Преглед НЦД 10 (2007), 68–70

Zoran P. Perišić
(Mathematical Institute SANU, Belgrade)

TOO MUCH … INFORMATION

Abstract: Scientists and scientific institutions use modern information technologies to communicate their ideas and research results worldwide. Distributed and grid computing are progressing very fast. The time and cost savings in disseminating and sharing information are huge, not to mention the ease of archiving with integrated search systems, both of which are essential for the digitalization. But nowadays nothing, even information, is safe as it might be, so there comes an urgent need for continuous improvements in the management of information security.
Keywords: information, grid computing, open access, security management

If you remember John Lennon's '' Imagine …all the people …'', be sure that all the 6,5 billion people from our planet Earth today, name by name, can be stored in just one standard-type hard disc of the PC on your table. This is a pretty simple illustration of how huge the capability of advanced electronic devices is, with them available at hand at very reasonable prices.

The new information world, where archives and libraries are becoming more and more virtualized, started developing less than 30 years ago when the first specialized databases were created. It was a long time before today's Internet. The first databases were rather large registries, but searching throughout them was much faster and more flexible than searching throughout printed lists of keywords. Soon this new computer based information world has started replacing the old paper world, making a completely new aspect of the information access.

Contemporary time yields a vast amount of data and information, especially in natural sciences and technology, making the global science archive multiplying at a huge dynamic rate every year. It also makes it difficult to extract the needed data even by using more and more sophisticated filters and tools. Instead of becoming happy with all these highly performing devices at hand, we encounter troubles with ongoing processes in the wide field of digitalization, in particular. To illustrate this recall that only half a century ago science and technology produced about 2000 new publications daily, but nowadays there are tens of thousands of scientific papers published every day! Exactly the same happens with the number of scientific journals in total: in the mid of the 19^{th} century there were roughly 1000 journals published and nowadays the number is estimated to be bigger than 200000!

Huge databases offer a wealth of information but searching throughout them is a demanding task. Something that they have to provide is a quick search of user interfaces, reasonably priced relative to the complexity of the search, the reason being that information has acquired strategic importance, particularly in the world of science and technology. But, was it all like that in the past?

Let me tell you a short story: when I was very young every building in my street had an attic. Each of them was a spacious, single room accessible to all the tenants who have used it to store all the 'useless' things like old books, magazines and photographs

are. They were really a kind of what we call nowadays the share-ware. Every building had a basement, too. But each basement was divided into separate cellars, which were carefully locked by their owners because some 'useful' stuff, like coal and woods for winter days, was stored in them. Oh yes, it was something very worthy at that time.

But, as Bob Dylan sings, the times they are a changing. Today we are trying to connect attics and cellars, aimed at sharing worthy information and data from books, magazines and journals on a quite new base – thinking all the time about their reliability and their security.

Today there are lots of easily accessible sources of information on Internet but many scientists, even those very experienced, overestimate their reliability. In order to ensure the availability of scientific information over a long period of time, networks of non-commercial, internationally controlled institutions should take over responsibility for archiving data, for standardization of data formats and interfaces, and should provide open source-based access software. Some of the aims of new electronics policies are to comprise complete documentation of the scientific and cultural material (including the heritage), to enable the use of multimedia tools, to archive the original data and to provide online-networking of the sources used. The technical accessibility of systems must allow the broadest possible dissemination and usage of the information and data, and all that together must make the scientific work and results available to the public, with no restrictions imposed.

As we have already mentioned before, one of many other complex problems in practice is that the volume of the stored data is so massive that even the best performing personal computers are not able to process it. This problem has forced scientists to develop the emerging field of grid computing, with uniform standards for applications on the grid. Grid computing is aimed at strengthening multiple computer resources for solving complex problems, like those concerning huge data management, or like those concerning joining together computers, file systems and networks. The development of grid technologies enables coordinated resource sharing and problem solving in dynamic, multi-institutional, virtual organizations through advanced software layers that unify resources and simplify their use.

As we saw before, people know how to lock their doors, their cellars and how to close the blinds. However, most of them, including many scientists, haven't developed the same habits in the rapidly expanding digital world of data and information, which should be also properly protected and secured, because science processes can no longer operate without a continuous supply of information. The information security is an important activity, intended to control the provision of information and also to prevent unauthorized use of information; it is considered to be one of the main management challenges for the upcoming years. Although its importance was largely ignored in the previous period, the security of information is in our focus now; in particular this is due to the growing use of Internet. The number and the use of local computer networks have also grown very much, not only within scientific centers but also between them, as well between the science and the outer world. The increasing complexity of information technology infrastructure implies that the science, in the whole, is now more vulnerable to technical failures, computer viruses, human errors or intentional human acts. The growing complexity requires a unified approach in the security management, which should encompass the development, introduction and evaluation of security measures.

Information security is also aimed at ensuring the safety of information in the sense of making it invulnerable to known risks, and of avoiding unknown risks

whenever possible. A tool to provide this is the security, and its main goal is to protect the value of the information. This value depends on confidentiality, integrity and availability. Here the confidentiality includes protection of information against unauthorized accesses and use, the integrity includes accuracy, completeness and timelessness of the information, while the availability means that the information should be accessible in any agreed time. This depends on the continuity provided by the information processing systems. Other aspects include privacy, anonymity and verifiability (ability to verify that the information is used correctly and that the security measures are effective).

If the consciousness of the need for having accurate information and data management exists, certain security requirements for data management should be clearly established and documented, whilst responsibilities for that should also be assigned to the data owners, users and customers. If the need for security of data and for understanding of all required actions exists, the knowledge exchange about them must become a standard practice between scientists and all other users. Data and all kinds of information including software, which are available in systems with public or open access (such as grid computing is) and which are also assumed to be at a high level of integrity – should be protected by appropriate measures, like digital signatures are. But the use of digital signatures tends to a wide variety of legal acts, particularly towards the aspect under what circumstances are the digital signatures obligatory.

In scientific systems with open or public access the management of data and information, including optimization of the use of information and ensuring that the information is available as it is required by users, is focused on maintaining the completeness, the accuracy and the availability of the data, as well as on the protection of the archived data. This goal is achieved by proper archiving, by proper backing-ups, by testing restorations, by adequately managing the onsite and the offsite storage of data, and by providing secured systems for disposal of data and information. The measure of success in the managing is measured by users satisfaction with availability of data, by the percentage of successful restorations, and by the number of incidents where sensitive data was retrieved after being disposed to the media.

To conclude: in broadest terms, we have discussed about the thin line between the open and free disposition of information and the growing quests to secure it, both in favor of the best possible flow of information and data throughout the whole world of Science.

References

1. Roland Wenegenmayr, *The Global Archive of Science*, Max Planck Research, 3/2006;

2. Iva Helms, *Catching a Wave with the Research Net*, Max Planck Research, 3/2005;

3. Zoran P. Perišić, *Digging into the future of the past*, Review of the National Center for Digitization, 7/2005;

4. Zoran P. Perišić, *Open Sesame*, Savetovanje u NBS, 2005, unpublished;

5. APS News, April 2005, Vol 134, No 4;

6. Robert Schlogel and Theresa Velden, *The Freedom of the Internet*, Max Planck Research,3/2004;

7. Reto M.Hilty, *Copyright in a Digital Dilemma*, Max Planck Research,3/2003.