The data deluge has made the scientific method obsolete – at least that's what the editor of Wired magazine, Chris Anderson, announced last year.

“The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world,” he claimed. “Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.”

Big data, big science: whether you agree with Anderson or not, the nature of scientific research is changing dramatically. Governments around the world have poured huge amounts of money into what the British call ‘e-science’. In Australia we call it ‘e-research’; the Federal Government’s NCRIS program has invested about $400 million since 2005. The $1.1 billion Super Science initiative announced in 2010 will add a lot more.

Similar things are happening in the social sciences. The Australian Social Science Data Archive (ASSDA) is collecting large quantitative datasets and making them available for analysis over the Web. ASSDA is also moving into the area of qualitative social science, though this raises some challenging questions about re-using data from interviews and surveys.

E-research in the humanities is more problematic. Much of the digital work in the humanities to date has been funded as part of specific research projects, and has focused on building scholarly Web sites. The humanities have received very little funding from programs like NCRIS. Nevertheless there are signs that fundamental change is on the way.

For a start, humanities research projects are getting bigger. In Australia, the ARC's Network for Early European Research (NEER) has spent the past five years building a national framework for research in medieval and early modern studies. We’ve done this partly by funding Research Clusters to encourage international research partnerships and partly by setting up a Web-based collaborative environment using software called Confluence. We’ve linked up with our European equivalent, CARMEN, to develop proposals for large-scale projects.

An increasingly large amount of source-material for humanities researchers is available in digital form. Cultural institutions have digitized a significant proportion of their unique holdings, making many artworks, manuscripts, archival documents and historic objects available over the Web. Early European researchers are particularly well-served. Almost every English book published between 1473 and 1800 is now available in digital form.

But this does not add up to e-research in the scientific sense. All these sources of data need to be joined up, to enable researchers to pose large-scale questions across the whole corpus of material. Boutique web sites focused on scholarly interpretation need to make room for sites which assemble large amounts of data so that they can be re-used by other researchers.

Earlier this year, I was at a workshop in Birmingham (UK) sponsored by the European Science Foundation, which was attended by medieval manuscript experts from all over Europe. We set ourselves this challenge: how can we build a digital environment which enables research questions to be asked about all the medieval manuscripts in public collections around the world?

At the moment we don’t even know how many manuscripts there are. Our first task was to put together a road map of the digital infrastructure services which will be required. We will be using this as the basis for large-scale funding applications.

The crucial question is how to organize and present the data. Traditional databases are too rigid and limited, while converting existing data to a single format is a pipe-dream. The best way forward involves the use of so-called Semantic Web technologies: uniquely identifying objects, people and concepts; constructing graphs to describe and navigate the relationships between them; and linking them to all kinds of relevant digital data.

This approach is being used for Europeana – the European Union’s digital library of cultural heritage objects. The Humanities Research Institute at the University of Sheffield has been using text-mining software to identify, extract and encode personal names found in more than 190,000 digitised pages of the Old Bailey Proceedings Online.

Designed and conceptualized properly, e-research holds out great promise for the humanities. ‘Big humanities’ may actually be feasible after all!