SCIRES-IT
SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 1, Issue 1 (2011), 3-12
e-ISSN 2239-4303, DOI 10.2423/i22394303v1n1p3

INFORMATION AND TECHNOLOGY ARE NOT ENOUGH: WE NEED ALSO KNOWLEDGE AND SCIENCE.

*Ferdinando Boero**

*DiSTeBA, Università del Salento: CNR-ISMAR (boero@unisalento.it)

Abstract

The era of Information and Technology has changed our life, allowing those who can access an internet connection to consult an almost unlimited number of sites covering almost all topics. It is difficult, however, to distinguish reliable information from unreliable one. The data posted in the web might be very precise while lacking accuracy. It is very important that information is based on solid knowledge, like the one guaranteed by peer-reviewed scientific journals. It is argued, with some examples, that the current policy of Information availability through the internet is not based on sufficient knowledge. More reliability is available at the sites of scientific journals, but they often require a subscription and enforce copyright restrictions, so hampering the spread of knowledge which is the basis of scientific "publication" (i.e. the delivery of the results of science to the public).

Introduction

There is a big difference between information and knowledge. In this historical period, information is easily available to all those who have an internet connection and a computer. Writing a key word in a searching machine leads to hundreds, or thousands or even millions of hits, where information on that word is available at the click of a button. So, information is available to everybody and is, thus, truly democratic, if we disregard for a while the billions who do not have access to the internet.

Easy access to information is of course confusing, because informations are often contrasting and can be interpreted in various ways. There are innumerable examples in support of this statement, so I will choose an extreme one, to show the difficulty of making up own mind when exposed to information not supported by sufficient knowledge.

Evolution, right or wrong?: The National Research Council of Italy vs the Pontificial Academy of Science

In the year 2009, a Darwinian year, because Charles Darwin was born on February 12th 1809, and published the Origin of Species in 1859, a very important scientific institution gave

a financial contribution to publish a book just on evolution. The materials of that book were the proceedings of a meeting that was held in the premises of that institution, following an initiative of its Vice President. The title was: *Evolutionism, the dismissal of a hypothesis*. The book delivered a very clear message: Evolution is not a valid scientific hypothesis anymore. Following that bold statement, the President of another important scientific institution reacted and said that there was no scientific basis for such a statement and that evolution is still the main pillar of biology.

Who is right? Having the information about the two positions might be confusing if there is no further knowledge. We cannot be knowledgeable about all the fields of investigation, so one might rely on the authority of those who sustain one of the two conflicting theses.

The two institutions I am talking about are The National Research Council (the largest public body dedicated to scientific research in Italy) and the Pontificial Academy of Science (an emanation of the Vatican, supporting the Pope and the rest of the Catholic Curch in the various fields of science).

Both institutions are important. The surprising fact is that the National Research Council hosted the meeting, and paid for the publication of the anti-evolutionist book, stemming from the action of its Vice President, prof. Roberto De Mattei, whereas the defender of evolution was the late prof. Nicola Cabibbo, president of the Pontificial Academy of Sciences.

If you were religious, and ignorant about evolution, whom would you listen to? And the same question might be posed to an atheist.

The internet is replenished of Creationist sites, and they propose their ideas as if they were supported by solid facts, criticizing the proposals of evolutionists. And of course there are many Evolution sites, ridiculing those who claim that evolution is an hypothesis surpassed by creationsm. Science is mixed with religion (as it happens so often, since the beginning of a distinction between the two).

The final result might be that one can find support to own thesis, in the internet, whatever the thesis. There are sites on extraterrestrial intelligence that provide solid proof about the esistence of Martians. Other sites can prove the existence of Guardian Angels. I am sure that if you want to believe that the Earth is flat, and that it was created 6.000 years ago, you will surely find support for your belief.

There are higher levels of validation, though.

If scientific papers are consulted, in fact, it becomes immediately evident that the president of the Pontificial Academy of Sciences is in line with current sceintific knowledge, whereas the

vice president of the National Research Council is expressing an opinion that has nothing to do with scientific research!

Peer review

What is the difference between a free site and a validated scientific site? The answer is very simple, in the free site you can write whatever you want, you have complete freedom. In a validated scientific site, for instance a site containing articles published by a scientific journal, any posted item is subjected to peer review. An editor appoints a group of reviewers who have demonstrated expertise on the topic covered by the proposed item, and they inspect the paper. Their expertise is demonstrated by a publication score on the same topic, from tribunes that are subjected to peer review. Peer means equal. People with equal knowledge to that of the author check the facts that s/he is proposing and try to find any possible fault. If there are faults, the reviewers list them and the author is asked to answer to the critics. If the answers are not satisfactory, the article is not published.

In other words: you are free to say whatever you want while sitting in a bar, you can even say that water is made of iron and lead. But if you want to publish an article in a scientific journal and you write that water is made of iron and lead, then you will not have your paper published. Freedom does not imply that you can divulgate wrong information, at least from some tribunes.

Serious journals and magazines, even those that are not necessarily scientific, do have fact checkers that control any statement made by the journalists.

If Prof. De Mattei would propose an article to a leading scientific journal, stating that evolution is not a valid explanation for the interpretation of the organization of matter in a living form, and he would propose his alternative explanation (i.e. divine intervention), then his article would not pass the reviewing process, because the support to his statements is very weak.

The incredible fact is that such a book is published with the label of the National Research Council, and this might induce the belief that the content of the book has been carefully checked by peer reviewers.

This extreme case is not isolated, though. There are many other examples of how the information can be very misleading, for instance by sustaining a series of acceptable

statements, while disregarding others that are equally accepted but that are left in the shade. Maybe the best case regards the demise of taxonomy in the era of biodiversity.

Information on biodiversity

The exploration of biodiversity is considered, since the issue of the Rio de Janeiro Convention on Biological Diversity, in 1992, the most important challenge for our species, since we can survive only if we preserve the goods and services that nature is providing us through the functioning of the ecosystems. Nations invested money on the study of biodiversity, lots of money. Big projects were launched to collect biodiversity information, so as to make it available to the scientific community and to the public at large. The investments, however, went all in one direction: technology and information, and zero resources were dedicated to improving the knowledge on biodiversity.

The result is that the internet is full of precise information about biodiversity, but we do not know if that information is accurate. Of course, if contrasting opinion are presented, the authority of the sources should help in decerning the right from the wrong answers. But the quarrel about evolution shows how reaching a reasonable consensus is a very hadrd task.

Precision and accuracy

The difference between the two is obvious to those who make measurements in a professional way, but the concepts expressed by the two words might be confused (as knowledge and information often are). A variable can be measured with great precision, and one might say that the temperature at a given place is 23,2536444569 °C. There are lots of figures after the comma, so the precision is great. But... that measurement might not be accurate. If measured with a better instrument (or with a set of better instruments) the temperature might result to be 24,1 °C. This value is not as precise as 23,2536444569 °C but what is the use of a very precise measure if, then, it results not accurate? What is the use of all this precision if the value is simply wrong? We have precise information about something (in this case temperature) but we end up not knowing its real value.

Back to biodiversity

We have described and named about two million species, our estimates are that there are at least ten million of them, but probably they might be even 100 million if we consider also the Procariotes. Obviously, our knowledge about biodiversity is still largely incomplete and we miss a huge amount of information (regarding the still undescribed species). Furthermore, species can be real or can be nominal, and there is a big difference between the two. Real species are biological entities and comprise groups of individuals sharing a common evolutionary history, not shared with similar groups descending from a common ancestor. Nominal species are just names, given to groups of individuals that might be real species or not. Furthermore, a real species can be given different names, when morphological differences are not properly evaluated. This originates the problem of synonymies, and the names we gave to organisms are often very redundant, i.e. the same species might have many names, each referring to one of the many morphologies a species can express (for instance males vs females, or larvae vs juveniles or adults). But it can also happen that under the same nominal species there are several real species, and we do not recognize them because they look similar. The distinction between the chimp and the bonobo required some time to be recognized, for instance.

The current knowledge on biodiversity is very incomplete (many species are still waiting to be discovered) and the precision of the names in identifying real biological entities is far from being accurate.

Putting the available information into a computer and making it available for everybody is a laudable initiative ONLY if the information is validated by knowledgeable people, who can distinguish nominal from real species or that, at least, are aware of the problem. A pre-condition in accepting an Information and Technology approach to biodiversity, thus, should be that the investments destined to the exploration of biodiversity are distributed fairly, respecting the contribution of the various segments of the scientific community.

This is not happening, and the portion of the scientific community that describes species (the taxonomists) are becoming extinct for lack of resources, just when their work should be crucial.

A capital mistake

The demise of taxonomy is due to lack of distinction between nominal and real species, between precision and accuracy, and between information and knowledge. Nothing wrong, if the choice would have been directed towards real species, accuracy, and knowledge. Instead, the choice fell on nominal species, precision, and information, and disregarding the alternative. The result is that the web is replenished with incomplete, inaccurate and sometimes wrong information about biodiversity and, furthermore, the knowledge required to recognize these faults is disappearing, so that our ignorance is perpetuated (which is bad) and is confused with knowledge (which is even worst).

A wise IT program on biodiversity

Information and Technology are not bad, but their are not the solution to all problems. They are an instrument and, as all instruments, they can be used with profit but, if used with lack of wisdom, they can be dangerous, just like a bisturi. As argued before, IT, for instance, contributed to the serious distress of taxonomy. It did so by attracting most funds dedicated to biodiversity research, leaving taxonomy in complete poverty in a period that should be characterized by its triumph.

Just like a bisturi in the hands of a serial killer, instead of those of a skilful surgeon, Information and Technology became lethal to taxonomy because its powerful approach has been used with lack of wisdom (and maybe also with some intellectual dishonesty) but, in fact, IT is of vital importance for taxonomy.

The pillar of taxonomy is the principle of priority: you cannot give a name to a species (i.e. you cannot propose a new nominal species) if the real species you are naming received a name already. If you do so, the name you propose will become a junior synonym of the older name, when your mistake will be recognized. Of course nobody erects nominal species while knowing that an older name has been given to the species in question. Nominal species are erected in the hundreds every day and, eventually, some of them might turn out to be junior synonyms of older names. The revision of supraspecific taxa, usually genera or families, is of vital importance so as to “clean” taxonomy from the junior synonyms and to label species with sound names.

Due to the law of priority, no taxonomic paper can be left behind, and we must take every description into account. The beginning of modern animal taxonomy dates back to Linnaeus' Systema Naturae, namely to the tenth edition of it, published in 1758. Hence, in order to have sound taxonomy, huge libraries are needed, containing all taxonomic literature. Such libraries are present in a handful of places, such as the Natural History Museum of London, the Musée National d'Histoire Naturelle of Paris, or the Smithsonian Institution at Washington. When making a revision of a genus, thus, a taxonomist must gather all the papers containing accounts on all the species that have been referred to that genus, form original descriptions to revisions to simple records. The list of the nominal species must be made, and then they must be compared, the type specimens must be inspected, new material is to be collected, allowing also for molecular inspection. But the basis of all this is the literature.

Hence, the basic building block of an information and technology aid to the study of biodiversity is the digitalization of taxonomic literature and its ordering so as to allow the extraction of information according to precisely posed questions. The simplest one involves the writing of the name of a species in the search engine and the subsequent extraction of all the information referred to that species from the taxonomic literature. Or the writing of the name of a genus, and the ensuing extraction of all the information regarding all the species referred to that genus. And we can widen the scope of our search, writing the name of a family or of an order, etc., and extract the list of the species and of the higher taxa and all the published information regarding them. Of course the whole body of scientific literature should be digitized and made freely available on the internet.

IT specialists know HOW to do this job, but they do not know WHAT to do to accomplish it. Taxonomists do know what to do, so they are essential, they must guide the building of the database, validating the entries so as to substantiate information with sound knowledge.

The making of such a database requires an enormous effort and, since it is not available, in spite of the enormous investments in biodiversity information, we obviously wasted our money. The available parts, with some noticeable exceptions, are often lacking accuracy. There are SOME papers reported, but not all. The synonyms are not worked out well, and the difference between names and species is disregarded. Furthermore, with the advent of molecular tools, many species are being fragmented into groups of twin species. They look similar, but their genes tell us that they are different from each other. These species should be named. But, in the past, nominal species might have been erected and then disregarded while corresponding to the newly recognized species by using molecular names. And these

apparently new species might have received a name already, albeit for different reasons. Taxonomy is like a flea market, nothing is wasted and everything might acquire value.

The Zoological Record

Since 1864, the Zoological Record gathers all the papers (and books, and any other kind of publication) published on animal species. These papers are listed and then there is a detailed subject index that lists all the names that are cited therein. The names of the new species are prominent, but also the names of all the other species are reported, if they have been cited in a paper, or book. Of course, we need to go back to 1758 to have a complete review of the available information on species: luckily, some authors of the past accomplished this task for several groups. All this information is just to be put into a computer, indexing it as much accurately as possible. The accuracy can be tested ONLY by expert taxonomists who have to work closely with the IT specialists.

Once this is assembled, it is not so difficult to extract the list of ALL the species described so far, distinguishing synonyms from accepted names. Synonyms must be linked to the presently accepted names. In this way, besides having all the citations of the species under its current name, also the previous names are available and the information provided with them, in the papers that used them to refer about a species that now has a different accepted name.

Organized information

The first product of this enterprise, thus, is the list of all the species described so far and the link of each name with the available information on that species. Such a platform should contain organized knowledge about biodiversity, providing vital information to taxonomists. The optimal solution would be to split all the literature into single entries regarding each species, so as to assemble these entries and build monographs for each species, containing all the information published on it, in pdf format.

The information about each species should be organized so as to provide information about crucial aspects such as:

adult morphology

complete life cycle
distribution
preferential habitat (and other habitat types where the species occurs)
trophic position
location of type specimens
type locality
timing of reproduction and other aspects of phenology
genetic fingerprinting or barcoding
and any other information that might be contained in the available accounts.
Such a data base would be of invaluable support to research on biodiversity and it is still mysterious why it is not available yet.

Interactive participation

The availability of organized information on all species might become upgradable (upon scrutiny by professional taxonomists) by any contributor willing to share own knowledge about that species. The most probable upgrades regard species' distribution but, for some popular groups (such as molluscs and insects) even non professional specialists can be of very high level, producing much valuable contributions on any aspect of the biology of species. The publication of geographic records, for instance, is not so rewarding in terms of career output, causing a lack of availability about where species do occur. Providing spaces to post such information (preferably supported by photographs and reference to deposited specimens) in integration to the already available knowledge, would allow a continuous upgrade of our knowledge about biodiversity. In a period of global change, in fact, it is much informative to monitor the answer of biodiversity to change in terms of species occurrence and timing of reproduction.

Copyright problems

There is a conflict between the need of making the results of scientific research public (since most research on biodiversity is financed with public money) and the need for scientific

publishers to earn their living, and those of their employees. Scientific publications are subjected to copyright and are usually available upon subscription. This means that if those who need scientific information do not subscribe to the journal where that information is published as a scientific article, then they are precluded from having access to what they need. This situation is common for research institutions of many emerging countries (but not only) that cannot afford subscriptions that, often, require thousands of dollars. Open access journals are a welcome novelty in scientific publishing, but it is undeniable that a large portion of the articles and books on biodiversity are still covered by the copyright and cannot be made public.

The solution of this problem has several facets. One might regard the policy of publishing houses, at least for areas that regard the study of biodiversity. Another might be the shared decision, by those who publish in these areas of research, NOT to publish in journals that do not allow the posting of the pdfs of biodiversity papers in specific repositioires and in the database described above.

The solution of the copyright problem is crucial for the success of such enterprises, and it is not a trivial issue.

Conclusion

The use of Information and Technology tools is crucial for a true democracy in science. Knowledge share throughout the world wide web is a handy reality, but there are still many issues that need clarification, the first being the accuracy of the posted information and its reliability.

These obstacles, however, seem easily surmountable. Information and knowledge are extremely important and should support each other. Knowledge is useless if nobody knows about it, and this is obtained through information. But information is dangerous if not based on solid knowledge. The third component in this game is wisdom! The totalizing importance given to Information and Technology led Knowledge and Science into distress, as exemplified by the taxonomy crisis in the era of biodiversity. Playing with acronyms, the solution of the problems highlighted in this paper resides neither in IT nor KS: we need ITKS!