SCIENCENODE™

Home   Archive   Contribute   Sponsor

About   Give Now

# Big data in the humanities and social sciences

Peter Doorn, director at Data Archiving and Networked Services in the Netherlands, tells iSGTW how big data is revolutionizing research in the humanities and the social sciences.

iSGTW *speaks to Peter Doorn, director at* Data Archiving and Networked Services *in the Netherlands. Through its electronic archiving system, the institute provides sustained access to digital research data. It makes thousand of datasets available, primarily from social sciences and humanities research.*

Why is big data an important issue for the humanities and social sciences?

The analysis of big volumes of data opens up new avenues of research and makes it possible to answer questions that were previously

Posted on FEB 5 2014 7:30AM
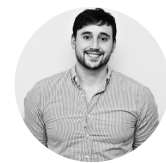
Andrew Purcell
European editor

Share this story

Tags

arts

big data

data

digitization

Humanities

"The data itself does not define what you can do with it," says Doorn. Image courtesy Peter Doorn.

Front page image courtesy Jared Zammit, Flickr (CC BY-SA 2.0). Image has been cropped and rotated.

unanswerable. In social science research, there is a great tradition of survey methodology with people doing interviews about all kinds of ideas people may have. However, a new approach is to do things like a sentiment analysis on Twitter posts, for example. This is a totally new way of getting knowledge about what is going on in society.

I've heard you talk before about data being 'born digital', versus data that was originally analogue but which has subsequently been

digitized. Could you perhaps explain more about this idea?

If we look where the big data in the humanities and social sciences comes from, there are two broad categories. First of all, there is that which is 'born digital' and, secondly, that which is created through mass digitization. These can, of course, be subdivided into a number of categories. In the 'born digital' category, for example, I have already mentioned data that is produced through social media, but there's also data from administrative processes, financial transactions, brokers, *etc*. A good example of a mass digitization project is the online database Delpher. Delpher allows millions of digitized pages from Dutch books, newspapers and magazines to be searched through a central location.

Aren't you blurring the lines between humanities and business analytics a little though?

You could say that the data itself is neutral; the data itself does not define what you can do with it. It's what you do with it that makes it data that is relevant for social science research or for humanities research.

Even data from the natural sciences can also be used in different ways. If we look at the data produced by, say, particle accelerators or telescopes, it can still be used in a variety of different ways by different specialists, depending on what kind of questions they want to pose.

So, just how big exactly is 'big data' in the humanities?

It's important to note that big data itself is relative. What we consider big now is not the same as what we considered big five years ago or what we will consider big five years from now in the future. I really like what Sayeed Choudhury of John Hopkins University, Maryland, US, has to say about this: you should not just look at volume of data, you should also look at methods. Big data is when your method breaks down, when you need a completely new method to analyze the data that you have available.

Indeed, the volumes we are dealing with in the humanities and social sciences are generally an order less than in the natural sciences, but the analysis of multiple terabytes of data still creates these methodological problems whereby you simply cannot use your traditional methodology anymore. You may, for instance, need to parallelize your analysis and run it on a grid instead of on your laptop or personal computer.

What about the complexity of the data?

In many cases, researchers in the natural sciences are dealing with data coming from a particular measuring device or piece of equipment. This means that there are massive streams of data, but that it's all roughly the same. By contrast, in the humanities the problem is more often that we are dealing with data from multiple sources, which are all different. The challenge of harmonising and

comparing that data is certainly an area where big data techniques can help.

One piece of research that comes to my mind is an attempt by a group of economic historians to bring together data on economic growth and distribution of wealth on a global scale for the past 200 years. That's literally thousands of data sets. But let's say from India they perhaps had data on the price of bread for a decade in the 18th century and from Egypt they had grain prices from another period. They had to find ways to recalculate all of this into some new standard measurement that makes it possible to compare these countries economically over time and that's where the challenge lies.

So the challenge for the humanities and social sciences is really less about 'big data' and more about 'complex data', right?

Yes, but in many definitions complexity is taken as one of the characteristics of big data. But perhaps I'd be willing to say that... in the majority of cases it's true.

Do you feel that research in the humanities and social sciences is taking full advantage of the opportunities that big data currently presents?

No, I think that it is so far still only a very small group that is intrigued by these new possibilities, as well as the new challenges. The majority, however, are not. We can only speculate why. Perhaps it's because their research questions are more

traditional ones that they can solve with just a small data set and their own laptop. Personally, I think there needs to be more demonstrator projects, which can serve as examples to the rest of the community of what can be done with big data. The more projects that are carried out, the more others will see the enormous advances that are being made.

Also, there's currently a lack of training and skills in using these tools and methods, which is holding researchers back. It is true that a small number of training courses already exist in this area, but it is still very, very limited. There needs to be more done to train students in the humanities and social sciences to help them take full advantage of the exciting new opportunities big data presents.

Join the conversation

Contribute ☰

Do you have story ideas or something to contribute? **Let us know!**

OUR UNDERWRITERS

Thank to you our underwriters, who have supported us since the transition from International Science Grid This Week (iSGTW) into Science Node in 2015. We are incredibly grateful.

View all underwriters

CATEGORIES

Advanced computing

Research networks

Big data

Tech trends

Community building

CONTACT

Science Node
Email:
editors@sciencenode.o
Website:
sciencenode.org

IU GIVE NOW

Copyright © 2022 Science Node ™ | **Privacy Notice** | **Sitemap**

Disclaimer: While Science Node ™ does its best to provide complete and up-to-date information, it does not warrant that the information is error-free and disclaims all liability with respect to results from the use of the information.