

Kim H. Veltman

“Framework for Long-term Digital Preservation from Political and Scientific Viewpoints,”
Digitale Langzeitarchivierung. Strategien und Praxis europäischer Kooperation, Deutschen Nationalbibliothek, anlässlich der EU-Ratspräsidentschaft Deutschlands, 20-21. April 2007, Frankfurt: National Bibliothek, 2007.

Abstract

The range, depths and limits of what we know depend on the media with which we attempt to record our knowledge. This essay begins with a brief review of developments in a) media: stone, manuscripts, books and digital media, to trace how collections of recorded knowledge expanded to 235,000 in 1837 and have expanded to over 100 million unique titles in a single database including over 1 billion individual listings in 2007. The advent of digital media has brought full text scanning and electronic networks, which enable us to consult digital books and images from our office, home or potentially even with our cell phones. These magnificent developments raise a number of concerns and new challenges.

An historical survey of major projects that changed the world reveals that they have taken from one to eight centuries. This helps explain why commercial offerings, which offer useful, and even profitable short-term solutions often undermine a long-term vision. New technologies have the potential to transform our approach to knowledge, but require a vision of a systematic new approach to knowledge. This paper outlines four ingredients for such a vision in the European context. First, the scope of European observatories should be expanded to inform memory institutions of latest technological developments. Second, the quest for a European Digital Library should be expanded to include a distributed repository, a digital reference room and a virtual agora, whereby memory institutions will be linked with current research. Third, there is need for an institute on Knowledge Organization that takes up anew Otlet’s vision, and the pioneering efforts of the Mundaneum (Brussels) and the Bridge (Berlin). Fourth, we need to explore requirements for a Universal Digital Library, which works with countries around the world rather than simply imposing on them an external system. Here, the efforts of the proposed European University of Culture could be useful. Ultimately we need new systems, which open research into multiple ways of knowing, multiple “knowledges”. In the past, we went to libraries to study the recorded world. In a world where cameras and sensors are omnipresent we have new recording worlds. In future, we may also use these recording worlds to study the riches of libraries.

Universal Digital Library

For the first time in history, all the significant literary, artistic, and scientific works of mankind can be digitally preserved and made freely available, in every corner of the world, for our education, study, and appreciation and that of all our future generations.

Up until now, the transmission of our cultural heritage has depended on limited numbers of copies in fragile media. The fires of Alexandria irrevocably severed our access to any of the works of the ancients. In a thousand years, only a few of the paper documents we have today will survive the ravages of deterioration, loss, and outright destruction. With no more than 10 million unique book and document editions before the year 1900, and perhaps 100 million since the beginning of recorded history, the task of preservation is much larger. With new digital technology, though, this task is within the reach of a single concerted effort for the public good, and this effort can be distributed to libraries, museums, and other groups in all countries.

Indian Institute of Technology, 2007¹

Table of Contents

1. Introduction
 2. Media
 - 2.1. Tablets and Stones
 - 2.2. Manuscripts
 - 2.3. Books
 - 2.3.i From Private Collections to National Repositories
 - 2.3.ii University and Research Collections
 - 2.3.iii Germany as an Exception
 3. Tools for Access
 - 3.1. Secondary Literature and Universal Bibliographic Control
 4. Electronic Media and Networks
 - 4.1. Titles
 - 4.1.i North America
 - 4.1.ii Europe
 - 4.2. Full-Text Scanning
 - 4.2.1 Memory Institutions
 - 4.3. Digital Reference
 5. Challenges and Concerns
 - 5.1. Privatization of Knowledge
 - 5.2. Access to Content
 - 5.3. Citation Indexes
 - 5.4. Systematic Access to Content
 - 5.4.1. Multilingualism
 - 5.5. Born Digital
 - 5.6. Long Term Needs and Short Term Funding
 6. Ingredients of a New Vision
 - 6.1. Awareness and Integration of Emerging Technologies
 - 6.2. Distributed European Electronic Resource (DEER)
 - 6.2.1. Distributed Repository
 - 6.2.1.i Integrated Authority Files
 - 6.2.1.ii Integrated Full Contents
 - 6.2.2 Digital Reference Rooms
 - 6.2.2.i. Levels of Knowledge
 - 6.2.2.ii Worlds of Knowledge and Scales of Knowledge
 - 6.2.3 Virtual Agora
 - 6.3. Knowledge Organization Institute
 - 6.4. Recording Worlds
 - 6.5. Towards a WONDER
 - 6.5.1 European University for Culture
 7. Conclusions
-
- Appendices
 1. Examples of the Range of Major Collections
 2. List of Image Collections available in the California Digital Library
 3. UCSF Subject Headings
 4. Copyright
 5. Rob Bruijnzeels: Libraries 2040: The Netherlands Public Library Association
Notes

1. Introduction

Elephants may have long memories, but only humans have memory institutions which ensure a cumulative memory beyond the lifetimes of individuals. This quest has a grand tradition. In the Ancient world, there was the library at Pergamon and the most famous attempt to collect the whole of human memory in a single institution was the Museion, remembered as the Library of Alexandria, which first burned down c. 48 B.C.² If the claims that it contained five million items including 500,000 scrolls be true, then it took humans nearly two millennia to recover.

This human quest for collective, cumulative memory entails two fundamental principles: long-term preservation and communication. The evidence of the past millennia reveals a curious tension. The most secure media are not the best for communication. To appreciate this requires a brief review of developments in a) media: stone, manuscripts, books, digital media; b) tools for access and c) electronic networks. This will lead to a survey of concerns and a new vision.

2. Media

McLuhan suggestively claimed that the medium is the message. The medium in which we store knowledge is, in fact, much more than a storage method. It determines the limits of what can be accessed; it shapes our definitions and visions of knowledge. The shift from clay and stone to manuscripts, books, and electronic media is thus much more than a quest for “light reading”. It is implicitly a story of the progress of knowledge.

2.1. Tablets and Stones

Long before the hard disk, the Sumerians created cuneiform tablets some 5,000 years ago.³ Famous examples such as the Code of Hammurabi (1760 BC)⁴; the Achaemenian Treasure Inventory (Ganjnameh, 6th c. B.C.) or the Rosetta Stone (196 BC)⁵ confirm that this medium stands the test of time. They also confirm the vital importance of multilingualism. The treasure inventory is in Old Persian, Elamite and Babylonian.⁶ The Rosetta Stone is in two Egyptian language scripts (hieroglyphic and demotic) and classical Greek. Without these early examples of multiple languages, Champollion could not have re-opened our understanding of Egyptian culture. In terms of policy, this points to the dangers of assuming that one language could ever be sufficient for our databases.

China was particularly fascinated in this process of what they called stone books (*shishu*). This inspired two of the most impressive projects ever in long-term preservation. One was the so-called Forest of Stele, at Xian, a collection of 1700 tablets that began c. 1070 A.D. and still remains today.⁷ The second project set out to record in stone the (Great Buddhist Scriptures (Dazangjing). The project began in the Sui Dynasty (581-618) and concluded c. 1644. The good news was that the project was successful. It resulted in a series of more than 14,000 tablets⁸ that are likely to survive for at least the next couple of millennia. The less good news was that the project took c. 850 years; a time scale that goes well beyond contemporary funding agencies. One lesson might be that the quest for true long-term preservation requires more patience and longer term planning than we had suspected.

The cliffs of the Treasure Inventory in Persia were solid, but they could not be moved. Even the Chinese stele and tablets did not permit easy transportation. In retrospect, we see that the

long-term advantages of stone were outweighed by three disadvantages: a) time of production; b) the sheer weight of the product and c) limitation to a single colour, namely, the colour of the stone used. The next major medium of civilization improved on all three of these shortcomings, but at the price of being more fragile.

2.2. Manuscripts

The complex history of how manuscripts gradually evolved from papyrus, to parchment and then to mediaeval forms of paper has been well documented. Even a fleeting glance at a mediaeval example such as the Vienna Dioscorides⁹ reveals the new worlds of colours that were introduced by the new medium and why it relegated earlier methods literally to the stone age. Until the invention of printing, manuscripts became the obvious medium for communication of knowledge.

The paradox of manuscripts is that they did not stop with the advent of printing in the West in 1454 (cf. 2.3. below). King Matthias Corvinus was not enthusiastic about printing and commissioned humanists to make manuscript copies of key texts. Hence, some of the most lavish manuscripts ever were produced after the invention of printing.¹⁰ The Trivulziano Library in Milan confirms that these Corvini were not an isolated case. Even in the early 17th century, John Donne (1571-1631) wrote: “What printing presses yield we think good store / But what is writ by hand we reverence more.”¹¹

If the production of manuscripts decreased after 1454, their systematic collection increased. When Nicholas V formally founded the Vatican Library¹², in 1451, it had some 1,160 books.¹³ “At the end of the 15th century, more than three thousand manuscripts (Latin, Greek and Eastern), were added and placed in an orderly way in four rooms.”¹⁴ Today, the Vatican library has over 75,000 manuscripts; the State Library in Berlin has 18,300 occidental manuscripts and 40,000 oriental manuscripts¹⁵ and the Bibliothèque Nationale de la France has 250,000 manuscripts.¹⁶ The British Library has 314,914 Western manuscripts and also administers the collection of the India Office with a further 392,000 manuscripts.¹⁷ The Library of Congress, in its manuscript collection, which counts pages rather than documents, has “nearly forty million items contained in ten thousand separate collections, include some of the greatest manuscript treasures of American history and culture.”¹⁸ The University of Chicago claims to have 30 million manuscripts and archival pieces.¹⁹ As a result, the 20th century has collections of manuscripts, which are far richer and comprehensive²⁰ than the mediaeval or Renaissance periods when they were produced. In terms of policy for long-term preservation, this introduces another important point: the advent of a new medium does not render obsolete the preservation of earlier media.

2.3 . Books

The invention of printing in Korea just after the year 800 A.D. heralded a further crucial chapter in the development of communication. It reached China soon afterward such that the oldest extant printed book, the Diamond Sutra,²¹ (868 A.D.), comes from China. Printing was also used in China to publish their laws. As Michael Giesecke²² has shown in his fundamental study on the history of printing, the great innovation of Johannes Gutenberg, lay not in a new technology, but rather in the idea of using printing for the sharing of knowledge and for the common good. Hence, the “re-invention” of printing in Germany was really about discovering its potentials for public use, in the very same decade (1450-1460) that Pope Nicholas V was redefining the Vatican into what became the oldest public library in Europe.

The results were slow. Gutenberg went bankrupt. It took 80 years before scientific publications could deal seriously with complex illustrations. It took roughly 200 years (1450-1650) until the major publications were mainly in print form. And just as this was being achieved there was a redefinition in the scope of printing. Until the mid-17th century, printing had focussed on treatises. Correspondence among scholars had continued in handwritten form. The advent of the *Journal des Savants* and subsequently the *Göttinger Gelehrten Anzeigen* meant that correspondence now ended up in print. This time it took only a century for the idea to become established. By 1750, what we now call secondary literature (journals, serials etc.) had become established as a basic category of scholarly knowledge.

This transformed both the nature and size of memory institutions.²³ By 1630, the Herzog August Bibliothek (Wolfenbüttel), with 130,000 books was reputed to be the largest in the world. That honour passed to the Bibliothèque du Roi (Paris) in the 18th century and to the British Museum Library in the 19th century, which had 235,000 books in 1837.

2.3.i From Private Collections to National Repositories

The next fundamental advance was a shift in mind-set rather than a breakthrough in technology. The idea of collections shifted towards public collections and then became linked with the idea of a National Library with a right to legal deposit. Since 1661, the Swedish Royal Library has been entitled to a copy of all works published in Sweden.²⁴ The Bodleian Library at Oxford claims that they had the notion as early as 1610. At the beginning of the 20th century their librarian:

calculated that the purchase cost of materials which the Bodleian received by legal deposit in 1900 amounted to about £2,000. Nowadays, more than 100 years later, the Bodleian is receiving something in the region of 70,000 monographs and 100,000 periodical parts annually via legal deposit, altogether amounting to a value probably approaching £5 million.²⁵

A crucial step in this transformation came from an unexpected source: an Italian nationalist, turned British citizen, Antonio Panizzi. During his tenure as Keeper of Printed Books at the British Museum Library (1837-1856) “its holdings increased from 235,000 to 540,000 volumes, making it at the time, the largest library in the world.”²⁶ By 1911, the British Museum was officially a national library and a legal deposit for books published in Great Britain. National libraries became the fashion. Today three of the five greatest libraries in the world are outside Europe.

Library	Books	Items
Library of Congress ²⁷	17 million	130 million
Russian State Library ²⁸	16.5 “ ”	41 “ ”
British Library ²⁹	13 “ ”	150 “ ”
Bibliothèque de la France	12 “ ”	
Beijing National Library	24 “ ”	

	82.5 million	

Figure 1. Estimates of contents of five of the world’s largest libraries.

Harvard ³⁰	15.5 million
University of Toronto ³¹	15
Yale ³²	12.1
University Illinois (UIUC) ³³	10

	52.6 million

Figure 2. Estimates of contents of four of the largest university libraries in North America

In the process, libraries expanded far beyond book collections to include manuscripts, maps, photographs, images, music, film (cf. Appendix 1), effectively pointing the way towards contemporary notions of memory institutions. Major libraries such as Göttingen in addition to books, provide access to 300 online and CD-ROM databases. Today, the Library of Congress, which claims to be the world's largest library has 17 million books.³⁴ Meanwhile, the National Library in Beijing, which claims to be the world's fifth largest, has 24 million books (figure 1).³⁵

2.3.ii University and Research Collections

From 1850-1950, national libraries were the focus points for the greatest collections worldwide. Since 1950 there have been two important shifts, especially in North America. First, the New York Public Library (NYPL) has developed into a unique structure with 20 million books in 3000 languages.³⁶ Secondly, a handful of the university libraries grew to proportions that surpass most national libraries in terms of collections (figure 2). This trend is spreading and some collections continue to grow by about a million books each year.³⁷ For instance, Chicago which had 7 million in 2005 will have 10.5 million by 2009.³⁸ These developments, significant in themselves, gained in weight when, in December 2004, five of these collections announced that they would work together with Google³⁹ (namely, The New York Public, Harvard, Stanford,⁴⁰ Michigan,⁴¹ Oxford). These five collections together represent 56.9 million books,⁴² even if the announcement was that they would scan the full texts of "only" 15 million in the near term.

2.3.iii Germany as an Exception

One striking example to these transformations of the knowledge landscape was Germany with its historical traditions, whereby education and many powers remained at the state level (the *Länder*). This had three consequences. First, libraries continued to see themselves as local collections. Second, there was a reticence even to share information with institutions in neighbouring states. Third, there was no national library as such. The problems this posed for Germany's position in international scholarship were brought into focus in a lucid study by Bernhard Fabian (1983).⁴³ Since then there have been a number of fundamental advances.

First, a German National Library (Deutsche Nationalbibliothek) was established in 1990 although its present name, Die Deutsche Bibliothek, was only introduced in 2006.⁴⁴ Second, a vision arose of virtual National Library,⁴⁵ through networked co-operation of the libraries in Munich, Wolfenbüttel, Frankfurt, Göttingen and Berlin. Third, as this conference confirms, Germany began to explore its role in international networks. Fourth, the size and scope of the greatest libraries has expanded dramatically (figure 3). A comparison with the global scale (figures 1 and 3) reveals that six of the most important German libraries combined are comparable in size to the combined holdings of four of the greatest libraries of the world. On the other hand, if we compare the university scenes in Germany and the US, we see that the

National Library ⁴⁶	Leipzig	13 million
	Frankfurt	8
Berlin State Library ⁴⁷		10
Bavarian State ⁴⁸		8.8
FU Berlin ⁴⁹		8
Göttingen ⁵⁰		4.5
Wolfenbüttel ⁵¹		1.1

		53.4 million

Figure 3. Basic statistics about Germany's libraries.

largest American academic collection (Harvard) is nearly double that of Germany's largest academic collection (Freie Universität, Berlin). In this sense, the serious concerns of Bernhard Fabian (1983) have not yet been resolved.

These numbers provide a first glimpse of the scale of developments. From the advent of printing in 1454 to 1837, the largest collections rose from a few thousand to 235,000. New technology took nearly four centuries to change the size of collections from a few thousand to a quarter million. Since the mid-19th century, changes in policy increased the size of the largest collection 68-fold. Today, the largest single collection has over 17 million books and this is but the tip of the iceberg. A recent study estimated that Russia alone now has 1 billion books⁵² in its public collections. Size matters. Even so, to understand the revolution that is underway, we need to look also at two other dimensions: a) transformations in tools for access to recorded knowledge and b) networks.

3. Tools for Access

In theory, the advent of printing implied a challenge of translating handwritten knowledge into printed form. In practice, what happened was rather different. First, as Giesecke has shown, there was a new formalization of oral knowledge. Second, as McLuhan, Ong, Yates⁵³ and others have noted, there was a new fascination with the organization and presentation of knowledge. The mediaeval Tree of Porphyry blossomed into a forest of lists for displaying categories and branches of knowledge, leading to catalogues, classification schemes, taxonomies and gradually into fields such as library science, information science and knowledge organization. In 1876, for instance, Melvil Dewey, developed the Dewey Classification System, a pragmatic solution that has since spread around the world.⁵⁴ A method for organizing collections of books by the publisher, Prosper Marchand, led to the *Système de Paris*, which was adapted and evolved into the Library of Congress Classification (1897).⁵⁵ But the most dramatic changes began rather suddenly in the last five years of the 19th century (1895-1900).

3.1 Secondary Literature and Universal Bibliographic Control

As noted earlier, another of the unexpected consequences of printing was to translate written correspondence into printed form, which led to the emergence of secondary literature: articles in journals discussing and commenting upon original texts (primary literature). Germany made pioneering contributions in this domain. In 1896, it launched the *Internationale Bibliographie der Zeitschriftenliteratur*.⁵⁶ At the turn of the 20th century it played a significant role in the idea of a global brain (*Gehirn der Welt*).⁵⁷ This idea broached by Naumann (1907), was developed by Wilhelm Ostwald and led to the foundation of The Bridge, the International

Institute for the Organization of Intellectual Work (1911), which included many famous international members: Andrew Carnegie, Marie Curie, Selma Lagerloef, Henri Poincaré, Ernest Solvay, Bertha von Suttner and Paul Otlet who became honorary president. In 1948, the Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI), was founded. In 1971, the Committee on Conceptual and Terminological Analysis (COCTA)⁵⁸ was recognized as Research Committee number one of the International Political Science Association at its World Congress in Munich.⁵⁹ In the same decades, Ingetraut Dahlberg founded a) the journal *International Classification* (1974), which later became *Knowledge Organisation* (1993); b) the Gesellschaft für Klassifikation e.V. (1977) and c) the International Society of Knowledge Organization (ISKO, 1989). Since then a series of classification societies have emerged in a number of countries.⁶⁰ Berlin founded a Deutsche Bibliotheksinstitut (1978), which was then closed in 1999.⁶¹ Germany has led the world in the development of technical subject libraries (*Fachbibliotheken*); which has led to *Die Virtuelle Fachbibliothek*⁶² (2001) and more recently Vascoda.⁶³ It remains active in projects.⁶⁴

Meanwhile, one of the most significant innovations again came from an unexpected quarter. Two Belgian lawyers, Paul Otlet and Henri La Fontaine⁶⁵ launched efforts towards universal bibliographic control. In 1895, Otlet wrote to Melvil Dewey requesting permission to develop his classification system. Permission was granted so they founded a) Museum of the Book,⁶⁶ b) Union of International Associations (UIA) and c) International Institute of Bibliography (Brussels), which entailed a Répertoire Bibliographique Universel (RBU) and led to the Mundaneum.⁶⁷ Between 1904-1907, they launched the Universal Decimal Classification (UDC).⁶⁸ Their RBU grew to nearly 16 million entries by 1930⁶⁹ and became the International Federation for Documentation (FID, 1937), which moved to The Hague, where it evolved into the International Federation for Documentation and Information (FID)⁷⁰ and remained as the curator of UDC until the FID was dissolved in 2003. Today UDC continues as a small consortium without proper funding.

What distinguished these efforts at the turn of the 20th century was a comprehensive plan. The Mundaneum in Brussels served as a think-tank and co-ordination centre. The theoretical aspects of terminology moved to Vienna to become linked with ISO, and later Termnet, Infoterm etc. The practical application of Universal Decimal Classification shifted to the FID in the Hague, linked with the International Federation of Library Associations (IFLA) at the Koninklijke Bibliotheek. Meanwhile, the UIA in Brussels served as a global observatory for new trends, concepts, problems and words, which required more systematic treatment. What made these projects remarkable, was that they were clearly linked to a much ampler vision. As Anthony Judge and others have noted, Otlet's work was both visionary and prophetic. A full decade before Vannevar Bush wrote "As we may Think"⁷¹, Otlet (1935) foresaw the future development of networked communication and described it lucidly:

Man would no longer need documentation if he were assimilated into an omniscient being - as with God himself. But to a less ultimate degree, a technology will be created acting at a distance and combining radio, X-rays, cinema and microscopic photography. Everything in the universe, and everything of man, would be registered at a distance as it was produced. In this way a moving image of the world will be established, a true mirror of his memory. From a distance, everyone will be able to read text, enlarged and limited to the desired subject, projected on an individual screen. In this way, everyone from his armchair will be able to contemplate creation, as a whole or in certain of its parts.⁷²

One of the sad trends of the past century has been that Europe, which pioneered the ideas and institutions for systematic access to knowledge on a global scale, has closed its key places or neglected them to a point where they longer play a role in the world scene: notably, the FID (The Hague), the DBI (Berlin) and Mundaneum (Brussels and since 1998, Mons). One of the challenges for a new vision is to revive these initiatives with a solid European basis and an eye to the whole world.

4. Electronic Media and Networks

Otlet's vision of global networked knowledge took time. During the 1930s and 1940s, alternative models for programmable computers emerged in Germany, England and the U.S. After the war the American model gained dominance. By 1949, Shannon and Weaver had a model that favoured bits and Boolean logic over Bayesian and other approaches. By the 1960s, the practical realities of computers and networks began to merge. For the world of recorded knowledge, the most evident developments have been with respect to a) titles; b) full-text; c) new methods.

4.1 Titles

4.1.i North America

With respect to electronic titles and automated cataloguing, Canada⁷³ was a pioneer, but its efforts were bought up by the United States,⁷⁴ where there were two major developments. One was the Research Libraries Information Network (RLIN, 1971). Working closely with major research libraries in Europe and around the world, this grew to become the Research Libraries Group (RLG) Union Catalogue with records representing 400 languages including Arabic, Chinese, Cyrillic, Hebrew, Japanese, and Korean scripts.⁷⁵ For a time, this became the largest international network, with some 48 million unique titles and some 150 million listings in individual copies.⁷⁶

Second, there was the Ohio College Library Center (OCLC, 1967), in Dublin, Ohio, which began as a local venture and became the Online Computer Library Center (OCLC, 1981).⁷⁷ In 1988, OCLC bought the Forest Press and gained copyright over the Dewey Decimal System (DDC). In 1994, OCLC, inspired by the ideas of the late Juri Rubinsky, became famous for the Dublin Core Metadata Initiative (DCMI).⁷⁸ In the 1990s, they also developed the WorldCat database.⁷⁹ This is now a "worldwide union catalog created and maintained collectively by 9,031 institutions... with... more than 76 million bibliographic records and 1 billion individual holdings."⁸⁰ By 1998, OCLC covered 26,000 libraries in 64 countries.⁸¹ In 1999, OCLC acquired the Western Library Network (WLN).⁸² In 2002, OCLC acquired the PICA (Project for Integrated Catalogue Automation) network based in the Netherlands.⁸³ In 2005, OCLC PICA acquired Sisis Information Systems (Germany) and the Fretwell Downing Informatics Group (UK). In 2006, OCLC acquired RLIN.⁸⁴ In 2007, OCLC describes itself as a worldwide library co-operative with 41,555 libraries in 112 countries and territories.⁸⁵

OCLC is much more than a database of titles. It provides computer-based cataloguing, reference, resource sharing, eContent, preservation services and research to 54,000 libraries in 109 countries,⁸⁶ and offers a Catalogue of Art Museum Images Online (CAMIO) with 90,000 examples.⁸⁷ Bill Gates has recently contributed \$9 million to OCLC for a project that involves

Germany	GBV ⁸⁸	27.1
United Kingdom	COPAC ⁸⁹	30
France	CCFR ⁹⁰	20
Netherlands	NCC ⁹¹	14
Austria	Aleph	8.5
Spain	REBUIN ⁹²	7.8
Italy	ICCU ⁹³	7.7
Sweden	Libris ⁹⁴	5
Hungary	MOKKA	6
Belgium	CCB ⁹⁵	4

		125.1

Figure 4. Some of the chief national and international networks in millions.

10,000 public libraries in the US.⁹⁶ It has a Net Library E-Books division to which the Bayerische Staatsbibliothek is subscribing.⁹⁷ These developments reveal that the OCLC is engaged in much more than a modern library system for North America. Their database uses 11 scripts⁹⁸ and deals with 400 languages. To an outsider this database with over 1 billion books might seem nearly all inclusive. We begin to understand the immensity of the challenges ahead when we recall that Unicode now lists 239 scripts⁹⁹ to deal with the world's 6,500 languages. In reality, WorldCat still represents only a very small sample of the world's recorded knowledge.

4.1.ii Europe

During the 1970s and 1980s, Europe's efforts were largely focussed around national libraries and their networks. These efforts have led to many databases typically ranging from 1-27 million entries (figure 4). In 1994, the idea of a (GAteway and BRIdge to Europe's National Libraries (GABRIEL) was broached at a meeting (Oslo) of the Conference of European National Libraries (CENL).¹⁰⁰ In 1995, G7 Pilot Project 5, Bibliotheca Universalis,¹⁰¹ introduced concrete attempts to integrate the authority files for author names in some of the European National libraries.

By 2001, the idea of The European Library (TEL) emerged. In 2005, the results of GABRIEL and Bibliotheca Universalis were merged with TEL. An initial portal entailed 11 million books from 19 National Libraries.¹⁰² There are now 150 million entries and in 2007 the scope is scheduled to expand to all 46 European national libraries.¹⁰³ Meanwhile, the Karlsruhe Virtual Katalog (KVK) provides access to 500 million titles in library and book catalogues.¹⁰⁴

Looking back, in the period 1837-1950, a policy commitment to national libraries raised the quantitative size of the largest collections from 235,000 to c.10 million books. Since then a commitment to networked collections has increased the number of unique titles available to well over 150 million with access to over 1 billion listings in a single database. In simple terms, the number of recorded unique titles has increased 15 fold and the number of listed copies accessible in one database has increased 100 fold in the past fifty years. Taking into account the extraordinary rise in electronic databases, so-called born-digital knowledge, the actual rise of recorded knowledge is much greater than the rise in books. The good news is that we now have access to a wide range of media including music, films, databases, CD-ROMS. The less good news is that we now often have to search in a great number of different catalogues to find what we want. More integration is needed.

Dates	Volumes	Project	Individual/Institution
-------	---------	---------	------------------------

1949-1974	52	Index Thomisticus	Father Busa
1972 -2001	12,000	Thesaurus Linguae Graecae	Marianne McDonald, T.F. Brunner
1971-	20,000	Project Gutenberg ¹⁰⁵	Michael Hart
1990s-	125,000+	Pollard	Chadwyck >Proquest
2000	1.5 million	Dissertations	Proquest
2001	1 million	Books	Raj Reddy, Gloriana St. Clair
2004	15 million	Books	Google
2005-10	6 million	Books,	EC via BNF and TEL

Figure 5. Some important full-text projects in the past half-century.

4.2 Full-Text Scanning

As electronic methods evolved one of the first full-text projects was the *Index Thomisticus*, which began as a doctoral thesis of Father Roberto Busa, S.J. (Rome).¹⁰⁶ This project, which took a quarter century (1949-1974) led to a publication of 52 volumes. The 1970s saw two new projects which entailed the scanning of 12,000 and 20,000 volumes respectively. The 1980s saw one of the first commercial scanning projects as Chadwyck-Healey began to produce electronic full-text versions of all the books in the English Short Title Catalogue (ESTC). The Early English Books Online (EEBO) now contains some 100,000 books of the 125,000 books before 1700.¹⁰⁷

In the past decade, the scale of full-text scanning has grown enormously (figure 5).¹⁰⁸ The Internet Archive working with the Open Content Alliance has scanned 100,000 books.¹⁰⁹ OCLC's e-books has scanned 210,000¹¹⁰ books so far. Microsoft, in conjunction with the British Library is scanning 25 million pages.¹¹¹ These trends are now worldwide. In 2000, before the US and Europe had an explicit policy, China had scanned the 800 million characters of the Emperor's Library in Unicode. The Chinese *Apabi D-Lib project* offers 130,000 books today.¹¹² In the past five years these projects have taken a quantum leap forward with a vision of a World Library (see § 6.5 below).

4.2.1 Memory Institutions

The scope of digitization has gone far beyond libraries to include museums, and archives.¹¹³ The Canadian Heritage Information Network (CHIN, 1972) was pioneering in this domain. Today, the Virtual Museum of Canada (VMC) includes 420,000 images; more than 150 interactive games and over 500 Virtual Exhibits and Community Memories Exhibits.¹¹⁴ In Europe, projects such as the MEDICI Framework and E-Culture Net¹¹⁵ catalogued some of the pioneering projects in this context. Today, there is a significant German project, Bibliotheken, Archiven und Museen (BAM),¹¹⁶ which addresses directly this theme of combined memory institutions. If we choose Leonardo da Vinci we find results under all three categories. Even so, we are still a long way off from having a list of everything Leonardo created with locations under each media etc.¹¹⁷ with cross references between primary literature, secondary literature, reviews and citations.

4.3 Digital Reference

Traditionally when scholars arrived at a great library, they required a period of orientation as they learned to use the resources available. In Germany, these materials are typically called the *Handapparat*. In English, these materials belong to the Reference Room, which is sometimes synonymous with the main reading room. At the old British Library, for instance, reference

books amounted to 300,000 volumes. The time required to master these tools, depended on the size of the library. In some places it was a few weeks. In the British Library it was two years. In a world where we make tens and even hundreds of millions of titles available online, readers need digital reference rooms.

On this front, there is both good news and bad news. The good news is that publishers have made many dictionaries, encyclopaedias and other standard reference works available in electronic form. Modern libraries now typically have an online section on Electronic Reference Sources.¹¹⁸ Special licences with publishers mean that some of these works are available free of charge at libraries and universities. Companies such as XReferplus now offer access to 100 or 150 standard reference works.¹¹⁹

The less good news is that the electronic versions of these reference works are frequently so expensive that they are beyond the reach of individual scholars. Meanwhile, there has been a trend for such reference works to be owned by a few key companies. In Germany, the pioneer in this field was K. G. Saur, which publishes “nearly 2000 print, microfilm, and electronic formats.” In 1987, Saur was acquired by Reed International. In 2000, it became part of the Gale Group owned by Thomson.¹²⁰ In the United States, Dialog,¹²¹ which was founded in 1967, and “provides access to over 9 terabytes or more than 6 million pages of information“, was acquired by the same Thomson Company in 2000.¹²² Meanwhile, Bowker¹²³ founded in 1872, which publishes Ulrich’s International Periodicals Directory (1932); and Books In Print¹²⁴ (1948-) was acquired by Xerox (1967) then Reed International (1981), then by Cambridge Information Group (2001), which has recently also acquired ProQuest Information and Learning (2006).¹²⁵ Today, works such as Books in Print, are available only to institutions and are no longer available to individual subscribers. Fifty years ago, only the richest libraries could hope to achieve near comprehensive coverage of secondary literature. Today, practically no library can hope to be comprehensive and most collections are retreating. For instance, Göttingen, which had over 70,000 serials in the 1970s, now covers 30,000 serials. The California Digital Library has 21,000 electronic journals, which is impressive until we recall that Ulrich’s Periodicals Index lists 250,000 journals and serials. Meanwhile, at the University of California San Francisco, we find another modern catalogue that looks objective until we look closely and discover that of the 20 headings nine are traditional subjects and the remainder are branches of medicine (Appendix 3).

5. Challenges and Concerns

5.1 Privatization of Knowledge

Ever since Gutenberg went bankrupt from the first printing, it has been obvious that publishers need to be attentive survival. For a very few companies this is not a problem. For instance, in 2004, Reed Elsevier¹²⁶ listed an operating profit of £1126 million and profit attributable of £675 million.¹²⁷ Somewhat disturbing is a trend whereby the world of long-term recorded knowledge is increasingly being framed in the terms of short-term business propositions, as if the whole of the public sphere was open to business exploitation.. At a recent international conference in Brasilia one of the keynotes spoke of the need for business intelligence in libraries. A recent report called Library 2007,¹²⁸ with input from the American firm Booz Hamilton, known for its links to the CIA, calls for the development in Germany of a Library Development Agency the BEA, (BibliotheksEntwicklungsAgentur), and a Competence Network for Libraries (Kompetenznetzwerk für Bibliotheken, KNB). Particularly striking is one sentence:

Finally: more than 60% of all German citizens are customers of a library. This makes the libraries one of the most widely used educational institutions and means that they have reached a top level in international terms.¹²⁹

The assumptions a) that citizens are only interesting as customers and b) that memory institutions and libraries in particular are only relevant to 60% rather than 100% of citizens run directly counter to Europe's quest for a knowledge society.¹³⁰

5.2 Access to Content

These trends towards privatization pose serious problems with respect to access to content. For our purposes two examples will suffice. In 1938, University Microfilms (later UMI) began an important project: a microfilm version of the books in the English Short Title Catalogue (ESTC).¹³¹ UMI evolved into Proquest. As noted above (figure 5), in the 1980s, Chadwyck Healey began with an electronic version of the ESTC which led to EEBO (Early English Books Online), which has now produced 100,000 works in full-text. Chadwick Healey also gained the rights to projects such as the *Patrologiae Graecae*¹³² (130 vol) and *Patrologia Latina* (221 volumes). Proquest then acquired Chadwyck Healey. As a result access to the complete writings of the Church Fathers and effectively every book published in England before 1700 is only accessible via Proquest. An English student or schoolchild without access to the databases of an American company, cannot have electronic access to the early editions of Shakespeare, Milton and other great authors of their own country.

This is not an isolated case. A second example is a significant project involving the digitization of periodical journals called Journal Storage or JSTOR. The Scholarly Journal Archive,¹³³ which began in 1995 and has scanned in some 750,000 pages of journals published prior to 1990. As long as one is a member of an institution subscribing to JSTOR all is well. Individuals need to seek permission with the original publisher for every article they wish to consult.¹³⁴ JSTOR is linked with Art Storage or ARTstor,¹³⁵ which now has 500,000 images available. ARTstor is not open to individuals. Both JSTOR and ARTstor are linked with ITHAKA,¹³⁶ which has major publishers and a former head of American Express on its board. ITHAKA has since "incubated" Aluka¹³⁷ --resources for Africa and the developing world; Nitle¹³⁸ --liberal education in the digital age, and Portico¹³⁹ --concerned with an archive of electronic materials.

ABI/Inform (Business)
Academic Search
Engineering Village
JSTOR
Lexis/Nexis Academic
Proquest Research
PsychInfo
PubMed/Medline
Web of Knowledge
Xreferplus

Figure 6. Top Ten Databases at Wisconsin.¹⁴⁰

An optimist will be happy that problems of long-term preservation are being seriously addressed. A pessimist will worry that these short-term packaged solutions cannot lead us to a

vision of comprehensive access for ordinary citizens. Meanwhile, such new initiatives, while attractive on the surface, give us no idea precisely how their packaged solutions available only to subscription paying institutions, relate to the corpus of recorded knowledge as a whole. Two further examples, will illustrate why this is a genuine problem. The University of Wisconsin is well known for its role in new information technologies. Their library has a convenient list of their top ten databases (figure 6). The quality of these databases is not in question. It is striking however, that these are exclusively American. There is not one database based in Europe, Asia, South America, Africa or Oceania. How is a student at Wisconsin, who has only this sample at their disposal, to become aware that 80 % of the Internet and 95% of the world's population lies outside of America.

Once more, this is not an isolated example. The California Digital Library has rightly gained attention as one of the most modern versions of digital libraries. It boasts tens of thousands of electronic books, and more than 250 article and reference databases.¹⁴¹ It has Image Service Collections,¹⁴² which provide access to 13 databases. On 6 February 2007, this added up to 313,713 images, but the collections are predominantly American. There is no mention of the Virtual Museum of Canada (VMC) with 420,000 free images, no mention of a single European collection or any of the thousands of museums listed on the ICOM site. If leading universities and public institutions provide access only to insular, packaged bits of the world's recorded knowledge, how can we educate students for a world in which the United States and even Europe represent small minorities?

What Americans choose to do is, of course, entirely their concern. Our interest in citing these cases is simply with respect to their possible applicability to Europe where there exists a tendency to assume that America is bigger, better and more modern; that they will have solutions that are better than our own. The cases cited reveal that the latest trends in America do not address the richer linguistic and historical dimensions of Europe's memory institutions. Indeed, they suggest that short-term business propositions cannot guarantee the educational range, research depth and long-term preservation needs of society. This implies, that in order to maintain its multi-lingual and cultural diversity, Europe needs to create its own vision.

5.3. Citation Indexes

The difficulties of knowing how to weigh the significance of these packaged samples of the world's recorded knowledge have been compounded by a further development. In 1960, "Eugene Garfield's Institute for Scientific Information (ISI) introduced the first citation index for papers published in academic journals, starting with the *Science Citation Index* (SCI), and later expanding to produce the *Social Sciences Citation Index* (SSCI) and the *Arts and Humanities Citation Index* (AHCI)."¹⁴³ In the years that followed, Derek J. de Solla Price,¹⁴⁴ a British scholar, who emigrated to the United States invented the field of bibliometrics. By the 1970s, citation indexes had become a tool for research. The tool performs a valuable service:

It indexes more than 8,000 peer-reviewed journals, providing complete bibliographic data and author abstracts. Every item of significance is listed: articles, reviews, letters, notes, corrections, and editorials. A unique and useful feature: this index can be searched by first cited author and to find articles sharing one or more cited references.¹⁴⁵

These citation indexes were bought by Thomson Scientific and became an online academic database called Web of Science as part of a larger product called Web of Knowledge¹⁴⁶.

Thomson also has the Science Citation Index, which “provides access to current and retrospective bibliographic information, author abstracts, and cited references found in 3,700 of the world's leading scholarly science and technical journals covering more than 100 disciplines” and an expanded version which covers “more than 5,800 journals.”¹⁴⁷

Meanwhile, Elsevier offers three products: 1) ScienceDirect claims to offer “more than a quarter of the world's scientific, medical and technical information online”...involving “over 2,000 peer-reviewed journals, [and] hundreds of book series, handbooks and reference works.”¹⁴⁸ The price per article for Elsevier journals is \$30.¹⁴⁹ 2) Scirus claims to be “the most comprehensive science-specific search engine on the Internet. Driven by the latest search engine technology, Scirus searches over 300 million science-specific Web pages.”¹⁵⁰ 3) Meanwhile, SCOPUS is described as “the largest abstract and citation database of research literature and quality web sources,” which entails a wide a wide range of materials (figure 7).

- 15,000 peer-reviewed titles from more than 4,000 publishers
- Over 12,850 academic journals including coverage of 500 Open Access journals
- 700 conference proceedings
- 600 trade publications
- 29 million abstracts
- 265 million references, added to all abstracts
- Results from 250 million scientific web pages
- 18 million patent records from 4 patent offices

Figure 7. Key content offered in SCOPUS.¹⁵¹

Ulrich's Periodicals, founded in 1932 now covers 250,000 serials.¹⁵² The important commercial products cited above provide citations and abstracts from 2,000, 5,800, 8000+ or as many as 15,000 of these serials. In simple terms, this means that the largest packages still omit 235,000 periodicals. The smallest packages omit 248,000 periodicals. How is a scholar ultimately to know how representative their findings are? How much research is simply being overlooked or forgotten because it is not part of the popular journals of the day? These might readily seem rhetorical questions except for two considerations. The popularity of journals in these packages is largely determined by their being in English. Tenure committees, even outside the United States, in the case of scholars whose native language is not English, are using these citations as a factor determining whether scholars will be promoted.

Scholarship has some popular writers, but scholarship is ultimately not a popularity contest. In a rapidly changing world, these citation indexes are extremely useful as a stopgap measure. In the longer term, we need to develop more systematic approaches, which also give us some idea of how the sample that we are searching relates to recorded knowledge as a whole. In the past decade, there have been various efforts to automate the citation process: e.g, autonomous citation indexing¹⁵³ used in CiteSeer; Dynamic Contextualization¹⁵⁴ used in the HyperJournal Project and efforts towards a Self-Annotating Web.¹⁵⁵ These point to the possibility of citations becoming a regular dimension of future catalogues of recorded knowledge.

In terms of the big picture, it took from 1450-1650 to gain some bibliographic control over primary literature. The period 1650- 1900 gave us some control over secondary literature. At the turn of the 20th century, there was a vision of a global brain (*Gehirn der Welt*). The period 1900-2000 improved universal bibliographic control and added abstracts and citations. From a global viewpoint, America offers solutions which are quantitatively impressive, but most reflect

neither the scope nor the depth needed for European collections. Meanwhile, commercial companies offer a series of competing packages, which are neither interoperable nor cumulative. Ironically, as a result of these interim solutions, at the begin of the early 21st century, we were in some senses further away from the visions of Otlet, Laumann, and Oswald, than a century earlier. We probably need at least the next century to create further methods of bibliographic control, which integrate these tools seamlessly in a single system (cf. § 6.2 and Appendix 3 below) and then extend bibliographical control into full-text..

5.4 Systematic Access to Content

In addition to obvious problems of access and permissions to use content, are more subtle challenges of systematic access to content. One problem is precision. The Proquest Corporation,¹⁵⁶ has a Digital Commons which generously provides free access to resources. A search under the term “linear perspective” provided over 600 hits, but not one of the first 50 was actually about linear perspective in a technical sense and only one dealt with concepts of space, which are indirectly connected with the subject. Google Scholar lists 50,600 titles under Plato, but not one of the first 20 titles has anything to do with the historical Plato who was the student of Socrates and teacher of Aristotle.

Most websites are constructed on the assumption that more hits are better. Suppose a user wants to find a specific title: *GesamtKatalog der Wiegendrucke*. If they search in GBV, they get 211 locations. In TEL they find 115 locations. Such quantitative information about locations is sometimes vitally important. Normally, however, a user wants to know which copy is closest to where they are, instead of receiving 325 additional locations. In future, we need to integrate goals of users into our interfaces. Of course, when we are searching for titles under a specific subject, we want as many hits as possible. For instance, if we search for linear perspective in GBV we receive 396 hits. In TEL the same search gives us 118 hits.

	Bavarian Union Catalogue	1902
	Union catalogue of UK and Ireland & British Library	27794
	Common Library Network	15223
	Hesse Union Catalogue	1211
	Library of Congress (USA)	2493
	Northrhine- Westphalian Union Catalogue	1000
	Berlin-Brandenburg Cooperative Union Catalogue	1000
	South-West German Union Catalogue	2690
	Bavarian Union Catalogue	70
	Union catalogue of UK and Ireland & British Library	1673
	Common Library Network	664
	Hesse Union Catalogue	292
	Library of Congress (USA)	84
	Northrhine- Westphalian Union Catalogue	125
	Berlin-Brandenburg Cooperative Union Catalogue	69
	South-West German Union Catalogue	102

Figure 8a. Hits in GBV for Information Technology; 8b the same query when one limits the search to a specific year, in this case, 2000.¹⁵⁷

The dilemmas of using traditional listing methods in combination with new technologies come into focus when we choose a contemporary topic. If we use Google and search for “Information

Technology” we get 725,000,000 hits.¹⁵⁸ The same search in GBV produces 53,209 results (figure 8a). In strictly quantitative terms, the Google results are staggeringly superior. From the viewpoint of a computer scientist concerned with information retrieval both results are successful. However, in terms of a scholar who genuinely wants to learn about information technology, the Google list is hopeless. No one has time to check 725 million titles. The GBV list is revealing. It shows at a glance that the GBV’s European databases record many more sources than the Library of Congress. Even so, the details of the GBV list still remain almost unusable. How many students or even professionals have time to browse 53,204 titles?

Per se the lists are fine. It is not the search algorithms of Google or GBV that need changing. Needed are new tools to make the lists manageable. In cases ranging from 10 to say 500 titles, a simple ability to view the alphabetically by author or title, chronologically and geographically would be extremely useful. Also useful would be an ability to combine these options with an ability to filter by medium: e.g. only films on the subject in Germany between 1950-1960. Using the advanced features today one can already limit the search for Information Technology to a specific year, say 2000. This immediately produces a list which is much more manageable (figure 8 b). Increasing the initial search fields to include Who, What Where When, How and Why, would help make the equivalent of multiple searches in one query.

In the case of longer lists, users need subject headings relating to the term being searched, in order to help them choose which subset of the field, Information technology, actually interests them. In the past, subject headings and authority names belonged to the world of cataloguers and indexers, who were akin to a secret priesthood. Only they knew the true authority names, the keys to knowledge, a language which uninitiated users could never hope to learn. In one sense they were right. However, in a world where knowledge increases at an estimated 7 million new pages a day, and where even the leading experts at places such as the Library of Congress need to meet weekly to discuss ever changing shifts in terms we all need help. If we use tools such as Roget’s *Thesaurus*, and these lists of the experts themselves, our vague words at the beginning of a search can be refined into precise queries, which give us a result that is small enough to be useful.

Scholars, students and citizens who are beginning research in a field would be advised to gain an initial orientation in a new field by starting their search in subject catalogues and classification systems. The good news is that TEL already allows such searching via subject categories. At present, “Painting and Paintings” has four hits including one art museum for the whole of Europe. Such examples remind us that there is still a gap between the TEL’s goals and its practical value for everyday users. At present systems, such as GBV use the Basis Classification (BK). Europe has access to the Universal Decimal Classification (UDC) which continues to evolve. In future, this and alternative classification systems should be available for advanced research. Classification systems are ultimately snapshots into the history of sense - making and ordering the world.

Union catalogues and single portals are all the rage, but even here there are enormous gaps between claims, everyday realities and what is desirable. Contemporary search systems around the world start from an underlying assumption that more is better. Hence, in purely quantitative terms the 725,000,000 hits¹⁵⁹ for Information Technology from Google are much more successful than the 53,209 from GBV. Quantity is necessary but not sufficient. For scholarship and the advancement of learning, we need indicators of quality and some sense of how thorough and comprehensive something is. Does the Google list represent 90% of what exists

in the U.S., 50% of all that exists on information technology; or perhaps only 10%? As long as we use random queries as our standard we can never have answers to such questions.

Traditionally, scholarship has approached these dimensions of quality and thoroughness at the micro-scale. First, articles and books have external reviewers to ensure that it has a required level. Second, there are abstracting services to identify the key points of books and larger articles. Third, there are reviews to assess the contributions of books and sometimes thematic review articles to assess the impact of a number of contributions in the periodical literature. Fourth, there are bibliographies which list and annotated Bibliographies which survey and assess major contributions in a field. All this was so successful that by 1955 Theodore Besterman could write a large *World Bibliography of Bibliographies*.¹⁶⁰ The apparatus for finding the materials to find the sources had become a profession in itself, which is one of the reasons why information science has seen such a dramatic growth in the past half century. Needed are ways to integrate these tools of abstracts, reviews, and bibliographies into our union catalogues (cf. § 6.2.3).

5.4.1. Multilingualism

Most discussions of multilingualism are in terms of having databases in multiple languages which is, of course, a fundamental step. Traditionally, there was an assumption that a scholar learned the languages needed for his research. The European Union now entails 25 languages. Worldwide, there are over 6500 languages, mastery of which is beyond any individual's abilities. Increasingly there is a need to be able to type in a query in one's own language and receive multilingual results in multiple languages.¹⁶¹ This has begun to happen in the GBV but more is needed in being able to move seamlessly from standard titles to multilingual variants. As we approach a union catalogue that lists translations as a matter of course we shall be much better served in finding useful translations. In future, as electronic translation methods become more reliable, we need tools to provide near automatic translations of sources that interest us.

5.5. Born Digital

Another concern is the term, "Born Digital", which has rapidly become one of the new buzzwords. A number of computer scientists and information professionals assume that born digital resources are the main concern and some even pretend that born digital is effectively the sole concern with respect to long-term preservation. In 2001, Raj Reddy an influential computer scientist at Carnegie Mellon University, cited the work of Lyman and Varian to claim that the world produces 10^{18} bytes of new information annually and that 90% of new information is stored digitally.¹⁶² This is highly misleading for at least four reasons.

First, physical books show no sign of disappearing. In 1994, Frank Ogden, published a book entitled *The Last Book you will Ever Read*.¹⁶³ Research by Michael Ester at the now defunct Getty AHIP programme explained why this is less likely to happen than was predicted. The human eye sees approximately 3,000 lines per inch when light is reflected from the page of a printed book. The eye sees approximately 2,000 lines per inch when light from a computer screen comes into it directly. This means that we see one third less on a screen than on a printed page, which explains why doing corrections on a screen is inevitably more difficult than in traditional proofreading. As a result, in cases where we want to read a sustained piece of text, the book is not likely to go away, and electronic readers are not likely to make the enormous inroads that salesmen have predicted.

Second, having digital surrogates may mean that we return to the originals less often, but it does not lessen the need to include physical originals in our visions of long-term preservation. Even if we stopped using printed books in everyday life, then the need for libraries as repositories of our collective memory does not go away. Indeed, as noted earlier, the shift to printing increased needs for long-term preservation of manuscripts. New technologies are not replacing them. They are leading us to re-catalogue existing collections in ever greater detail: e.g. the Harley Manuscripts in the British Library. As a result new technologies are helping us achieve an overview of the contents of manuscripts superior to that of any scholarly monk of the Middle Ages.

Third, scholarship, as a profession of making claims that can be checked and tested, requires that we create ever more efficient and reliable ways of returning to the sources (*ad fontes*) of our claims. In the past centuries footnotes¹⁶⁴ marked an important step forward, but these remained pointers to sources beyond the cover of a given book and often required great efforts to track down the source in question. Electronic media allow us to go further. In the longer term, only those references which take us back to its official surrogate and potentially the physical copy on which it is based are of enduring interest. Creating an electronic born-digital network de-coupled from the objects it represents would be a step backwards.

Fourth, the latest developments in technology point beyond the non-physical dimensions of born-digital. In 2005, the International Telecommunications Union (ITU) published a provocative report on *The Internet of Things*,¹⁶⁵ claiming that sensors implanted ubiquitously in the physical world were opening a new chapter in the development of the Internet potentially much more significant than its appearance on the screens of desktop computers. In the past, every object in a memory institution had its own call number or identification number. In the next decades, technologies such as RFID could lead to each recorded object having its own electronic identification. RFIDs can be called with a computer or a cell-phone. This points to new worlds of interaction, which will be considered later (§ 6.4). Ultimately, the importance of Born Digital materials lies in the extent to which we can couple them with other worlds that are print-born or born otherwise.

5.6 Long Term Needs and Short Term Funding

One of the fundamental problems remains that major projects are only possible in the long term. The Chinese stone books project that took 850 years is an extreme case. More recently, the *GesamtKatalog der Wiegendrucke* which started at the letter A in 1925 has reached the letter H in 2007. The *Catalogo dei manoscritti Italiani*, begun in 1897, has published over 100 volumes but is not yet finished. Eric Schmidt¹⁶⁶ of Google has stated that a 300 year framework is needed for a complete digitization of knowledge. Meanwhile, most projects are in the three to five year range. To address long-term preservation seriously, we need to create long-term frameworks.

6. Ingredients of a New Vision

The above survey makes clear that the past centuries have brought remarkable advances both quantitatively in terms of recorded knowledge (books, manuscripts, music, films etc) and qualitatively in terms of tools for finding them (classification systems, thesauri, bibliographies, catalogues etc). We are thus faced with two options. One possibility is to abandon attempts at cumulative knowledge. There is, for instance, a school of thought which claims that: “The new theories of growth (known as ‘endogenous’) stress that development of know-how and

European Science and Technology Observatory	(ESTO) ¹⁶⁷
European Information Technology Observatory	(EITO) ¹⁶⁸
European Audiovisual Observatory	(EAO) ¹⁶⁹
European Observatory for Digital Literacy	(EODL) ¹⁷⁰
European Observatory on IL Policies and Research	(ENIL) ¹⁷¹

Figure 9. Five examples of existing European observatories relevant to memory institutions

technological change - rather than the accumulation of capital - are the driving force behind lasting growth.”¹⁷² If so, one need not be concerned by dangers of being overwhelmed by new Information, a scenario explored by Pierre Levy in *The Second Flood. Report on Cyberculture* (1996).¹⁷³ A radical version of this approach would state that memory institutions as systematic repositories of recorded knowledge no longer have a central role in the future.

Since this is not our view, there is a second option: a new vision to bring increasing mastery over the contents of memory institutions. This requires at least four ingredients: 1) a greater awareness of existing and emerging technologies; 2) a new approach to distributed resources that goes considerably beyond existing plans for a European Digital Library; 3) an institute on knowledge organization; 4) research into recording worlds. In addition Europe, needs to decide whether it wants to continue playing an important role in the international scene, in which case new research on multiple ways of knowing is needed.

6.1 Awareness and Integration of Emerging Technologies

Almost daily there are developments in new technologies. By way of example, in the world of memory institutions, there are trends to use Remote Frequency Identification (RFID) to identify individual books¹⁷⁴; and new tools ranging from chatbots¹⁷⁵ to information mavens. When Shannon and Weaver published their *Mathematical Theory of Communication* (1949), limitations in technology led them to choose a binary system. The incredible advances in new technologies mean that other choices are possible. Within the next generation, advances in nano- and bio-technology could lead to DNA becoming a basis for long-term storage. Librarians and museum curators cannot be expected to follow all these developments themselves.

Fortunately, Europe, in the French tradition of *veille technologique*, has established a series of observatories (figure 9) in addition to specialized research groups¹⁷⁶ and national research councils.¹⁷⁷ Our first suggestion is that the scope of these observatories be adjusted to provide advice to memory institutions via existing structures.¹⁷⁸ The European Audiovisual Observatory already has a section on Intellectual Property,¹⁷⁹ which could be extended to cover copyright (Appendix 4). It would be very useful if this could evolve into an umbrella organization co-ordinating specialized needs and requirements from different memory institutions and related bodies.¹⁸⁰ Some technical aspects are already being addressed.¹⁸¹

6.2 Distributed European Electronic Resource (DEER)

In terms of long-term preservation, there are effectively two parallel challenges in memory institutions. First, there is the long-term preservation of the original objects. Here, it makes sense to continue restoring and conserving them in separate institutions: books in libraries, paintings in museums, archives in archives etc. Indeed, the enormous success of these existing

structures in the past centuries, suggests that a wise way forward lies in a phased approach, whereby we expand the scope of what we have in place, while exploring how to create a new framework.

Second, and that is our main concern here, there is a challenge of new, long-term systematic access to the digital surrogates of those originals, which allow us to return to the originals when necessary. This entails three challenges or requirements: a) bibliographic control of full-contents; b) integration of bibliographic control over reference materials; c) integration of enduring knowledge of recorded works with personal and collaborative knowledge of recording works. In our view, these requirements can be met through a Distributed European Electronic Resource (DEER) with three main ingredients: 1) a distributed repository; 2) a virtual reference room and 3) a virtual agora.¹⁸² To avoid confusion with existing initiatives we refer to this vision as DEER. Whether this is eventually called The European Library (TEL), European Digital Library (EDL) or some other name does not affect the vision.

6.2.1. Distributed Repository

Current library holdings give us surveys of titles in individual collections. We need a distributed repository to give greater bibliographic control for a) more systematic surveys of titles in all collections through integrated authority files and b) integrated full- contents.

6.2.1.i Integrated Authority Files

The great advances towards union catalogues especially at the national level are an excellent step towards a day when we shall have an European, and some day a World, union catalogue of authority names (with variants), terms, places and standard titles (with variant titles of all translations). Today a search by author provides us a list of all titles including those of translations with no clear overview of manuscripts, editions, translations, reviews, abstracts, citations, secondary literature, and bibliographies.¹⁸³ We have the makings of such lists at the level of individual authors in the form of *catalogues raisonnées (opera omnia, oeuvres complètes)*. These need to be integrated (figure 10) not least in order to maintain an overview of what has been scanned and translated. Absence of such control will lead to unnecessary duplication of efforts in an area where available financial support is already meagre.

6.2.1.ii Integrated Full Contents

Scanning millions of books in full text versions makes little sense if our access methods are not refined to provide new levels of bibliographic control over these materials. For some levels of research, scholars will need to search not only all titles by Aristotle but also all references to

Author	Standard Title (Primary Literature)
	(Manuscripts)
	Editions
	Translations
	Reviews
	Abstracts
	Citations
	Secondary Literature (monographs, articles).
	Bibliographies

Figure 10. Elements of bibliographic control that need integration.

these names in the full texts of monographs and/or articles. To do so efficiently requires integrating the standardized authority names, terms places, etc. (and their variants) of union catalogues when searching through full-text versions of literature.

At first sight, all this could seem to be excessive attention to detail.¹⁸⁴ On reflection we see that it offers unexpected answers to elusive questions of quality and the quest for a web of trust. Once we have a union catalogue in this deeper sense that it allows us to see all the works, their citations, bibliography and literature concerning them, we can begin to answer questions about how thorough or comprehensive a work is. What percent of a topic or a field does it cover; how oft is it cited; by what audiences is it cited? When combined with classifications of knowledge we can potentially begin to map which topics have been unexplored in the past decades, which are the fashion and suggest which directions look promising. Major corporations have such strategies in exploring areas where to seek patents. We need similar methods for knowledge generally.

At the end of the 19th century, the rise of secondary literature required new levels of bibliographic control and new methods. In the early 21st century, as an European Digital Library (EDL) expands the scope of a distributed repository from libraries to include a) memory institutions and b) current research as part of the vision of a European Research Area, two more elements will be necessary: digital equivalents of reference rooms with levels of knowledge and virtual agoras.

6.2.2 Digital Reference Rooms

In the past decades, companies such as K. G. Saur have scanned a couple of thousand reference works. Companies such as Xreferplus, offer packages of 100 or 150 of the most popular reference works. However, as we begin making the full texts of tens of millions of books online, we need electronic equivalents of the reference rooms in great libraries. One way forward might be to begin with the c. 300,000 books of a famous example such as the reading room of the old British Library and then move towards new multi-national hybrids that combine features of reference rooms in major collections.¹⁸⁵ Such a virtual reference room would probably grow to about one million books. Choosing these books as a priority in scanning campaigns might be a wise step.

Once established, such a reference room would function much the same as their physical equivalents. We typically begin with an everyday dictionary. If we have a problem in English we go to the shorter Oxford. In more detailed research, we use the complete Oxford Dictionary and other etymological dictionaries. Just as Xreferplus now offers very simple packages, in future one can imagine undergraduates beginning with a standard package of sources and then expanding their range as curiosity, energy and discipline permits. Graduates would have a starter package with more and post-doctoral researchers would tend to begin with the entire apparatus, which would then be customised to fit their research needs.

One important dimension of these reference rooms will be to provide systematic access to integrated versions of reference materials. For instance, today a researcher needs to consult Dissertation Abstracts for theses written in North America. In Europe, individual countries each have their own catalogues. Research projects need to be searched by organization. In future, these should be integrated so that a better overview of current and recent research is possible.

Virtual Reference Room

1. Terms	Classifications, Thesauri
2. Definitions	Dictionaries
3. Explanations	Encyclopaedias
4. Titles	Catalogues, Bibliographies
5. Partial Contents	Abstracts, Reviews, Citation Indexes

Distributed Repository a: Primary Literature in Digital Memory Institutions

6. Full Contents	Books, Paintings, Films, Music
------------------	--------------------------------

Distributed Repository b: Secondary Literature in Digital Memory Institutions

7. Texts, Objects	Analyses, Interpretation
8. Comparisons	Comparative Studies, Parallels
9. Interventions in Extant Object	Conservation
10. Studies of Non-Extant Object	Reconstructions

Virtual Agora: Emergent and Future Primary and Secondary Literature

11. Collaborative Discussions	Comparisons, Interventions, Studies
12. E-Preprints	Collaborative Primary, Secondary Literature.

Figure 11. Virtual Reference Room, Distributed Digital Libraries and Virtual Agoras with different levels of reference and secondary literature.

6.2.2.i. Levels of Knowledge

One of the fundamental problems of search engines today, including Google, is that they do not distinguish between levels of knowledge: i.e. a search for a term results in definitions (dictionaries); titles and full text passages being listed indiscriminately. In the language of computer science, different layers of granularity are offered randomly. Traditional libraries, by contrast, have reference rooms with specific sections for terms (classification systems, thesauri, terminology lists); definitions (dictionaries); explanations (encyclopaedias); titles (catalogues, bibliographies) and partial contents (abstracts, reviews, citation indexes) as well as the collection proper.

In addition, libraries today already have some distinction between primary and secondary literature, between conservation/restoration studies and reconstructions in the main part of the library. Implicit in these levels is a basic progression of detail. It is useful to have a topic before searching for titles on a topic; it is useful to have a title before deciding to search for the full contents of that title. Needed in a future Distributed European Electronic Resource (DEER) is an ability to move systematically through these various levels such that the functions of virtual reference room and digital repository are integrated. Next steps will be to include different worlds and scales of knowledge and to integrate these catalogues of enduring knowledge with personal and collaborative knowledge in virtual agoras.

6.2.2.ii Worlds of Knowledge and Scales of Knowledge

Related to levels of knowledge is the notion of worlds of knowledge. Philosophers have long distinguished among such different worlds, typically ranging from three or five to 33 worlds. These worlds can be seen as different levels of reality: e.g. Metaphysical, Mental, Physical, Man Made and Social which, in turn, can be mapped to basic disciplines: e.g. religion and philosophy are metaphysical; mathematics, grammar, dialectic (logic), literature are primarily

mental; geography has a basic physical dimension; architecture a man-made dimension, while sociology and psychology are primarily social. Once we have a systematic union catalogue linked with classification systems, we would potentially be able to use levels of reality as a filter for searching. In future, this would mean that when searching for lion, we could choose between physical lions in zoos, literary, artistic and other lions.¹⁸⁶

Today's Geographical Information Systems (GIS), distinguish between different scales in the physical world. A systematic treatment of levels and scales of knowledge would mean an ability to navigate seamlessly between different levels of the micro-, meso- and macro-cosm. In the long-term we need the equivalent of a GIS of the mind and imagination; of mental and social worlds; not all possible worlds, but at least those that have been recorded in some form.

6.2.3 Virtual Agora

In the past, memory institutions were focussed on enduring knowledge and this needs to remain a priority. Even so, in a world where personal and collaborative knowledge play an ever greater role, there is an ever greater need to link enduring knowledge with ongoing research through a virtual agora. This will have several dimensions. One entails a need to go beyond libraries, museums and archives to include resources from national research institutes and other organizations in higher education. Here, the Joint Information Systems Committee (JISC) has played a pioneering role.¹⁸⁷ The role of Max Planck in the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) was useful.¹⁸⁸ Projects such as Mapping of Research in European Social Sciences And Humanities (MORESS)¹⁸⁹ have provided excellent preliminary work. Within France, there has been a certain consolidation within CNRS and the network of Maisons des Sciences de l'Homme. Even so, integration with corresponding bodies in other European countries is still very much needed

There are other dimensions of the virtual agora. Knowledge, as we stated at the outset is not just about recorded thoughts and claims. It is equally about communication. In pre-history this remained as part of an oral collective memory. From roughly 4,000 B.C. until 1650, such communication was largely in the form of letters and correspondence. From 1650-1960, this written correspondence evolved into scholarly letters, reviews, discussions, debates and other journal activities. Since 1960, these print forms have begun moving to electronic equivalents: e-mail, audio conferences, video conferences, blogs, chat rooms, etc. In April 2006, David Sifry, who has been tracking the growth of blogs noted that there were "35.3 Million weblogs, and the blogosphere we track continues to double about every 6 months."¹⁹⁰ Just as no student or scholar can be expected to browse 750 million hits on information technology found by Google, no one can be expected to check which of 35+ million weblogs are relevant for their interests. Thus far we have search engines and finding aids to get basic hits. In future, we need filters to help us find those subsets, which match our interests professional or personal. A Virtual Agora will permit us to combine and integrate these new communication tools with traditional library reference tools such as thesauri, classifications and catalogues. The combined effect of the virtual agora will be to integrate collaborative discussions and E-Preprints as two further levels of the knowledge process (figure 11).¹⁹¹

6.3. Knowledge Organization Institute

Early classification systems, including those of Dewey and the Library of Congress were excellent pragmatic tools without a theoretical basis. The early visions of Paul Otlet at the end

of the 19th century foresaw a theoretical basis. A first result was an expansion of the scope and detail of the Dewey Decimal Classification (DDC) System to become the Universal Decimal Classification (UDC) a process which evolved at the FID and now continues as an adjunct of the Koninklijke Bibliotheek. (KB, The Hague). Subsequent attempts at a more systematic classification include the contributions of Bliss and Ranganathan.

In the 1960s, Perreault suggested that a way forward might be to distinguish between subsumptive, determinative, and ordinal relations. Since then, there has been enormous interest in relations which has inspired a veritable explosion in work on thesauri and ontologies. Even so, Perreault's work remains of seminal importance because it offers a fresh approach to the age old debate of universals versus particulars (figure 12).

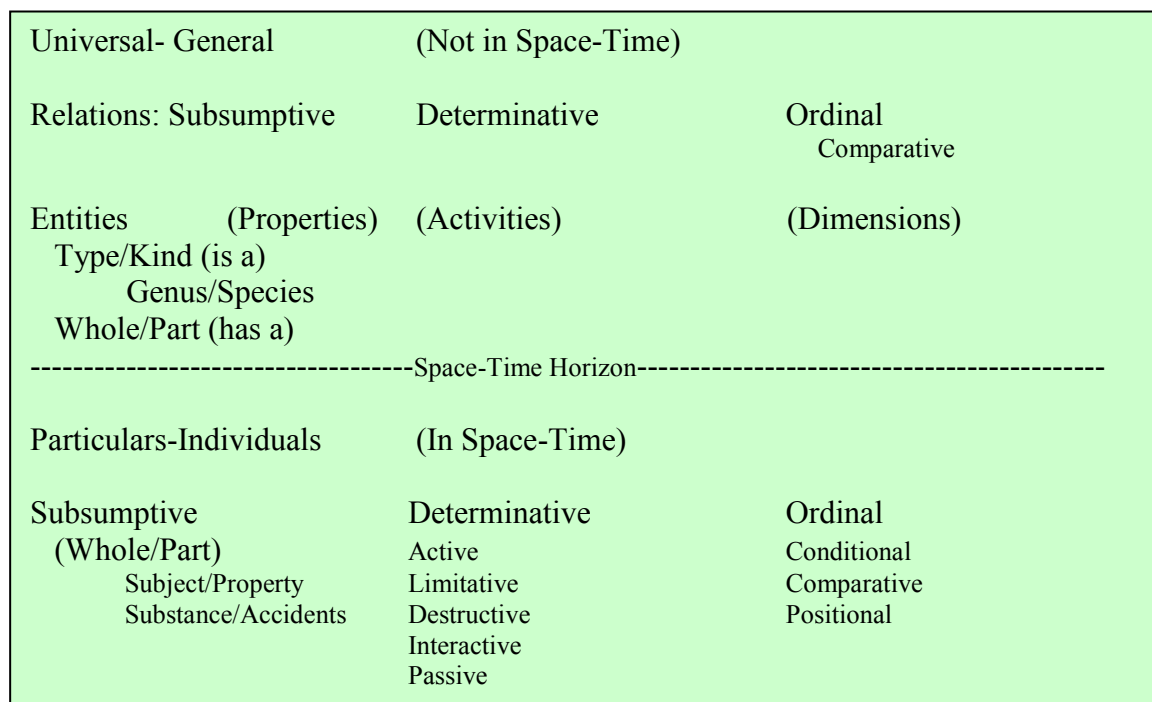


Figure 12. Universals and Particulars in Light of Perreault's Relations.

Classification systems and most ontologies today focus on subsumptive relations of entities and properties. As Dahlberg has shown, entities and properties provide us with information. Knowledge¹⁹², by contrast, entails entities, properties, activities and dimensions. In the long term, we need to create a framework for knowledge organization that includes all these elements. We need bridges between the universals in abstract systems of concepts above the space-time horizon, and the particulars below this horizon.¹⁹³ To take a simple example. We have taxonomies in zoology to classify kinds of lions in abstract terms. At the same time we have many thousands of photographs, films and other records of physical lions. Needed in future, is a framework that allows us to go seamlessly from the classes of universal knowledge to the particulars of specific examples.

Most systems in information science of the past century have started from the premise that all their information is assumed to be true. In branches of science and logic this is a useful and some would say necessary premise. In the case of enduring knowledge, especially in the realms of metaphysics, the truth of many assertions cannot ultimately be tested: all we can hope to do is report truthfully what is claimed and to provide reliable links back to original

sources. We need new ways for reflecting variant versions of the same story or theme. These are further dimensions, which the institute should explore.

To achieve this will require an institute that functions partly as a think-tank, that will take up afresh at a European level aspects addressed by earlier pioneering efforts such as the Mundaneum, the Bridge, FID, DBI and ISKO.¹⁹⁴ This institute needs to combine talents from the library world, computer science, logic and philosophy; should have a small permanent staff; have some visiting fellows and some posts for post-doctoral students.

6.4 Recording Worlds

In the traditional paradigm scholars sometimes retreated from the world to record their thoughts; sometimes went into the world and recorded thoughts. The products of their efforts were recorded knowledge, which became part of a cumulative corpus in memory institutions. Hence, these memory were primarily the end-station of the knowledge process, even if their reading rooms provided new inputs for a subsequent cycle of knowledge.

The rise of the Internet seemed at first sight to signal the creation of a new World Wide Web of knowledge, which was effectively autonomous and separate from the rest of reality. It appeared that one would need a simple scanning process from printed to digital media so that materials from outdated print could be “translated” into the new form. Accordingly, as noted earlier (5.5) computer scientists have emphasized the rise of born-digital materials as a world apart, typically represented as a cloud detached from surrounding things.

Just as printing had unexpected “side-effects”, the rise of digital technologies is bringing a series of unforeseen developments, which are increasingly linking the born-digital cloud with the physical world. At least four developments have begun in parallel and are beginning to converge. First, there is a process of capturing and mapping the world with cameras and sensing devices ranging from views from space to local surveillance cameras. This began with former Vice-President, Al Gore’s vision of a Digital Earth (1998),¹⁹⁵ linked with a vision of integrating all recorded knowledge. Microsoft’s Virtual Earth in conjunction with the Department of Defense (DOD), with its Local Live features is one of the most striking examples. This new digital earth will be a 1:1 scale copy of the physical earth. Second, there is a wave of reconstructing the world, whereby scholars are reconstructing historical buildings, complexes and even whole cities in various interpretations. Third, there are dramatic developments in technologies for recognizing images. To date these developments in image recognition have largely been used for surveillance, law enforcement and security. But they can equally be applied in the world of knowledge. Fourth, developments in nanotechnology mean that we have begun embedding the world with sensors of various kinds including Radio Frequency Identifications (RFIDs). This trend, which the ITU has called the Internet of things, means, that the world is no-longer simply a passive set of objects to be recorded. The physical world is increasingly imbued with so-called intelligent objects, which can be recording or communicating.

Capturing and Mapping	Photography, Remote Sensing, GIS, UMTS
Reconstructing	Computer Aided Design (CAD)
Recognizing	Image Recognition, Surveillance, Computer Vision
Embedding	Sensors, Nanotechnology

Figure 13. Four developments and their related fields which are converging.

Each of these four developments is impressive in itself. In combination, they point to remarkable new possibilities. The physical world can now become an interface for searching. By way of illustration, two scenarios will suffice. We are walking in the woods and we see an interesting tree which we do not recognize. We take a picture with our mobile camera-phone, send the image to a digital reference room.

A first stage of image recognition determines that this is botany, determines this is a tree, opens a taxonomical database of images of trees; matches the photograph from the mobile phone with the database and determines that this is an *Ilex Aquifoliaceae*. Such image recognition can vary enormously from simple curiosity exercises, to classroom fieldtrips to new tools for botanists exploring possible unknown species in a rain forest. In such cases where there is no known match in the database, the researcher would be offered nearest samples in the database as a context for classing the new species.

A second scenario entails the man-made world with applications ranging from tourism to archaeology. Projects such as Archeoguide have demonstrated how an augmented reality reconstruction of a no longer extant building can be superimposed on a physical landscape that we see in front of us. Projects such as the NUovo Museo Elettronico (NUME) have shown how a one can link a reconstruction of an historical centre of a city, Bologna, with recorded evidence from manuscripts in libraries and objects in museums and churches. The complete database is over 7 terabytes and much too large for viewing the entire database in real-time using contemporary broadband connections.

Projects such as Microsoft's Live Local Virtual Earth show how small subsets from enormous databases can be used to show specific views one at a time. In future, such a Virtual Earth demo could be co-ordinated with historical reconstructions such that one can see how a given square or specific building looked like in the 15th, 18th, 19th centuries indeed at any time in its history. This becomes especially useful in the case of famous buildings which have been continually re-adapted for other purposes such as Hagia Sophia, which began as a Christian Church, became a Muslim Mosque and is now a museum. In a world of embedded objects at the micro-scale and the nano-scale, historical buildings could have a series of tiny sensors that trigger local stories on portable pen type computers, potentially in combination with audio-guides.

This approach applies equally at the level specific statues, architectural details and ornaments. Today a typical tourist photographs items of interest and pastes them in an album or posts them on an internet site. In future, such images can become a starting point for further study when one has returned home. We send the image to a virtual reference room, which identifies the item, and traces its origins back to the symbol for Mother Earth as Ninhursag. We then receive a series of other examples. The extent to which we pursue this depends on the depth of our interest.

Such examples show how the new technologies allow us to take the power of reference rooms and the collective memory of memory institutions into the fields, streets and indeed all walks of everyday life. The corollary is that the we can equally take imagery of the everyday world into our studies within libraries or within our homes. In the past, we could consult an atlas. Soon we shall be able to consult any place on earth at different at different scales from views in space to 1:1 scale imagery and potentially also at the microscopic level.

- 1 book = 500 pp.
 - 1MB uncompressed – 300KB compressed
 - 108 to 3x 108 books = ~1014 bytes = 100 terabytes
- Over 100 million computers on the Internet
 - At 1 GB each, >100 petabytes now
- 1 GB of disk costs ~\$3
 - 100 terabytes < \$300 thousand to \$1 million

Figure 14. Statistics from Raj Reddy (CMU) to show that we can store everything.¹⁹⁶

Scholarship is, of course, about much more than simply identifying isolated decorations, objects, buildings or places. It is also about exploring contexts, versions of stories, traditions, possible reasons for events. In memory institutions of the future where image recognition techniques are used to search for copies, versions and related iconographical materials whole new chapters of studying our past and present are imaginable.

6.5. Towards a WONDER

A largely romantic memory of the original Library of Alexandria continues to inspire visions of a single building or complex which would house all records. Examples include: the new Library of Alexandria and the G7's Bibliotheca Universalis project (1995), for which the G7 was "unable" to find funding. In 1997, NEC announced the world's first virtual library: the Universal Digital Library.¹⁹⁷ This is linked with Japan's plans to scan all the 8.14 million books of the National Diet library (Tokyo)¹⁹⁸ and with a range of projects in e-content being coordinated by the National Institute of Informatics (NII). In April 2000, China's National Digital Library Corp, was digitizing 200,000 pages of information per day.¹⁹⁹ In 2000, Raj Reddy and Gloriana St. Clair (CMU) in conjunction with Beijing²⁰⁰ and Tsinghua University, launched a project a Million Book Project (MBP).²⁰¹ This project is part of larger vision for:

a Universal Library starting with a free-to-read, searchable collection of one million books available to everyone over the Internet by the year 2008. This first major project toward building a Universal Library is named the Million Book Digital Library Project (MBP). Within 10 years, it is expected that the collection will grow to 10 Million books. The result will be a unique resource accessible to anyone in the world, 24 x 7, without regard to nationality or socioeconomic background.²⁰²

In 2001, this became a vision of a Universal Digital Library (UDL),²⁰³ which is to have 1 million books by 2008 and 10 million books "within 10 years". By 2003, this included India²⁰⁴ and Egypt.²⁰⁵ By November 2005, over 600,000 books were scanned: 170,000 in India, 420,000 in China, and 20,000 in Egypt.²⁰⁶ Reddy starts from the premise that we can store everything. A closer look at his plans (figure 14) reveals that his figures greatly underestimate the potential scale of the venture. He calculates 100 terabytes for the full contents of a million books. In 2000, the Library of Congress was making daily backups of 40 terabytes mainly for the titles of its books. The Library of Congress is making a new high-level scan of the Gutenberg Bible at 767 MB per folio which implies about ½ terabyte for a single book. At that rate, the 100 terabytes that Reddy estimates for 1 million books would cover 200 books. Similarly the estimates that there were "no more than 10 million unique book and document editions before the year 1900, and perhaps 100 million since the beginning of recorded history"²⁰⁷ underestimate wildly the scale of the challenges ahead. A related vision by Kevin Kelly, a well known futurist at *Wired* underestimates the challenge even more dramatically:

The universal library should include a copy of every painting, photograph, film and piece of music produced by all artists, present and past. Still more, it should include all radio and television broadcasts. Commercials too. And how can we forget the Web? The grand library naturally needs a copy of the billions of dead Web pages no longer online and the tens of millions of blog posts now gone — the ephemeral literature of our time. In short, the entire works of humankind, from the beginning of recorded history, in all languages, available to all people, all the time.

This is a very big library. But because of digital technology, you'll be able to reach inside it from almost any device that sports a screen. From the days of Sumerian clay tablets till now, humans have "published" at least 32 million books, 750 million articles and essays, 25 million songs, 500 million images, 500,000 movies, 3 million videos, TV shows and short films and 100 billion public Web pages. All this material is currently contained in all the libraries and archives of the world. When fully digitized, the whole lot could be compressed (at current technological rates) onto 50 petabyte hard disks. Today you need a building about the size of a small-town library to house 50 petabytes. With tomorrow's technology, it will all fit onto your iPod. When that happens, the library of all libraries will ride in your purse or wallet — if it doesn't plug directly into your brain with thin white cords.²⁰⁸

As we noted earlier, the WorldCat catalogue is now 76 million unique titles and even this is but a small fraction of what exists. Ultimately, such miscalculations and under-estimates of technical details and scale of the process are of secondary significance. They have inspired a vision that the challenges can be met and have brought into play China, India, Egypt as well as the United States. By 2003, their combined vision had expanded to create a library of 20 million books by 2020.²⁰⁹

On 14 December 2004, Google in conjunction with five major libraries announced that they would scan the full texts of 15 million books. By 2005, this had become 20 million books²¹⁰. In May 2005, as Europe was deciding on a reply to Google, the Bibliotheca Alexandrina (BA) joined plans for a World Digital Library.²¹¹ In June 2005, the Library of Congress officially announced new plans for this new World Digital Library (WDL).²¹² In November 2005, Brewster Kahle announced his vision of a World Online Library through the Open Content Alliance.²¹³ Their aim is to scan 1 Million books. In Canada, the CBC's Alouette Project foresees the equivalent of 4 million books.²¹⁴ Also in November 2005, Google contributed \$3 Million to WDL.²¹⁵ On 1 December 2006, UNESCO and the Library of Congress hosted a meeting on the World Digital Library project.²¹⁶ The Library of Congress has avoided mentioning specific numbers of books in the project, but two features of the proposed project are particularly noteworthy. First, it aims at the cultures beyond Europe. Second, there is a vision to include the World Wide Web within the scope of the project:

Such a project could begin—but need not end—with multinational efforts through UNESCO to begin digitizing online “Memory” projects for three great cultures each of which lies beyond Europe and involves more than one billion people: Chinese East Asia, Indian South Asia, and the worlds of Islam stretching from Indonesia through Central and West Asia to Africa. ...

The Library of Congress and eight consortia involving 36 other American institutions, are well along in figuring out what to save from, and how to preserve the vast flood of ephemeral and unfiltered material on the World Wide Web. The Library of Congress has already harvested 26 terabytes of these evanescent Web sites; and our initial partners are expected to gather in 60 terabytes of “at-risk” digital content.²¹⁷

Universal Digital Library - Japan	8+
China Digital Library	2+
Universal Digital Library - US-India	20
Google	20
Open Content Alliance	1
CBC (Canadian Broadcasting Co.)	4

	55+

Figure 15. Estimates of books for full text scanning in the next 15 years in some major projects outside Europe.

While Raj Reddy's project foresees 100 terabytes for a million books in its first phase; the Library of Congress foresees 86 terabytes simply for the contents of the World Wide Web in its first phase. China was committed to 20 TB of online content by 2005.²¹⁸ The combined effect of these plans is incredibly impressive. On closer view, the US vision focusses on three economically wealthy areas of the world. It foresees a World Digital Library where Sub-Saharan Africa, South America²¹⁹, Oceania, Russia and Europe are effectively sidelined or forgotten. The Library of Congress vision is not for a geographical world library: it is a world library of areas that also reflects American political and economic priorities. .

Meanwhile, Europe's public awareness of these developments has focussed mainly on Google, whose announcement sparked a reaction from the Bibliothèque Nationale de France (BNF), and led to a commitment to scan the full-texts of 6 million books by 2010. Europe thus faces two alternatives. First, if it maintains its present commitment, which is dwarfed²²⁰ by the 55 million full-texts being scanned elsewhere (figure 15), Europe will quickly dwindle to being a follower on the global scene. Second, Europe could decide to build a Distributed Electronic European Resource (DEER) and also become a serious partner in a World Online Networked Digital Electronic Resource (WONDER).

6.5.1 European University for Culture

There are plans for a new European University for Culture with initial seats in Luxembourg, Berlin, Venice, Madrid and Paris. Such an institution could become a vanguard in experiments how a new multilingual approach to culture through networked memory institutions might work in practice. In the past, the great changes came partly from technology and largely through shifts in mindset, policy decisions that shifted developments to a new level.

- Global Brain (Gehirn der Welt)
- Global Intelligence
- Global Mind
- Global SuperBrain
- Global Superorganism
- Hive-Mind
- Mémoire mondiale
- Noospheric Brain (organ of collective human reflection)
- Planetary Brain
- Social Brain
- Super Minds
- Super-Brain

Figure 16. Some names for the quest for a Global Brain (*Gehirn der Welt*).

Today, in a world where Europe represents approximately 12% of the world population, our memory institutions have a new challenge: to help form awareness of multiple approaches to knowledge, to acknowledge the contributions of other great cultures, while not losing our own values.²²¹

7. Conclusions

There are many visions of the future. One recent exploration of libraries in 2040 A.D. explored seven alternatives: 1. alphabet hotel; 2. bibliothèque d'amis; 3. hormone library; 4. partisan library; 5. survival library; 6. virtual library of the future and 7. Brabant library (cf. Appendix 5). Our analysis has focussed only on the sixth of these possible scenarios. We began with a brief summary of media in the past two millennia, tracing a shift from stone, to manuscripts, printed books and more recently digital media.

We traced how the recorded world of memory institutions grew from a few thousand items in the 15th century, to a couple hundreds of thousands due to the advent of printing in the West. The next major shift came through a policy decision, which introduced a vision of national libraries. As a result, collections grew from a quarter of million to collections of 12-24 millions (figure 1). The advent of electronic networks has transformed the numbers of titles to as much as 78 million individual titles and over 1 billion copies in a single network.

We found that long-term preservation is not simply a matter of moving or “translating” knowledge from the previous to a new medium. The shift from manuscripts to printing entailed becoming much more aware of manuscripts than previously and developing better strategies for their long-term preservation. The advent of printing brought new kinds of communication in the form of secondary literature and led, gradually to new levels of bibliographic control. Similarly, the rise of born digital resources is bringing unexpected developments. The ability to scan full-text introduces needs for more detailed bibliographic control and greater integration of existing tools.

Paradoxically the rise of the born digital resources, which were initially viewed conceptually as a net, web or a cloud, are increasingly becoming embedded and inter-linked with physical and other worlds. This is bringing new importance to recording worlds: new mechanisms for capturing, mapping, surveying, sensing, analysing, recognizing objects and images of the physical world. In the past, we went to libraries to study the recorded world. In future we may also use recording worlds to study the riches of libraries.

If such visions are to become a practicable reality we need to expand our present notions of a European Digital Library into a Distributed European Electronic Resource with a) a distributed repository; b) a digital reference room and c) a virtual agora which bridges enduring knowledge with trends in personal and collaborative knowledge. We also need a Knowledge Organization Institute that takes up earlier visions of the Mundaneum and the Bridge and works towards a more systematic approach to relations as foreseen by Perreault.

Between 1950-2000, the idea of scanning texts evolved gradually from initial attempts to a couple hundred thousand books. Since 2000, the scale of projects for full-text scanning has shifted into the millions with a possible 60 million books in the next 10-15 years. While Google is the most publicized player, these visions now include UNESCO, the Library of Congress, Japan, China, India and Egypt. Europe is also involved, but if it continues to think in small

scale terms and remains content with present plans to scan only 6 million books, then it will be eclipsed by the realities of various large-scale projects around the world.

Europe's complex history has given it a double heritage. In the eyes of critics, Europe led many of the waves of colonialism and imperialism, whereby it often imposed its values on persons around the world. But there is also another Europe, whereby it has constantly studied cultures of others. In the 1140s this began with a conscious decision by thinkers such as Abbot Suger to launch a major translation campaign from foreign languages, especially Arabic into Latin. This included translating the Koran into Latin so that they could understand this religion. Suger's efforts at St. Denis provided the seeds of Gothic art and culture. Renaissance Italy saw efforts in translating and understanding Greek, Chaldaean and other cultures. Subsequently such efforts often created or revived awareness that the original peoples had forgotten: e.g. the efforts of Jesuit priests to catalogue native languages; Jean-François Champollion's translation of the Rosetta stone to recover Egyptian hieroglyphics; Paul Pelliot's studies of Anghor Wat to recover a forgotten civilization or Max Müller's incredible studies of the *Sacred Books of the East*.

Given these complex traditions Europe faces a choice. On the one hand, it can remain numbed by some earlier mistakes and retreat from the world scene. On the other hand, given its rich traditions and deep experiences in linguistic and cultural diversity, it is poised to take up anew the vision of a global brain (*Gehirn der Welt*) conceived exactly a century ago (figure 16). It could develop a Distributed European Electronic Resource (DEER) and potentially play an important role in the creation of a WONDER (World Online Networked Digital Electronic Resource).

Acknowledgements

I am deeply grateful to the German National Library for the honour of being asked to write this survey and vision statement. Some of its basic ideas, notably the idea of a DEER and WONDER reflect a vision developed by a community of cultural institutions in E-Culture Net. The author is conscious of sometimes taking a rhetorical and/or polemical stance on some issues. This stance, intended to provoke reflection, is entirely the personal responsibility of the author and does not expressly reflect the views of either the National Library or other bodies.

Appendix 1. Examples of the Range of Major Collections

Bibliothèque Nationale de la France

Significant Works

- Printed Papers Department owns about 12 millions books, among them two copies of Gutenberg Bible.
- Geographical Maps (1st collection in the world)
- World's richest collection of engravings (12 million) and photos (2 million).
- 250,000 manuscripts including the Dead Sea Papyrus and medieval psalm books with painted works such as Carolus Magnus and Louis IX's collections.
- 600,000 coins and medals.
- Antiques
- Music and performing arts.²²²

Berlin State Library:

- 10 million books
- 4,400 incunabula
- 18,300 occidental manuscripts
- 40,000 oriental manuscripts
- 250,000 autographs
- 66,350 music autographs
- 1,400 personal archives
- 450,000 print music editions
- 960,000 maps and atlases
- 38,000 subscription periodicals and monographic series
- 180,000 early newspaper volumes and 400 subscription newspapers
- Diverse electronic databases
- 2.3 million microfiches and microfilms
- 13.5 million images in the picture archive²²³

Bavarian State Library

- ca. 8.8 million books
- ca. 85,700 manuscripts, including:
 - Manuscript A of the Nibelungenlied
 - Freising manuscripts
 - the Carmina Burana
- more than 44,000 subscription periodicals and monographic series (Europe's second largest holding)
- 18,667 incunabula (the world's largest holding).²²⁴

Appendix 2. List of Image Collections available in the California Digital Library

Collection	Number of Images
1. AMICA	100,000+
2. David Rumsey Historical Map Collection	11,000+
3. Estate Project for Artists with Aids	3,000+
4. Farber Gravestone Collection	13,527
5. Hartill	16,660
6. Hoover Institution Poster Collection	53
7. Japanese Historical Maps	896
8. LUCI (Library of UC Images)	3,690
9. Museums and the Online Archive of California	77,193
10. Saskia	26,000
11. SPIRO: Arch+Arts+Places	65,000
12. Tebtunis Papyri	55
13. UCSF Demonstration Project	99

	313,173

Appendix 3 UCSF Subject Headings: Those in blue are subsets of health/medicine.

[Allied Health](#)

Arts & Humanities

Biology

Business

[Consumer Health](#)

[Dentistry](#)

[Disease](#)

Education

General Reference

[Health Administration & Policy](#)

[History of Health Sciences](#)

[Medicine](#)

[Nursing](#)

[Pharmacy & Pharmacology](#)

[Physical Medicine & Rehabilitation](#)

Physical Sciences

[Public Health](#)

Social & Behavioral Sciences

Technology

[Veterinary Medicine](#)²²⁵

Appendix 4. Copyright

A further concern is copyright. Publishers have developed a two-faced strategy. With respect to their investors they emphasize their profits. With respect to their authors they emphasize their lack of money and use this to leave authors with ever fewer rights. In one recent case, a famous publishing company at the oldest university in England, wanted not just permission to publish without charge, but all rights in perpetuity. When this was resisted, they refused to publish the author's keynote at an international conference. Companies such as Google

officially scan out of print works. In practice they scan large sections of contemporary works without consulting authors.

In 2018, current copyright laws will cease to apply in the United States. One plan is to have a Public Domain Enhancement Act,²²⁶ whereby authors would need to pay a minimal fee annually for each publication. In cases where authors fail or forget to do so, their copyright privileges would revert to a central organization. One of the champions of this proposed act is also on the board of the Creative Commons.²²⁷ Another member of the same board is Co-directing Project iCampus, the MIT/Microsoft Alliance for Research in Educational Technology Co-directing the MIT Educational Technology Council.²²⁸ WIPO offers an international viewpoint. Still needed is a European approach that reflects the realities of the Union beyond traditions in an individual country. Unless Europe, develops a clear position, conventions from abroad will tend to be applied, and further restrict permission to access knowledge.

Appendix 5. Rob Bruijnzeels: Libraries 2040: The Netherlands Public Library Association

1. alphabet hotel – “the memory;” books everywhere; the library is the memory, so it now becomes a hotel lodge; actually combined a hotel and a library; was the first 24/7 library in the Netherlands; was staffless - just books; books weren’t stolen, but the bookmarks were/
2. bibliotheque d’amis
3. hormone library – a library designed by young adults; asked them about the future; the kids said they are changing second by second because of their hormones; so created several spaces in the public library that express those feelings; no books in these “emotional interfaces;” it’s where you “love to go”
4. partisan library – library designed by even younger kids at a special event; libraries would replace the attraction of prohibition with the excitement of discovering hidden treasures; had wonderful librarians; they designed the library as a landscape
5. survival library – the idea of a hidden library; largest library in the Netherlands – 45 square acres, all outdoors; only has 30 books; could only find them by solving puzzles; book was related to the place where you found it; goal was to find all 30 books (which takes 3–4 days);
6. virtual library of the future - collective memory of all of the people, organized
7. the brabant library – MVRDV; like Seattle Public Library; will it still make sense to hold on to the existing concept? or is this the right moment for something new?²²⁹

Notes

¹ Universal Digital Library: <http://udl.iita.ac.in/>

² Wikipedia: http://en.wikipedia.org/wiki/Library_of_Alexandria; Some sources say 47 B.C. <http://www.bede.org.uk/library.htm>. The historical event continues to have a political dimension.; VA Mohamad Ashrof, "Who destroyed Alexandria Library?", Milli Gazette, New Delhi, 01, 12, 2002: <http://www.milligazette.com/Archives/01122002/0112200252.htm>

³ For a simple example. See: <http://www.athabascau.ca/emd/mags/Design/cuneiform.jpg>
Efforts such as the Cuneiform Digital Paleography Project are translating this medium into electronic form: <http://www.cdp.bham.ac.uk/>

⁴ Code of Hammurabi: http://en.wikipedia.org/wiki/Code_of_Hammurabi

⁵ Rosetta Stone: <http://www.thebritishmuseum.ac.uk/compass/ixbin/goto?id=OBJ67>

⁶ The texts in old Persian, Elamite and Babylonian are inscribed on the walls of a cliff in Hamadan, Iran, c. 6th c. B.C. near the city of Ecbatana, Persia:

<http://www.mojesafar.com/IMAGES/places/ganjnameh.jpg>

⁷ Forest of Stele: <http://www.cs.iastate.edu/~jia/pictures/2005/xian-tablets/b6.JPG>

⁸ Chinese Script: <http://www.crystalinks.com/chinascript.html>

⁹ Dioscorides, [Vienna](#), [Österreichische Nationalbibliothek](#), Cod. med. gr. 1., 6th c. A.D.: <http://en.wikipedia.org/wiki/Image:ViennaDioscoridesFolio483vBirds.jpg>

¹⁰ Corvinus, Eg.:

<http://www.corvina.oszk.hu/images/Codex/GEHAB/gehabcodguelf8511Aug2.jpg>

¹¹ Cited on the page of the Manuscripts Catalogue of the British Library.

<http://www.bl.uk/catalogues/manuscripts/>

¹² The Vatican collection began nominally in the 4th century

Cf. <http://www.infoplease.com/ce6/ent/A0850545.html>

¹³ The Vatican Library a Chronology:

<http://www.ibiblio.org/expo/vatican.exhibit/exhibit/History.html#chronology>

Cf. <http://www.ibiblio.org/expo/vatican.exhibit/exhibit/History.html>

Three decades later in 1481, when the librarian Platina made a catalogue, there were 3,800 books.

¹⁴ The Vatican Library: A Paper Treasure in Rome.

<http://www.romanguide.com/vaticancity/vatican-library.html>

¹⁵ Berlin State Library: http://en.wikipedia.org/wiki/Berlin_State_Library

¹⁶ BNF: <http://www.linkparis.com/Bibliotheque-Nationale-de-France.htm>

¹⁷ I am grateful to Rachel Stockdale, Department of Manuscripts, The British Library for these statistics and further helpful information on British developments.

¹⁸ Library of Congress, Manuscript Reading Room: http://www.loc.gov/rr/mss/mss_abt.html

¹⁹ University of Chicago Library: <http://www.lib.uchicago.edu/e/about/>

²⁰ Medieval Manuscripts: http://home.hetnet.nl/~otto.vervaart/manuscripts_me_eng.htm

²¹ Diamond Sutra, British Museum:

<http://www.bl.uk/onlinegallery/themes/landmarks/diamondsutra.html>

²² Michael Giesecke, *Der Buchdruck in der frühen Neuzeit - Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien.*, Frankfurt: Suhrkamp, 1991.

http://www.michael-giesecke.de/giesecke/menue/index_h.html

²³ Cassiodorus Senator, *An Introduction to Divine and Human Readings*, Book One, by (New York: Columbia University Press, 1946) for an early example of lists of library desiderata. James J. O'Donnell *Avatars of the Word: From Papyrus to Cyberspace*, Cambridge, Mass.: Harvard University Press, 1998, p. 30.

²⁴ National library, Wikipedia; http://en.wikipedia.org/wiki/National_library

²⁵ Bodley: <http://www.bodley.ox.ac.uk/dept/readerserv/history/british.htm>

²⁶ Antonio Panizzi, Wikipedia: http://en.wikipedia.org/wiki/Antonio_Panizzi

²⁷ Library of Congress, FAQs: <http://www.loc.gov/about/faqs/>

The Library of Congress and the World Beyond its Walls.

http://books.nap.edu/openbook.php?record_id=9940&page=144

Some claim it has 29 million books. Cf. The Wikipedia article on Library of Congress (http://en.wikipedia.org/wiki/Library_of_Congress#_note-about) claims:

“The Library of Congress holds about 130 million items with 29 million books against approximately 150 million items with 25 million books for the British Library. ^{[5][6]}

^{5^a b} [About the Library](#). Library of Congress. Retrieved on **2006-08-08**.

^{6^a b} [Did You Know?](#). British Library. Retrieved on **2006-08-08**.”

²⁸ Russian State Library, Infoplease: <http://www.infoplease.com/ce6/ent/A0909792.html>

Another source claims 40 million for the reasons outlined in note 3. cf. “Lenin Library Succumbs to Decrepitude; Crumbling Outside, Chaotic Inside, Historic Moscow Facility Is Ordered Closed,” *The Washington Post*, November 23, 1991”:

<http://www.highbeam.com/doc/1P2-1096704.html> . During the mid- and late 20th centuries the collection, renamed the Lenin library went through a temporary crisis. Those days are over and the library remains one of the greatest in the world.

²⁹ The British Library site claims it has: “13 million books, 920,000 journal and newspaper titles, 57 million patents, 3 million sound recordings, and so much more.”: <http://www.bl.uk/>. The Wikipedia article, cited in note 4 above, claims that the British Library has 25 million books.

³⁰ Harvard University Library: <http://hul.harvard.edu/hgproject/faq.html>

³¹ Devin Crawley, “The Infinite Library. Researchers are welcoming Google's plan to digitize millions of books, but the implications for libraries are profound,” *U of T Magazine* Toronto, 2005: <http://www.magazine.utoronto.ca/05autumn/library.asp>

³² Yale University, Wikipedia: http://en.wikipedia.org/wiki/Yale_University

³³ Andrea Lynn, “Where do you put 10 million books? Construction of library warehouse under way,” *News Bureau, University of Illinois at Urbana Champaign*, 10/14/03: <http://www.news.uiuc.edu/NEWS/03/1014warehouse.html>

³⁴ In an electronic age one would think that the question of how many books are in the leading libraries would be straightforward. It is not. One problem is to decide what constitutes a title. The Anglo Saxon tradition claimed that a periodical series constitutes one title. The Russian tradition, by contrast, considered every individual volume of a series as a separate book. As a result, during the 1970s the World of Learning listed the Lenin Library (Moscow) and Saltychov Schedrin (Leningrad now Saint Petersburg) as being at least twice the size of the British Library and the Bibliothèque Nationale in France. A second reason for discrepancies is nationalism.

³⁵ A report from China claimed that in 2001 their National library had over 24 million books. They claimed at the time to be number 5 in terms of numbers of books and number 3 in terms of physical size. See: “China to Enlarge National Library to the World's Third Largest. China's national library in Beijing will be enlarged to 240,000 square meters, the world's third largest in terms of size, and equipped with an online collection to supplement its regular books. People's Daily, Beijing, 19 December 2001:

http://english.peopledaily.com.cn/200112/19/eng20011219_87053.shtml

This claim of ranking fifth is also made on the site of the library itself:

<http://www.nlc.gov.cn/en/aboutus/history.htm>

Harvard claims to be 4th in the world with 15 million (http://en.wikipedia.org/wiki/Harvard_University_Library) ignoring that the New York Public claims it has 20 million. <http://www.nypl.org/press/2004/google.cfm>

³⁶ NYPL Partners with Google to Make Books Available Online. Full Text of Thousands of Books to Be Searchable on the Web, New York, December 14, 2004:

<http://www.nypl.org/press/2004/google.cfm>

The New York Public Library is a system of 89 libraries, and is unique in combining major research facilities and a branch library system within one overall structure. The four Research Libraries have collections in more than 3,000 languages and dialects.... NYPL has over 50 million items in its vast collections, 20 million of which are books.

³⁷ The University of Michigan has 7.8 million books. Stanford which claimed to have 7.6 million in 2004 claimed to have nearly 9 million in 2005. UCLA has 8 million books. Cf. “Stanford and Google Book Search Statement of Support and Participation December 2005:

http://wwwsul.stanford.edu/about_sulair/news_and_events/stanford_google_project.html

³⁸ University of Chicago: <http://www.lib.uchicago.edu/e/about/>

Our resources total more than seven million printed works, increasing at the rate of 150,000 volumes per year. Over thirty million manuscripts and archival pieces, 420,000 maps and aerial photographs, and large sets of microform materials complement the printed collections.... expand the University Library by more than 3.5 million volumes and create by June 2009 one of the nation's largest university collections of materials under one roof.

³⁹ Scott Carlson, Jeffrey R. Young, “Google Will Digitize and Search Millions of Books From 5 Leading Research Libraries,” *Chronicle of Higher Education*, 14 December 2004:

<http://chronicle.com/free/2004/12/2004121401n.htm>

⁴⁰ Cf. “Stanford and Google Book Search Statement of Support and Participation December 2005:

http://www-sul.stanford.edu/about_sulair/news_and_events/stanford_google_project.html

⁴¹ Google Library Partnership, University of Michigan New Service, 2005:

<http://www.umich.edu/news/index.html?BG/google/index>

⁴² New York Public Library 20 million

Harvard 15

Stanford 7.6

Michigan 7.8

Oxford 6.5

56.9 million

Estimates of Contents of Five Research Libraries in Google Project.

⁴³ Bernhard Fabian, *Buch, Bibliothek und geisteswissenschaftliche Forschung : zu Problemen der Literaturversorgung und der Literaturproduktion in der Bundesrepublik Deutschland*, Göttingen: Vandenhoeck & Ruprecht, 1983

<http://www.ib.hu-berlin.de/about/gesch/bfabian.htm>

⁴⁴ German National Library, Wikipedia:

http://en.wikipedia.org/wiki/German_National_Library.

This came by a merger of “the Deutsche Bücherei Leipzig (founded 1912, later the national library of East Germany) and the Deutsche Bibliothek Frankfurt (founded 1947, later the national library of West Germany).”

⁴⁵ Gerd-J. Bötte, “A virtual National Library for Germany – the Sammlung Deutscher Drucke [Collection of German Printed Works],” *World Library and Information Congress: 69th IFLA*

General Conference and Council, 1-9 August 2003, Berlin.

http://www.ifla.org/IV/ifla69/papers/140e_trans-Boette.pdf

⁴⁶ 13.6 million are held in Leipzig, approx. 8 million in Frankfurt am Main and roughly 1.3 million by the Deutsches Musikarchiv: Deutsche Bibliothek:

http://www.ddb.de/eng/wir/ueber_dnb/dnb_im_ueberblick.htm

⁴⁷ Berlin State Library, Wikipedia: http://en.wikipedia.org/wiki/Berlin_State_Library

⁴⁸ BSB: <http://www.bsb-muenchen.de/>

Bayerische Staatsbibliothek, located in Munich, Germany, is the second largest academic library in German-speaking countries with a collection of just under 9 million books, 85,000 manuscripts and more than 44,000 printed and electronic journals.

⁴⁹ PORT: Libraries in German-speaking countries:

<http://port.igrs.sas.ac.uk/GERlibrariesGER.htm>

⁵⁰ SUB Göttingen: http://de.wikipedia.org/wiki/SUB_G%C3%B6ttingen

⁵¹ Herzog August Bibliothek: <http://www.hab.de/>

Cf. Robert L. Kusmer, "Transatlantic Connections: a Tour of German Libraries and Publishing Houses," *WESS Newsletter*, Spring 2003, Vol. 26, no. 2 :

<http://www.dartmouth.edu/~wessweb/nl/Spring03/Kusmer.html>

⁵² Irina L. Klim, "Libraries and Library Collections in Russia: what they will be like in 2005," 2000: http://digital.casalini.it/retreat/2000_docs/klim.doc

⁵³ For bibliographic references to these works cf. the Author's New Media:

<http://sumscorp.com/kavai/newmedia/>

⁵⁴ Dewey Decimal Classification, Wikipedia:

http://en.wikipedia.org/wiki/Dewey_Decimal_Classification

⁵⁵ C. Berkvens-Stevelinck, Prosper Marchand, *la vie et l'oeuvre (1678-1756)*, Leiden 1987. (Studies over de Geschiedenis van de Leidse Universiteit, 4) .

⁵⁶ IBZ: <http://www.ub.fu-berlin.de/literatursuche/datenbanken/titel/KOB11958.html>

Since 1983, an online version, also known as *Der Dietrich*, indexes over 10,700 English and European periodicals, books, and dissertations.

⁵⁷ Friedrich Naumann, *Das Gehirn der Welt*, *Süddeutsche Monatshefte*, 4, 1907, 759-764.

For more literature on the world brain see the author's Understanding New Media under the subject World Brain at: <http://sumscorp.com/kavai/newmedia/>

⁵⁸ Research Committee on Conceptual and Terminological Analysis RC35 :

<http://www.ucm.es/info/isa/rc35.htm>

⁵⁹ Henry Teune, "Committee on Conceptual and Terminological Analysis", Prepared for the 2000 Research Council Conference of the International Sociological Association, *Social Transformations at the Turn of the Millennium: Sociological Theory and Current Empirical Research*, Montreal, July 28-30, 2000: <http://www2.hawaii.edu/~fredr/teune8.htm>

⁶⁰ For a survey see the author's New Media under Classification:

<http://sumscorp.com/kavai/newmedia/>

⁶¹ DBI: http://de.wikipedia.org/wiki/Deutsches_Bibliotheksinstitut

⁶² Virtuelle Fachbibliothek: <http://www2.sub.uni-goettingen.de/cgi-bin/ssgfi/anzeige.pl?db=meta&nr=000524&ew=SSGFI>

⁶³ Vascoda: <http://www.vascoda.de/>

⁶⁴ E.g. EUBAM: <http://www.eubam.de/> and BRICKS

⁶⁵ Paul Otlet: <http://www.db.dk/bh/Core%20Concepts%20in%20LIS/articles%20a-z/otlet.htm>

⁶⁶ Mundaneum: <http://www.mundaneum.be/index.asp?ID=340>:

Musée du livre, association qui avait pour objet l'étude du livre au sens large (caractères, encre, papier, composition, diffusion, ...), l'Union des Associations Internationales (UAI), créée au départ de 170 associations et enfin le célèbre Office

International de Bibliographie (OIB), dont l'objectif était d'établir un Répertoire Bibliographique Universel (RBU) composé de fiches mentionnant chacune la référence bibliographique d'un ouvrage paru, ainsi que l'indice CDU définissant au moyen d'un langage composé de chiffres et de signes, la nature de l'ouvrage désigné. Regroupées dans des meubles spécialement conçus, les fiches (12 000 000) étaient généralement classées alphabétiquement ou par thèmes.

⁶⁷ Mundaneum: <http://www.mundaneum.be/index.asp?ID=240>

⁶⁸ UDC: <http://www.udcc.org/about.htm>

⁶⁹ What is Information Science and How is it Related to Library Science?

<http://www.cas.usf.edu/lis/lis6260/lectures/infosci.htm>

70. W. Boyd Rayward, "The International Federation for Information and Documentation (FID)", *Encyclopedia of Library History*, edited By Wayne A. Wiegand and Don G. Davis, Jr. (New York: Garland Press, 1994), pp. 290-294.

<http://people.lis.uiuc.edu/~wrayward/otlet/FIDHIST2.htm>

Cf. P. Peris Bonet, *La Constitucion de la Documentacion (1): Otlet*:

[http://www.uv.es/Ciencia/Esquemas%20Cap-2\(Int%20Inf%20Cient%20Tecnica\)2004-05.pdf](http://www.uv.es/Ciencia/Esquemas%20Cap-2(Int%20Inf%20Cient%20Tecnica)2004-05.pdf)

⁷¹ Vannevar Bush, "As we may Think", *Atlantic Monthly*, July 1945, Volume 176, No. 1; 101-108: <http://www.theatlantic.com/doc/194507/bush>

⁷² Paul Otlet, *Monde: essai d'universalisme -- connaissance du monde; sentiment du monde; action organisée et plan du monde*. Brussels, Editions du Mundaneum, 1935 cited in:

<http://www.laetusinpraesens.org/docs/otlethyp.php>

⁷³ In Canada, Bob Macdonald implemented one of the first library circulation systems at a North American academic library in 1965.⁷³ In Ontario, GEAC⁷³ began in 1971.

GEAC: <http://www.answers.com/topic/extensity-company-private>

The University of Toronto Library Automated System (UTLAS) began in 1974.

UTLAS: <http://www.libsci.sc.edu/BOB/istchron/ISCNET/ISC1970.HTM>

It was sold to CARL systems (Denver, 1991).

Carl Systems Acquires Utlas T/Series 50 Libraries, May 1991:

http://findarticles.com/p/articles/mi_hb3329/is_199105/ai_n8039664

CARL was then acquired by TLC:

Marshall Breeding, The Library Corporation Acquires CARL Corp.:

<http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=17783>

Cf. <http://www.library.ubc.ca/libsys/history.html>;

<http://www.libsci.sc.edu/BOB/istchron/ISCNET/ISC1970.HTM>

TLC History: <http://www.tlcdelivers.com/tlc/tlchistory.asp>

⁷⁴ See, for instance, W. Boyd Rayward, *A History of Computer Applications in Libraries: Prolegomena*, *IEEE Annals of the History of Computing*, New York Volume 24, Issue 2, April 2002:

<http://portal.acm.org/citation.cfm?id=570399&dl=acm&coll=&CFID=15151515&CFTOKEN=6184618>.

Cf. Marshall Breeding, "Computers in Libraries [September 2003]", *Library Technology Guides* 74: <http://www.librarytechnology.org/ltg-displaytext.pl?RC=10647>

Attached to this is: This graphic shows the history of mergers and acquisitions in the library automation industry: <http://www.librarytechnology.org/automationhistory.pl>

Notable about this very informative chart is how it moves the founding of GEAC from 1971 to 1977; UTLAS from 1974 to 1981 and does not mention a single contribution from Europe.

⁷⁵ RLG Features: http://www.rlg.org/en/page.php?Page_ID=327

⁷⁶ RLG: http://www.rlg.org/en/page.php?Page_ID=327; Big Bibliographic Services: OCLC and RLG: <http://www.librarytechnology.org/ltg-displaytext.pl?RC=12100>

RLG union catalog http://www.rlg.org/en/page.php?Page_ID=174
150,571,870 / over 48 million: 1st number—individual bibliographic records; 2nd number—
titles covered

⁷⁷ OCLC: <http://www.oclc.org/about/history/default.htm>

⁷⁸ History of the Dublin Core Metadata Initiative: <http://dublincore.org/about/history/>

⁷⁹ In 2001 Worldcat had 45 million unique records and 750 million library location listings
<http://www.oclc.org/research/grants/reports/perrault/intro.pdf>

⁸⁰ Worldcat Database: <http://www.oclc.org/worldcat/database/default.htm>

⁸¹ OCLC Sitesearch 4.0: <http://www.oclc.org/research/publications/archive/releases/1998-04-02b.htm>

⁸² Paula J. Hane, “OCLC and WLN Begin Negotiations to Merge”, 16 November, 1998:
<http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=17872>

⁸³ OCLC Pica: http://en.wikipedia.org/wiki/OCLC_PICA

⁸⁴ RLG to Combine with OCLC: http://www.rlg.org/en/page.php?Page_ID=20942;
http://www.ringgold.com/biblio-tech/si_pd.cfm?pid=10&articleid=1306&issueno=73

⁸⁵ OCLC: <http://www.oclc.org/about/default.htm>

⁸⁶ “OCLC and RLG to Merge”, *Biblio Tech Review*, 6 May 2006:

http://www.ringgold.com/biblio-tech/si_pd.cfm?pid=10&articleid=1306&issueno=73

OCLC, Wikipedia: <http://en.wikipedia.org/wiki/OCLC>

⁸⁷ OCLC, CAMIO: <http://www.oclc.org/camio/default.htm>

⁸⁸ GBV: http://www.gbv.de/vgm/info/biblio/01VZG/01ueber_die_VZG/index

⁸⁹ COPAC: <http://copac.ac.uk/faq/#what>

⁹⁰ CCFR http://www.ccf.fr/bnf.fr/rnbc_visu/acc1.htm

⁹¹ NCC http://www.utwente.nl/ub/help/glossary/glossary.doc/dutch_union_catalogue.html

⁹² REBUIN [http://rebiun.crue.org/cgi-](http://rebiun.crue.org/cgi-bin/abnetop/X16180/ID1852780176/NT1?ACC=299&HELPIID=proy)

[bin/abnetop/X16180/ID1852780176/NT1?ACC=299&HELPIID=proy](http://rebiun.crue.org/cgi-bin/abnetop/X16180/ID1852780176/NT1?ACC=299&HELPIID=proy)

⁹³ ICCU: 7.795.187 million: <http://www.iccu.sbn.it/genera.jsp?s=110>

⁹⁴ LIBRIS http://websok.libris.kb.se/websearchhelp/help_eng/allmant.htm

⁹⁵ CCB: http://lib1.ugent.be/cmsites/default.aspx?ref=ABABAA&lang=EN_CA

⁹⁶ “OCLC to create computing portal for public libraries. Interactive Web site to support those
who provide public access to technology,” Bill and Melissa Gates Foundation, May 2, 2002:
<http://www.gatesfoundation.org/UnitedStates/USLibraryProgram/Announcements/Announce-020502.htm>

⁹⁷ OCLC: <http://www.oclc.org/news/releases/200626.htm>

Bayerische Staatsbibliothek purchases NetLibrary eBooks, DUBLIN, Ohio, USA, 12 June
200:

NetLibrary, a division of OCLC Online Computer Library Center, Inc., and a leading
platform for eContent to libraries worldwide, will provide eBooks to Bayerische
Staatsbibliothek.

⁹⁸ Searching WorldCat Indexes: Currently, Connexion and FirstSearch support the following
non-Latin/non-roman scripts: Arabic, Bengali, Chinese, Cyrillic, Devanagari, Greek, Hebrew,
Japanese, Korean, Tamil and Thai. Other scripts will be supported in the future.

<http://www.oclc.org/support/documentation/worldcat/searching/searchworldcatindexes/>

⁹⁹ Unicode Online Data, Languages and Scripts: [http://unicode.org/onlinedat/languages-](http://unicode.org/onlinedat/languages-scripts.html)
[scripts.html](http://unicode.org/onlinedat/languages-scripts.html). A few of the languages listed are not yet covered by Unicode.

¹⁰⁰ Juha Hakala, Gabriel - Gateway to Europe's National Libraries,

<http://www.lib.helsinki.fi/tietolinja/0199/gabriel.html>

Cf. Michael Smethurst, “European National Libraries a Review, 1999-2000,” *Liber Quarterly*
2000: <http://webdoc.gwdg.de/edoc/aw/liber/lq-2-01/08Smethurst.pdf>

¹⁰¹ Authority files are a central and more elusive challenge than is generally recognized. Cf. Gabriele Meßner, "The German Name Authority File (PND) in the Union Catalogue": <http://www.google.com/search?hl=en&sa=X&oi=spell&resnum=0&ct=result&cd=1&q=first+library+network+germany&spell=1>

¹⁰² The European Library. FAQ: <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/06/102&format=HTML&aged=0&language=EN&guiLanguage=en>

Cf. : <http://dbs.hab.de/linkdb/linkdb.php?kopf=ort&ind=fndo>:

Zurzeit bietet "The European Library" eine integrierte Suche über Sammlungen und Kataloge von 19 der 45 europäischen Nationalbibliotheken, die in der Konferenz der Europäischen Nationalbibliotheken (CENL) organisiert sind. Bereits 2007 werden zwei Drittel aller CENL-Bibliotheken voll in "The European Library" integriert sein.

EU: Pan-European library information service launched:

<http://ec.europa.eu/idabc/en/document/4648/330>

¹⁰³ EDL Project Blog: <http://edlproject.blogspot.com/>

¹⁰⁴ KVK: <http://www.ubka.uni-karlsruhe.de/hylib/en/kvk.html>;

<http://library.cern.ch/HEPLW/2/papers/4/>

¹⁰⁵ Project Gutenberg: http://www.gutenberg.org/wiki/Main_Page

¹⁰⁶ Father Busa, Wikipedia: http://en.wikipedia.org/wiki/Roberto_Busa

¹⁰⁷ EEBO: <http://eebo.chadwyck.com/home>

¹⁰⁸ In the 1990s, IBM pioneered a number of projects. Stephen S. Hall, "Digital Libraries: Libraries without roofs," *Think Research*, IBM, 1996:

http://domino.watson.ibm.com/comm/wwwr_thinkresearch.nsf/pages/libraries196.html

Cf. Henry M. Gladney, Fred Mintzer, and Fabio Schiattarella, "Safeguarding Digital Library Contents and Users. Digital Images of Treasured Antiquities," *D-Lib Magazine*, July/August 1997: <http://www.dlib.org/dlib/july97/vatican/07gladney.html>

¹⁰⁹ Open Content Alliance: <http://www.opencontentalliance.org/>. This alliance works with commercial firms such as Adobe and Xerox and has close links with the key figures in RLIN.

¹¹⁰ <http://www.oclc.org/support/documentation/firstsearch/databases/dbdetails/details/Ebooks.htm>. This was again a sacred text. When printing began in Korea (c. 805 A.D.) the first book was a sacred text. When printing began in Germany (1454), Gutenberg used a sacred text, the Bible.

¹¹¹ Microsoft and the British Library work together to make 25 million pages of content available to all: <http://www.bl.uk/news/2005/pressrelease20051104.html>

¹¹² Chinese Collection starts to go digital

http://www.ucs.ed.ac.uk/bits/2005/november_2005/item10.htm

¹¹³ A decade ago there was a spectre of fear that large corporations would acquire major amounts of cultural content and thus control what can be seen. Fortunately, museums around the world are beginning to put their resources online free of charge which means that the arguments for fee based access to culture are retreating. Precisely how these vast amounts of materials can be made accessible online in a useful way remains one of the large challenges for the coming generations.

¹¹⁴ CHIN VMC: <http://www.virtualmuseum.ca/English/About/index.html>

¹¹⁵ E-Culture Net: <http://www.eculturenet.org/>

See also under Projects in the electronic version of the author's *Understanding New Media*:

Augmented Knowledge and Culture, 2006: <http://sumscorp.com/kavai/newmedia/>

¹¹⁶ BAM: <http://www.bam-portal.de/>

¹¹⁷ There is also an FP6 project, BRICKS, (Building resources for Integrated Cultural Knowledge Services): <http://www.brickscommunity.org/> which is equally lacking in this respect.

¹¹⁸ Rutgers: http://www.libraries.rutgers.edu/rul/rr_gateway/e_ref_shelf/e_ref_shelf.shtml

¹¹⁹ Xreferplus:

http://www.xrefer.com/index.php?option=com_content&task=view&id=16&Itemid=34

¹²⁰ Gale Group Acquires K.G. Saur Verlag From Reed Elsevier - Company Business and Marketing - Brief Article, *Online Newsletter*, Nov, 2000:

http://findarticles.com/p/articles/mi_m0BNO/is_2000_Nov/ai_66297320

¹²¹ Dialog: Key Dates: <http://www.dialog.com/about/keydates/>

¹²² Thomson Corporation Completes Acquisition Of Dialog - Company Business and Marketing, *Online Newsletter*, June, 2000

http://findarticles.com/p/articles/mi_m0BNO/is_2000_June/ai_62263365

¹²³ Bowker: <http://www.bowker.com/company/history.htm>

¹²⁴ OCLC, Books in Print:

<http://www.oclc.org/support/documentation/firstsearch/databases/dbdetails/details/BooksInPrint.htm>

¹²⁵ Cambridge Information Group Acquires ProQuest Information and Learning, Posted Dec 19, 2006: <http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=18813>

¹²⁶ "Reed Elsevier- Making the Death Fair Run Smoothly", *Corporate Watch*, 07 September 2005: <http://www.corporatewatch.org.uk/?lid=2008>

¹²⁷ Reed Elsevier Annual Review 2004: <http://www.reed-elsevier.com/media/pdf/t/b/International%20Financial%20Reporting.pdf>

¹²⁸ Bertelsmann Foundation and Federal Union of German Library Associations (publishers), *Library 2007. Strategy concept* :

http://www.goethe.de/kug/pro/bib/Bibliothek2007_englisch.doc

¹²⁹ Ibid, p. 24. A child born in Germany in 2007 will, during his or her life, be able to access reliable information: to learn how to read and to acquire language, to learn and study and for everyday use, for leisure and for every job that the child will ever have. This easily accessible information, designed for specific target groups and processed to provide a good overview, will accompany that child into its old age.

¹³⁰ In 1995, During the G7 Conference and Exhibition in Brussels, Europe rightly insisted that we need an Information and Knowledge Society rather than an Information Highway.

¹³¹ Proquest: <http://www.proquest.com/brand/umi.shtml>

When University Microfilms (later known as UMI) first opened its doors in 1938, founder Eugene Power was racing against time. War clouds threatened the treasures of scholarship in the British Museum, and Power was determined to preserve them on microfilm. The result was UMI's first (and only) product that year: Early English Books, the ongoing microfilm edition of the printed works listed in Short-Title Catalogue I (Pollard & Redgrave) and Short-Title Catalogue II (Wing).

Soon afterwards, Power began gathering, indexing, filming, and republishing doctoral dissertations in microform and print. Today, our Dissertation Abstracts database has archived over 2.3 million dissertations and master's theses. Some two million of them are available in full text in print, microform, and digital format.

¹³² Migne, *Patrologiae Graecae*: <http://rosetta.reltech.org/reltech/Pg/>

Electronic version cost \$3,400 in year 1 + \$400 in subsequent years for institutions and 1,100 in year 1 + \$100 in subsequent years for university professor/secondary school.

¹³³ JSTOR: <http://www.jstor.org/about/board.html>

¹³⁴ JSTOR: <http://www.jstor.org/about/individual.html>

-
- ¹³⁵ ARTSTOR: <http://www.artstor.org/info/about/letter.jsp>
- ¹³⁶ ITHAKA: <http://www.ithaka.org/about-ithaka>
- ¹³⁷ Aluka: <http://www.ithaka.org/incubated-entities/aluka>
- ¹³⁸ Nitle: <http://www.ithaka.org/incubated-entities/nitle>
- ¹³⁹ Portico: http://www.portico.org/about/advisory_committee.html
- ¹⁴⁰ University of Wisconsin: <http://www.library.wisc.edu/>
- ¹⁴¹ CDL: http://www.cdlib.org/programs/shared_collections.html
- ¹⁴² Access the UC Image Service Collections: <http://imageservice.cdlib.org/>
- ¹⁴³ Citation Index, Wikipedia: http://en.wikipedia.org/wiki/Citation_index
- ¹⁴⁴ Derek J. de Solla Price, Wikipedia: http://en.wikipedia.org/wiki/Derek_J._de_Solla_Price
- See also: E. Garfield, "In tribute to Derek John de Solla Price: a citation analysis of little science, big science", *Scientometrics*, Volume 7, Numbers 3-6 / March, 1985. Akadémiai Kiadó, co-published with Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V., ISSN: 0138-9130 (Print) 1588-2861 (Online)
<http://www.garfield.library.upenn.edu/essays/v8p232y1985.pdf>
- For a recent contribution: Henk F. Moed, *Citation Analysis in Research Evaluation* (Information Science and Knowledge Management), Berlin: Springer 2006:
<http://www.amazon.com/Citation-Evaluation-Information-Knowledge-Management/dp/1402037139>
- ¹⁴⁵ Wisconsin: <http://www.library.wisc.edu/guides/europeanhistory/elec/elec1b.htm>
- ¹⁴⁶ Thomson, Web of Knowledge, <http://www.isiwebknowledge.com/>
- ¹⁴⁷ Science Citation Index: <http://scientific.thomson.com/products/sci/>
- ¹⁴⁸ Elsevier, Science Direct:
http://www.sciencedirect.com/science?_ob=HomePageURL&_method=userHomePage&_btn=Y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=6b8f8a0bb11c4eb e676f09e6d7da52c8
- ¹⁴⁹ Science Direct: <http://www.info.sciencedirect.com/licensing/individual/ppv/>
- ¹⁵⁰ SCIRUS: <http://www.scirus.com/srsapp/aboutus/>
- ¹⁵¹ Scopus: <http://info.scopus.com/overview/what/>
- Other projects include getCITED and GoogleScholar.
- ¹⁵² Ulrich's: <http://ulrichsweb.com/ulrichsweb/>
- ¹⁵³ Steve Lawrence, C. Lee Giles, Kurt Bollacker, "Digital Libraries and Autonomous Citation Indexing", *IEEE Computer*, Volume 32, Number 6, pp. 67-71, 1999.
<http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html>
- ¹⁵⁴ HyperJournal: <http://www.hjournal.org/features>
- ¹⁵⁵ Steffen Staab, "Towards the Self Annotating Web", *ILASH Seminar*, Sheffield, 2004.
<http://www.dcs.shef.ac.uk/research/ilash/Meetings/ilash-staab.pdf>
- ¹⁵⁶ Proquest: http://www.umi.com/products_umi/digitalcommons/; Digital Commons
http://digitalcommons.proquest.com/cgi/query.cgi?field_1=full_text&connector_2=and&field_2=ancestor.link&op_2=in&value_2=http%3A%2F%2Fdigitalcommons.proquest.com&hidden_2=1&value:1=linear%20perspective&sortby=publication_date%3Ddescending&x_start=26
- ¹⁵⁷ GBV: <http://p7.gbv.de/iPort>
- ¹⁵⁸ This was on 7 February 2007.
- ¹⁵⁹ This was on 7 February 2007.
- ¹⁶⁰ Theodore Besterman, *A world bibliography of bibliographies and of bibliographical catalogues, calendars, abstracts, digests, indexes, and the like*, New York: Scarecrow Press, 1955.

¹⁶¹ It is striking that systems such as Google do not do this. If one types in *veille technologique* one gets only French websites on these words. Similarly if one searches for Nantechnology Germany, the first hundred titles are exclusively in English. In future we shall want filters to provide us results in the languages which we read.

¹⁶² Raj Reddy, Universal Library Applications:

<http://www.rr.cs.cmu.edu/gatech.ppt#263,14>, Universal Library Applications

¹⁶³ Frank Ogden, *The Last Book You Will Ever Read*, Toronto: Macfarlane Walter & Ross, 1994.

¹⁶⁴ Anthony Grafton, *The Footnote. A Curious History*, Cambridge, Mass.: Harvard University Press, 1997: <http://www.hup.harvard.edu/catalog/GRAFOO.html>

¹⁶⁵ ITU: Internet of Things: <http://www.itu.int/osg/spu/publications/internetofthings/>

¹⁶⁶ David Wright, "Google's 300 Year Plan," *Sydney Morning Herald*, 1 July, 2005: http://blogs.smh.com.au/razor/archives/search_engines/001426.html

¹⁶⁷ ESTO: <http://esto.jrc.es/>

¹⁶⁸ EITO: <http://www.eito.org/index-eito.html>

¹⁶⁹ EAO: <http://www.obs.coe.int/index.html.en>

¹⁷⁰ EODL: <http://www.bentli.net/obs/>

¹⁷¹ ENIL: www.ceris.cnr.it/Basili/EnIL/gateway/gatewayhome.htm

¹⁷² Pierre Levy, Green Paper on Innovation, The European Commission, Brussels 1995 Connected: <http://www.connected.org/pandora/publi.html>

¹⁷³ Pierre Levy, The Second Flood. Report on cyberculture, Brussels, 1996. Under the auspices of the Council of Europe.

Cf Review: www.culturelink.org/review/21/cl21levy.html

¹⁷⁴ TAGSYS Implements Item-Level RFID Infrastructure At Hamburg Public Library And Sets New Performance And Capability Standard, Oct 25, 2006

<http://www.tagsysrfid.com/html/rfid-news-334-2-1.html>

¹⁷⁵ "Stella, Library Chatbots in Electronic Reference", *Access Conference*, Ottawa, ON, October 13, 2006

http://216.239.59.104/search?q=cache:fVqJftjmC1oJ:www.access2006.uottawa.ca/christensen_ppt+Hamburg+library+millions&hl=en&ct=clnk&cd=13

¹⁷⁶ E.g. JRC, Max Planck, Fraunhofer and in the German context.

¹⁷⁷ E.g. DFG, CNRS, CNR, NWO etc.

¹⁷⁸ E.g. CENL, European Branches of IFLA, ICOM (especially CIDOC etc).

¹⁷⁹ Intellectual Property: <http://www.obs.coe.int/medium/intl.html>

¹⁸⁰ E.g. VG Wort, co-ordinating with WIPO where appropriate.

¹⁸¹ Part of the challenge is a mainly technical one, which becomes political when permissions are required in crossing national and other boundaries. These are being addressed by various efforts. In terms of content, memory institutions have traditionally focussed on enduring knowledge. As we attempt to record collaborative and personal knowledge, E.g. GEANT, SERENATE, Grid Projects, and more specialized efforts such as Göttingen's Digital Repository Infrastructure Vision for European Research (DRIVER) DRIVER:

http://www.dl-forum.de/deutsch/projekte/projekte_2813_DEU_HTML.htm

¹⁸² Preliminary efforts to define the elements of this second phase were outlined in E-Culture Net's vision of a Distributed European Electronic Resource (DEER). See particularly the report by Suzanne Deere. E-Culture Net: <http://www.eculturenet.org/>. Cf. also the author's *Augmented Knowledge and Culture* (Calgary, 2005): http://www.sumscorp.com/kawai/newmedia_cd/

¹⁸³ Thus far libraries have aimed at union catalogues of titles. In most cases, this meant that if an author wrote a book with editions in many languages, each foreign translation was simply treated as a further title. (In terms of the library's internal records there is a reference to a

standard title of the original but this is usually not visible to users in major databases.) As a result if a user wants to find the work *De pictura* of Leon Battista Alberti, searching for that title gives only Latin titles on *De pictura* many not by Alberti. Searching under Alberti theoretically gives us the titles if we realize that we need to look not only for *De Pictura*, but also *Della Pittura*, *On Painting*, *Traktat der Malerei*, *O Malarstwie* etc. Needed is a new level of bibliographical control that creates for each author a list of standard titles and allows users to see at a glance various a) manuscripts; b) editions, viewable alphabetically by place, chronologically by date, languages etc.; c) secondary literature linked with, to the extent that they exist; d) abstracts; e) reviews; f) bibliographies and g) citation indexes.

¹⁸⁴ The reasons for more systematic overviews of knowledge becomes clearer if we recall that the bibliography for Shakespeare alone for the period 1962-2006 contains over 110,000 annotated entries. *World Shakespeare Bibliography Online* (1962 and 2006):

<http://www.worldshakesbib.org/>

¹⁸⁵ E.g. the British Library, Bibliothèque Nationale, Wolfenbüttel and the Vatican

¹⁸⁶ For a more thorough discussion see the author's: "The Lion and the Unicorn," *Convegno internazionale: Il Castello Svelato, Direzione Regionale per i Beni culturali e paesaggistici della Campania e Istituto di Cibernetica E. Caianiello of C.N.R.*, Castel del Ovo, Naples, 28 September -2 October, 2006, Naples, 2007 (in press).

¹⁸⁷ JISC: <http://www.jisc.ac.uk/>

¹⁸⁸ Max Planck:

<http://www.mpg.de/english/illustrationsDocumentation/documentation/pressReleases/2003/pressRelease20031016/index.html>

¹⁸⁹ European University Association (EUA), MORESS Final Report, 2006:

http://www.cfa.au.dk/Publikationer/Rapporter/Final_MORESS_Report.pdf

¹⁹⁰ Sifry' Alerts, 17 April 2006: <http://www.sifry.com/alerts/archives/000432.html>

¹⁹¹ For another discussion of these themes see the author's: Access Claims and Quality on the Internet: Future Challenges, *Progress in Informatics*, Tokyo, no. 2, November 2005, pp. 17-40. http://www.nii.ac.jp/pi/n2/2_17.pdf. This is a much expanded version of the article produced for the Milan Open Source conference.

¹⁹² One of the fundamental contributions of Germany in this past century has been a basic distinction between information and knowledge. If we think of the six basic questions (who?, what?, where?, when?, how? and why?) then information entails an answer that is assumed to be true to a single question. Knowledge by contrast is about claims which entail an answer to multiple questions, the veracity of which can then be tested. For instance: 90 year old man is information; the 90 year old man left 50 volumes of rare books in the library of the Travellers club at 5 pm on Friday 2 May, 2006 potentially involves knowledge because we are able to check whether the claim is true. Continuing in the tradition of Leibniz, Kant, Hegel etc.

¹⁹³ For a more detailed discussion see the author's: "Towards a Semantic Web for Culture," *JoDI (Journal of Digital Information)*, **Volume 4, Issue 4**, Article No. 255, 2004-03-15. Special issue on New Applications of Knowledge Organization Systems. See:

<http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/>. See also: *Understanding New Media: Augmented Knowledge and Culture*, Calgary: University of Calgary Press, 2006.

¹⁹⁴ The Hermann von Helmholtz-Zentrum für Kulturtechnik at the Humboldt-Universität (Berlin) is addressing some of these themes and might be expanded in its mandate:

<http://www2.hu-berlin.de/kulturtechnik/weber.php>

¹⁹⁵ Vice President Al Gore, "The Digital Earth: Understanding our planet in the 21st Century", *The First International Symposium on Digital Earth*, San Francisco, June 5-7 1998: http://www.isde5.org/al_gore_speech.htm

¹⁹⁶ Raj Reddy, We Can Store Everything: [http://www.rr.cs.cmu.edu/gatech.ppt#268,16,We Can Store Everything](http://www.rr.cs.cmu.edu/gatech.ppt#268,16,We%20Can%20Store%20Everything)

¹⁹⁷ NEC develops the world's first virtual library
~ Universal Digital Library ~ , NEC, 8 January 1997:
www.nec.co.jp/press/en/9701/0801.html

¹⁹⁸ Jill Hurst-Wahl, "National Diet Library working on digitization of books", Friday, April 22, 2005 (<http://hurstassociates.blogspot.com/2005/04/national-diet-library-working-on.html>):
In Japan, "the National Diet Library is wrestling to digitize 8.14 million books to keep pace with the age of the Internet and to prepare against major earthquakes and other natural disasters. The Diet library, the only archive of the legislative branch of government in Japan, has been collecting publications issued in the country since its opening in 1948." Approximately 55,000 books are currently available on the Internet.

¹⁹⁹ Overview of Digital Libraries in China: http://www.white-clouds.com/iclc/cliej/cl16zhang_files/frame.htm

²⁰⁰ For other Chinese databases see: Janno E. Lecher, Full-text Databases:
<http://www.sino.uni-heidelberg.de/igcs/igtexs.htm>

²⁰¹ Recent Development in China-US Million Book Digital Library Project:
<http://www.new.dli.ernet.in/UDL-Talks-May2004/Progress%20in%20MBP%20in%20China040529.pdf>

Gloriana St. Clair, Million Book Project (MBP):
http://www.library.cmu.edu/People/gstclair/MillionBook_JHU.ppt

²⁰² Digital Library of India, <http://www.ieee-tcdl.org/Bulletin/current/balakrishnan/balakrishnan.html>

²⁰³ Raj Reddy, "Computing the Next 10 Years: Infinite Memory and Bandwidth: Implications for Universal Access to Information," Georgia Tech, 2001:
<http://www.rr.cs.cmu.edu/gatech.ppt>

²⁰⁴ In India this is linked with a Universal digital library which has the vision of "A Million Books To The Web Assembling - The World's Biggest Library On Everybody's Desktop":
<http://udl.iिता.ac.in/about%20udl.htm>

²⁰⁵ The Bibliotheca Alexandrina, has founded an International School of Information Science (ISIS) which is now linked with this and a number of other projects worldwide. BA ISIS, Million Book: <http://www.bibalex.org/isis/ProjectDetails.aspx?Status=ongoing&id=11>

²⁰⁶ Million Book Project, Wikipedia: http://en.wikipedia.org/wiki/Million_Book_Project

²⁰⁷ Universal Digital Library: <http://udl.iिता.ac.in/>

It is fascinating to see how widely variant claims re: statistics are. For example: Estimates show that the amount of new information produced in the year 2002 alone -- 5 billion gigabytes -- is 37,000 times larger than the holdings of the US Library of Congress, the world's largest library.

"Europe Ponders Digitalization of Cultural Heritage", *Deutsche Welle*, 21, 06, 2006:
www.dw-world.de/dw/article/0,2144,2063271,00.html

²⁰⁸ Kevin Kelly, "Scan this Book", *The New York Times*, New York, 14 May 2006:
<http://www.nytimes.com/2006/05/14/magazine/14publishing.html?ei=5090&en=c07443d368771bb8&ex=1305259200&pagewanted=print>

²⁰⁹ N. Balakrishnan, "Million Books to the Web. An Example of Indo-US Collaboration. Lessons Learnt and the Road Ahead", Indo US Workshop on Open Digital Libraries and Interoperability, Washington, 23 June 2003:
[http://fox.cs.vt.edu/IndoUSdl/Balakrishnan.ppt#894,1,Million Books to the Web An Example of Indo-US Collaboration Lessons Learnt & The Road Ahead](http://fox.cs.vt.edu/IndoUSdl/Balakrishnan.ppt#894,1,Million%20Books%20to%20the%20Web%20An%20Example%20of%20Indo-US%20Collaboration%20Lessons%20Learnt%20&%20The%20Road%20Ahead)

²¹⁰ David Canton, "Google book scan plan tests copyright laws", *London Free Press*, 22 October 2005: <http://www.canton.elegal.ca/archives/2005/10/>

-
- ²¹¹ BA Joins the World Digital Library Project, 11 May 2006: <http://www.bibalex.org/ISIS/ProjectDetails.aspx?id=51>
- ²¹² World Digital Library: http://www.loc.gov/about/welcome/speeches/wdl/wdl_6-6-05.html
- ²¹³ Heidi Benson, "A Man's Vision: World Library Online. Brewster Kahle hopes to realize his 25-year dream of an international book archive", *San Francisco Chronicle*, 22 November 2005: <http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2005/11/22/MNGQ0FSCCT1.DTL>
- ²¹⁴ "Canadian libraries join race to digitize books," 29 December, 2005, CBC: <http://www.cbc.ca/arts/story/2005/12/29/canada-libraries.html>
- ²¹⁵ Google contributes \$3 Million to "World Digital Library": www.whatistheword.com/story/SciTech_277.html
- ²¹⁶ UNESCO and US Library of Congress host meeting on World Digital Library project http://portal.unesco.org/ci/en/ev.php-URL_ID=23521&URL_DO=DO_TOPIC&URL_SECTION=201.html
- ²¹⁷ Proposal for a World Digital Library: http://www.loc.gov/about/welcome/speeches/wdl/wdl_6-6-05.html
- ²¹⁸ The Development of the China Digital Library: http://southernlibrarianship.icaap.org/content/v03n03/Yang_g01.htm
- ²¹⁹ To be sure the United States already has projects with Brazil and South America. Our point is that this important continent is not explicitly named as part of their new vision.
- ²²⁰ Mariona Vivar Mompel "Google print outshines the European Digital Library", *CafeBabel.Com*, Paris, 4.10.2006: <http://www.cafebabel.com/en/article.asp?T=T&Id=8276>
- ²²¹ In the 1970s, a remarkable Italian produced some books which had an impact far beyond their unlikely contract in South America funded by the CIA. The man was Edward de Bono. His books such as *Lateral Thinking*. In the future, we need new approaches that help us to see that major schools of philosophy in different cultures are also stepping stones to lateral thinking in a more diverse sense.
- ²²² BNF: <http://www.linkparis.com/Bibliotheque-Nationale-de-France.htm>
- ²²³ Berlin State Library, Wikipedia: http://en.wikipedia.org/wiki/Berlin_State_Library
- ²²⁴ Bavarian State Library, Wikipedia: http://en.wikipedia.org/wiki/Bavarian_State_Library
- ²²⁵ UCSF: <http://www.library.ucsf.edu/db/>
- ²²⁶ Eric Eldred: http://www.eldred.cc/ea_faq.html
- ²²⁷ Creative Commons: <http://creativecommons.org/>
- ²²⁸ Hal Abelson: <http://swiss.csail.mit.edu/users/hal/doing.html>
- ²²⁹ ALA: Europe Looks to Public Libraries in the Future, *The Shifted Librarian*, Monday, June 27, 2005: http://www.theshiftedlibrarian.com/archives/2005/06/27/ala_europe_looks_to_public_libraries_in_the_future.html