VERBRUGGEN LAB DATA MANAGEMENT GUIDELINES

V3.0 7/1/2015

Version	Summary of changes	Person responsible	Last amended
2.0	Second draft based on UK Data Archive's best practices, but tailored as much as possible to the needs of the CCAL research group.	Myriam Mertens & Frederick Verbruggen	02/12/2012
2.1	Incorporation of suggestions Open Access Team Exeter and Marine Renewable Energy Group. Omitted centralised CCAL back-up system until more resources have been identified.	Myriam Mertens & Frederick Verbruggen	21/01/2013
2.2	Correction of spelling errors and sources added. Comment about R added in section 2.1.3.	Myriam Mertens	17/07/2013
3.0	Scope of guidelines reduced to the Verbruggen lab. Revision and extension of existing guidelines (roles & responsibilities; data storage, documentation and formatting). Addition of guidelines on data sharing & archiving.	Myriam Mertens & Frederick Verbruggen	7/1/2015

Sources: the text and content of this document are based on and adapted from:

- the UK Data Archive's data management guidelines:
  - Brochure ‘Managing and Sharing Data. Best Practice for Researchers’, May 2011 (http://www.data-archive.ac.uk/media/2894/managingsharing.pdf)
  - UKDA website, ‘Create & Manage Data’ (http://data-archive.ac.uk/create-manage)
  - L. Corti, V. Van den Eynden, L. Bishop and M. Woollard, *Managing and Sharing Research Data. A Guide to Good Practice* (Los Angeles, 2014)
- the Marine Renewable Energy Group's data management policy (I. Ashton, H. Lloyd-Jones and A. Cowley, ‘Developing Research Data Management Policy at Research Group Level’, July 2013. Available at: https://ore.exeter.ac.uk/repository/bitstream/handle/10871/12107/MRCaseStudyFinal.pdf?sequence=5)
- the University of Exeter’s Research Data Management pages (http://as.exeter.ac.uk/library/resources/rdm/)
- G. Cole, J. Evans and H. Lloyd-Jones, ‘Checklist for depositing data on Open Research Exeter (ORE)’, 2013 (Available at: https://ore.exeter.ac.uk/repository/bitstream/handle/10871/9255/ChecklistforOREdeposit13_05_09.pdf?sequence=2)
- the University of Exeter’s IT pages (http://as.exeter.ac.uk/it/)
- the University of Exeter’s Records Management Services pages (http://www.exeter.ac.uk/recordsmanagement/)
- the University of Exeter’s Intellectual Property Policy pages (http://www.exeter.ac.uk/research/toolkit/sharing/ip/ippolicy/)
- the University of Cambridge’s Managing Research Data pages (http://www.lib.cam.ac.uk/dataman/index.html)
- British Psychological Society, ‘Code of Human Research Ethics’, 2010 (Available at: http://www.bps.org.uk/sites/default/files/documents/code_of_human_research_ethics.pdf)
- Dropbox’s help page (www.dropbox.com/help)
- Time Machine’s support page (https://www.apple.com/uk/support/timemachine/)
- suggestions made by the University of Exeter's Open Access and Data Curation Team (http://as.exeter.ac.uk/crs/ae/open_access_&_data/) and Records Manager (http://www.exeter.ac.uk/recordsmanagement/)

Authors: Dr Myriam Mertens and Prof. Frederick Verbruggen

Summary

Proper research data management is integral to good research practice: it ensures that the data generated by lab members are stored securely, will be reusable in the future, and can be shared easily amongst collaborators. Moreover, it is an increasingly important part of funder and institutional requirements regarding open access to research.

Below you will find guidelines detailing how to manage data, and assigning roles and responsibilities. They are largely based on the UK Data Archive's best practices, but tailored as much as possible to the needs of the Verbruggen lab.

Contents

1. Roles and responsibilities

Research data management is the joint responsibility of the Principal Investigator (PI) and other researchers in the lab (postgraduate research students (PGRs), research assistants and postdocs).

1.1. The PI of the lab is responsible for:

- Setting up new research projects. This involves:
  - Discussing the data management procedures with new staff and PGR students at the beginning of projects
  - Establishing project acronyms for file and folder names (for non-PGR projects)
- Assisting lab researchers with data documentation where necessary
- Centrally storing and backing up definitive/final versions of experiment files
- Securely storing paper forms of completed studies, and safely destroying them after a 5-year retention period
- Depositing data that support publications on which he is lead author
- Approving lab researchers' data files before they deposit
- Depositing remaining data sets and project-level documentation at the end of research grants
- Overseeing data disposal on lab computers at the end of PhD and grant-funded research projects
- Reviewing data management guidelines annually and in response to emerging issues and changing institutional or funder policies
- Keeping a lab data inventory, based on the centrally stored experiment files
- Maintaining a data management resources library with further guidance, templates, and key policy documents relevant to the lab

1.2. Lab researchers are responsible for:

- Creating documentation files for their experiments
- Storing and backing up data and other experiment files in the right format
- Securely storing signed consent and other paper forms of ongoing studies
- Handing over signed consent and other paper forms of completed studies to the PI
- Providing the PI with the definitive/final versions of their experiments' data, software and documentation files for central storage
- Depositing data that support publications on which they are lead author

1.3. PGR students are in addition responsible for:

- Depositing remaining data sets and project-level documentation at the end of their doctoral research project

2. Data documentation, formatting and storage

2.1. File structure and data documentation

For each experiment, lab researchers should store data and other files in one folder. Within each experiment folder, there can be several sub-folders (depending on the nature of the experiment):

- Data
- Software
- Documentation

The PI and PGRs can group together the folders of experiments that are part of their larger research projects (i.e. grant-funded and PhD projects) in a 'project' folder.

2.1.1. Data

The data folder of an experiment will have two subfolders:

- Raw data: raw files collected during the experiment. No other files should be present in this folder.
- Processed data: any data that have been subjected to automated or manual processing routines. It is important to document the processing routines in an analysis document or in R (see section 2.1.3).

2.1.2. Software

The software folder of an experiment will have two subfolders:

- Experiment software: the software that was used to run the experiment
- Analysis software: the software that was required to analyse the data

2.1.3. Documentation

Detailed documentation of the data collection and data analysis process helps other researchers understand your data and can provide further evidence of data quality.

Lab researchers should create sufficient *experiment/study- and data-level documentation* for each of their experiments, so that others can reuse, replicate or reanalyse the data. The PI can assist with experiment and analysis documentation, for example by providing lab researchers with templates of research project summaries or by sharing R analysis scripts.

The documentation folder of an experiment should always include:

- Experiment Documentation: a text document with experiment and data information as outlined in **Appendix B**
- Analysis Documentation: a text document explaining how the data were processed and analysed, and including the information listed in **Appendix B**. If you use R, you can create the analysis document within R and include a summary of data analysis outcomes.

In addition, the documentation folder of an experiment *could* also include:

- Lab notes: any notes taken during the experiment that are relevant for understanding and interpreting the data
- An 'access and use conditions' document: may be required to specify under what particular (i.e. restricted) conditions data can be accessed and used by other researchers (see Section 3, 'Data sharing and archiving' for more details)
- Blank consent forms: with information/debriefing sheets
- Other documents: experiment instructions, blank questionnaires, and any other documents used in a study/experiment. When depositing data, research outputs such as presentations and publications can also be included as a form of documentation.

Together with their project folders, the PI and PGRs should also create *project-level documentation* giving more detailed information about their research projects as a whole and thus the broader context of data collection: history, aims, hypotheses, etc. This can take the form of research reports (e.g. ESRC End of Award Reports) and PhD theses, so it shouldn't require much additional work.

2.2. File formats, folder and file naming, and version control

2.2.1. File formats

Choosing appropriate file formats ensures longer-lasting digital data.
Appropriate file formats for long-term usability of data are typically standard and open, non-proprietary formats. Lab researchers can use the formats and software most suitable for their analyses, but–when practically possible–should convert their files to open or standard formats before offering data to repositories. Open or standard formats should also be considered for backups.

Recommended file types:

- Documentation files: rtf and/or pdf
- Behavioural data: where possible, create a txt or csv copy
- Questionnaire data: csv file with the scores/ratings
- Digital image data: TIFF version 6 uncompressed (.tif)
- Digital audio data: Free Lossless Audio Codec (FLAC)

2.2.2. Folder and file naming

Experiment folders and files should be named appropriately and consistently.
Experiment folder and file names should include the acronyms of both the larger research project and the specific experiment, and file names should in addition include any relevant specification. You should be able to tell what a file contains on the basis of its name rather than its location on the computer.

For example, the file named 'ECG_GambleChange10_ExpDoc.rtf' contains the experiment documentation of the Gamble Change 10 experiment in the 'Executive Control of Gambling' project. For more folder and file name examples, see **Appendix A**.

2.2.3. Version control

Version control should be used when there is, or likely will be, more than one version of a file, for example because it is edited or stored in multiple locations.

Version control helps you keep track of changes, and identify and locate the *master file* (the original from which working copies are created), *milestone versions* (significantly changed versions to keep), or the most recent version of a file. In this way, you know the correct version to work with and can revert to an older one if necessary. This is especially important in collaborative research.

Appendix C lists some *version control* strategies that the PI and lab researchers can adopt.

Important points to keep in mind:

- Lab researchers should keep a single master file of raw data that should be left untouched. For sharing with collaborators, editing, analysis etc. they should use working copies instead, so that the original files are not affected.
- Lab researchers should never delete raw data files, even when experiment subjects are excluded or something went wrong with data collection. The only acceptable exceptions are raw data generated when they run themselves through a few trials to check whether the computer program works, or to give a demo.

2.3. Short-term data storage and backup

Consistently storing and backing up data keeps them safe and recoverable.

2.3.1. Local storage

Lab researchers should store uncompressed experiment files on their computer's hard drive, in their personal filespace on the University's central server (i.e. the U:Drive, restricted to a maximum of 20 GB), or in their Dropbox account.

The advantage of Dropbox is that, as an online storage service, it gives access to your files from any computer connected to the internet, and synchronises them across your various devices. Files stored in Dropbox will be automatically stored on the Dropbox cloud servers as well as on your computer's hard drive (if you have installed the Dropbox desktop application there). A disadvantage of Dropbox is that it is not suitable for storing *personal, confidential* or *sensitive data*, as discussed below.

Lab researchers should regularly check the completeness, accuracy and integrity of all stored files for the experiments they conduct (e.g. checking that raw data files are complete, that there are no duplications, that all required documentation files are present and accurate, that there have been no inadvertent file deletions or modifications…).

2.3.2. Central storage

The collaborative nature of the lab's research also calls for central storage of data, software

and documentation. Therefore, the final/definitive versions of all experiment files (data, documentation, software) will be centrally stored on the PI's business Dropbox account. Lab researchers should provide him with these files as soon as they have finalised them (see also Section 3 on data sharing).

2.3.3. Backups

To prevent the loss of data and other files, both lab researchers and the PI should back up their own computer systems (including all office and personal computers where experiment files are held) at regular intervals. The backup process can be automated by using backup software, for example Apple's built-in Time Machine for Macs.

If you use Time Machine, keep in mind that it makes *incremental backups* to a backup disk, and deletes older backups in favour of newer backup files once the disk is full. As it is best not to overwrite old backups with new ones, have Time Machine notify you when the backup disk is full and old backups are deleted, and then select a new backup disk.

Backup files should be stored offline, on external hard drives (avoid USB sticks for backups). Removable storage media should be properly organised and labelled (indicating content and date) to facilitate restoration when needed.

It is recommended that the PI and lab researchers regularly verify their backup files by trying to restore them.

2.3.4. Data security

Computers
To improve security, all office, lab and personal computers where experiment files are stored should be locked with a password and protected against viruses and malware.
For more information, see the University of Exeter pages on Information Security: http://as.exeter.ac.uk/it/infosec/virusesandmalware/

Dropbox
All lab members using Dropbox are strongly encouraged to use two-step verification for additional protection to their accounts. For more information, see https://www.dropbox.com/en/help/363

Lab researchers should not install the Dropbox desktop application nor set the login screen on the Dropbox website to 'Remember me' on lab computers. This is to prevent other users of the lab computers from accessing lab researchers' Dropbox account, and also to avoid syncing while running an experiment.

The PI and lab researchers should NEVER store *personal, confidential* or (commercially or otherwise) *sensitive data* on Dropbox. Crucially, only *anonymised research data* that cannot reveal the identity of living individuals, or be linked to individual participants by data recipients outside the lab, should be stored in the cloud.

See the **Glossary** for more information about *personal, confidential* and *sensitive data*, and *anonymised research data*.

External hard drives
External hard drives should be *encrypted* and stored securely (e.g. in a room that can be locked). When not using their computers for a longer period (e.g during holidays), the PI and lab researchers should not store their hard drive in the same location as their computer system.

As physical storage media have a limited life, files should be transferred to a new hard drive every 2 to 5 years (with the files on the old drive properly erased, i.e. by overwriting them).

Paper-based files
Paper-based files containing participants' details, such as signed consent forms or brain stimulation screening questionnaires, have to be handled securely to avoid disclosure of personal data. Moreover, it is important to keep them separate from raw research data files to prevent data recipients outside the lab from linking research data to individual participants.

Lab researchers should store completed questionnaires of ongoing studies separately and securely (i.e. in a locked filing cabinet). Upon completing their study, they should accurately convert the questionnaire responses into digital tabular data (which should not contain any personal identifiers) and subsequently hand the completed paper questionnaires to the PI.

Lab researchers should store signed consent forms, payment sheets, and brain stimulation screening questionnaires of ongoing experiments/studies separately and securely (i.e. in a locked filing cabinet), and hand them to the PI upon completing a study.

The PI should in turn safely store the paper forms received, and destroy them after a 5-year retention period. Based on the lab's data inventory, he can track the date of origin and status of experiment data, and identify experiments with forms that need to be securely disposed of. This can be done with a shredder conforming to appropriate security standards.

3. Data sharing and archiving

3.1. Legal and ethical issues regarding data sharing

Data sharing touches upon a number of important legal and ethical issues involving *intellectual property rights (IPR)*, *personal, confidential* and *sensitive data*, and informed consent that should be taken into account.

The PI and other lab researchers should NEVER formally or informally share:

- Data and other files for which they don't own *copyright* or don't have permission to share from the copyright holders
- Data and other files that contain *personal*, *confidential* or *sensitive information* without research participants' written consent
- *Anonymised research data* without research participants' informed consent

In other words, data sharing requires researchers to know who owns the *IPR* of their data and how the UK Data Protection Act, which covers the use of personal data, applies to them. It also has important implications for the way in which informed consent forms should be drawn up.

See the **Glossary** for more details about *IPR*, *personal, confidential* and *sensitive data*, and *anonymised research data*.

For more information about data sharing and informed consent, and a template consent form, see **Appendix D**.

3.2. Sharing active data

Lab researchers and the PI can use Dropbox to informally share appropriate *active* (i.e. not yet archived) data and other files with each other, and with trusted researchers beyond the lab, such as reviewers and grant collaborators.

Lab researchers should consult the PI first before sharing *active data* with external researchers.

When informally sharing *active data* and other files, it is important to prevent accidental changes to or deletion of those files. If you use Dropbox for sharing, its website lists two main ways to avoid unauthorised file changes or deletions:

- Don't invite people to a shared Dropbox folder, but send them a link to the folder or file you want to share, which they can then click to view and download a copy of the file(s). Remove the link if you no longer want the folder or file to be accessible from the link (with a Dropbox business account, you can set passwords and expiration dates to control access to links).
- If you do invite people to shared Dropbox folders, give them view-only permissions. In this way, members of your shared folder will be able to see the latest version of the files in the folder, but will not be able to add, delete or edit files.
- For more information on how to do this, see the Dropbox website: https://

www.dropbox.com/en/help/5823

3.3. Archiving completed data sets

It is lab policy to formally share completed research data sets by depositing them in a data repository. This raises the data's profile and enables their proper citation by other researchers. Moreover, depositing data ensures their long-term preservation and fulfils growing funder and institutional requirements regarding open access to research.

3.3.1. Who?

Lead authors, whether they be PI, PGR students or other lab researchers, should deposit the completed data sets (with accompanying software and documentation) supporting their scientific publications (journal articles, book chapters...).

The PI and PGR students should in addition deposit project-level documentation and any remaining data identified for archiving from their grant-funded research and PhD projects respectively.

3.3.2. Where & when to deposit?

Deposits supporting publications
When publishing papers, lead authors from the lab should simultaneously make the underlying data available by depositing them in ORE, the University of Exeter's own backed-up repository.

Deposits at the end of PhD and grand-funded projects
Upon completion of their grant-funded or PhD research projects, the PI and PGR students should make sure that they comply with funder and/or institutional requirements regarding data deposits.

- PGRs should offer any remaining data sets that need to be archived as well as project-level documentation (their PhD thesis) to ORE.
- The PI should also offer any remaining project data selected for archiving to ORE.
- In addition, if their funder requires them to make their data available via a particular repository, PGRs and the PI should register their projects, and provide (links to) project-level documentation as well as links to the data on ORE in this funder-designated repository within the required time frame (the procedure of providing links to data already available on ORE was approved in May 2014 by the UKDS ReShare repository, for example, where ESRC grant holders are expected to deposit their data within three months of the end of their award).

3.3.3. Before depositing data

Before offering data and accompanying files to ORE, lab researchers should seek the PI's approval. The PI grants permission after reviewing the contents of the files to be deposited. If necessary, he can ask lab researchers to edit files before offering them to ORE. If changes are made, the files in question should also be shared with the PI for central storage.

The PI and other lab researchers should consider the following questions before offering data for archiving:

- **Do all collected data sets need to be preserved?**
  - The PI and other lab researchers should in the first instance focus on archiving the data that support their publications.
  - For deposits at the end of grant-funded or PhD research projects, the PI and PGR students should select remaining data for preservation in compliance with funder and/or institutional requirements.

- **Is it legally, ethically and commercially appropriate to share the data?**
  - The PI and other lab researchers should ensure that there are no issues regarding intellectual property rights, disclosure of (commercially or otherwise) sensitive, confidential and personal information, or informed consent that would preclude them from making their data open access.

- **Do the data require a restricted level of access?**
  - If there are legal/ethical/commercial issues, repositories can restrict and regulate access and use of data, so that these data can still be deposited and shared under certain conditions.
  - The PI and other lab researchers should also consider whether they want the repository to delay making their data publicly available for a certain period, for example until they have published the research outputs associated with these data.
  - If applicable, the PI and lab researchers can indicate that restricted access is required in an 'access and use conditions' file (see Section 2.1.3., 'Data documentation').

- **Are the data and documentation files properly structured and formatted, and are data appropriately documented?**
  - The PI and other lab researchers should check whether the data sets selected for archiving comply with repository requirements regarding file structure, formatting and documentation (if they have followed the lab's data management guidelines, this should normally be the case), and whether file contents are accurate.
  - Depositors should ensure, in sum, that other researchers are able to understand and reuse their data.

3.3.4. After depositing data

Once it is confirmed that data and accompanying files have been archived in ORE and associated research outputs have been published, lab researchers may choose to dispose of the corresponding files on their computer hard drives and in their Dropbox accounts, and thus only hang on to active and unpublished data if they wish.

If they need or want to dispose of data, it is recommended that the PI and PGR students refrain from deleting experiment files on their computer hard drives and in their Dropbox accounts *at least* until they have fulfilled all institutional and/or funder data deposit

requirements at the end of their grant or PhD project, and have published the associated research outputs.

The PI will decide on and oversee the disposal of data on lab computers at the end of grant-funded and PhD projects.

4. Glossary

Data documentation:
Data documentation enables users to understand research data, as it documents their creation, meaning, content, structure and manipulations. It also forms the basis of catalogue metadata compiled by data repositories in order for users to find and cite research data. Data documentation includes information about research data at different levels: the project level, the experiment or study level, and the data level. (*by 'study', we mean a meaningful (sub)group of experiments, e.g. an experiment and follow-up experiments).

Experiment/study-level and data-level documentation:
This is the responsibility of lab researchers conducting experiments. Elements of data-level documentation can often be embedded within a data file itself (e.g. variable labels within a data file). Experimenters should nevertheless also prepare supporting documents to provide sufficient experiment-/study- and data-level documentation. These files should enable other researchers to replicate the study or reanalyse the data.

Project-level documentation:
Applicable to research that is part of a larger project. Project-level documentation includes information about the broader context of data collection: the project history, aim, objectives, hypotheses, etc. It can take the form of research reports (e.g. ESRC Website Abstract or End of Award Reports), so it shouldn't require much, if any, additional work. This is the responsibility of PIs and PGRs in charge of grant-funded and PhD research projects respectively. If project-level documentation is not applicable (e.g. for pilot experiments), data documentation is restricted to Experiment/study-level and data-level documentation.

Version control:
Version control entails a variety of strategies that researchers can adopt to keep track of different versions of files. Also see **Appendix C**.

Master file and milestone versions:
A master file is an original file that you always want to be able to go back to, and from which working copies are created for editing. For example, if you are editing a digital image, you would not want to lose the original picture. Milestone versions are versions of a file that are changed to such an extent–compared to previous versions–that you want to keep them (e.g. the first draft, revised draft, final draft of a file). Also see **Appendix C**.

Incremental backup:
Time Machine makes incremental backups to a backup disk. This means that it makes an initial full backup of your system, and subsequently backs up the files that have changed since the last backup. It saves hourly backups for the last 24 hours, daily backups for the past month, and weekly backups beyond that. Once the backup disk is full, Time Machine will delete the oldest backup file to make room for the most recent one. You can have Time Machine notify you when the disk is full and older backups are deleted by turning on this notification in Time Machine Preferences. It is best not to overwrite old backups with new ones, e.g. by selecting a new backup disk.

Active data:
Experiment data that have not yet been archived, i.e. that have not yet been deposited in a data repository for long-term preservation.

Intellectual Property Rights, including copyright:
Before sharing data, researchers should know who owns the intellectual property rights (IPR) of these data. Copyright is an important IPR that prevents unauthorised copies and publishing of original work that exists in written or recorded form. Researchers should therefore never share data or other materials via Dropbox, data repositories, or other means unless they own copyright or have permission for sharing from the copyright holders.
Under UK Law, universities usually own the IPR arising from research by their employees in the course of employment. In practice, as is the case at the University of Exeter, copyright of research materials and publications is often assigned to researchers, unless the latter have signed IPR agreements (e.g. with funders) stating otherwise. As PhD students are not considered to be university employees, they own the copyright of data and outputs from their research (unless they have sponsors claiming IPR). For collaborative research, all researchers or institutions involved own copyright.
More details on Exeter University's Intellectual Property policy can be found at http://www.exeter.ac.uk/research/toolkit/sharing/ip/ippolicy/#d.en.238396

Personal, confidential and sensitive data:
Research with human participants can generate data containing personal, confidential or sensitive information, the sharing of which may breach legal and ethical obligations.
In a UK context, personal data refers to data that, on their own or linked to other information, can reveal the identity of a living individual. Sensitive personal data in addition contain information about a person's race, ethnicity, religion, political beliefs, physical or mental health, sex life, trade union membership, (alleged) criminal offences… . Confidential data are data containing identifying information that an informant gives in confidence, that two parties agree to keep confidential, i.e. secret.
Researchers should not share personal and confidential data without research participants' written consent. Disclosure of (sensitive) personal and confidential data without consent constitutes a breach of the UK Data Protection Act (DPA) and psychologists' duty of confidentiality towards research participants under the British Psychological Society's ethics code.
It is best that lab researchers do not collect personal and sensitive data that are not necessary for their research. They should not disclose or share participants' personal details on the signed consent forms and payment sheets, nor experimental data/questionnaire responses in which participants can be individually identified.
More details on the DPA and research can be found here: http://www.exeter.ac.uk/media/level1/academicserviceswebsite/it/recordsmanagementservice/Research_DP_Guidance_200912.pdf

Data can also be sensitive from a commercial or institutional point of view. Researchers should be careful with storing/sharing valuable or sensitive University information via Dropbox or otherwise. The University of Exeter's IPR policy (http://www.exeter.ac.uk/research/toolkit/sharing/ip/ippolicy/#d.en.238396) includes a section on confidentiality, which governs the disclosure of intellectual property with potential commercial value.

Anonymised research data:
Researchers can share anonymised research data, where (in)direct personal identifiers (such as names) revealing the identity of living individuals have been removed, and which data recipients are unable to link to individual participants. Anonymised research data are exempt from the UK Data Protection Act, and anonymising data maintains confidentiality. Researchers should inform subjects about the sharing of anonymised research data when seeking their consent to participate in experiments.

Encryption
Encryption is used to safely store or move files, as it prevents unauthorised access. You can encrypt individual files, folders or entire storage devices.
Lab researchers should encrypt the information on their external hard drives, as these can easily get lost or stolen. When using Time Machine, this can be done by checking the 'Encrypt backups' box for the backup disk. There are also external hard drives available with built-in encryption (Exeter University's IT recommends Freecom Tough Drive: http://as.exeter.ac.uk/it/infosec/devices/). **Caution: if you forget your password, you will not be able to recover the files on your encrypted backup drive!**

5. Appendices

Appendix A. Folder structure and file naming

Examples from the 'Executive Control of Gambling' research project:

Folder names:
- Project folder: *ECG*
- Experiment Folder: *ECG_BarEye*

This would create the following structure:
- *ECG*
  - *ECG_BarEye*
    - *Data*
    - *Software*
    - *Documentation*
  - *ECG_SlotStop*
    - …

Data files:
- Behavioural data: *ECG_BarSA_Behavioural_s1.txt*
- GSR data: *ECG_BarSA_GSR_s1.txt*
- Eye-movement data: *ECG_BarSA_eye_s1.edf*
- …

Program file names:
- Final, bug-free version of program file: *ECG_BarEye.m*
- During debugging stage, one can use a suffix; e.g. *ECG_BarSA_d1.m*. Usually, you can delete the d-files.
- Analysis program: e.g. *ECG_BarEye_ChoiceAnalysis*.R
- …

Documentation folder:
- *ECG_BarEye_ExpDoc.rtf*
- *ECG_BarEye_LabNotes.rtf*
- *ECG_BarEye_AnalysisDoc.rtf*
- *ECG_BarEye_InformedC.rtf*
- *ECG_BarEye_Debrief.rtf*
- …

Appendix B. Experiment and Analysis documentation

The documentation folder of an experiment should always include:

- **Experiment Documentation**: a text document with the following information:
    - Project information: a short summary of the larger research project of which the experiment is a part, so as to explain the wider context of data collection
    - Study/Experiment information, including:
        - Experiment code(s) as mentioned on informed consent forms
        - Research question
        - Experiment context (full name of experimenter, location, date and time, credit or paid participants)
        - Data collection methods (description of experiment, data collection protocols; instruments, software, and hardware used…)
    - Data file information, including:
        - Structure of data files (overview of all data files--e.g. behavioural, Eyelink, SCR…--and the relationship between them)
        - Content of data files: names, labels and descriptions for all variables and their values
    - Quality control measures: e.g. piloting of experiments, timing checks, calibration procedures (e.g. Eyelink), checking of manual data entries (e.g. in case of questionnaires), etc.

- **Analysis Documentation**: a text document that explains how the data were processed and analysed. It should include the following information:
    - How many subjects were included? Did you exclude subjects in the final analysis. If so, why?
    - Which variables were analysed? Did you exclude variables? If so, why?
    - Did you exclude trials when analysing the data (e.g. incorrect trials, trials following an error, mean +2.5 SD, etc.)?
    - Did you modify data or create derived/constructed variables from the original data? If so, explain the logic.
    - When R is used, the analysis document can be created within R and include a summary of data analysis outcomes. Examples can be provided by the PI.

Below is an example of an experiment documentation file:

Project information

People often need to take decisions that involve some element of risk. This project investigates what factors determine risk-taking behaviour when making monetary choices. In particular, it focuses on how the need to exercise control over one's motor actions can influence risk-taking when making monetary decisions. Recently, the researchers have found that being prepared to cancel a motor response reduces risk-taking in a gambling task (Verbruggen, Adams, Chambers, in press). Furthermore, they have demonstrated that learning to stop actions reduces gambling even when the tasks are performed up to two hours apart. These findings suggest that action control and high-level decision-making are intimately related. This offers exciting prospects for the development of interventions in e.g. gambling, substance abuse, and overeating. In a series of studies, different aspects of control will be examined to increase our understanding of the conditions in which low-level control of action transfers to the control of higher-order decisions. This could lead to the development of new behavioural training programs. More generally, this work has the potential to contribute to a better understanding of self-control and the ways in which deficits in control can lead to impulsive behaviour, poor restraint over urges, and increased risk-taking.

Funder: ESRC (PI: Verbruggen)

Study/experiment information

Research question:

In this experiment, we investigate whether mixing a speed/accuracy task with the gambling task (used by Verbruggen, Adams, & Chambers, 2012, PsySci) influences gambling.
Main question: Is gambling reduced in the accuracy blocks compared with the speed blocks (viz. the signal vs. no-signal blocks).

Experiment context:

Code: XXXXX (to be added)
Who: Myriam Mertens
Where: G12
Credit/paid: Credit (subject code < 100) & paid (subject code > 100)
When: November-December 2012

Brief description of method (provide all info required to understand the headers):

On every trial, participants were presented with 6 bars. Each bar was associated with a certain amount they could win; however, they were informed that the higher the amount, the less probable a win. Trials started with the presentation of the amounts and the corresponding keys. The choice amounts depended on the stake (low, medium, or high). After 3.5 sec, the bars started rising, & participants had to respond when they reached the top line. At the end of a trial, participants were told how much they had won or lost, and what their current balance was. If they were too late, responded too soon, or if they pressed an incorrect key, they lost a fixed amount (see Verbruggen, Adams, & Chambers, for a detailed description of the bar task).

After the feedback was presented, two large grey rectangles appeared on the screen. Subjects

had to determine whether the brighter rectangle appeared on the left of right of the screen. In half of the blocks, they were instructed to respond as accurately as possible; in the other half of the blocks, they were instructed to respond as quickly as possible. Feedback at the end of the trial was adjusted accordingly.

Apparatus: DELL PC, TFT screen, Matlab (Psychtoolbox)
Calibration protocol: None

Data file information:
File Types: Behavioural data files only.
Variable labels (headers) and variable coding:

- block: the block number (0 = practice)
- condition: 1 = accuracy; 2 = speed
- height: 1 = low bars; 2 = heigh bars
- stake: low, medium, high
- choice: 1 = highest amount; 6 = lowest amount
- outcome: NaN (no outcome), win, loss, anticipatory response (i.e. too soon), miss, wrong (i.e. incorrect key; e.g. 'v')
- resp: response 1-6 = e, r, t, y, u, i respectively
- rt: reaction time in ms
- balance: the current balance
- location:
  - 1 = the brighter rectangle is presented on the left
  - 2 = the brighter rectangle is presented on the right
- brightness: RGB [127 127 127] + 'brightness' for the brightest one (darker one = 127 - brightness). Minimum is 1.
- deadline: deadline to determine whether response was quick enough
- AccCount and SpeedCount: the counters for Acc-tracking and Speed-tracking respectively.
- cResp: response in the S/A task
  - 1 = left (v)
  - 2 = right (n)
- cRT: RT for the brightness task
- acc: accuracy. Was the response correct (1) or not (0) ?
- speed: was the RT shorter than the deadline (1) or not (0)?
- the order of the amounts

Quality control measures
The experiment was programmed in Matlab using Psychtoolbox. Timing accuracy of the Psychtoolbox routines has been tested frequently by the developers for both Windows and Mac operating systems (http://www.psychtoolbox.org). We conducted a pilot (N = 5) to check whether the S/A tracking worked properly; some minor adjustments to step-size were made.

Appendix C. Version control strategies

Listed below are version control strategies that lab researchers can adopt:

- To uniquely identify and distinguish between different versions of a file, incorporate version information in the file name:
  - Include a version number (number each successive version of a document sequentially, indicating major and minor changes by whole and decimal numbers respectively: e.g. v1, 1.1, 1.2, v.2...), or a date (in a format that allows easy sorting: 'yymmdd'). For example:
    - ECG_GambleChange10_ExpDoc_v1
    - ECG_GambleChange10_ExpDoc_140930
  - For files edited by multiple authors, also include author initials as an identifier. For example:
    - ECG_GambleChange10_ExpDoc_v1fv
    - ECG_GambleChange10_ExpDoc_140930fv

- Save files under a new file name (i.e. with a new version number or date, and if applicable, initials) *before* editing them.

- If you need to keep track of a file's history, include a version control table within the file (such as the one at the beginning of this document) or in a separate document. Record for every version of the file what changes were made, when and by whom, so that you know how different versions differ from each other and who contributed what.

- For the files you are responsible/lead author for, you might have to manually merge edits from multiple co-authors into a new version (with a new file name).

- Decide which versions of a file to keep, and how to organise them. Things to keep in mind:
  - Keep a single master file, especially for files where loss of the original would present a problem, such as raw data. Store them separately and leave them untouched (so do not overwrite or delete!).
  - For sharing with collaborators, editing, analysis etc. use working copies of these master files instead, so that the original files are not affected and you can always go back to them, if necessary, to start from scratch.
  - Lab researchers are responsible for the master files of experiments they conduct: only they should have write access to these master files.
  - You may not have/want to keep all old versions of files, but identify milestone versions to hold on to.
  - You can store non-current versions of files separately from current versions.

- Synchronise files held on different computers: e.g. by using Dropbox, which will automatically sync files on devices connected to the cloud storage service.

Appendix D. Data sharing and informed consent

It is important that informed consent forms, while discussing personal data protection and confidentiality, don't preclude data sharing.

Taking the nature of the lab's research into account, consent forms should inform participants that:

- The information and data provided will be used for statistical research and research administration purposes only.
- No data or responses will be published in which they can be individually identified.
- No personal details of individual participants will be revealed to persons outside the research project; signed consent and payment forms will be securely stored and disposed of after a 5-year retention period.
- Anonymised research data will be registered and archived in an institutional/specialist data repository so as to make them available to other researchers.

Below is a template of a consent form:

INFORMED CONSENT

EXPERIMENT NAME: EXPERIMENT CODE:

[Add description of experiment]

I understand that:

- Participation in this study is entirely voluntary and I can withdraw from the study at any time without giving a reason.
- Non-participation or withdrawal from the study will not be of any disadvantage to my university career.
- If I prematurely withdraw from the study I will still be appropriately compensated for my participation.
- During the data-gathering phase, I can instruct the researchers carrying out this study to delete (part of) the data obtained from me from their records.
- I am free to ask any questions at any time and I can arrange to discuss any concerns with the lead researcher (see below).
- The information and data provided by me in this experiment will be used for statistical research and research administration purposes only.
- No data or responses will be published in which I can be identified individually.
- My personal details will not be revealed to people outside the research project. Signed consent and payment forms will be securely stored and disposed of after a 5-year retention period. In accordance with the Data Protection Act, I can have access to my personal data on these paper forms within the 5-year retention period.
- Anonymised research data will be registered and archived at the University of Exeter's ORE repository, the UK Data Service's ReShare repository and/or another data repository in order to make them available to other researchers in line with current data sharing practices.

If you require more information about this study, please contact: ... (Lead Researcher)

Email: