filters. In keeping with our comments in section one, any references to sources would connect directly with those sources. Thus traditionally static books will become dynamic repositories of organised knowledge.

Today, when a scholar focuses on bibliographical aspects, a special category of printed books called a bibliography results. These lists are usually separate from more discursive, historical accounts of a subject. In future, there can be new links between such different kinds of publications. While I am reading a discursive history of perspective which mentions a name such as Leon Battista Alberti, I can click on the name and be taken to the bibliographical entry under this name in order to see which books he wrote, where extant manuscripts exist, what editions, translations etc. have been made. Similarly I could click on the same name and choose instead to call up biographical sketches or even complete biographies of Alberti. Indeed, it could link to all reference materials about Alberti.

Historians such Professor Robert Darnton have played with related ideas. He has a fantasy about an e-book in the form of a pyramid:

An "e-book," unlike a printed codex, can contain many layers, arranged in the shape of a pyramid. Readers can download the text and skim through the topmost layer, which would be written like an ordinary monograph. If it satisfies them, they can print it out, bind it (binding machines can now be attached to computers

and printers), and study it at their convenience in the form of a custom-made paperback. If they come upon something that especially interests them, they can click down a layer to a supplementary essay or appendix. They can continue deeper through the book, through bodies of documents, bibliography, historiography, iconography, background music, everything I can provide to give the fullest possible understanding of my subject. In the end, readers will make the subject theirs, because they will find their own paths through it, reading horizontally, vertically, or diagonally, wherever the electronic links may lead.<sup>21</sup>

The differences between our visions are simple. Darnton is interested in links at will. I am interested specifically in structured links to the recognized works about a person or a place. He is interested in a specific kind of e-book. I am interested in applying this approach to all publications and equipping even printed books with new augmented facilities whereby they too can link to distributed, electronic collections via virtual reference rooms. Darnton is concerned with a single new product: I am concerned how the traditional categories of books such as biography, bibliography and monograph will continue to have their unique functions but be enhanced in their usefulness by new sets of links.

### **Collaborative Knowledge**

In the previous section we outlined the problems which occur through having different words to describe the same object, concept or process. These problems increase dramatically in the context of collaborative knowledge for at least two reasons. First, when teams of dozens or hundreds of persons are working on the design of a new machine or building, these individuals will often have their own personal terms for objects and processes. Here the new kind of dictionary with its variant names can serve as a common thesaurus for bridging individual usage. Instead of forcing everyone in the team to learn someone else's vocabulary, the variant thesaurus would allow individuals to add as many personal usages of terms as possible without impeding co-operation on major projects.

A second problem arising in large collaborative projects is that individual companies frequently employ separate parts lists with different names for a given part. In international projects this problem is made more complex because the parts names are also in multiple languages. Here too a thesaurus with all the variant names can ensure that the collaborative process advances as planned without the tower of Babel like confusion arising from differing names for the same items.

### **Enduring Knowledge**

The problem of variant names for authors and variant terms for subject headings and classification systems is an age old problem in the realm of enduring knowledge, and particularly in the library world. Here various methods have been tried. Already at the time of the Library of Alexandria, there was a vision of collecting all known materials in a single place and uniting them through a single catalogue. This method was revived in

Method Used	Where Applied
1. Everything in one central place using a single method	Library of Alexandria, British Library, Bibliothèque Nationale
2. Everyone uses same template using same rules	Machine Readable Card (MARC) Dublin Core
3. Templates which map to variants in different templates	PROMETHEUS <sup>22</sup>
4. Formalized metadata walkthroughs	IMASS

Figure 4. Four approaches to library catalogues.

the 19<sup>th</sup> century with the vision of a universal library and led to great collections such as the British Library and the Bibliothèque Nationale. In the course of the twentieth century, it became increasingly clear that no single building, however impressive could hope to house the entirety of these collections. Even more problematic, within the same building there arose a variety of catalogues and cataloguing methods for different specialized collections: e.g. manuscripts, rare books, prints and drawings, maps, newspapers.

The 20<sup>th</sup> century saw the emergence of a second dream which, it was hoped, would solve the problems of standardisation: the idea of a single template. If everyone used the same template then everyone would have catalogues using precisely the same rules. An obvious candidate seemed to emerge in the form of the Machine Readable Cataloguing (MARC). Two factors hindered this view. First, a number of countries continued to pursue their own cataloguing rules. Hence, while the Anglo Saxon world developed the Anglo-American Cataloguing Rules (AACR), the Northern Germans continued with their Prussian Instructions. Second and partly because of this, the next decades saw a series of subtle shifts whereby a great number of national variants emerged. New Zealand's version was slightly different from Australia's AUSMARC, which in turn was different from USMARC, UKMARC etc.

Notwithstanding these obstacles, some would say failures, to establish a single standard, the vision lived on and more recently inspired the Dublin Core Metadata Initiative,<sup>23</sup> an attempt to find a lowest common denominator of fields which should be acceptable to everyone. Not unlike the MARC effort, this attempt overlooks the enormous impetus of existing legacy catalogues and collection management systems. Replacing these may sound attractive to those outside the institutions but is almost doomed to be rejected by owners of the collections.

Hence, at the outset of the 21<sup>st</sup> century, a third method is emerging. In order to have a unified approach one needs to maintain the quest for a single template at some centralised location. However, instead of trying to impose this blindly onto the entire range of institutions at the local, regional, national and international levels, one develops methods of linking this standardised template at a centralised location with the many variant templates distributed throughout the world. This approach was broached in the European Project REANRDUS and is being developed in the German project PROMETHEUS

(figure 4). Accordingly terms such as linking, mapping, bridging, and walkthroughs, have become increasingly central to discussions.

A next logical step is to recognize that these mappings are less obvious than one might have hoped: i.e. often there is no one to one correspondence between terms especially as one switches languages. Accordingly one creates a graphing structure whereby equivalences are weighted: e.g. term A has a 90% equivalence with term B. This leads to a mapping of distributed conceptual schemas compliant with modeled metadata standards via a set of mapping tables: In simple terms a set of formalized metadata walkthroughs.

This approach is being explored in European projects such as IMASS (figure 4).<sup>24</sup> Unlike earlier systems, this allows one to keep intact the diversity and complexity of local and regional variants without needing to rewrite the existing data-structures.

Meanwhile in America there is talk of dramatic developments in the Internet. In a first phase, the Arpanet dealt with packets. In a second phase, the Inter-net dealt with objects. Soon, we are told, there will be a third phase or wave, the Inter-space, which deals with concepts:

The Third Wave will bring the level of analysis, of correlation of knowledge. It will move past search of individual repositories, beyond federation across repositories, to analysis of diverse groups of information across sources and subjects....In this third wave, the Interspace, there will be distributed services to manipulate concepts across domains just as the ARPANET had distributed services to transfer files across machines and the Internet is having distributed services to transfer objects across repositories. The Interspace environment supports fundamental manipulation of concept spaces: indexing and retrieval, grouping and sharing.<sup>25</sup>

On the surface this third wave or phase might seem to be the same as the fourth approach just outlined above. It is important to recognize, however, that they are fundamentally different. The American vision of the future speaks of semantic mapping but assumes that concepts can be mapped to each other seamlessly. Accordingly no differences are recognized between national, regional and local uses. The model is essentially uni-lingual and uni-cultural.

By contrast the European approaches (especially 3 and 4) allow for a multi-lingual and multi-cultural approach whereby local, regional and national variants are maintained. Instead of using standards as an excuse to remove heterogeneity, these approaches use standards in technology to maintain heterogeneity in content: to preserve differences which are essential for cultural diversity. This linking and mapping in the European approach is achieved using the same principles as those outlined for personal knowledge, with one significant difference, namely, scale. In the case of personal knowledge the mapping process typically includes hundreds of terms. In the case of enduring knowledge from libraries, museums and archives, thousands and sometimes hundreds of thousands of terms are involved.

Kind of Knowledge	Institution	Kind of Classification	Kind of Subject Heading
Enduring	Large Library	Library of Congress	Library of Congress (LC)
Enduring	Smaller Library	Sears	Subset of LC
Collaborative	Company	Specialized Thesaurus	Sub-subset of LC
Personal	Individual	Personal Thesaurus	Sub-sub-subset of LC <sup>26</sup>

Figure 5. Relation of personal knowledge with other knowledge systems.

Hence, a centralised template contains one set of terms, which serves as a standard and records all variants such that it can effectively translate between its preferred term and those used elsewhere. The centralized headings at a large library thus represent a (super) set of which smaller libraries, companies and personal collections represent various levels of subsets (fig. 5).

Here the principle of variant names applies not only to containers (field names) but equally to the contents of fields via standardized subject headings such as those of the Library of Congress or subsets thereof. As a result catalogues at the local, regional and national levels can be integrated without being rewritten entirely. Through this process the complexity of the system becomes hidden from the typical user because they can continue to have the luxury of idiosyncratic terms and yet have all the privileges of a disciplined use of terminology. A new systematic treatment of the variant names of categories of knowledge thus marks a second fundamental change introduced by the shift from analog to digital.

### 3) Differing Claims

Traditionally an article or book developed one claim, while notes referred to possible alternative interpretations. Review articles and review monographs compared conflicting claims and theories. Nonetheless the static, linear nature of print meant that it was difficult to make visible the dynamic, non-linear nature of claims.

In digital form, an author's claim and other claims can be linked in the manner of standard and variant names. This introduces new ways of representing the history of claims and attributions and leads to new views of dynamic knowledge. Indeed when combined with the new approach to sources, this means that one can confront sources in new ways. For instance, if author A attributes a painting of the Leonardo school to Bernardino Luini and author B attributes the same painting to Andrea Solario, then one can call up all paintings attributed to Luini and Solario to reconsider the attribution.

In a traditional print publication there is one static list of paintings by Leonardo and his school. Another print publication usually has another version of such a list. A digital version permits us to create one master list, which serves as a standard and then include all alternatives as subsets as if they were variant names. By these means we are able to trace the history of attributions in new, dynamic ways.

Because every claim is still linked with the sources this approach does not throw us into a complete relativism as some of the constructivists and deconstructionists would have us believe. Indeed we would suggest that the complaints of the deconstructionists arose from a legitimate malaise with the limitations of print culture. It may be that their worries were more about the limitations of a given medium, than an accurate account of the epistemological limits of scholarship. If so then scholarship in the digital age shifts from being a single, static claim or attribution to comparative claims and attributions based on the combined evidence of the sources. This marks a third fundamental way in which the nature of scholarship is being transformed.

#### **4) Scope**

It is a truism that whereas the sciences focus on quantitative aspects, the humanities and scholarship in general focus on qualitative dimensions. Size and number are less important than quality. There are, however, cases where quantitative changes are on such a scale that they transform the context for qualitative insight and understanding. The enormous increase in scope of the sample on which one bases one's judgement is perhaps the most striking example in this context.

In the past this sample was almost inevitably based on the resources of the city where one lived. If one had access to a library such as the Herzog August Bibliothek there was much more that one could do than if one were limited to a typical library in a small town. In 1490 Leonardo da Vinci's collection of 119 books was considered a major collection. In 1630 the Herzog August Bibliothek was one of the largest libraries in the world with 130,000 books. In the 1960's the British (Museum) Library was one of the largest libraries of the world with over 10 million titles. Today the catalogue of the Research Libraries Group (RLG) which links major American libraries such as the New York Public, the Newberry and the Getty with at least 20 of Europe's largest libraries including the British Library, the Bibliothèque Nationale and the Vatican contains over 100 million unique titles and continues to add at least a million new titles annually. Future digital libraries projects are likely to be considerably larger in scope, as they move towards reflecting collections on a global scale.

While we can readily insist that not all books are equal, that some classics are more valuable than others, it would be difficult to claim that the insights possible from a study of 119 books are the same as those possible from a sample of 100 million books or more. Needless to say such a change in scale requires completely different methods. Reading every line of 100 million books would scarcely be possible. New search and summary methods, statistical overviews, possibly the intermediary of agents will be required. Ultimately this will lead to new levels of insight not possible on the basis of radically smaller samples of evidence. This new level of sample thus represents a fourth way in which ICT in the form of networked computers are transforming our approach to knowledge.

## 5. Why?

One could of course ask: Why go to such trouble? Are these not excessive measures, luxuries, which are not really needed in times when economic realities call for many cutbacks? On the contrary, we would suggest that these consequences of ICT should be seen as important new tools to help bring about a swing in the pendulum back in the direction of truth.

The notion of getting “back to the sources” (*ad fontes*) was an important catch-phrase connected with the rise of humanism in the 15<sup>th</sup> and 16<sup>th</sup> centuries indicating: “a scholarly commitment to ancient authorities, both secular and religious. In the religious realm, this meant that Renaissance and Reformation scholars worked to revive the direct, original-language study of the Bible and other key writings at the source of the Christian tradition.”<sup>27</sup>

Implicit in this goal was a commitment to getting to the source of things, to the bottom of things and hence the truth. This led to much work in the fields of exegesis, hermeneutics and the rise of interpretation as a systematic field of study. The early nineteenth century saw enormous advances in science and the whole of human knowledge. Coupled with the new wealth and success of the industrial revolution this led to an extreme optimism, which was epitomized by positivism:

a philosophical doctrine that denies any validity to speculation or metaphysics. Sometimes associated with empiricism, positivism maintains that metaphysical questions are unanswerable and that the only knowledge is scientific knowledge. The basic tenets of positivism are contained in an implicit form in the works of Francis Bacon, George Berkeley, and David Hume, but the term is specifically applied to the system of Auguste Comte, who developed the coherent doctrine. In addition to being a dominant theme of 19<sup>th</sup>-century philosophy, positivism has greatly influenced various trends of contemporary thought. Logical positivism is often considered a direct outgrowth of 19<sup>th</sup>-century positivism.<sup>28</sup>

The second half of the 19<sup>th</sup> century saw much new optimism through the unification of Italy and Germany, the Victorian age and the rise of Darwinism. By the second half of the nineteenth century this excessive optimism became tempered by a new awareness of complexity and a renewed quest to return to the sources (*ad fontes*) as a means of regaining a basis for serious claims to truth. For instance, in art history, Eitelberger von Edelberg, produced a series of *Quellenschriften für Kunstgeschichte* (Source Writings for the History of Art). Heiberg set out to produce a new critical edition of all Greek and Roman Mathematics. Loeb produced a new translation of all the Roman classics.

The great thrust towards objectivity in the West, epitomized by scientists such Hermann von Helmholtz, led in the 1890s to a reaction in Russia. Beginning in artistic and literary circles, this quest to reassert the importance of subjectivity, feelings and other non-objective human dimensions, led to constructivism in art and literature. In Russian the words for constructivism and post-modernism<sup>29</sup> are etymologically linked (*postronie*).

Via Prague, this school of thought spread to Paris where it became linked with psychological, psychiatric and psycho-analytical thought: e.g. Foucault, Lacan, Derrida. It also spread to America, notably Yale, Berkeley and later the University of California at Los Angeles (UCLA).

In light of these trends, the history of science became increasingly eclipsed by philosophy of science, social history of science, sociology of science<sup>30</sup> and cultural science. This brought numerous insights with respect to individual disciplines and especially with respect to the role of the observer in all things observed, thus challenging the simplistic subject-object distinction the history of which philosophers such as Cassirer sought to trace. At the same time, however, the pendulum which had swung too far in the direction of a simple, verifiable, certain truth in the nineteenth century has now swung so far in the opposite direction that there sometimes seems nothing left but vague contexts which swirl like fog around ever more uncertain claims, as if truth were outmoded, as if everything were open to debate, as if there were nothing but subjectivity.

At first sight, the advent of ICT only heightens this invisible crisis of scholarship. The estimated seven million new pages which come each day through the Internet bring such an avalanche of data, information and knowledge that no individual can possibly keep up. There are so many facts, there is so much context that one feels suspended between the sensation that everything has been done, said and the fear that there is no way to gain a serious overview of what has been said and done. Here the new tools for dealing with sources, names, claims and scope offer new horizons and renewed hope in the age old quest for genuine learning and understanding.

## **6. Conclusions**

The advent of ICT has many consequences. It is leading to digital libraries, digital archives, virtual museums and many other innovations. At a more basic level we have shown that there are four areas in which ICT is transforming the nature of scholarship as we now know it. First, it is changing the role of sources. In traditional print publications the text of books and articles made claims, the evidence for which was then referred to in notes and bibliography: i.e. they pointed to sources outside of the publication itself. By contrast, digital publications permit us to link every claim with the source on which it is based. Such a direct coupling of claim in the digital publication and evidence in the hyperlinked source might well become a new measure for serious scholarship. References are not enough. Certified digital versions of original sources are needed.

Second, ICT is transforming the way we deal with variant spellings and names for persons and objects as well as variant terms and concepts under which to class them. In manuscript and printed works, these variant names especially in the case of classes typically introduced confusion because we would forget under which variant we listed something. The creation of standardised lists with variants using ICT means that we can continue using variant terms and still have the advantages of authority files with standardised terms.



Third, ICT is transforming the way we deal with differing claims and attributions. Whereas print publications give us a static snapshot of one claim, or one list of attributions, digital versions permit us to present various earlier claims as subsets of a single list much like the method of standard names in an authority file linked with variant names. This dynamic approach provides new possibilities for understanding conflicting claims and evidence with respect to attributions.

Fourth, digital versions bring a radical increase in scale, which dramatically alters the scope of scholarship. Whereas early printed collections dealt with hundreds of items, and the largest libraries range from 10-15 million books, digital collections extend to over 100 million individual titles. If one includes all known copies then that figure exceeds 750 million titles as is the case with the OCLC database, which links the catalogues of 26,000 libraries throughout the world. ICT is not a passing fashion. ICT is transforming scholarship, learning and education through new approaches to sources, names, claims and scope. All this leads to augmented knowledge and culture.<sup>31</sup> While these new tools offer no simplistic formulae for truth, they do offer new approaches for truthfulness, which must remain a basic ingredient of any quest for learning in its deeper sense. Unlike the Bible which promised “ye shall know the truth, and the truth shall make you free” (John 8:32), we can at least hope to search for paths that lead in the right direction.

Arlington and Maastricht 17-24 June 2002.

### **Acknowledgements**

I am grateful to my colleague, Dr. John Becker, for kindly reading the manuscript and suggesting improvements specifically with respect to the section on Why? My colleague Drs Johan van de Walle kindly suggested developing the points connected with figure 4. I am grateful to Dr. Chris Zielinski for carefully reading the paper and proposing a number of points which are cited in the notes.

---

## Notes

<sup>1</sup> Marshall McLuhan, *The Gutenberg Galaxy. The Making of Typographic Man*, Toronto: University of Toronto Press, 1962.

<sup>2</sup> Derrick de Kerckhove, *Gekoppelde intelligentie. De opkomst van de WEB-maatschappij*, Hoolberg: Stichting Maatschappij en Onderneming, 1996. English Translation: *Connected Intelligence, The Arrival of the Web Society*, Toronto: Somerville Books, 1997.

<sup>3</sup> Nicholas Negroponte, *Being Digital*, New York: Knopf, 1995.

<sup>4</sup> Michael Dertouzos, *What Will Be, How the New World of Information will Change our Lives*, San Francisco: Harper Edge, 1997.

<sup>5</sup> Manuel Castells, *The Information Age: Economy Society and Culture*. Vol 1: *Rise of the Network Society*, Malden, Mass: Blackwell Publishers, 1996; vol. 2: *The Power of Identity*; vol. 3: *End of Millennium*, 1998.

<sup>6</sup> Michael Giesecke, *Der Buchdruck in der frühen Neuzeit. Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien*, Frankfurt am Main: Suhrkamp, 1991.

<sup>7</sup> See: <http://www.humi.keio.ac.jp/treasures/incunabula/B42-web/b42/html/index01.html>.

Cf. <http://www.gutenbergdigital.de/gudi/eframes/index.htm>.

<sup>8</sup> Cf. <http://www.ccel.org/fathers2/>.

See: [http://www.brepols.net/publishers/corpus\\_christianorum1.htm](http://www.brepols.net/publishers/corpus_christianorum1.htm).

<sup>9</sup> The science fiction writer, Michael Crichton in his Remarks to the Commonwealth Club, San Francisco, September 15, 2003, noted:

“I have been asked to talk about what I consider the most important challenge facing mankind, and I have a fundamental answer. The greatest challenge facing mankind is the challenge of distinguishing reality from fantasy, truth from propaganda. Perceiving the truth has always been a challenge to mankind, but in the information age (or as I think of it, the disinformation age) it takes on a special urgency and importance.”

See: [http://www.crichton-official.com/speeches/speeches\\_quote05.html](http://www.crichton-official.com/speeches/speeches_quote05.html)

This reference was kindly provided by my friend Dr. Eric McLuhan.

<sup>10</sup> The Donation of Constantine and the critique of Lorenzo Valla.

See: [http://www.tertullian.org/rpearse/donation/donation\\_of\\_constantine.htm](http://www.tertullian.org/rpearse/donation/donation_of_constantine.htm)

Cf. Medieval Sourcebook: The Donation of Constantine (c.750-800).

See: <http://www.fordham.edu/halsall/source/donatconst.html>.

<sup>11</sup> Chris Zielinski has noted: “Where the prior art only exists in digital form, without any analogue version (this is the case with millions of “pages” of human genome data, for example), there is plenty of scope for tampering with the content, and (often more important in prior art) the date at which it was created. I think text watermarking is the solution, but first the problem has to be recognised.”

<sup>12</sup> Chris Zielinski has kindly drawn attention to this innovation: “thanks to one of the IMPRIMATUR partners) the possibility of watermarking text – text in Word, Acrobat and HTML/XML formats in the first instance. This is not a steganographic technique such as those explored by Xerox, AT&T and IBM (coding hidden information by varying the size of the full stop, or of spaces between letters, for example), which were very vulnerable to format changes, but something quite different and patentable. It is unique,

---

and is not in public circulation yet. Clearly this has application to authenticity – you don't trust a text if the watermark is broken." The Lab of Professor Benoit Macq, Université Catholique de Louvain, is working on analogous techniques.

<sup>13</sup> As Chris Zielinski has noted: "Metadata clearly have implications for authenticity, and there may be moral rights aspects involved as well. With the prohibitions on decryption of technical measures to protect copyright in the European Copyright Directive and WIPO Copyright Treaty, the question of technical measures to protect metadata arise. For example, by changing the rights information contained in metadata you can steal the content. Since metadata may be stored separately (geographically, or in a separate file), you may be allowed under these directives to do just that. This illustrates the role of metadata in authenticity questions: to guarantee the authenticity of digital content, you will have to guarantee the authenticity of associated metadata as well."

<sup>14</sup> Chris Zielinski, "Academic Authors and Moral Rights,"xx.

<sup>15</sup> In the physical world it has proven necessary to create a Better business Bureau to check companies with suspicious credentials. Undoubtedly such a mechanism will need to be incorporated into the Internet especially with what ICANN calls sponsored candidates in the .com field etc.

<sup>16</sup> "Challenges for a Semantic Web," *Semantic Web Workshop 2002. Proceedings of the International Workshop on the Semantic Web 2002 (at the Eleventh International World Wide Web Conference), Hawaii, May 7, 2002, Honolulu, Hawaii*, pp. 16-22. Position paper also published electronically.

See: <http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Position/veltmann.pdf>. Reprinted in Cultivate Interactive, Issue 7, June 2002.

See: <http://www.cultivate-int.org/issue7/>). A more detailed discussion is found in the author's "Syntactic and Semantic Interoperability, New Approaches to Knowledge and the Semantic Web," *New Review of Information Networking*, Cambridge: Taylor Graham, Volume 7, 2001, pp. 159-183.

<sup>17</sup> Such name dictionaries obviously need to be personalized. If I am studying developments in Germany in the eighteenth century I would not usually need the names of persons in many other countries and at different periods.

<sup>18</sup> Even so if the lists are long I will often need to go through all three lists when I am searching for something.

<sup>19</sup> Chris Zielinski has kindly noted: "There is an excellent implementation of this in Collexis software, which is vector-based (rather than Bayesian) conceptual software that makes use of existing thesauri to create "conceptual fingerprints" of texts [unfortunately named, since a Collexis "fingerprint" is a weighted index listing of concepts in a text and has absolutely nothing to do with watermarking, or "fingerprinting" as it is often known]...I have used Collexis in consultancies, including one on creating an ontology/taxonomy for "health policy and systems research" at WHO."

See: <http://www.collexis.com/>

<sup>20</sup> This is a topic on which there are, of course, many opinions. Some have predicted the end of the book. Cf. Frank Ogden, *The Last Book You'll Ever Read and Other Lessons from the Future*, Toronto: Macfarlane Walter & Ross, 1993. 212 p. + 1 computer disk. Many assume the rise of the electronic or e-books. Elsewhere we have explored the possibility of augmented books whereby such e-books complement the features of traditional printed books. See the author's *Augmented Knowledge and Culture*.

---

<sup>21</sup> Robert Darnton, "A Historian of Books, Lost and Found in Cyberspace" *Chronicle of Higher Education*, March 12, 1999.

See: <http://www-sul.stanford.edu/siliconbase/darnton.html>.

<sup>22</sup> See: <http://www.prometheus-bildarchiv.de/>.

<sup>23</sup> A more significant attempt at industrial-strength generic metadata schemes is found in the <indec> project ([www.indec.org](http://www.indec.org)). <indec>, a post-IMPRIMATUR activity, "which brought together a number of IMPRIMATUR partners with quite a broad group (including some of the key brains behind the DOI and publishing applications of EDI) to formulate a generic metadata model and methodology: The <indec> model which we tested against Dublin Core, MARC, MPEG, DOI and a number of other naming efforts and found that they all were consistent with the <indec> standard." Note courtesy of Chris Zielinski

<sup>24</sup> See: <http://www.i-massweb.org/public/workpackages/project%20presentation.doc>

<sup>25</sup> See: <http://www.canis.uiuc.edu>. Bruce R. Schatz: "The Interspace: Concept Navigation Across Distributed Communities." *IEEE Computer* 35(1): 54-62 (2002). I am grateful to Mr Knud Lonsted, EU, for drawing this to my attention.

Cf. <http://citeseer.nj.nec.com/chung99semantic.html>.

<sup>26</sup> With alternative Terms.

<sup>27</sup> Ad Fontes website.

See: <http://www.ad-fontes.com/aboutus.asp>

<sup>28</sup> Encyclopedia.com.

See: <http://www.encyclopedia.com/html/p1/positivi.asp>, which refers to Kolakowski, *The Alienation of Reason* (tr. 1968) and *Positivist Philosophy* (tr. 1972); C. Bryant, *Positivism in Social Theory and Research* (1985).

<sup>29</sup> Cf. F. L. Jackson, Post-Modernism and the Recovery of the Philosophical tradition.

See: <http://www.mun.ca/animus/1996vol1/jackson.htm>

<sup>30</sup> Here the major thinkers of the day include Steve Woolgar.

See: <http://www.brunel.ac.uk/research/virtsoc/people/woolgar.htm>

and Bruno Latour. Cf. T. Hugh Crawford, An Interview with Bruno Latour, *Configurations*, 1993, 1.2:247-268

See: <http://muse.jhu.edu/demo/configurations/1.2crawford.html>

<sup>31</sup> See the author's, *Understanding New Media: Augmented Knowledge and Culture*.