1. Definition of “document”
What is a document? It is of course a fundamental question - if not even the funding question *par excellence* - of every possible science of bibliography, documentation, librarianship, information, or however else we choose, or will choose, to call it. And just like for every funding questions, it is not taken for granted that it had been handled and answered in a satisfactory and definitive way at the dawning of the historical evolution of the respective discipline; in fact, during the last years it has been so much discussed, and it will still be discussed during this meeting.

In this multimedia and hypertext age I would take for definite that a document could be not only necessarily textual or alphanumeric and that both the material supports – be they analogical or digital - that will constitute document’s substratum and the techniques used to produce, distribute, preserve, organize and use them, will be innumerable and diversified.

In so doing we removed, or at least weakened, the limits in which sometimes in the past the concept of document had been artificially enclosed; however trying to trace new and broader limits is even harder, although this limits are necessary in order to avoid that the concept of document could expand excessively to the point of coinciding with the one of the Universe itself (of course this is not what Theodor Holm Nelson was mentioning with his “Docuverse”) as some current, more canonic definitions would entail. For example:

Document: every physic entity, in whatever form and material, in which information are recorded.

Now, aren’t the most various information available in every physical entity, to the one who is able to read them? Perhaps the point lies in the intentionality of the recording? In fact the definitions mentions “recorded” and not “contained” information, and in so doing allows us to understand as follows:

Document: every physic entity, in whatever form and material, in which information have been recorded.

Underlining the intentional, purposive and communicative dimension leads us to a second well known definition that, although very broad, seems to put clear and definite borders to the concept of “document” tracing those borders in the expression of human creativity:

“Any expression of human thought” was a frequently used definition of “document” among documentalists.

Nevertheless this definition lacks in excluding *a priori* possible - and realistic, according to many people - producers of quite relevant documents, that is from the para-human, extra-
terrestrial ones - typical document: the monolith of *2001: A Space Odyssey* - up to the highest divinities of every respectful cosmogony - typical document: the Universe -.

But this is not the most relevant limit of the “expressionist” definition: it would be sufficient to add “para-human” and “super-human” to “human”; rather, this limit is in the inadequacy against the developments of “social sciences in the twentieth century [that] deeply re-elaborated the concept of “document” [dividing it in] materials that support information that have been “knowingly” recorded in order to transmit them through the time (*ad memoriam*), and materials that, once they have been interpreted, become useful sources independently from the awareness of the people who used them.

Every object, be it natural or artificial, is then liable to be interpreted as a document, that is as a recipient of information, a vehicle of signs. We are then again in trouble in distinguishing the Docuverse (the universe of documents) from the Universe *tout court*. We are therefore condemned to tautological, circular definitions...

The term “document” were used in a specialized meaning as the technical term to denote the objects to which the techniques of documentation could be applied…

... or subjectivistic, relativistic definitions...

Material becomes documents to those who are able to recognize in it the piece of a *segnico* element.

Document: any source of information, in material form, capable of being used for reference or study or as an authority.

Actually the “tautological” approach (document is what documentalists deal with, using the techniques and methods proper of documentation) and the “relativistic” approach (document is the one in which someone is able to trace signs or information) join and lead us to the solution of the enigma. Let’s take the classic example of the antelope:

*An antelope running wild on the plains of Africa should not be considered a document [...] But if it were to be captured, taken to a zoo and made an object of study, it has been made into a document.*

Everything, potentially, is liable to be considered a document; however only when a specific will to utilize, benefit from, study, preserve a specific entity as a source of information comes up, this entity, observed *sub specie documenti*, really becomes a document. Universe and Docuverse tally, but only for a facet.

We said that every grain in the Universe contains lots of information for those who are able to read them. But *who* is able to read them? It is a manifold choice. Physicists, for instance, properly read the whole universe and each of its parts as matter and energy, ruled in their relationship by physic laws. Chemists do the same with elements and laws of their own discipline, and in a different but similar way do biologists, mathematicians, jurists, economists, historians, philosophers, etc. The same object (*noumeno*, in itself unknowable) is liable to be caught by the limited human cognizance only if it is filtered by a discipline or by any other form of organized knowledge - or by their commingling,
hybridization and vulgarization, among which lies the so called “common sense” - which organizes, pigeon-holes, classifies a specific facet making it seizable and then thinkable\textsuperscript{11}.

The researcher is then the real originator of the document in relation to his interest and culture: ten analysis, ten documents but only one “object”\textsuperscript{12}.

Finally, a little but significant amendment to the current canonic definition of “document” can thus be deduced:

Document: every physical entity, in whatever form and material, \textit{since} information are recorded in it

Just like to Borges every writer creates his own forerunners\textsuperscript{13}, so every observator creates his own documents. It happens both in everyday life and in the most sophisticated scientific research. The different disciplines read the same materials with different glasses, and each pair of glasses visualizes a different document. There is then a discipline that deals with the glasses themselves rather than with the material observed, focusing on the methods with which documents are produced, distributed, preserved, organized and used:

If historical-technological disciplines related to \textit{books} are part of the science, and represent one of its regions, proper bibliographic disciplines are on a meta-level if compared to the level of investigating science – a sort of Science of Science – since they deal with the structures and organization of the knowledge itself\textsuperscript{14}.

A further matter is the relationship between those disciplines and the possible others - like epistemology or hermeneutic - that would claim a similar role of “meta-disciplines” in the cognitive field. Easier instead is the coexistence of documentary-bibliographic sciences with metaphysics, \textit{par excellence} the “meta-discipline” in the ontological field: should we settle the question with a joke, we could say that metaphysics claims to organize Universe, documentation contents herself with something limited, that is organizing Docuverse.

\textbf{2. Local Electronic Resources (LER) and Remote Electronic Resources (RER)}

The term “Electronic Resource”, or “Digital Resource”, is more and more often used to mean a wide range of products going from electronic periodicals to CD-ROMs, from e-books to websites, from mailing lists to data banks, all of them having the common feature of being used - and sometimes modified - only thanks to the medium of a computer or however an electronic device.

In fact inside this all-inclusive terminology lies a great variety of entities which are extremely different one-another. The first distinction to be done is the one between off-line resources (called “local access electronic resources” by ISBD, “direct access electronic...
resources” by AACR2 and, more concisely, “local electronic resources” here) and on-line resources (called “remote access electronic resources” by ISBD, “distant access electronic resources” by AACR2 and “remote electronic resources” here).

In fact, inside LER itself it is not easy to trace the borders between further under-categories. The real forest of available supports (thousands kinds of CDs and among all their probable heir, the DVD) should not lead us to mechanically overlap the differences among physical objects with the differences among typologies of documents, as it used to happen in the age we often define “age of Gutenberg” and that should more properly and generically be defined “analogical age”.

It is easy to distinguish a music passage “to be listened”, registered on a vinyl disk, from the transcription of the same passage “to be read” in the pages of a printed publication. It is banal to distinguish a book printed in a printing office from the original manuscript, or typescript - even if electronic -, that however vehicles the same work. In the analogical world it takes just one look to understand what document are we dealing with; in the digital world the same object might contain the most heterogeneous entities, more and more often mixed among them.

By the point of view of a bibliographer, of a librarian, of a scholar of publishing history or of information science, what is the difference between a multi-media encyclopedia on CD-ROM, full of sound documents, educational games and software for data research and manipulation, and an audio-CD to be normally listened on one’s own HI-FI but also enriched with only PC usable videos? Why, in the most updated histories of publishing, do we consider at limit only the first kind of product? Why, in the libraries, are they managed in the most different ways? Why, in the bibliographies, are they mentioned in different lists?

The reason why even in the most familiar off-line environment it is increasingly difficult to find a way through, and to differentiate among, products that used to be clearly distinguished, can be identified in the concept of “convergence to digital” well explained by Ciotti and Roncaglia. “Information of various nature can be all reduced to the same basic code, the long chains of 0 and 1 proper of digitalized information. This “convergence in codification”, as we could call it, becomes a real “convergence of technology” at the moment the PC becomes the tool able to effectively manage great quantities of information in a digital format; then, instead of tools based on different technologies (typewriter, TV set, radio, telephone, cine-camera, cine-projector, camera...), new tools turn up; they are often very differentiated as to functions and interface, but they all have the same “heart” made of a microchip, whose function is in acquiring, manipulating and distributing information in a digitalized format. And here happens that markets traditionally different one another (publishing, cinema, television, telephony...) integrate each other (convergence of market) and with the one that has historically been the first form of information market in a digitalized form, that is software market. Of course all this will have consequences by the point of view of styles and languages of communication, leading to a step forward in the possibilities of integration among different codes inside unitary information products (digital integration).

In the RER field further complexities add to this matter, which is implied in the concept of “electronic resource” itself. Theodor Holm Nelson, the visionary originator of the never-realized planetary hypertext Xanadu – that could be considered at least a forerunner of the World Wide Web - several times stated that “publishing CD-ROM is not electronic publishing, rather it is publishing plastic”. It is hard to say he is wrong if we consider the cultural and technologic abyss that lies between printing and spreading worldwide a certain number of objects almost similar one another, containing certain information - and this procedures is the same as to printed books, CD-ROM and other digital supports to be used off-line that is with an isolated, not linked to a net, device -, and locating the same

---

15 GUERRINI [1999] asks similar question by a properly catalografic point of view.
17 See, for example, NELSON [1992] or Xanadu official site <http://xanadu.com>.
information in the memory of only one PC linked in a net with all the PCs of people who will not need to leave home or office to use on-line those contents.

Most of the interpretative equipment that have been created with the purpose of analyzing the printed book by a bibliographic/documentary, economic, sociologic and historic point of view, are liable to be arranged to the LER with just a few adjustments; instead, as far as it regards RER the usual critical tools may sometimes not be able to realize the "Copernican Revolution" – that is no longer documents moving towards the reader, rather readers going to the document - that has been created by computer nets and, most of all, by the internet, the net that gathers all the others.

Besides the three categories which are, generally speaking, more homogeneous if compared to their equivalent on paper, that is electronic periodicals, electronic books and data banks, a wide range of sources is available online and it is not so obvious to consider them as “documents” in the sense we usually give the word.

Virtual communities exchanging messages by mailing lists, newsgroups, chats or forums whose retrospective archive are only seldom available on the web. TV and radio stations broadcasting also, or only, on the web. Software and music records sold or distributed as a test by the companies, or by the citizens, that created them. Games of the most varied kinds sometimes downloadable and often directly, online usable by individuals or groups. Commercial publishers providing only bibliographic reference or excerpts of their books and periodicals on their website. “Repository” of articles and contributions lacking of that partition in instalments we are used to associate to the concept of periodical, be it paper or electronic. Text, images, videos and sounds made available by individuals in their personal homepages. Institutional and companies’ sites offering a variety of “grey-literature” which in the previous, paper version was not diffused through the normal channels of commercial publications and that therefore was hardly identifiable and accessible. Sites dedicated to electronic commerce of the most varied digital and “real” goods, or to the provision of services that lie at the border between “factual” and “documentary” like news, stock-exchange listing, weather forecast, horoscopes, street directories, etc. Search engines, directories by subject and virtual reference desks helping find what is available in the net and portals that, in addition to this, give access to a variety of services as those listed above.

Which of those RER may be listed in the universe of the proper documents, and which of them lie outside it since they belong to the not documentary, although virtual, realities? Is the website of a body – institution or company - to be considered, in its complex, as a macro-document produced by the body itself, a cyber-equivalent of the body as a whole or is it to be considered as a new office or virtual sector of the body, adding to those already available in the real world? Is all the material that rise to the surface of the web to be considered as a document that the libraries must preserve and catalogue? Do projects like the Wayback Machine of the Internet Archive http://www.archive.org, that download and preserve everything has a URL without grieving with subtle philosophic-documentary distinctions, have a librarianship feature? Is cyberspace – or at least the World Wide Web - an exclusively documentary space?

As we stated in the first section of this speech, it is already theoretically possible – even if pragmatically eccentric - to extend the concept of Docuverse in order to make it tally with the concept of Universe; in this frame it is even more natural, even if more on a theoretical dimension than on a practical one, to state the – potential - identity of Docuverse and Universe when the second term of this equation becomes virtual, digital, electronic, converting itself in Cyberspace.

Real Universe and virtual Universe – this last being represented by cyberspace – are not symmetric by a documentary point of view. Since all electronic resources known by us are produced by human beings -even if sometimes thanks to the medium or thanks to the collaboration of more or less automated entities-, then even the “expressionist” definition of document (“any expression of human thought”), which we rejected in the first section due to its being restricted, is applicable to them. The “convergence to digital” represents a second strong issue on behalf of the documentary homogeneity of the whole of the electronic resources, whose greater part is however normally considered as being part of
the documentary range. To these topics who are common to both LER and RER, in the case of Cyberspace must be added the strong skills of integrability – that is indefinite extensibility – and interactivity – that is malleability – which, together with the skills of multi-linearity and multi-mediality exalt its hypertextuality, typically documentary feature.

In fact, it is not a pure chance that the term “resource” is considered too generic to denote analogical documents and it is coupled in these cases to adjectives like “informative” or “documentary”, while in a digital domain – and particularly in the net – “resource” and “document” are basically used as synonyms. Therefore we should not waste our time in asking what a document is, or is not, in the cyberspace. Everything can be a document if we believe it is convenient to consider it so.

3. RER in OPAC

Considering a RER an entity of documentary interest, since it vehicles information and therefore is a possible document, does not necessarily mean that we decided a priori that libraries must preserve, catalogue and manage the access to all past, present and future RER. Rather, one of the results of this “documentary attention”, articulated in times of theoretical general reflection – as meetings like the one we are participating today- and times of practical professional activity in every library – hopefully always enlightened by theory - will be to decide which resources, and with which means, will need to be preserved, catalogued and made available.

It is true that it would be meaningless to waste time enquiring in the abstract what is a document and what is not in the cyberspace; even more true is that every library will spend some time more efficiently to evaluate practically what typologies of RER are worth investing resources for costly activities like cataloguing and preserving, coherently with the mission of the library itself in order to improve service to users without penalizing other and equally useful services.

The main conceptual options available to the library, that desires to provide its remote or local users with the most easy and efficient access to a selection of RER, are the following:

1. Widening the traditional OPAC also including links to RER selected on the basis of the mission of the library; nevertheless, this could mean to betray the catalogographic nature of the OPAC changing it to a hybrid between a catalogue and a bibliography which includes also documents that could in every moment change, move or disappear due to the choice of their authors or publishers, without any possibility of preserving them by the library – and with great difficulties in updating the links.

2. Keeping the OPAC traditionally restricted to the only local documents - analogical resources and LER -; it is necessary to physically go to the library in order to consult

---

18 The restrictive definition of “Cyberspace” limits it to the whole of RER connected among them – today internet, tomorrow Xanadu –; however, a broader definition could also include LER since they are potentially connectible to other electronic resources which are available in the net (canonic examples are CD-ROM works whose updating are available in the net and data banks that can be, at the same time, LER to one library and RER to another).

19 “Integrability and interactivity are not completely independent one another, since the only concrete possibility an hypertextual system has to be open to the outside, then to grow indefinitely, lies in the enrichment brought by always new readers-authors. However, both these features emerge more evidently in those hypertexts that, although devoid of Xanadu’s totalizing ambition, tend to configure themselves as virtually “planetary” structures that without pre-definite limits open from a specific text to the whole Docuverse. This kind of hypertext, apart from being the most interesting to librarianship and information science, is also the one that embodies the essence itself of hypertextuality, which in “closed” applications is only a merely aesthetic indulgence to current fashion” RIDI [1996] p. 13.

20 “The researcher working in the net often meets new objects that are usually defined “resources” due to their being changeable and hybrid if compared to traditional types” ZORZI [2000] p. ii.

21 See, among the others, BANERJEE [1997], CRUPI – PARLANTI – SIMONE [1999], GNOLI [2000], GORMAN [2001] and particularly the work by Gambari and Guerrini that will soon be published and that has been mentioned in the introduction to the bibliography.
the documents, of which the library guarantees long term preservation and access. However:

2a. Including selected RER in the OPAC, but only after having changed them in LER, that is after having placed them in virtual local shelves of the single library or, better, of the network of libraries, in order to guarantee long term access and inalterability;

2b. Delegating access to RER to lists or archives separate from OPAC, managed with faster cataloguing rules and easier enquiry modalities, often connectable to the category of virtual reference desk (VRD)\(^{22}\);

2c. Only providing users with lists of general tools for internet search (engines, meta-engines, directories, virtual reference desks, etc.)\(^{23}\).

It is not banal at all to decide whether to sympathize in abstract, schematically, for the inclusion of RER in OPAC or not. In order to easy the opposition between catalog of owned material and bibliography of existing material, a supporter of the inclusion may remark that it is only common experience to effectively find in the collection of the library what has been listed in the catalogue, since at the moment one moves to the shelf in order to take the book he may discover that it is out of place, lent, stolen, damaged; in the same way, while following the remote link one may discover that the catalogued RER has been in the meantime modified, moved or deleted. In both cases the catalogues indicates what is normally accessible from the library, unless particular circumstances.

However, a supporter of the exclusion may replicate that being a book out of place or damaged is in fact a particular case since the whole organism of the library is devoted to prevent, or at least to limit, these episodes; nevertheless, the changing of a RER or of its URL is not a particular case, an exception or an accident, rather it is normal and is implied in the same nature of RERs which are remote, that is out of the control – and metabolism – of the library; the library, in fact, delegates the control of all – electronic and paper - remote resources to other entities called bibliographies, which in the last 250 years have been distinct by the libraries and their catalogues.

My personal proposal relating to the organization of access to RER by the electronic space of the library seems to satisfy both the practical needs of users and the coherence with bibliographic tradition, and consists of restructuring the normally available kinds of access to OPAC in order to allow the user to choose among the following kinds of consultation:

A- Traditional OPAC, including only analogical documents and LER - both original and produced by the library by stocking RER on its virtual shelves;

B- VRD - in the form of a list or archive, however automatically created by extraction from the unique cataloguers’ working archive of all the types of documents - including only RER selected by the library;

C- Enriched – or broaden – OPAC, sum of A and B, including all the documents available and selected by the library.

After this reorganization, a last question relates to the classification of the full-text access to two kinds of electronic resources which are widely diffused particularly in the academic libraries, and which are not freely accessible by any surfer that uses the library’s web and who is not its institutional user:

1. Electronic periodicals and commercial data banks, that have been produced and are maintained by external hosts and which are accessible by the library’s institutional users only as long as the library corresponds some tariff to the host;

2. CD-ROMs or other commercial data banks locally and indefinitely stocked by the library thanks to special agreements – or tariffs – with hosts. CD-ROMs and data banks which are produced locally - by the library itself or by the body to which it

---

\(^{22}\) For a definition and examples see, for example, RIDI [2001].

\(^{23}\) The case [2c] is, in the strict sense, only a particularly simplified and reduced sub-case of [2b]. For definitions and examples of general tools for internet search see, for example, RIDI [2001].
relates – and to which for any reason is not given free access to everybody (for example costly didactic material produced or acquired to train uses or staff). Digital reproductions of documents owned by the library on paper (even according to copyright laws on the basis of agreements with rights owners).

Both [1] and [2] should be included in the enriched OPAC [C], which would be the most all-inclusive access tool in the library – possibly dowered with cross-filters to narrow the search to the only analogical sources, or to the only LER or RER, or to the only remote sources usable by everybody or only by registered users. However a twofold choice remains:

X. Supporting a choice of preservation – that is of **possession** -, including then [1] in [B] and [2] in [A], following the traditional distinction between catalogue [A] of possessed and bibliography [B] of documents selected, suggested and of which consultation is eased but whose long term preservation is not guaranteed. In this case, both in [A] and B[] during the enquiry the not registered remote user would also retrieve the description of electronic resources that have been catalogued but to which he will not be allowed to directly access unless he physically goes to the library.

Y. Supporting a choice of usage – that is **access** -, therefore including both [1] and [2] in a new OPAC’s segment [D], which would retrieve from the general archive the only electronic – local or remote – resources usable also remotely by the registered users, and only by physically going to the library for those who are not registered. In this case the registered user would be aware that by enquiring both [B] or [D] – even unifiable in a special research for registered users – he would get to the full-text document, while not registered user would be aware that he will be allowed to get this just for B.

In this scheme of options – hopefully easier to be applied than explained – which are offered to the single user and library, I hope it would be possible to find a little contribution in order to overcome the dialectics (RER inside or outside OPAC) that if too abstract and schematic would risk to divert us from the more concrete and important matter of the selection of RER suitable to the specific users and compatible with the real resources, and the selection of the most efficient methods for their cataloguing.

I do not have time enough in this occasion to enquire these matters, which will be however eminently discussed by the other speakers today and during the next days; however, I will only provide one brief advise, that is in fact as much easily deducible from general principles of librarianship as banally applicable even to this new “electronic-remote” sector.

The advise is: Cooperate! Even in the work of selecting and cataloguing RERs, as in all the other library activities, cooperation is more and more crucial and it will be achieved for example by following standards and criteria as much as possible shared and formalized, that will also allow to import and export records and authority files from and to other archives, accumulations and meta-enquiries. Cooperation will also be fundamental in order to apply to collective projects related to selection, cataloguing, preservation and supply of electronic documents. Let’s try, finally, to learn from the past and to do, in our digital future, less mistakes then we did in our analogical past.
A special thank to Stefano Gambari and Mauro Guerrini that allowed me to read the draft of their book on electronic resources cataloging, that will soon be published by Editrice Bibliografica. On the same subject (however limitedly to electronic resources) I have usefully read the Tesi di Laurea recently discussed by Maria Cristina Bassi at University of Studies in Venice (relatively Chiara Simonato Rabitti). Some of the subject of this speech have been anticipated with an inquiring form in the editorial RIDI [2000]—successively discussed with Michele Santoro that I warmly thank—and in the “virtual round table” GNOLI [2000]. From both texts of mine, excerpts have been taken and often modified.


SALARELLI, Alberto [2000] L’informazione digitale, in La biblioteca digitale, di Alberto Salarelli e Anna Maria Tammaro, Milano, Editrice Bibliografica, 2000, p. 17-100.

