Taming the vast -- and growing -- digital data-sphere

September 22 2009

(PhysOrg.com) -- European researchers are making an impressive effort to link up digital repositories to create a vast network of easy to search online data. The DRIVER project work - one of the largest efforts of its kind - aims to make some sense and better use of the growing online digital world, the 'data-sphere'.

Europe has thousands of digital libraries and archives located in different institutions across the continent. These repositories house information or data which has been converted into digital format for safekeeping. The sort of information that traces civilisation or the sum of humankind's knowledge, from science and language, to the literary works of great scholars.

But while keeping digital records has its advantages, it also raises some problems of its own. The platforms or technologies used to create and maintain these digital databases may change or become obsolete. The software or systems used by different repositories in, for example, Italy may differ from those used in Belgium or Sweden. And as time passes, the ability to retrieve this data - this valuable documentary of humankind - may be compromised, otherwise known as a database 'legacy' problem.

The sheer volume and variety of information, broad technological differences and linguistic diversity are yet more barriers to accessing knowledge contained in digital repositories. Only those with access to the right technology or who know how to create complex database queries gain entry.

This is likely to change for the better. The DRIVER project is a coordinated, multi-phase effort by European information scientists to create a cohesive, robust and flexible, pan-European infrastructure for digital repositories. The researchers have already created a search engine that regroups over a million 'open access' articles from 260 of Europe's leading institutions.

Open access is a major theme of the project. A lot of publicly funded research ends up locked in proprietary journals, accessible only with often-expensive subscriptions. The Open Access movement is reversing that trend, to make published research readily available especially over the internet.

Where possible, Europe now mandates that publically funded research be made freely available to the general public and researchers, but technological bottlenecks remain. The DRIVER project is tackling those bottlenecks step by step, and it now has a stable version of its technology available to both casual surfers and serious scientists alike.

The infrastructure, of course, links diverse repositories of knowledge, but it will also offer sophisticated services and functions for researchers, administrators and the general public. It is the largest effort of its kind in the world and it intends to enhance repository development worldwide.

Customised experience

The EU-funded DRIVER project has performed a vast programme of research on a modest budget of just €2.5 million for phase 1 and €3.4 million for phase two, which runs until November this year.

Their work embraces everything from fundamental technology issues to interfaces (the way the user interacts with the technology) and

international standards. The main technological focus is D-NET v1.0, which is a suite of open source software - meaning anyone can work with the underlying code. The system allows users or institutions, no matter where they are in the world, to customise their experience. What's more, it is extensible which means new services can be created on top of the system.

D-NET is really a pioneer. It allows users to collect together open access content from diverse institutional repositories and presents the content in a uniform and openly accessible way.

The software has been released under the Open Source Apache licence and is available at www.driver-repository.eu/Downloads . D-NET can be used by anybody who wishes to set up a similar portal providing services like the DRIVER search portal, for example national or thematic organisations starting their own initiatives.

Moreover, institutional repositories can use the software to plug in and allow their content to be accessible from DRIVER or any of the other deployed portals.

So far, DRIVER's technology work has created a stable platform that can access any kind of text document, but the researchers will continue development until the system can access content from any media.

"Really, this is a project that will never end, because there will always be something else to do, or new standards and technologies will emerge that need to be added to D-NET," concludes DRIVER's coordinator Professor Yannis Ionannidis. "We have achieved a milestone, but it is the first of many."

The DRIVER project received funding from the Research Infrastructure priority of the EU's Sixth Framework Programme for research.

More information: DRIVER II project: www.driver-repository.eu/

Provided by ICT Results