Face-to-Face with Wolfram Horstmann
CIO, Bielefeld University Library
Expert in the DL.org Quality Working Group
Programme Chair of Open Repositories 2010 – The Grand Integration Challenge, July 2010, Madrid

Digital Repositories are being developed in a number of fields, spanning education, research, science and cultural heritage, on national, regional, institutional, lab and personal level. In this evolving landscape, repository platforms are transforming the nature of scholarly communication. Universities, research laboratories, libraries and commercial organisations are developing innovative repository-based systems that address the entire life-cycle of information, from supporting the creation and management of digital content, to enabling use, re-use and inter-connection of information and ultimately ensuring long-term preservation and archiving. In this interview with **Wolfram Horstmann** (pictured), Programme Chair of the Open Repositories 2010 conference and an expert serving in on the DL.org Quality Working Group, we explore the issues around integration that were addressed during Open Repositories with reference also to interoperability from the perspective of quality within the context of DL.org.

How would you describe the main aims and focus of Open Repositories 2010?
The Open Repositories conference series, now in its fifth edition, has grown to become one of the most important conference on repositories, particularly in terms of ensuring academics are well placed to offer rich information sources. This year's Open Repositories brought together individuals and organisations engaged in the conception, implementation and management of digital repositories, as well as stakeholder communities to address theoretical, practical and strategic approaches and ensure Digital Repositories offer value-add now and in the years ahead.

The burning issues at this year's conference are summed up as "The Grand Integration Challenge", as repositories become increasingly complex and dispersed, not just linking to Google but to a very broad provider scenario encompassing research publications, research data, bibliographic data and administrative data linked to the world wide web and diverse applications.

The Open Repositories conference series primarily targets the infrastructure community made up of developers and service providers who focus on machine interoperability as well as the end-user community. One trend that has emerged through Open Repositories is an increasing focus on developing repositories that cater to specific academic domains, including customised services meeting specific end-user requirements in fields as diverse as Biology, Chemistry or the Humanities. But while more end-user oriented services are emerging, there is also a great deal of interest in infrastructure development and issues that pertain particularly to developers, researchers, and service providers.

What specific burning issues were addressed?
This year's Call for Papers defined integration as a grand challenge that encompasses diverse facets along a multidimensional spectrum with the aim of shedding light on the most important facets by ensuring effective knowledge exchange on the different trends and perspectives that need addressing to ensure integration in the future. "**Web and repositories**" was one such challenge addressed, where the aim was to define the optimisation of repositories as not every repository is suited to the web. Other issues are concerned with "**knowledge and technology**" and "**wild and curated data**", that is, the tension between what is seen as valuable from a user's perspective and what developers believe is important. In other words, some groups believe well curated content and high quality are key, whereas others believe people should be left to do what they want.

The keynote by David De Roure from the University of Southampton addressed the challenges of "**isolated and linked data**" with special reference to MyExperiment and why and how linked data has been embraced to bring value-add to the services provided. "**Disciplinary versus institutional systems**" was also explored as an important theme moving forward. Here the contrast is between community-centric approaches and approaches that put Quality of Service (QoS) and long-term access high on the agenda, along with opposing perspectives, such as "**scholars and service providers**" and "**ad-hoc versus long-term access**". Open Repositories 2010 also explored "**ubiquitous and personalised environments**", where we have open environments such as Google in the first instance or a combination of open and restricted access in the second instance, which is particularly important for many types of research data. Finally, the conference looked at emerging trends like "**the cloud and the desktop**". While an increasing number of resources are moved from servers to a cloud, it is also integral to reach the desktops of researchers so building this bridge is important.

Summing up the main conference achievements, how do you see the Open Repositories conference developing in the future?

I see a number of developments taking shape over the next few years. One such development is an increasingly professional approach. This year's edition was the largest to date with 400 attendants from 36 countries. Open Repositories has always been an international effort, but global involvement is likely to grow in the years ahead. From a quality perspective, this year there were 6 times as many paper submissions, totalling 250, with a rejection rate of around 60%, which is an important achievement to take forward. The diversity of topics explored is a new trend that will probably continue. We can expect to see a shift away from open access and institutional perspectives as the main topics towards the world of research data and administrative data, particularly on building high-quality data resources in a landscape of diverse systems. It will be interesting to see how this differentiation is taken forward next year.

Wolfram, you serve as an expert on DL.org's Quality Working Group. Why is interoperability crucial for digital repositories and what is the role of quality?

With regard to repositories, interoperability is a decentralised paradigm that poses the question of how to link very heterogeneous and dispersed resources from all around the world. Interoperability is thus a crucial theme and the ultimate goal we are striving for. The role of quality in DL.org's Working Group is twofold. On the one hand, there is reliability of services. When building systems and operating on data in a distributed infrastructure, you need to be able to rely on every part. Precision of data is the second challenge. For example, in a major repository service, the Bielefeld Academic Search Engine (www.base-search.net), there are currently 1700 repositories with 25 million records for academic resources harvested on a daily basis. Considerable effort is needed to arrange all the filters for individual repositories to ensure that the end user has an homogeneous experience in working with such diverse sources. Quality must thus be provided in a decentralised manner, which requires standards.

Nicholas Ferguson & Stephanie Parker, Trust-IT Services Ltd

Wolfram Horstmann is CIO Scholarly Information at Bielefeld University, affiliated to the Bielefeld University Library. He is a Biologist by training, has an experimental research background in Computational Neuroscience and received his PhD for work in the Theory of Science. He has been involved in information management developments since 1997, e.g. distributed systems for academic publishing (across science and humanities, DIPP), for complex media in eLearning and electronic textbooks (educational simulations for the neural and cognitive sciences, MONIST) and for generic information management infrastructures (DRIVER: Digital Repository Infrastructure Vision of European Research).