To Measure what Information is Measurable.

*Antonio M. Ruiz Mariscal. Centro de Información CIATEC, Omega 201, Frac. Industrial Delta, Leon, Guanajuato, 37545. E-mail: mariscal@ciatec.mx

Following Galileo's dictum: "the objective of Science is to Measure what is measurable and to make measurable what it is not yet so", in this presentation results are given of an ongoing project measuring every conceivable presentation of the information phenomena. Some 2,100 diverse units grouped in 66 categories are presented, with an original dozen criteria defined, such as information's *half-life, equivalencies*, *stored, transferred* and an *index for information activity by country*.

Among the sources for such an ambitious effort, besides the field work, are the reports of the 800 most important companies and organizations handling information worldwide, over 40,000 libraries holdings and an exhaustive perusal of the scientific and technical literature worldwide.

©2005 by MDPI (http://www.mdpi.org). Reproduction for noncommercial purposes permitted

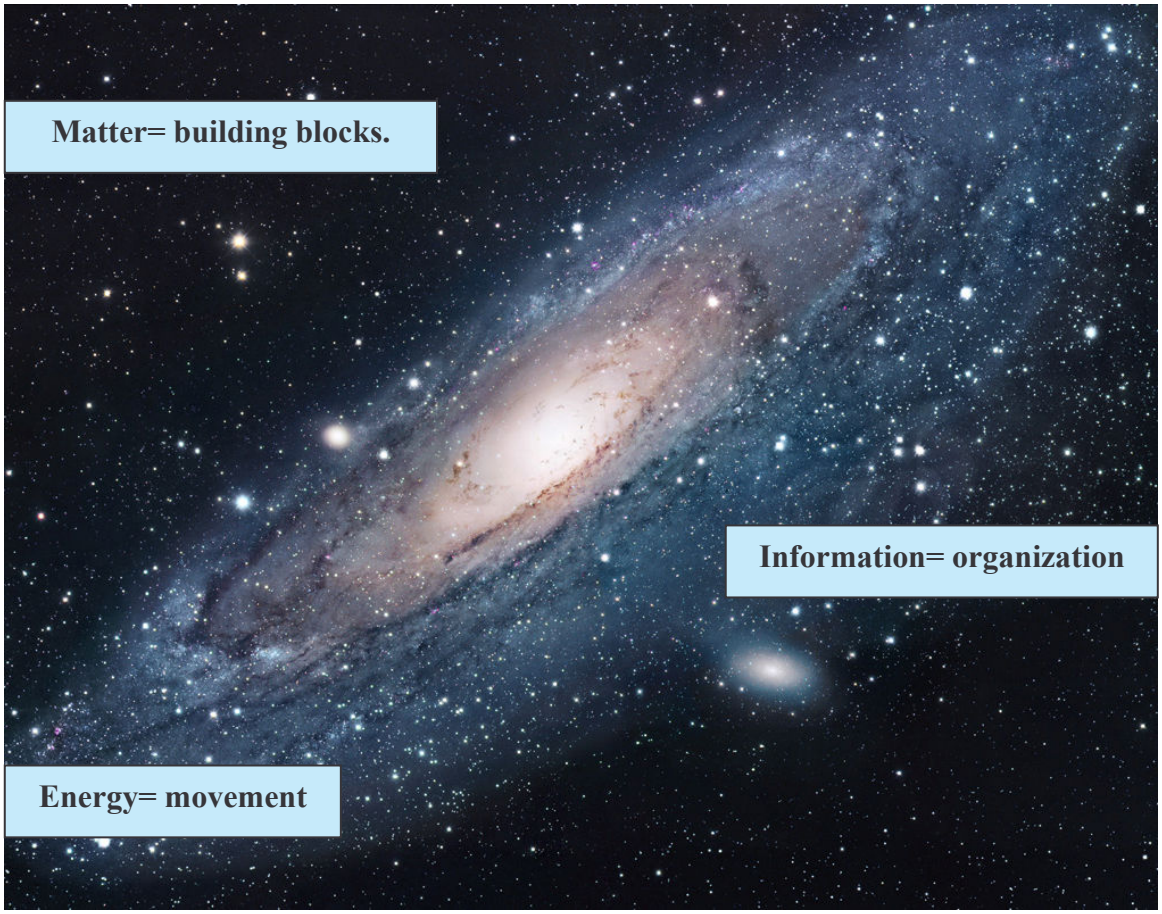
Introduction. Basic principles.

When thinking on the measuring of the information phenomena, some basic concepts should be firstly considered.

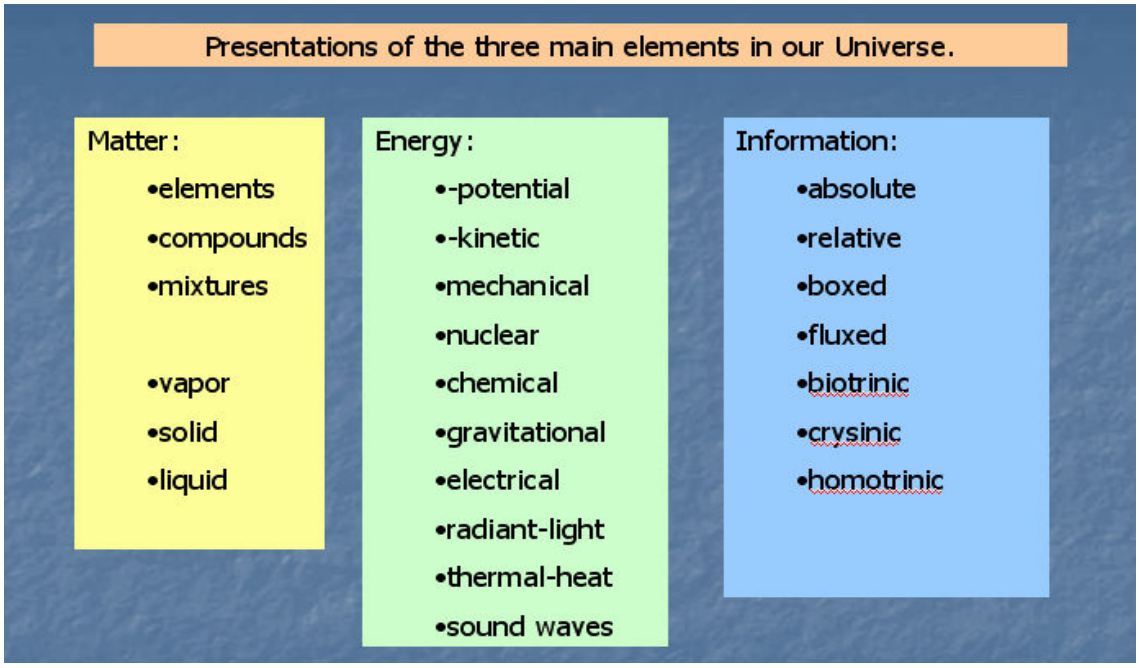
a) Information is one of the three elements that constitute our Universe. The other two being matter and energy.

Matter represents the building block of the Universe. Energy refers to the movement, the evolution or else the dynamics of matter. After considering the materials and the energy or work needed to move or transform it, then it comes into play the third element, *information* that gives pause to defining what kind of structure or matter organization we are going to have. Information tells us how something is organized. Information is an intrinsic characteristic of everything that has some structure or order to it.

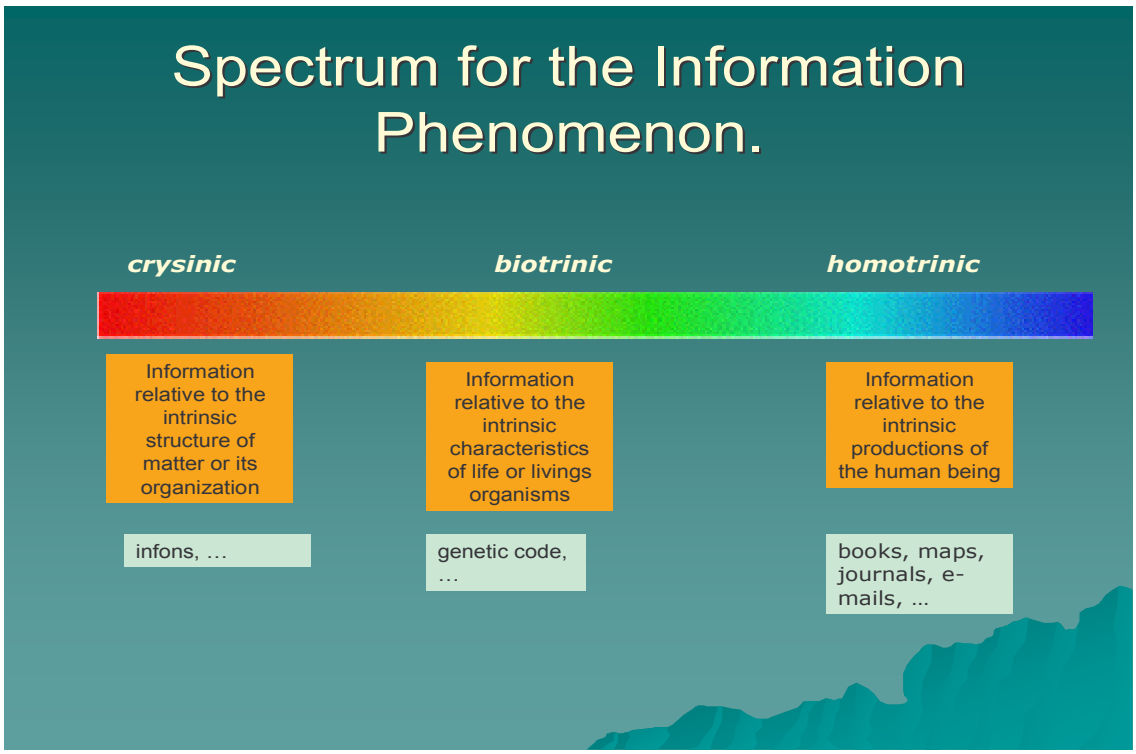
Elements of our Universe



b) There are many ways in which energy, for example, manifest itself. So after three centuries of studies on energy we accept that there exist mechanical and chemical energy as well as sound waves. The same can be said about information. It has diverse presentations and every one of them is in truth information in some form:



c) We could simply say that there is a spectrum for the information phenomenon, in which we find:



To Measure what Information is Measurable. Foundations of Information Science.

I decided some years ago to concentrate in the process of measuring the *homotrinic* information section of the spectrum. The reason for measuring being the definition of science set forth by Galileo Galilei, when he said that: *The objective of Science is "to measure what is measurable and to make measurable what it is not".*

Following this dictum, if you want to have some true foundation for any science, new and better measures of the phenomena should be available. This is in fact the path that other sciences have had to follow, for example chemistry, when some centuries ago it started with some amateurs measuring some gas volumes.

Regarding *homotrinic* information measuring, we find in the past century and a half several efforts, among the most noticeable, we have, in chronological order:

a) Hubbard's Publications table.
b) Shannon's Information theory.
c) Garfield's Bibliometric and scientometric developments.
d) Lyman & Varian's *How Much Information* reports

Hubbard published in 1882 a table [2] giving the total number of periodicals published worldwide. With a considerable effort it was possible for him to tell the exact number of publications, journals and newspapers published then in every part of the world:

Consolidated Statement		...	Aggregate of All Classes (daily, weekly,...)				
Grand Divisions	Population	...	Number	Total circulation per issue	Total Annual Circulation	Average circulation of each per issue	Number of Copies Annually Distributable to Each Inhabitant
Europe	301,356,369	...	19,557	76,431,447	7,344,956,805	3,909	24.330
North America (including the West India Archipelago)	76,033,776	...	12,400	36,995,197	2,787,842,262	2,983	36.660
Asia	1,007,128,657	...	775	1,099,114	195,010,924	1,418	0.010
South America	29,988,509	...	699	772,675	117,520,340	1,105	3.920
Australasia	3,670,850	...	661	851,132	112,417,322	1,287	30.630
Africa	205,000,000	...	182	250,445	31,751,795	1,376	0.010
		...					
Grand Total	**1,623,178,161**	...	**34,274**	**116,400,010**	**10,589,499,448**	**3,396**	**6.520**

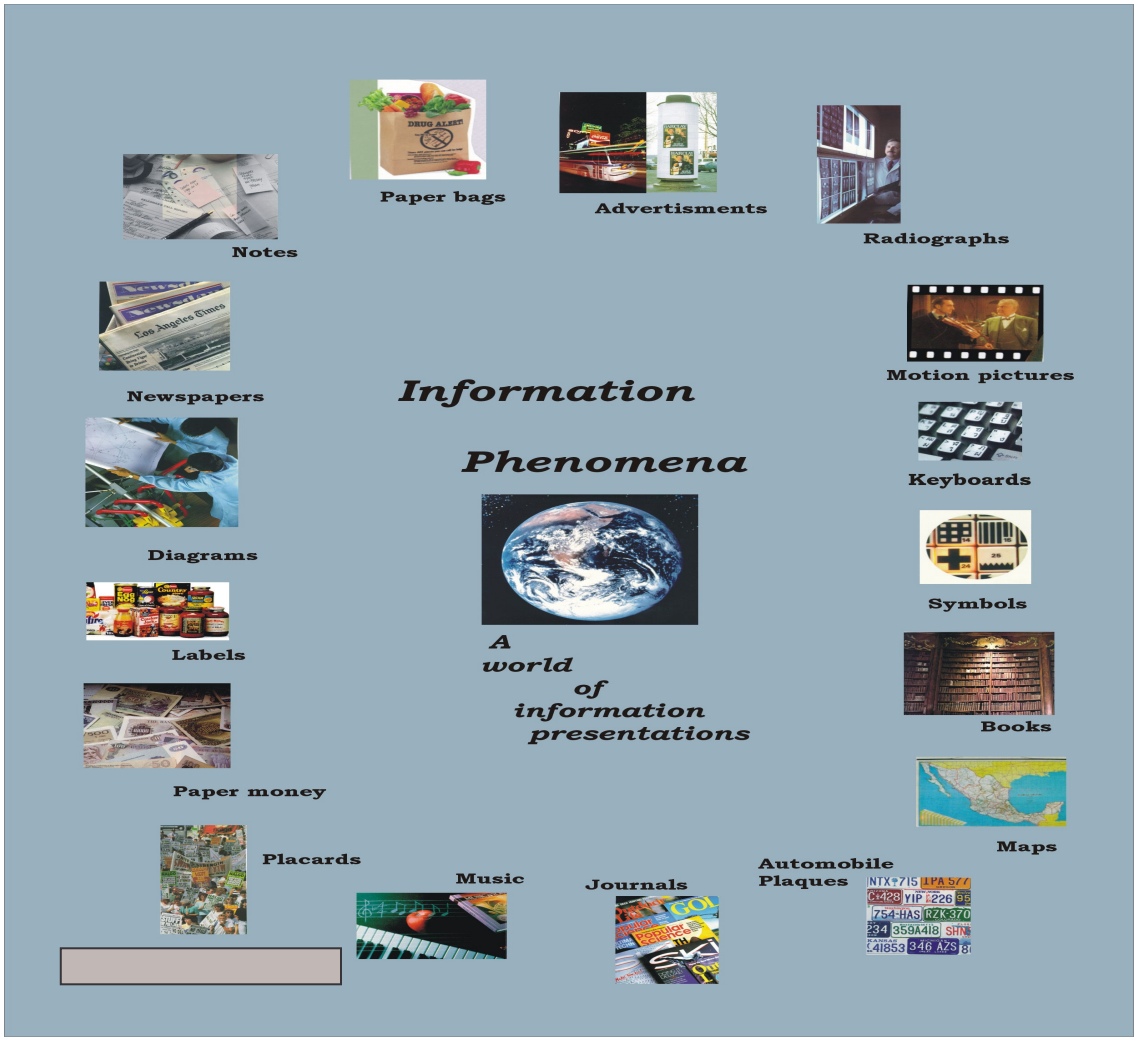
(A section of Hubbard's' "Exhibit of the Newspaper Publications of All Nations"

Claude Shanon is a milestone in the information studies for his 1948 publication [7] in which he defined the problems associated with moving information down a communication channel, distortion of channels, messengers and their measuring.

Eugene Garfield with the *Institute for Scientific Information* started forty years ago, the so called bibliometric and scientometric studies in which, through the citation counting process, measures for authors' production, impact of journals and research institutions are given.

Finally a special mention is due to the Lyman & Varian studies titled *"How much Information?"* [4]. Here they counted information produced in one year in four storage media: print, film, magnetic and optical and also transmitted through four channels: telephone, radio, television and the Internet. Overall they estimate the total output in 23 exabytes, 5 for the storage media and the rest for the information channels. One important consideration for these studies is that they were supported by Computer Companies, like EMC and not surprisingly most of the information produced worldwide was concluded to be stored in computers, and only 0.01% of the storage media total in paper. Other companies' estimations, for example by Eastman Kodak put the total information in the world as being over 85% in paper.

It is my contention that these efforts and some other less known ones are a very poor attempt at the comprehensive measuring of information. Citations and newspapers for example, are only two amid hundreds of presentations of information: books, films, patents.



LIBRU project.

So the LIBRU project aims at the measuring or quantification of information as overall phenomena. Thinking in terms of the wide spectrum already referred to I had to concentrate in the *homotrinic* section first. Here every item included under the broad term "information" is duly considered: maps, gravestones, slides, papyri, ostraca, video recordings, Christmas cards and several scores more.

Collection of data:

Different sources were used to get basic data for libru.

- Library/archives holdings: data from over 40,000 libraries, information centers and archives all over the world.
- Companies annual reports: 850 reports over a 4 years span period from the most representative information companies and corporations: publishers, telecommunications services, online services and similar.
- Books and journals: some 300 books and articles which give various statistics of relevance.
- On-line databases: OCLC-EPIC, ALA-directory, DBI-ZDBD and several more were analyzed

Consolidation:

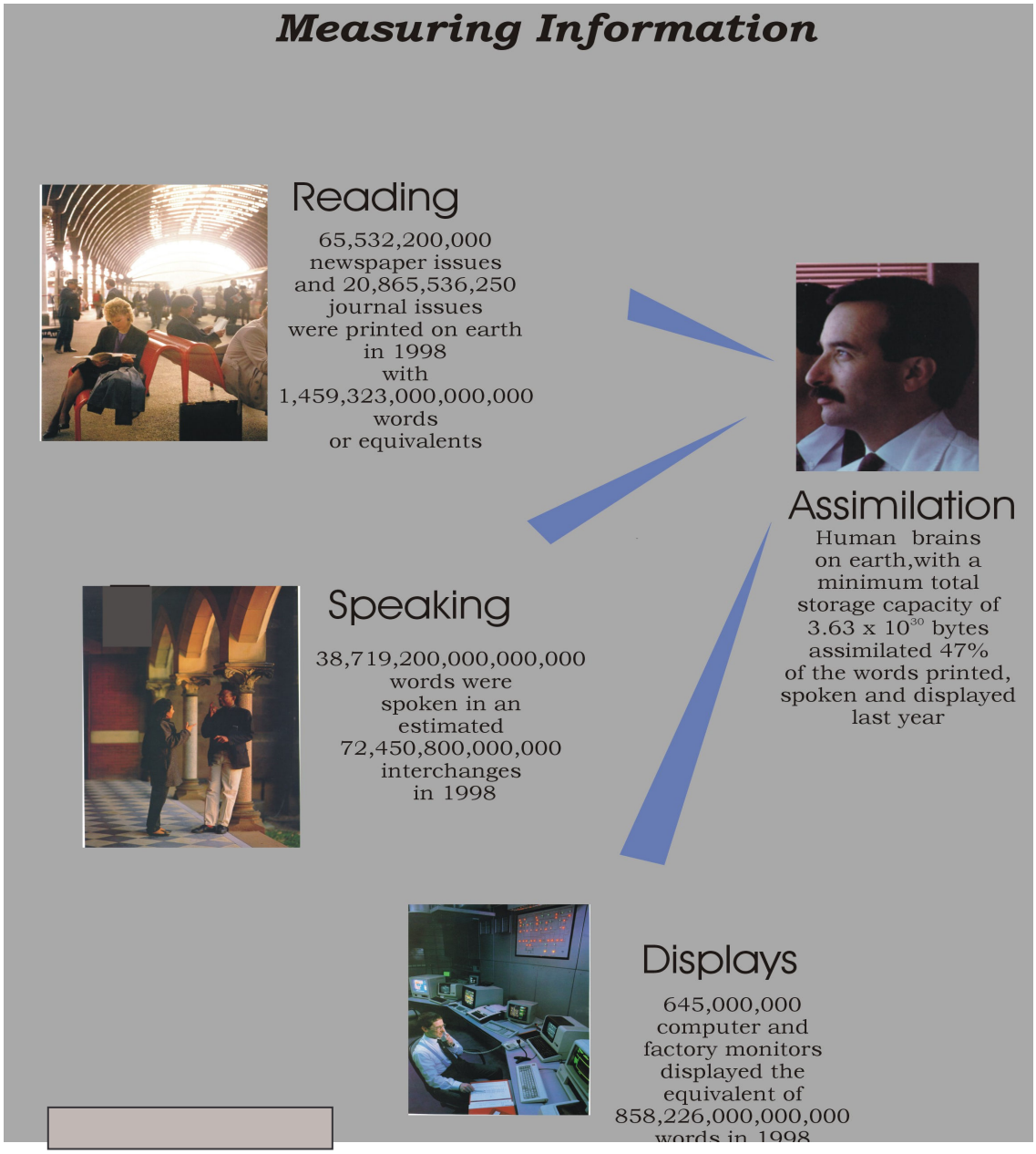
In the course of this project thousands of different types of information were studied, after a proper examination 66 categories or groups were identified:

Groups of Information Units.

1)	Advertisments	36)	Mss
2)	Av-mat	37)	Museums
3)	Books	38)	Music scores
4)	Cards	39)	Notes
5)	Catalogues	40)	Numbers
6)	Checks	41)	Paintings
7)	Cinema	42)	Pamphlets
8)	Clippings	43)	Paper
9)	Cloth	44)	Papyri
10)	Computer software	45)	Parchment
11)	Coupons	46)	Patents
12)	Curr per	47)	Photocopies
13)	Diplomas	48)	Photographs
14)	Diss/theses	49)	Post
15)	Docs	50)	Postcards
16)	Drawings	51)	Posters
17)	Electronic mail	52)	Prints
18)	Fax	53)	Radio
19)	Films 8/16/35 mm	54)	Receipts
20)	Games	55)	Reports
21)	Govt Docs	56)	Sculptures
22)	Gravestones	57)	Signs
23)	Information Data Systems	58)	Sound-rec
24)	Illustrations	59)	Stamps
25)	Incunabula	60)	Standards
26)	Inscriptions	61)	Telephone
27)	Invoices	62)	Television
28)	Labels	63)	Telex
29)	Machines	64)	Tickets
30)	Maps	65)	Video-recordings
31)	Mag/opt material	66)	Vols
32)	Memos		
33)	Microforms		
34)	Models	67)	Human beings
35)	Money	68)	Animals

Some of these groups were defined as such because they covered a set of materials with clear and distinctly characteristics; others like papyri because of their long tradition as an information spreading media. Taking for example the docs (documents) group we would find there: aircraft data, articles, charters, church records, court records, deeds, preprints, separate and many more specific items. Under numbers you would find, the serial numbers printed, stamped or engraved in every part of every machine; numbers for houses and buildings in the streets and a lot more. Categories 67 and 68 are the beginning of an attempt to measure information in the *biotrinic* sector.

(Some data calculated for 1998:)



Equivalences:

There are here two broad lines followed in this research. First there is the problem one faces when confronted with things like:

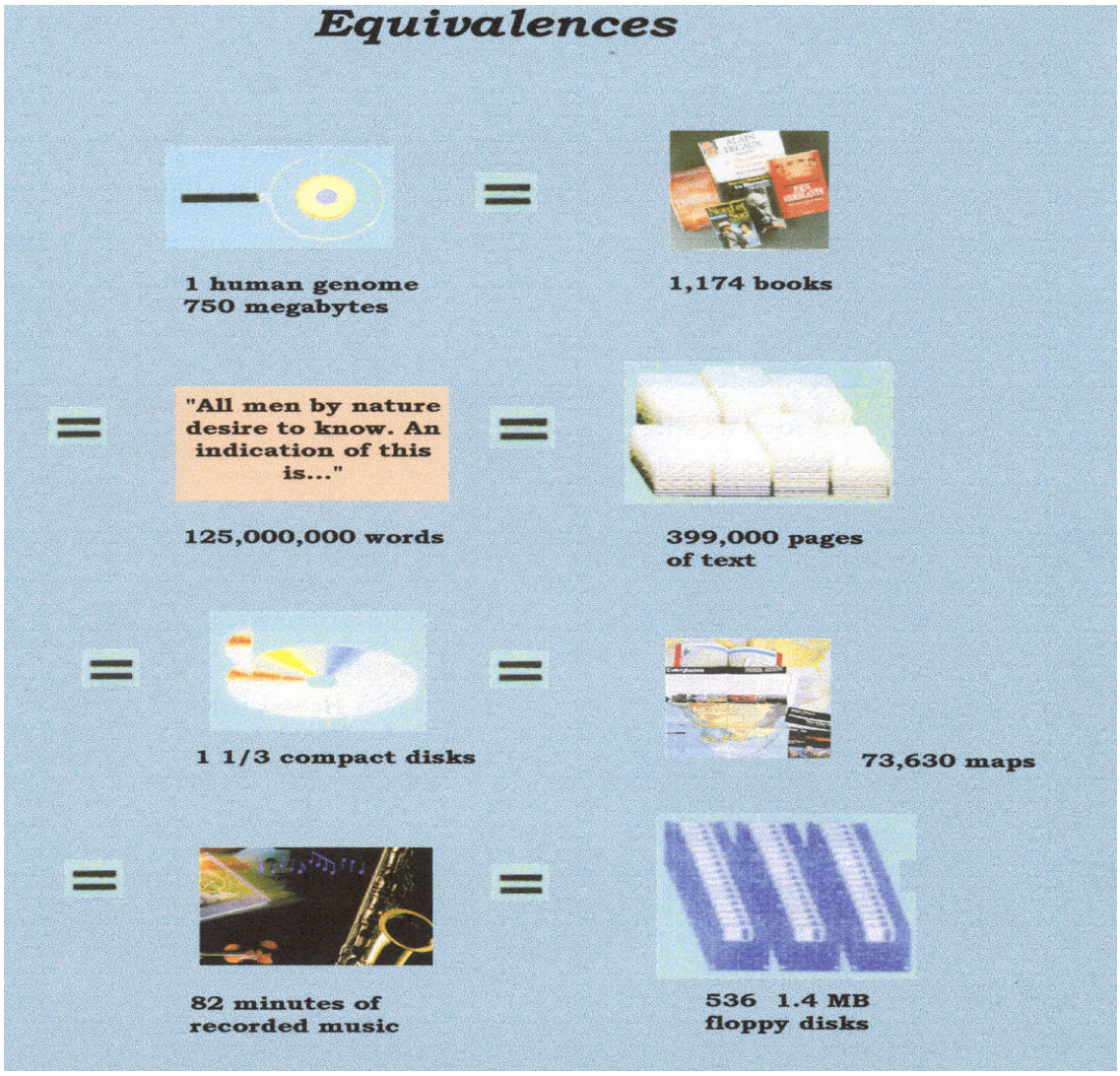
- Manuscripts
- Linear feet of manuscripts
- Boxes of manuscripts, …

How many manuscripts fit into a "box" and what kind of "box" to be precise?

Taking into consideration the most usual kind of boxes, drawers and similar in the market and then establishing the average size of each type of material; this last through some reports and studies already performed and through the actual measuring of different materials; tables like the following were elaborated:

Equivalences

unit	box	drawer	feet	meter
brochures	46	462	234	769
catalogues	13	130	66	217
diplomas	232	2,300	1,168	3,833
diss/theses	3	30	15	50
docs/indentures	50	492	25	820
govt docs	17	171	87	285
illustrations	263	2,600	1,320	4,333
leaflets	263	2,600	1,320	4,333
letters	789	4,000	2,032	6,667
maps	78	768	390	1,280
microfiches	364	3,600	1,829	6,000
mss	165	1,634	830	2,723
music scores	6	55	28	92
newsletters	56	552	280	920
pamphlets	22	220	112	367
patents	43	430	218	717
photographs	232	2,300	1,168	3,833
pictures	232	2,300	1,168	3,833
preprints	43	430	218	717
press clippings	525	5,200	2,640	8,667
programs	86	867	440	1,445
reports	12	120	61	200
serials	18	174	89	291
sheet music	47	460	234	767



Second, there is the problem of having equivalence for the different units between themselves and global equivalences.

Among the units usually found when referring to information as a whole are:

a) bits b) images c) minutes d) pages e) words f) infons.

Results: Some of the results of this project are tables giving:

- Amounts of a certain information group per country.
- For a given country how much information it has in every group defined.
- Number of libraries/archives and similar founded by year covering the last 1,600 years

As an example of these partial results we have the following table for current periodicals or subscriptions to journals per country:

CURR PER

Country	Amount
Afghanistan	30
Algeria	18,023
Angola	3,200
Argentina	214,892
Australia	930,720
Austria	155,042
Bahrain	1,392
Bangladesh	2,161
Barbados	3,585
Belgium	1,171,623
Belize	40
Benin	259
Bermuda	783
Bolivia	6,965
Botswana	1,311
Brazil	732,130
Bulgaria	112,787
Burkina Faso	1,037
Burma	7,120
Burundi	1,450
Cameroon	3,580
Canada	1,076,997
Central African Republic	603
Chad	311
Channel Islands	202
Chile	73,043
China-Peoples Republic	744,120
China-Republic	75,991
Colombia	128,993
Congo	1,905
-	-
-	-

New definitions:

What role plays a watch in the quantification of information problem? Is it an "information technology" item or not? If you have a photograph taken of a car, then you count one photograph, but if that same photograph appears in the newspaper circulating 2,000,000 copies, how many photographs do you have: 2,000,001 or just one or 1 original and 2,000,000 copies? A good copy has in many aspects the same weight as the original when in front of the users (newspaper readers and so on).

When you speak of having "all the works in one field or about some subject" you are usually, but not always, talking of "one copy of every item on the subject" rather than every copy of every item on the subject". Several dozens of such considerations exists and have to be properly studied when measuring information. As a result you have to define some concepts:

- *Absolute information*. These are the total amounts that there exists notwithstanding if they are originals or copies.

- *Relative information*. These are the amounts that exist once you relate them to a person or group of persons.
- *Mean-life of information*. This is the time it takes every item before it is transformed into another thing or else dies.
- *Half-life of information*. This is the time it takes half the production of one item, for example, copies of a book to be utterly transformed or destroyed.
- *Boxed information*. The one that is possible to keep in a physical medium, such like a box, shelve, library.
- *Fluxed information*. The one usually present in communications media. For example: telephone calls or minutes or conversation, postal items, TV, broadcast hours and so on.

The *crysinic, biotrinic* and *homotrinic* information parts of the spectrum have already been defined above.

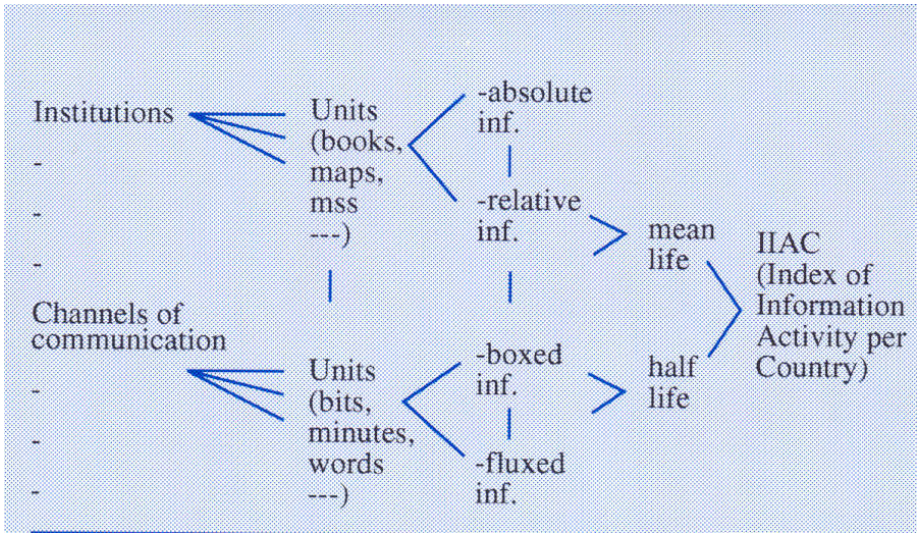
Updating.

As information is an evolving phenomenon it is very important to have channels for updating the data collected. One of the techniques used here is the conformation of tables of indexes of change. To calculate these indexes or factors much attention has been paid to the annual reports of some 800 companies worldwide, where many tendencies can be detected.

Index of information activity.

Among the results of this research an index of information activity is being proposed. This index is made of:

Index of Information Activity



This index will help the market of information in as much a way as the Dow Jones or S & P Indexes help other sectors, and this is by itself very important.

Final remarks:

Other things are being considered, for example: if we know that in terms of matter and energy, there exist the corresponding "conservation" principles, is it not reasonable to consider that the third element in our universe, information, would also have a similar principle?

The work towards the measuring of information will provide the more tangible elements contributing to the establishing of these kinds of principles, while building up *the information science*, as a first class discipline, with laws, paradigms, and models as specific and valuable as the ones in physics and chemistry are nowadays.

References:

1.-Anderla, George*: Information in 1985. A Forecasting Study of Information Needs and Resources,* OECD, Paris, 1973
2.- Hubbard, Harlan Page: *Hubbard's Newspaper and Bank Directory of the World, with gazetteer and atlas combined*. 3 vols, New haven, Connecticut, 1882-1884
3.- Licklider, J.C.R.: *A Crux in Scientific and Technical Communication*. Amer. Psychologists, 21, 1044-1051, 1966
4.- Lyman, Peter & Varian, Hal: *How Much Information?* SIMS, Berkeley, (2000, 2003).
5.- Price, Derek J. de Solla: *Little Science Big Science*, Columbia University Press, New York, 1963
6.- Price, Derek J. de Solla: *Science since Babylon*. Yale University Press, New Haven, Conn, 1961.
7.- Shanon, Claude: *The Mathematical Theory of Communication*, Bell System Technical Journal, 1948.
8.- Stonier, Tom: *Beyond Information*. *The Natural History of Intelligence*. Springer-Verlag, 1992
9.- Stonier, Tom: *Information and Meaning. An Evolutionary Perspective*. Springer-Verlag, 1997
10.- Stonier, Tom: *Information and the Internal Structure of the Universe*. Springer-Verlag, 1990