The University Of Sheffield.

Research & Innovation Services.

Doctoral Times.

Issue 12 | Summer 2016

Data Management.

The Newsletter for Doctoral Researchers

Introduction

From Dr Andrew Furley

There is increasing public interest in the probity of research across the globe. One of the major concerns is the gathering and analysis of data that researchers use to advance knowledge in their discipline. In a recent article in The New Scientist, entitled The Unscientific Method, it was claimed that 'dubious results are alarmingly common in many fields of science.'

As professional researchers it is incumbent upon us to ensure not only that our research data is accurate and our conclusions well founded but also that this can be seen to be the case. This involves not only expertise in the gathering and analysis of the data but also transparency in the use of the data in order that it can stand up to the scrutiny of others in our field.

At the University of Sheffield our Doctoral Researchers are fortunate enough to have access to a wide variety of training and support to enable them to practice their research in an ethical and open manner. This approach is essential if the research that we do is to have the impact that it should.

This edition of the Doctoral Times explores the latest issues in data, with contributions from current students, academics, alumni and the professional services that support our efforts to ensure that the research we do makes a difference.

Dr Andrew Furley is a Senior Lecturer in the Department of Biomedical Science and the Faculty Head of Research Training.

In this Issue:

Editorial Team

Gavin Boyce, Katie Cahill, Fozia Yasmin

Issue 11 Crossword Solution

S	T	A	T	E	T	H	E	O	B	V	I	O	U	S	■
E	■	D	■	X	■	■	■	B	■	A	■	C	■	■	H
N	A	E	V	O	S	E	■	L	I	G	A	T	U	R	E
D	■	M	■	R	■	■	■	O	■	R	■	Y	■	■	R
I	M	P	E	D	E	D	■	Q	U	A	A	L	U	D	E
N	■	T	■	I	■	R	■	U	■	N	■	■	V	■	D
T	■	■	C	A	T	A	L	Y	S	T	■	P	U	L	I
H	■	L	■	L	■	G	■	■	I	■	■	■	L	■	T
E	■	A	■	■	■	O	■	■	D	■	T	■	A	■	A
C	A	P	O	■	E	N	G	U	L	F	E	D	■	■	R
L	■	S	■	■	L	■	I	■	E	■	N	■	A	■	Y
O	B	E	I	S	A	N	T	■	D	A	N	G	L	E	R
W	■	■	M	■	T	■	A	■	■	■	Y	■	K	■	I
N	E	G	A	T	I	O	N	■	M	U	S	T	A	N	G
S	■	■	G	■	O	■	E	■	■	■	O	■	L	■	H
■	A	G	E	I	N	G	S	C	I	E	N	T	I	S	T

Setter: Prolix

Integrating Quantitative and Qualitative Data for Improved Food Supply Chain Analysis

Patience Muchada

Researching Food Safety and Standards: The Case of Campylobacter in Chicken Meat

When most people hear the word 'data', numbers and statistics spring to mind. And rightly so, because they are indeed some forms of data – but certainly not the only kind.

Through my project ('Analysing food security and food safety in food supply chains'), I find myself working with multiple forms of data as I try to follow the journey of food products, from farm to fork. For example, in a case study on the bacteria Campylobacter in the broiler meat supply chain, I have been using laboratory reports and data from microbiologists to understand the characteristics of the bacteria, as well as testimony from chicken growers and veterinary specialists to learn how the birds are raised.

I also study the processing, retail, consumption and post-consumption stages of the chicken supply chain. This means analysing production line manuals, consumer survey data, interview transcripts, videos and photographs, social media data (Twitter, YouTube), and even epidemiological reports as the bacteria can cause illness in humans. All these forms of data are obtained through different means, in primary field work or as secondary data. They also come in various forms (numbers, text or graphics) as well as at different scales (for example, individual, regional or national).

Additionally, the different perspectives of all the 'actors' in the supply chain are also built into the data. For example, as Campylobacter can be found in chicken at all stages along the supply chain, an interview question on who should be responsible for reducing or preventing it spreading to humans would likely be answered differently by a farmer, a meat processor, or a consumer. Similarly, a molecular biologist, a medical doctor, and a veterinarian would have varying opinions, and evidence, on the need for, and efficacy of, using antimicrobials in animal farming.

There are also ongoing debates in academic circles about the weight, comprehensiveness, appropriateness or validity of certain qualitative or quantitative data forms. The challenge is therefore to find ways to combine these different data forms in a way that is academically robust and with integrity, to answer my research questions. Hopefully, at the end of my PhD, I will be able to write another article on how I have achieved this!

This knowledge can help to encourage, and improve interdisciplinary research, resulting in more effective research with implementable results. It is this kind of interdisciplinary research that the Grantham Centre for Sustainable Futures, which this project is part of, encourages to help inform ways of using and managing natural resources more sustainably.

Getting to Grips With Publication Metrics

The use of metrics in academia is currently a hot topic, stimulating debate and intrigue in equal measure. Metrics can cover many aspects of academic life, including research grant income, student recruitment and satisfaction, PhD submission rates and the impact of research. In this article we will focus on metrics for your research outputs - the journal articles, books, chapters, conference proceedings, datasets, compositions, designs and all other outputs of your research.

Why are metrics so topical at the moment? In the current climate of restricted government budgets and spending cuts, there are huge pressures on the research funders (such as HEFCE and the Research Councils) to demonstrate a return on their substantial investment in UK research. In turn, universities want to better understand the output of their research, to enable them to manage and develop their strategies to help them win future funding. Some organisations are also increasingly using metrics to determine the employability of job candidates or to assess funding applications. There is also plenty of ongoing discussion as to whether future REF exercises (the tool used by the Government to assess UK research) should make more use of metrics in the future. If you're interested in gaining a more thorough understanding of the metrics debate than we can offer in this short article, then our very own Professor James Wilsdon has recently led an independent review into the role of metrics in the sector (The Metric Tide) - it's definitely worth a read.

Key concepts

Advances in technology have made available better, more comprehensive, and robust metrics about outputs. Yet across the research community there are plenty of misgivings about them, and worryingly a lot of misunderstandings and misinterpretation. As you begin a career in research, it is important that you understand the basics: what different metrics there are, their opportunities and pitfalls, where to access them, and how to use them.

We should start by stating that most metrics are geared around publishing practices in journals and conference proceedings. Metrics for other forms of publication, such as books, are much less developed, although altmetrics are now beginning to make inroads into this area.

H-index - are you more than just a number?

The h-index, the best-known of an ever-growing multitude of metrics designed to assess individual researchers, aims to measure both the productivity and citation impact of an individual. Many academics, particularly in STEM disciplines, will keep an eye on their h-index as they progress through their career, perhaps even including them on CVs, personal web pages and funding applications. An h-index of 7 indicates an individual has 7 publications with 7 or more citations.

Although widely-accepted by academics, an individual's h-index is influenced by many factors, including discipline, length of academic career and career breaks and the citations data source used to calculate it. This means that the h-index can only fairly be used to compare individuals in the same discipline at the same career stage.

Article-level metrics - good old fashioned citations!

The number of citations an article receives has been used since the 1960s as a proxy for research quality, with the underlying assumption that being cited is a positive thing (although of course, one can be cited by articles pointing out one's flaws). Over the years, a range of more sophisticated and diverse measures have been introduced, including benchmarking of citations counts with similar articles, geographical range of authorship and citations and the number of times an article has been downloaded and read from digital repositories. Looking at a range of these metrics can inform and educate researchers about how their publications are being used and received, but they should always be compared with the norms for particular disciplines, which can vary significantly.

Journal-level metrics - are we judging a book by its cover?

Metrics about journals are available in most publications databases, but these are sometimes inappropriately used to assess individual researchers and publications. The Journal Impact Factor, the best-known of these metrics, is based on the average (mean) citations per article in a journal over the previous two years, and is easily skewed by a few highly-cited articles. Because the number of citations that each article in a typical journal receives varies wildly, the use of journal-level metrics to assess an individual paper is controversial and explicitly discouraged by many bibliometricians and The Metric Tide report. Given the abundance of article- and individual-level metrics available to us, there is also no need to rely on such crude measures.

Altmetrics - who's saying what about your outputs?

Altmetrics measure online attention to research in sources such as news, blogs, policy documents and social media. They are complementary to citations and tell another part of the story, helping provide a more coherent, well-rounded understanding of the way in which people are engaging with your research. Altmetrics enable you to see exactly who is talking about your research and what they're saying. You can start to track this information as soon as your research is published - no waiting around for citations. You also have the opportunity to engage directly with those interested in your work.

As we have seen, many of these systems rely on being able to uniquely identify you. Your name is central to your professional identity, but is it actually enough to reliably distinguish you and ensure that you get the credit for your work? Any Consuela Bananahammock's out there may be OK, but if you're a Jones, Smith, Kim, or Wang you could find dozens of namesakes even within your own discipline. Just because you don't have a namesake at the moment doesn't rule out one appearing in the future.

Distinguish yourself with ORCID

ORCID is a new initiative to help researchers establish a unique professional identity. After signing up for an ORCID ID, you provide this to publishers and funders with your manuscripts and grant applications - enabling them to uniquely identify you. An accompanying ORCID profile enables anyone clicking on your ID to see details of your papers, grants, career history and any other information that you choose to make available. What's more, it enables systems to talk to each other with absolute certainty that you are not being confused with someone else - providing more accurate and robust metrics.

The University encourages all researchers to sign up for a free ORCID ID - it takes just a few minutes - and to begin using it in their academic life. You own your ORCID ID for life, taking it with you whether you go. The University has no control or access to the information in your profile. Find out more at: www.sheffield.ac.uk/ris/systems/orcid

Tools to help you

myPublications is the University's publication management software. It enables you to record details of your outputs, track citations and altmetrics, and even upload a copy to our open access repository. In most cases, myPublications will even find your outputs for you using online sources. All postgraduate research students have access. Find out more at: www.sheffield.ac.uk/ris/systems/mypublications

Google Scholar, Scopus, Web of Science and PubMed are online databases of research outputs published in a vast array of journals, conference proceedings, and increasingly books. You can use these databases to easily find details of published outputs, as well as citations and a range of other metrics. You should be aware that the citations only relate to other papers within that database, and should not be considered comprehensive.

google.com/scholar - scopus.com - webofscience.com - pubmed.com

Altmetric.com - the University subscribes to www.Altmetric.com which each day tracks 44,000 mentions of research. You can easily see who talking about your research and what was said. You can view Altmetric data directly within myPublications. Find out more at: www.sheffield.ac.uk/ris/systems/altmetric

Dave Jones and Jon Griffiths

part of the Research Information Team in R&IS

Managing and Sharing Research Data

Jez Cope

Research Data Manager, The University Library

Until the invention of GPS, ocean-going ships navigated by dead-reckoning: estimating how fast they travelled, for how long and in what direction and plotting that information on a chart. To make this possible, shipmasters would keep detailed logbooks recording the ship's position and heading, often several times a day. This often included other useful information, especially about the weather. To the shipmaster, this information served only one purpose: to tell them where the ship was and which way to steer. Once the ship arrived it was useless.

This changed in the 1840s when Matthew Fontaine Maury, a US Navy officer, broke his leg in a stagecoach accident and turned his attentions to the study of naval meteorology. By instituting a standardised reporting system amongst the nation's shipmasters and collating this information on a large scale, he was able to put together the first detailed charts of the oceanic winds and currents. In 1848, this information enabled Captain Jackson of the ship W. H. D. C. Wright to cut a full 17 days of the trip from Baltimore to Rio de Janeiro, and led to a revolution in naval navigation.

Many of these logbooks have since been preserved and are now being digitised and transcribed by volunteers through projects such as Old Weather. The resulting rich data is being put to further uses never anticipated by those early navigators, especially to help validate climate models and teach us more about how the world's weather changes over long periods.

This is just one example, and we now live in information-rich times. Thanks to technological advances, we're surrounded by more of the stuff than we know how to handle and the problem is getting worse all the time.

So why am I happy about this? Because information is like oxygen for research. Most of us don't give it much thought, but we couldn't do our work without it. The purpose of research is to turn information into new knowledge, and it's never been a better time to be a researcher.

New tools and techniques, from high-definition video to optical character recognition to DNA microarrays, make it possible to gather and analyse huge amounts of data; often we end up discarding information that is incidental to our own research but could be much more valuable to other researchers. At the same time, improvements in storage and networking mean that it's much easier to share these large quantities of data. These two factors between them mean that we're currently undergoing a culture shift towards data being seen as a primary output of research in its own right.

What is/are research data?

So what is "data"? Different disciplines have varying definitions of data, but broadly speaking the term "research data" refers to:

material collected, created or analysed to support research conclusions

That's very broad, so you can imagine that data means quite different things to different people. It also includes some things that might also be described as primary and secondary sources or various other terms.

In science and engineering, the concept of data is quite familiar. Data may be observations and measurements of phenomena made under carefully-controlled conditions or in the world around us; they may be generated by mathematical models; or they may be derived by combining and transforming existing data.

In the arts and humanities, research data are harder to define and require different treatment, often comprising carefully-curated collections of materials gathered over long periods of time.

Data in the social sciences cover a wide range, including both quantitative data such as census results and qualitative data such as interviews and diaries.

Why is it important now?

Open data contributes to society

Open data is now enabling exciting new developments across all areas of research. In 2011, an outbreak of food-poisoning in Germany was traced back to a batch of cucumbers originating in Spain, ultimately spreading to other European countries and the US. Three days later, a draft genome of the strain of E. coli responsible was made available under an open data license and, thanks to international cooperation from bioinformaticians, the full genome was assembled within 24 hours after that.

One area where you can see the impact of data is through the work of data journalists. Sites like the Guardian Datablog build compelling, accessible stories around all sorts of diverse datasets. See, for example, this recent story about feminism in the US electoral campaign.

Our own Humanities Research Institute has been heavily involved in the publication of a range of influential qualitative datasets. One such is Old Bailey Online, which makes the proceedings of London's central criminal court from 1674 to 1913 freely available in digital form.

Open data boosts your career

Sharing your data isn't just about what other people can do with it though. It's a great way to increase the visibility of your research, and to build your profile as a researcher. Research work is increasingly evaluated not just on its validity but its impact on society, and having your data reused by journalists, commercial companies or other researchers is a compelling way of demonstrating impact.

Datasets can and should be referenced and cited in a similar way to other research publications, such as articles and monographs. That means that you can gather citations for your datasets themselves as well as your own publications. In addition, there is growing evidence that publications whose underlying data is available receive more citations themselves.

Making your data available is also a great way of attracting collaborators. It's much easier to see and understand how another's research fits together and relates to your own if you can examine their data.

Finally, data sharing is a great way of bringing more transparency to the research process. Results whose underlying evidence is open to scrutiny are inherently more trustworthy, because the analysis can easily be verified.

So what do I need to do?

Research data management, then, is what makes all of this work. It refers to the practices we use to look after research data: a) while actively collecting and analysing the data; and b) when archiving them for long-term preservation. This includes:

1. Documenting the data so they can be understood, and to ensure they are valid, accurate and complete
2. Organising them so that they can be found quickly and easily when needed
3. Protecting them appropriately so that they are not accessed by unauthorised people, lost or corrupted
4. Selecting and preserving a subset for the long term
5. Optionally sharing these preserved data, either publicly or under certain restrictions, after a period of exclusive use

University Data Catalogue and Repository

The Library, in collaboration with CICS and R&IS, are developing a data sharing platform for the university, powered by figshare. This will allow you to:

1. Openly share moderately-sized datasets on the web
2. Advertise restricted-access datasets (such as those which contain confidential information or are simply very large)
3. Obtain a Digital Object Identifier for your data so that it can be consistently cited by other researchers

It's currently in the pilot phase but will be rolling out more widely soon, so keep an eye on your inbox for more information.

ORDA powered by figshare

Top tips

- Decide now how you're going to name and organise your data files and folders, and describe that structure in a README file that you keep with the data at all times
- As you collect data, record key information about the context (who collected it, when and why), content (what was measured, what do variables/codes/abbreviations mean, what units were used) and structure in your README file
- Make sure your data is backed up regularly: the easiest way to do this is to use CICS filestore
- If you need help protecting sensitive data, consider using encryption
- Look for datasets that you could use to augment your own data: re3data.org is a good place to start

If you want to find out more about managing and sharing your research data, there are several things you can do:

- Visit the Library web pages on research data management
- Sign up for one of our research data management workshops, available as part of the DDP
- Contact the team at **rdm@sheffield.ac.uk** with any queries

The Open Data Science Initiative

Professor Neil Lawrence,

Computer Science

Data is now at the core of everything we do in science, social science and engineering. Although in fact, data always has been. Indeed, when scientific endeavour was first characterised it was known as 'the empirical method'. Learning by direct or indirect observation. So clearly data is at the core. So what has changed?

Data used to be expensive and time consuming to collect, so much of science, and in particular statistics, was focussed on collecting the right thing: experimental design. Early scientists found that humans could also be mislead by data. Seeing patterns that were not there, so statistics also specialised in confirmatory data analysis. So what's different for data science?

Data science covers the interface between statistics and computer science that has been trying to make sense of the new challenges in data analysis. Large unstructured datasets, that are often collected by happenstance rather than by design. Reduced cost of sensors and wide availability of computing resource has led to acquisition of data on an unprecedented scale. This is presenting new research challenges in data assimilation and interpretation. However, the implications and impact of data science go far beyond the research challenges and spread through all departments. As well as the core methodological ideas there are the implications of data in society, visualisation of data, the evolving need for regulation, issues of loss of privacy and implications for academic disciplines that can move from qualitative analysis to quantitative analysis through the increasing availability of data.

The motivation for the Open Data Science Initiative is to increase access to data processing technologies. For our scientific colleagues, our industrial collaborators and colleagues in health and the developing world. Our activities include summer schools, not only in Sheffield but as far away as Colombia, Uganda, Kenya and Australia. We have been organising monthly meet ups to discuss data. Not just the challenges of its analysis but the issues as varied as ethics and data storage. To support deployment of these ideas we've instigated the foundation of Sheffield's first ever Research Software Engineering Group. We've also supported the delivery of a course in programming and data analysis to undergraduates in Biomedical Science.

This work is then underpinned by international quality research in machine learning and statistical analysis techniques. This is to ensure that the next generation of scientific challenges can be addressed armed with the right set of analysis tools.

The Open Data Science Initiative is now at the core of an international movement to increase access to the tools we need to deliver modern research agendas. So how can you help? With our ambitious agenda there is no way we can deliver everything on our own.

We need everyone to pitch in and help out. Have a look at some of our tools and ideas and see if there is anything you can integrate into your teaching, or do you have ideas to feed back to us on how we can help?

To find out more about us go to our webpage: http://opendsi.cc or have a look at the position paper behind our formation: http://inverseprobability.com/2014/07/01/open-data-science/

CiCS Support

Dr Mike Griffiths

CiCS

Data is ubiquitous and touches our lives, often without our awareness. Nearly every researcher has data, whether it is the writings of an important philosopher, videos of animal behaviour, imagery from astronomical devices or the mountains of data generated by large collaborative experiments (e.g. the LIGO detector used to make the recent discovery of gravitational waves or the Large Hadron Collider at CERN used to discover the Higgs Boson).

Researchers at The University of Sheffield are provided a range of support enabling them to access, create and analyse data sets. To do this researchers need to be able to store, backup, share data and publish the output of that research. Guidance on the support provided by the Library and CICs. (see http://www.sheffield.ac.uk/library/rdm/storage)

Most researchers use computers, usually personal computers or Mac's these can be used for a wide variety of data analyses. CiCS provides access to a variety of software used for different sorts of data analyses. Software and Licenses for packages, such as Matlab, SPSS and NVivo can be obtained from CiCS. Many packages may be downloaded from the software download centre (see http://www.sheffield.ac.uk/cics/software/available). Researchers are free to use other packages including free ones such as R. CICS has a small team of staff that are providing specialist support to researchers using these packages through training and advice including individual consultation if necessary (http://www.shef.ac.uk/cics/research/training). In addition to training workshops research events are hosted such as the HPC@sheffield event showcasing research computing across The University, there are plans to broaden the scope HPC@sheffield to support research on more platforms on more scales and for all faculties. The analysis of research data is not the final part of the data lifecycle, the University of Sheffield provides Guidance on publishing and sharing research outputs (see http://www.sheffield.ac.uk/library/rdm/publish).

In The Constant Persuit of Data, No Experiment is Perfect

Liam Marshall

Researching peptides in the Department of Chemistry

Data is not a sexy subject. It isn't as immediately exciting as trying something new and having all your problems be solved. It is mostly rows and rows of numbers, which when looked at as a solid unfiltered mass are boring and uninspiring.

Both its presence and its absence has such a profound effect on PhD students. Everyone knows the feeling of reviewing the work done so far, and thinking of all the exciting experiments that can be done. My PhD is going to change the world if I can just do this one thing! Everyone is equally familiar with the other side, where the data isn't showing what you want it to show. You are doing all the right things, but what is the crucial factor for the experiments that makes them work better? How can you change what you are doing so it is giving you the results that will have such an effect on your field?

At this point, you should stop. Consider what you are thinking. If you are confident in what you are doing (which you should be. You are one of a limited number of people doing their PhD at a world class university. You've got this.) then the data must be coming through right. Once you've checked the equipment, and your method, there must be something else at play. Either you have introduced a factor that you haven't considered, or the experiment is doing precisely what it is supposed to be. The most important thing to remember is to keep getting that data, and making a note exactly of what you are doing. There is no correct or incorrect results as long as the experiment is done correctly. There are only the results that those experiments will give you!

Maybe it can be accounted for by experimental error! The equipment was misaligned or just playing up. Maybe there is a factor beyond your control you can't affect. That doesn't matter. Just keep getting the data. You can write a thesis after accounting for whatever has given you the unexpected result. You can't write a thesis if you haven't got any data after doing the experiment once and then thinking about how everything will be better when it gives you what you expect.

All you see of everyone else's data is what they publish, which has a clear and definite story to tell. That is the data that has been selected for that paper, and all the things you are going through are the same as they did. Keep your head down, keep getting that data, and enjoy being surprised when you look back and have your own story to tell.

Challenges in Managing and Sharing Qualitative Data

Daniel Turner

Alumni, Department of Geography

In my doctoral studies at Sheffield, I came across a fairly typical problem with data management: how do I store and analyse the large amount of qualitative data that I was generating? I was collecting data from focus groups and interviews in Botswana, where I was looking at the experiences of People Living with HIV/AIDS using peer support groups. Quickly I had hundreds of pages of transcripts, and dozens of themes across them I wished to extract.

Even in my undergraduate days in Geography at the University of Sheffield I was lucky to get training in using data management and analysis software like ArcGIS, SPSS and Nvivo: advanced tools that many students don't come across until their doctoral studies. However, at the time I was a Mac user, and none of these packages were then available for my beat-up and well travelled iBook. So I set about developing my own solution with custom spreadsheet and search software that would allow me to code and extract themes across my whole dataset.

After completing my thesis, I had a number of post-doctoral posts on different health issues, all using qualitative methodologies in some way. Every project had it's own data management challenges, whether it was triangulating mixed methods data, or making sure that team members in different collaborating organisations had access to the latest data. But each time the team found that the existing software for managing and sorting qualitative data was lacking in something we wanted to do. I'd always assumed that someone else would sort out these problems, and create an easy to use software package that would work across different platforms, but this never materialised.

So, for the last 3 years I have designed and developed a new tool for analysing and managing qualitative data, called 'Quirkos' (www.quirkos.com). Taking up software development has been a steep learning curve, but that's just what I loved about every new research project I started. The software is already used by researchers in more than 40 universities across the world, and is a career direction I could never have anticipated!

I have no doubts that the early start that I got in using data management software, first as an undergraduate, but expanded and reinforced in the doctoral Research Training Programme at Sheffield, gave me the skills to manage large amounts of data in many projects as a research fellow. But it also gave me the flexibility to adapt for different research needs and challenge the norm, eventually leading to a new way for the next generation of researchers to engage with their qualitative data.

Managing Your Research Data

Hafez Abdo

Alumni, The Management School

I obtained my PhD from the Management School of the University of Sheffield in 2007. My PhD topic is "Exploring the Rationales for Relaxations in the UK Petroleum Fiscal Regime 1980-2000". In order to conduct my PhD I needed data worth £10,000 at the time I was undertaking my study. Luckily, I was given access to this data by Wood Mackenzie and I must thank them for allowing me that access for free. Beside the appreciation of the monetary value of my data I had to treat the data with much care and a high level of responsibility, and here I am sharing with you some useful advice in the same context.

Data is an important key to answer your research questions, sort out your research problem and fill in your defined research gap. It is a significant contributor to your 'Contribution to Knowledge'. Managing your research data is an important issue to you as a researcher. This is because you need to adhere to your university's ethical code of conduct; protect the rights and dignity of both your university and your data providers (sources). As your university is keen to protect the privacy and rights of the community you have a duty to protect the University interest and implement its research policies to high standards.

Therefore, it is key that you make yourself familiar with the university ethical code of conduct as early as possible and conduct your research in an ethical manner. Particularly because you are conducting your research study in such a reputable university like the University of Sheffield.

In managing your data you need to document how you intend to collect, store and process your data. Decide in advance how you will destroy your data after you complete your research. And more importantly, what you've planned to do with your data in terms of collection, storing, processing and destroying.

Make sure you obtain an informed consent from your research participants to assert their voluntarily participation in your research and to protect yourself in case something goes wrong – keep the signed informed consents in a safe place.

It is key that if you decide to keep any data after the end of your research project to keep these in a fully anonymised form, so there would not be a possibility to trace back the participants and/or their organisations. At the same time, this anonymised data can be used by you or by other members of your organisation as secondary data.

All the very best wishes to you in your PhD journey

Data Management:

From Running Research Tests to Operating Power Plants

Lyubka Spasova

Alumni, Department of Materials Science and Engineering

Handling data and making contributions to science and engineering have been an essential part of my education and professional career. I completed a Doctorate degree at the Materials Engineering Department in 2008. My research studies were on exploring an interesting, but not so well understood application of non-destructive monitoring to determine the 'health' (or structural changes) of materials. Acquiring, analysing and storing large data sets were required to identity patterns and derive viable conclusions. The effectiveness in finding patterns within tens of thousands of ultrasonic digitised signals was challenging, but yet achievable via different methods of data processing and logical derivation of documented ultrasonic event - material's 'health' relation.

Building upon my experience as a research student I pursued a career in industry. For a number of years I acquired, processed and validated large volumes of data related to industrial processes to develop algorithms for emerging products. The main requirement for these products was to increase plant reliability and reduce energy waste. To present strong evidence and secure Intellectual Properties (IPs) I used systematic data filing and clear simple presentation of long term research and development discoveries. The real challenge was how to communicate and organise the data so it becomes accessible and usable by others. From my experience one of the best ways is through direct involvement, close collaboration and supervision of other engineers.

The data is only useful when it is accessible and relevant. Existing software tools and records (electronic or on paper) are not always sufficient and often too complicate to decipher. In recent years I have been working in the UK nuclear power industry. The volume, complexity and detail of accumulated engineering and nuclear safety knowledge are immense. This knowledge is captured and managed by a centralised document management system ticking like a clock to enable continued stations operation and keep our lights on. In addition, the protection of this data by following written procedures is of paramount importance when nuclear safety is overriding priority.

The often found issues with data are tractability and clarity of description. These apply equally to data produced by a single person for a specific task or by thousands of people working on complex systems and plants. The lessons learnt by me are that any data, for scientific or commercial driven purposes, should be organised and managed by different levels of access. The effectiveness in accessing and contributing to the data management process would provide the needed certainty for success.

Statistics Training

The Doctoral Development Programme (DDP) offers a range of statistical training in the form of modular and short courses along with online materials. Some of the statistical training modules are subject specific, whilst others are open to all. Full details of all of the statistical training avaiable can be found at http://www.sheffield.ac.uk/ris/pgr/ddpportal/reg

Examples of modular training

Advanced Statistics for biologists

This module provides advanced training in the use of statistical methods and computers to present and analyse biological data. The course will be based on the statistical program R, using a series of workshops and student-centred learning assignments to develop skills and proficiency.

Further Statistics for Health Science Researchers

The unit covers fundamental statistical concepts, and both simple statistical methods and the more widely used advanced methods of multiple regression, survival analysis and generalised linear models. It will be a practical module, including the teaching of the statistical software SPSS, equipping students with the knowledge and skills necessary to design and analyse a study to answer specific research questions; to understand and critically appraise the literature; and to present research findings in a suitable fashion.

Examples of short courses

First Steps with NVivo

This course is designed for users to be able to create an NVivo document, introduce their sources to NVivo and start annotating and coding their sources.

Simple Graphs and Statistics with SPSS

For users to be able to explore their data graphically, carry out simple descriptive analyses and bi-variate analyses.

Looking After Your Research Data

This workshop will help you to get the most out of your data by keeping it safe and secure, organising and describe it, archiving it and sharing it, and by augmenting it with pre-existing data from a variety of sources.

Examples of online courses and other resources

Introduction to Statistics and Critical Appraisal (Distance Learning)

The unit, which is delivered online, introduces students to basic concepts and techniques such as hypothesis testing and confidence interval estimation in statistics. Students will learn some simple statistical methods and the principles behind some advanced methods such as regression. It will equip students with the knowledge and skills necessary to understand and critically appraise statistics in research literature.

MASH - Maths and Stats Help

MASH offers a range of resources and support for students looking to improve their maths and statistics skills. Resources on offer include: one-to-one support from tutors; drop-in facility; study room with resources and tutors to help if necessary; online material for self-study, targeted provision and classes in key topics.

Cryptic Crossword

By Abacist

Correct entries will be named in next issue.

First past the post will not only have the glory of their name in print they will have the option to compile a future crossword should they wish to do so!

**Please complete if you are submitting your entry*

Name: .....

Department: .....

Return entries to GRC, Dainton Building, Brook Hill

Clues

Across

1. Dave and Tara are agitated around me a bit as any couples seeking relationship analysis (9,4)
7. Advanced Technology initially in California now down in Greece. (4)
8. Point at which enough data has been gathered? (6)
9. Understudy back from representative group for whole population. (3)
10. Lower limit from Finland will satisfy hunger. (4)
11. Uncontrolled fire consumes fifty after Science and Technology leaders create super suppressor. (7)
15. Alpha takes genetically engineered value for considerable period. (3)
17. Erudite editor follows student, both keeping near about. (6)
18. Business takes on minister, student and divorcee in arrangement that is involved. (7)
20. Cease to function when papers are taken back by head engineer. (3)
23. Lost in wild mob I take lad back to get 26 down twice in 8 across. (7)
25. Opening post in leading Higher Education threads set aspiration. (4)
27. Speculation for principals in Bioengineering team. (3)
28. Opt out in the European monetary union is about self-interest. (2,2,2)
29. Central line for taking six back to Acton? (4)
30. Aberdeen University gets seduced about academic's wild stab in the dark. (8,5)

Down

1. Arts graduate takes me along to sing, as a way of influencing my decision. (7)
2. It's possible by virtue of the Bloomsbury Learning Environment. (6)
3. Students change dates in Professor Square's calendar. (7)
4. Research group strives for robot with multiple sclerosis. (4)
5. Brilliant success of researcher in Quebec latency tests. (4)
6. Systematically ordered data bundle arranged by team leader without initial network diagram.... (9)
12. ...which I take before old girlfriend to make a point. (5)
13. Anger about the limits between which variation is possible..... (5)
14. ....leads to strike in Ghana for better representation. (5)
15. Medical officer ducks out of 23 across crazily to make impromptu speech. (2-3)
16. The team edits wrongly and is judged.... (9)
17. ...being easily understood when Lancaster University calls in the detectives. (5)
19. The agitated get gaseous without The University of Sheffield providing an analysis of the origins of F. Scott Fitzgerald's old money. (4,3)
21. The little devil deceives, hiding means. (7)
22. Human resources get involved in movement to distract bullish academic making remark. (6)
24. Dominant researcher rejects the truth? Probably! (5)
26. The Ministry of Education takes 500 for the sake of appearance. (4)

Got the print version?

Want the pdf? Link here