

# Analysing research data using third party software



- Amy has recorded the measurements from her long-running chemistry experiment in an electronic lab notebook, and used the R and MatLab software packages to analyse her results and produce graphs which are included in her published paper.
- Since R and MatLab are both commonly-used software packages, Amy is not required to preserve the software as long as the metadata describing her research data is sufficient, and her paper explains the techniques she used. It may be useful for Amy to deposit the R/MatLab scripts that she used to analyse her results in a repository and link to this in her paper, because this will let others reuse her data and methods more easily and it is not an onerous task to complete.



<http://dx.doi.org/10.6084/m9.figshare.1540765>

Software Sustainability Institute

# Building scripts to support a workflow



- Brian has written a script which converts data from one format to another to allow him to interface two separate codes which use different input and output formats. This script is used in research work, which results in some publications.
- Brian is not expected to make the script available, as long as he has made the data that underpins the research work available and he has provided the metadata that describes it, including the formats. In this case it is of benefit to both Brian and other researchers for him to simply make the script available under an open licence. This is particularly the case if the amount of code was small, and there was no expectation that Brian would support the script after release.



<http://dx.doi.org/10.6084/m9.figshare.1540765>

Software Sustainability Institute

# Dealing with commercially confidential objects



- Diya is undertaking research which simulates the airflow over a vehicle chassis, and has created an improved version of a commercial software model provided by an industry partner. She has then published a paper with the permission of the industry partner which broadly describes the revised model and presents the results of applying this model to a test dataset.

In the case of 'commercially confidential' research data (in this case the airflow model), where a business organisation has a legitimate interest, it is not expected that the improved version produced by Diya would be made openly available. However, it would be reasonable to investigate making the revised model available subject to a suitable, legally enforceable, non-disclosure agreement to enable other researchers to verify the results published in the paper.

Software Sustainability Institute



# Creating new software as part of a research project



- Colin has written a piece of software which implements a new algorithm for calculating a statistical index on a pre-existing dataset, and has published this algorithm in a paper along with results benchmarking it against other implementations of the statistical index.
- As the paper describes both the algorithm and compares it to other work, it is important that Colin deposits the software and makes it accessible. It will also be important for others to have access to the pre-existing dataset to enable validation of the results in the paper, which ideally will have a DOI and be openly accessible under a Creative Commons Attribution licence.



Software Sustainability Institute



# Faced with enormous amounts of generated data



- Feng is working on a large theoretical physics experiment which uses a piece of software to generate simulated data for an event. Each event data set consists of a very large amount of data, but a scientifically equivalent data set can be recreated as long as the initial parameters are identical.
  - In some cases, it may not be possible or cost effective to preserve research data. For example, in the case of simulated data or outputs of models, it may be more effective to preserve the means to recreate the data by preserving the generating code and environment, rather than preserving the data themselves. Provided that the ability to validate published research findings is not fundamentally compromised, a deliberate decision to dispose of research data at an appropriate time is acceptable in these cases.



Software Sustainability Institute

<http://dx.doi.org/10.6084/m9.figshare.1540765>



# Exploiting software with commercial potential



- In the course of her EPSRC-funded research Erin has written some code which she believes has real commercial potential in its own right. She has written up the work and wishes to publish, but the results can only be validated by the code and Erin does not wish to jeopardise its commercial potential by disclosing it.
- Erin should seek the advice of her University's commercialisation support office because under EPSRC's standard grant conditions the university owns, and has the responsibility for exploiting, the intellectual property arising from EPSRC research grants. Because it is acceptable for there to be a delay in publication while arrangements are made to protect valuable IP, if the support office agrees with Erin they should ensure that suitable protection is put in place before the paper is published. It is important that the code is available to anyone who wishes to validate Erin's research after it is published.



Software Sustainability Institute

<http://dx.doi.org/10.6084/m9.figshare.1540765>

