

Pedro Ferreira

The Knowledge Society Agency (UMIC), Portugal

### **New infrastructures and frameworks for knowledge creation and diffusion**

I believe it is always useful to recall briefly the importance of knowledge repositories and to trace back how men came to perceive the paramount role that DR can play in today's networked society.

Scholar research about the importance of knowledge is still a new topic. It was only about half a century ago that Robert Solow trying to explain the economic performance of nations showed us how the traditional models of economic growth based on labor and capital could not explain the empirical evidence before us and something else was at the root of the process of wealth accumulation. Some 20% of that growth could not be accounted for and it was then named the residual.

But this residual has been growing ever since we measure it and when Solow actually won the nobel prize in economics, two decades ago, the residual was much closer to 80% than the mere 20% that Solow could not explain a decade after the second world war. Today, we call this the effect of knowledge and technology and this is the extent to which it became important in modern society.

We thus are fully aware of how important it is to effectively manage knowledge. As such, concepts and tasks like knowledge creation, diffusion and absorption have developed, either at the individual level, at the institutional level and at the national level and have been exhaustively studied in the literature.

Now, Europe as a whole is tackling this issue at the continental level establishing routines for knowledge creation and sharing across traditional geographic and administrative borders. This is a new exercise that is already offering strong headaches to you, the visionaries of how this can be accomplished in a timely manner.

Yet, we should not forget that, broadly speaking, there is codified knowledge and tacit knowledge all around, codified knowledge being the knowledge that we can translate into some intelligible language and record somewhere and tacit knowledge being that special aptitude that humans have to accomplish certain tasks and objectives that we can only explain to others by doing and not by describing them in DR.

And thus there is the first downside of DR, or in a milder way, this is something that DR, per se, cannot do for us, which is to become substitutes for the creative process of knowledge creation. But they can certainly bring us extensive good as effective tools to support knowledge diffusion and knowledge preservation as a way to facilitate knowledge absorption.

Digital repositories are important to the extent that codified knowledge is. And in recent times, we all have come to understand codified knowledge as a fundamental asset to achieve wealth creation, shared prosperity and, in many cases, social cohesion through inclusive development and we have already seen presentations of various projects that today make this real around the globe.

But now one always needs to go a step further and understand how strategic DR will be to accomplish these overarching goals and how important it is for us to make sure that a proper system of incentives is in place that allows, or if you want convinces, all of us to deposit information and knowledge into global DR. And this is where policy decisions regarding the development of DR are much needed. I will get back to this in a moment.

I would like to note that recording information and knowledge into DR is, per se, a noble task that will allow us, if nothing else, to keep track of mankind's achievements. However, and in good truth, we want to do more with DR than just archiving.

We are not interested in promoting static DR. In turn, we need to define suits of services that DR can provide taking into account that particular sub groups of the population will require refined access, and I am talking about researchers, teachers, doctors and the like, but also society at large will also turn into the high-quality information that repositories might provide to manage an increasing number of aspects of today's daily life, much notably aspects related to dealing with public risks.

Perhaps biased by the staggering growth of WWW, we tend to view digital repositories as huge amalgamations of information and knowledge. But there is yet a long way to follow to reach such state.

First, DR must be accessible to both knowledge creators, who will fill them up with interesting content, we hope, and to knowledge receivers who want to search over DR to find what they look for. So, we are talking here about the

need to develop tools for populating DR and for navigating through DR using intelligent discovery functions.

But to support all of this we ought to have an infrastructure layer in place that allows us, on the one hand, to massively store information and knowledge, and on the other hand, to use the appropriate bandwidth to access DR. And, of course, we can only expect objects in DR to become ever more complex and thus heavier to transport over communication networks.

Fortunately, and in a continuum over the last few years, a vision has emerged in Europe that positions this continent in a privileged place to lead the world in what e-science is concerned. Previous efforts by the European Commission and by Member States, namely through previous FPs, allowed for putting in place the needed infrastructures.

GEANT, and its global reach, ensures widespread connectivity across all research institutions in Europe and provides high speed connectivity between Europe and other continents.

EGEE is day by day showing us how we can collaborate around the world using, in a distributed fashion, massive storage and processing capacity over grids that we could hardly conceive installing in one single place.

FP7 is now set to show us how digital repositories are yet another important part of this infrastructure that we lack today to an unwanted extent.

Now, this vision needs complementary policy action from policy makers across Europe. Developing DR and the tools and services associated with them needs direct long-term funding on a rolling basis. A system of aligned incentives needs to be put in place as to mobilize knowledge creators to deposit outputs into DR. Because knowledge creation cannot survive without knowledge diffusion, the processes of R&D must increasingly include publishing activities into DR and thus be funded accordingly.

While men have used many forms of repositories for centuries in a row to register and communicate knowledge across generations, the digital nature of the knowledge objects we want to save in DR today is yet something quite new to us. We don't know enough about digital preservation, but we must to.

There is thus a pressing need to fund preservation activities, in addition to funding projects aimed at improving discovery tools, authentication and authorization mechanisms.

In parallel, object counting mechanisms that can feed statistics of data citations and similar R&D indicators must be sought. However, they will only be relevant when DR become the most used platform for publishing R&D outputs and we must keep this in mind.

But yet more importantly than tracking R&D output indicators is perhaps to keep track of the social web around science making that will emerge on the top of DR, in line with the widespread spirit of today's networked society.

Special care must be placed into promoting interactions among institutions and people that stem from what we import into DR. In other words, DR must be the beginning of it all and not the bottom-line of R&D. Open DR should become a new avenue for science to re-invent itself once again.

As DR become key for knowledge management and e-science the need for training on how to work with DR will certainly arise, and Europe can consider defining careers for practitioners and managers of DR, in addition to promoting mechanisms for certification of institutional DR. IPR and harmonization of law also need to follow up the accelerated pace of the development of DR and legislation needs to face these newest challenges posed by the knowledge economy.

And to finish, guess what? After all, DR can actually also come around to help us beyond securing codified knowledge. This is because if we can add a semantic layer to DR that explains what DR are and what is inside them, it is likely that machines themselves can navigate through DR to create new knowledge from knowledge previously recorded in DR. This, perhaps still far fetched view of things, is just around the corner with recent developments on semantic webs and I believe that projects like neuro-commons at MIT might surprise us all in a relative short period of time.

However, this view that I have just shared with you puts a significant burden on early decisions that we need to make today about how DR must develop. And in good truth, most of what we need is a coherent set of principles that can guide us through effective implementations of DR.

Note that today we know that what was missing 30 years ago, when the Internet was born, was a clear set of principles that could have guided the growth of the Net.

Principles that are dubbed to be root for the success of the Internet, such as the hourglass and the end-to-end principles were perhaps more observations of fact once things were done than actually early goals of the development of the Internet and principles that we would have certainly liked to account for at that time include security and quality of service which were not considered then but are top priorities for anyone who today envisions building new communication networks.

But it is clear to me that today there are even suggestions for sets of principles for the development of DR in Europe and around the world. Besides the aspects of legal conformity, protection of IPR, besides interoperability, quality of service and accountability, let me stress the importance of the so much debated issue of open access.

It is clear to me that the open nature of science and of science making can only call for some sort of open access. Knowledge is only valuable to the extent that we can update and improve upon our current stock of it and that is best achieved by sharing and using other people's knowledge.

Furthermore, it should be also clear that the creation of some knowledge is only made possible due to public funding of certain R&D activities. Thus, it should be natural that some knowledge, perhaps the knowledge codified as outcomes of publicly funded R&D activities, can only belong to the public besides belonging to those who worked to develop it and the open access scheme with a short embargo period tackles exactly this dichotomy.

So, there is certainly a need to strike a balance between public knowledge and privately owned knowledge. And what we have observed in recent times is precisely a set of market forces that renders a worrisome imbalance on this relationship that lead a number of communities around the world to feel hurt, I would say, and to start envisioning new models for managing and sharing knowledge.

And my take is that this movement will be as acute as current publishing mechanisms promote such imbalance. I thus welcome the efforts aimed at

developing new business models for publishing and the many pilot projects that have been addressing this issue in recent times.

It's the understanding that knowledge diffusion is part of the R&D process, and an important part of any S&T and innovation system, that will lead to extending R&D funding to cover publishing. But of course, and put this way, such funds will naturally be allocated to the most efficient way to do so. If current practice translates to significant welfare losses to the S&T communities then it should be no surprise that new mechanisms appear as serious alternatives.

The Portuguese Presidency of the EU is very much attentive to these efforts and intends to further delve into this issue during the next couple of months following the results of the conference we will hold in two weeks time in Lisbon on the future of science.

Meanwhile, I think we should welcome the sets of principles that have been proposed to guide the implementation of DR across Europe and all the progress that has been made so far towards implementing DR at the European level and it is with renewed confidence that I believe that the recommendations that I have just highlighted will contribute greatly to the advancement of e-science in Europe and will establish a powerful platform for Europe to engage in fruitful collaborations with both East and West and look ahead towards South.

Finally, it is interesting to note that while all of this develops a huge digital repository has evolved just right under our skin. It is called the WWW. And it is interesting to note that Web 2.0 raises a new paradigm for knowledge dissemination.

Today, every user can upload content to online repositories and the most visited websites worldwide are those developed based on the user's content as opposed to those that provide us with top-down information defined by the newest newsprints. The web is thus a good spaghetti of information, developing in a quite an anarchy-like spirit.

This is good news for DR. While the Web is developing into a world of increased freedom of expression, DR can develop into providing information from trusted sources and peer reviewed knowledge, thus marking a noteworthy difference to the Web that people are likely to increasingly appreciate, probably starting from the workplace into their personal lives.