

# Research Data Management Education for Future Curators

Mark Scott, Richard Boardman, Philippa Reed, Dorothy Byatt  
and Simon Cox

## Talk Outline

1. Teaching Research Data Management
2. The “Introducing Research Data”  
Booklet
3. Evaluation of Approach Taken
4. Conclusions

# Teaching Research Data Management

Two parts to our approach:

1. A booklet to give to all first year research students
2. A lecture to introduce basic concepts

# The Booklet Layout

Figure 5: Viewing genetic sequence mutation data (Broad Institute, 2012)

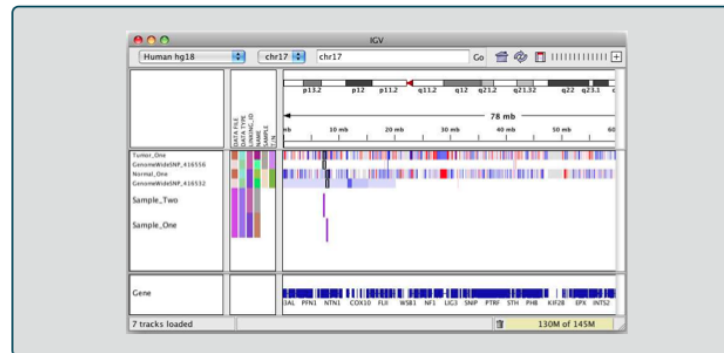
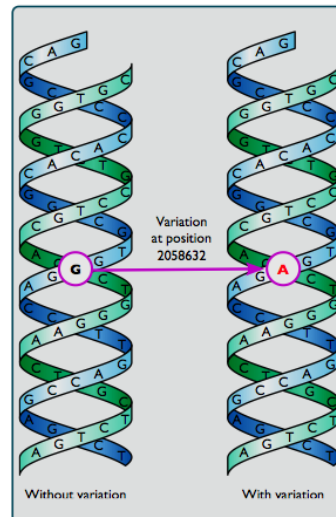


Figure 6: Variation in the *TSC2* gene, identified as one possible cause for *tuberous sclerosis*, a genetic disorder that causes tumors in organs around the body



## Summary

In this case study, the researcher collects DNA sequence data and, after preparing the data by aligning it, processes it to find regions that differ from a reference data set. Analysis of the data can involve comparing it against genome databases and manipulation in spreadsheet software. Publication will occur when patterns that relate to diseases are identified, which will then help improve the existing genome databases and analysis tools. This case study provided good examples of a wide range of research data including:

- Using reference data (the complete genome reference sequence, and the genome databases);
- Discipline specific data (the FASTQ files containing the DNA sequence reads).

# Sections in the Booklet

## Section 1:

“Five Ways to Think About Research Data”

## Section 2:

Case Studies

## Section 3:

Data Management Best Practices

# “Five Ways to Think About Research Data”

## 1. How it is created

*From (Research Information Network, 2008)*

## 2. Forms of research

*Adapted from (The University of Edinburgh, 2011)*

## 3. Electronic representation of research

*Adapted from (The University of Edinburgh, 2011)*

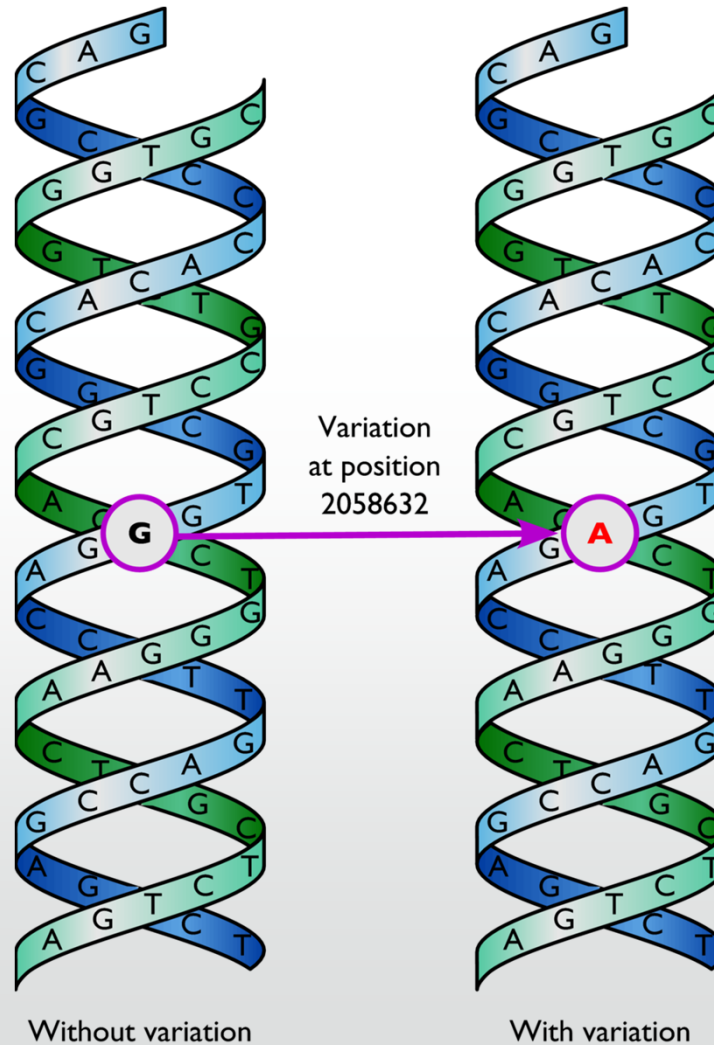
## 4. Size and structure of datasets

## 5. The data life cycle



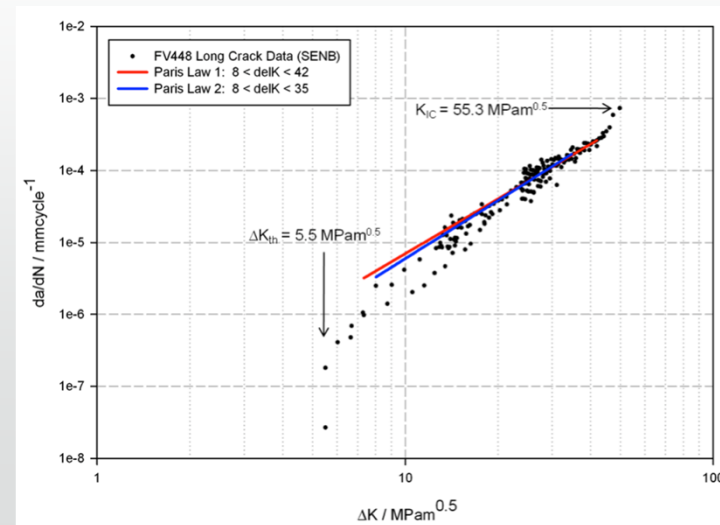
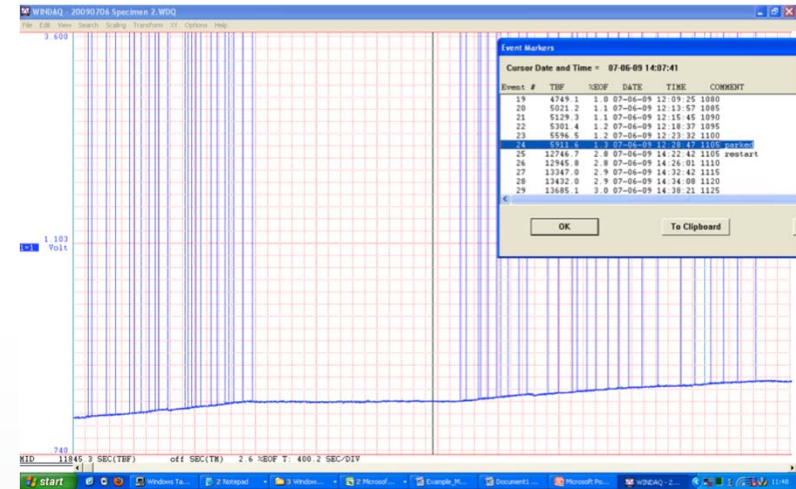
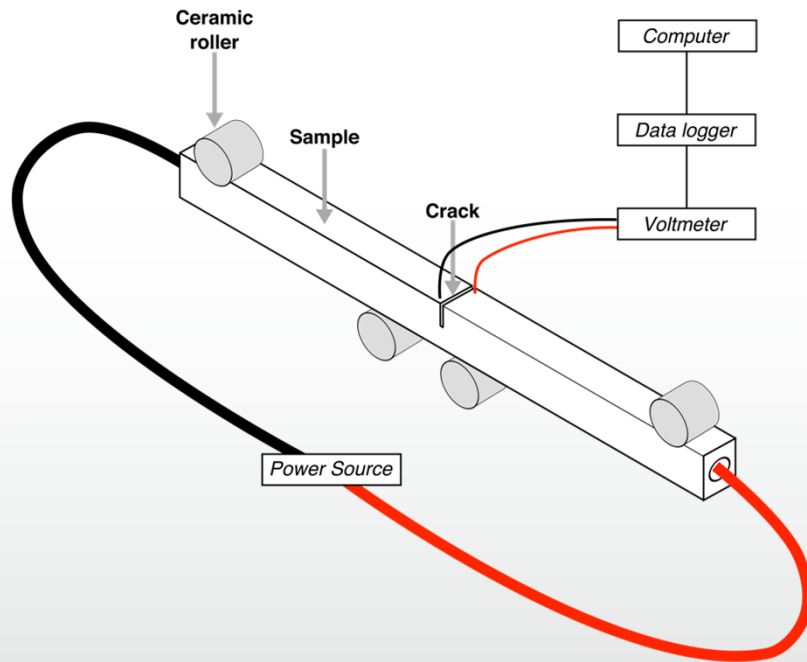
# Case Studies

# Human Genetics Case Study





# Materials Engineering Case Study



# Archaeology Case Study



# Evaluation of Approach Taken

# Evaluation of Approach Taken

## Section 1 (“Five ways to think about research data”)

- Useful as part of the booklet, but should be kept brief in a lecture to maintain the students' interest

## Section 2 (The case studies)

- Provided perspective and encouraged students to think about data in general

## Section 3 (Best practices)

- Most useful reference material and should be expanded in future versions

# Feedback From Lectures

Question	Lecture 1 (80 students at month 7)	Lecture 2 (30 students at month 2)
Would students change what they were doing as a result of the lecture?	60/80	15/30
Should the talk be given to future postgraduates?	78/80	20/30
Did any students feel it was a complete waste of time?	2/80	1/30

- Majority of feedback was positive
- Feedback was more positive when the lecture was given later in the year in month 7

# Conclusions

- The case study approach enables diverse students to engage and relate
- Delivered as part of research student training—and was well received
- Now part of the University's institutional data management strategy and 10-year road map  
*(University of Southampton, 2012)*

*Get people thinking about their data—and its value—EARLY!*

# Acknowledgements

- The categorisation of research data collection was defined in *Research Information Network (2008)*
- The forms of research data and categorisation of electronic storage of research data was adapted from *The University of Edinburgh (2011)*
- The main authors of the booklet and presentation were Mark Scott, Richard Boardman, Philippa Reed and Simon Cox in the Faculty of Engineering and the Environment
- The following people helped with the preparation of this work:
  - Andy Collins (Human Genetics case study)
  - Thomas Mbuya and Kath Soady (Materials Fatigue Test case study)
  - Gregory Jasion (CFD case study)
  - Simon Coles (Chemistry case study)
  - Graeme Earl (Archaeology case study)
- We acknowledge ongoing support from the University of Southampton, The DataPool project, Robert's funding, Microsoft, EPSRC, BBSRC, JISC, AHRC and MRC



# References

- Digital Curation Centre (2010), 'DCC Curation Lifecycle Model'. URL: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- Humphrey, C. (2006), 'e-Science and the Life Cycle of Research'. URL: <http://datalib.library.ualberta.ca/humphrey/lifecycle-science060308.doc>
- Research Information Network (2008), 'Stewardship of digital research data: a framework of principles and guidelines'.
- The University of Edinburgh (2011), 'Defining research data'. URL: <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/research-data-definition>
- University of Southampton (2012) 'Research Data Management web site' URL: <http://www.southampton.ac.uk/library/research/researchdata>