REPRODUCIBILITY BY DESIGN

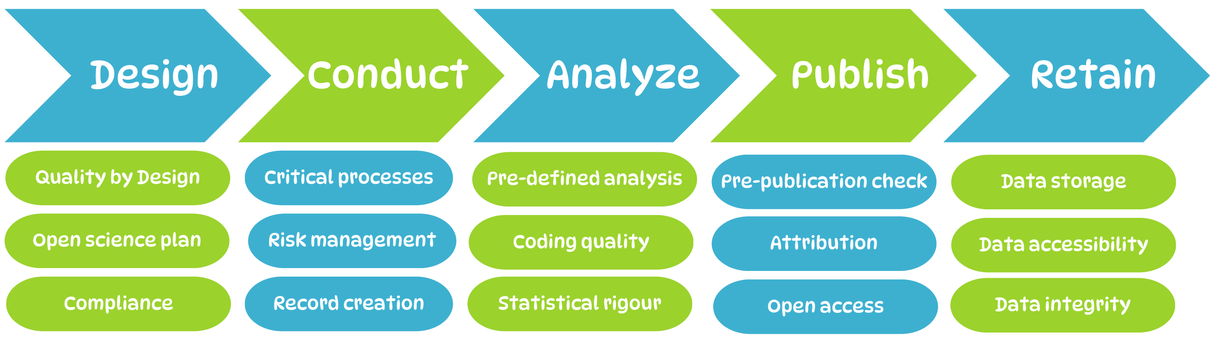
Reproducibility by Design: Creating a Quality Assurance Framework for Academic Life Sciences

Fiona Booth (ORCID ID 0009-0008-1617-9822), Nicholas Beazley-Long,
(ORCID ID 0000-0001-7342-2771) , Fabian Fischer (ORCID: 0000-0003-2325-9886),
Joe Flannagan (all authors are University of Bristol staff or students)

Introduction

The University of Bristol (United Kingdom) has devised a novel a quality assurance framework for life sciences which spans every stage of the research process. The “Reproducibility by Design” programme incorporates the concept of “Quality by Design”, a key principle of pharmaceutical research and development. Quality by Design is an approach based on gaining a deep understanding of a process where quality is critical to the output and is intrinsically linked to quality risk management and the application of controls in a risk-proportionate manner.

The University of Bristol’s Reproducibility by Design framework applies a systematic approach to the design, conduct, analysis, publication and retention of research. It has basic components which underpin all aspects of research and are common to the majority of life science disciplines, such as data management planning and bias minimisation. These are supported by more detailed discipline-specific guidance, training and tools. The design is modular to enable continuous improvement processes to be applied to ensure that the content can be evaluated and improved as needed.

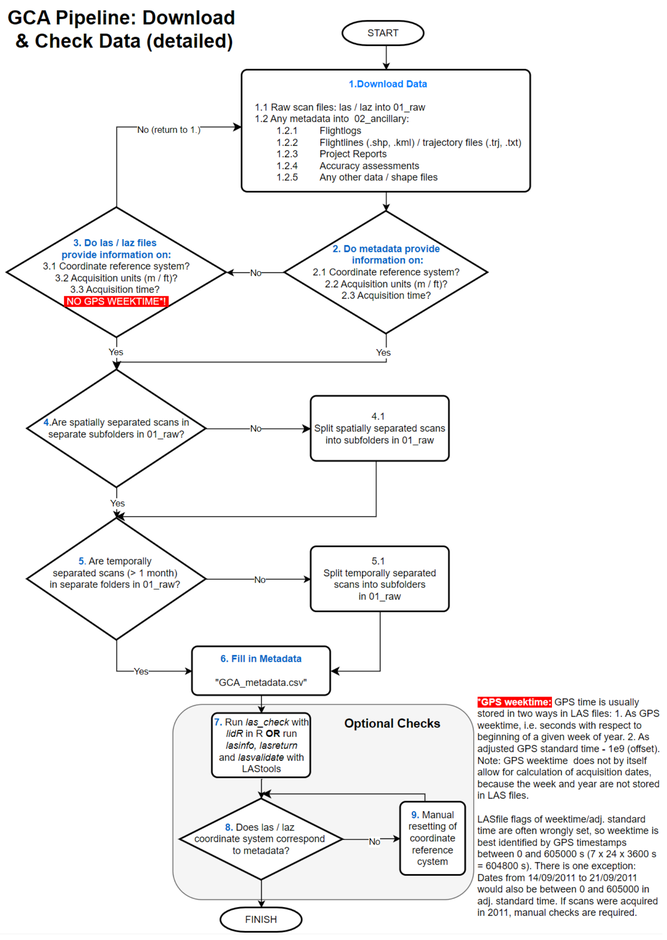


The framework has three key principles:

- Research data should conform to the high possible standards of data integrity.
- Quality by Design principles will be applied to ensure that research is designed to the highest possible standard, incorporating open science principles from the outset wherever possible.
- Quality risk management will be applied proportionately to identify and minimise risks to the integrity of data, the ethical conduct of research and to participants and the environment.

In this poster we present two case studies from different life science disciplines to demonstrate how Quality by Design, critical process and data identification, and risk management can enhance the overall quality of research outputs. The first case study demonstrates how process maps can efficiently and clearly describe the data quality checks and processing steps required to enable accurate interpretation of forest canopy data which is used in ecological research. The second case study describes the process flow for a surgical procedure performed on rats. This flowchart identifies areas for potential variability in how the procedure is conducted and identifies the metadata which is essential in providing contextual information on the experimental unit and the procedure itself.

Forest Canopy Case Study



The Global Canopy Atlas (GCA) is a forthcoming global database of airborne and drone-based laser scanning (ALS/DLS) acquisitions over forested ecosystems. It is curated specifically for ecological and climate-related research. In particular, it can be used to study the structure and carbon stocks of forests, understand the underlying dynamics (tree mortality, carbon gains and losses) and to validate satellite observations. To ensure global comparability, any dataset that is included in the GCA has gone through the same standardized and maximally robust pipeline (Fischer et al. 2024[1]). The pipeline was written in R and uses LAStools software.

Three flowcharts were created which describe manual steps to ensure that datasets meet minimum quality requirements and automated steps to derive analysis-ready products. The flowchart shown here describes a sequential set of steps to check raw scan data and the accompanying metadata, which typically includes flightlogs, flightlines, coordinate reference systems and scan quality reports.

Key checks include whether laser scans are properly georeferenced and whether the acquisition period is known. The latter is important to separate vegetation changes across years from vegetation changes across seasons, such as leaf fall in winter (boreal and temperate forests) or dry seasons (subtropical and tropical forests).

We expect the three process maps to greatly facilitate adoption of the pipeline by new users, highlight key quality requirements for laser scanning in global ecology and climate change research, and aid in the future development of the workflow, such as the identification of improvements and the translation to new software environments.

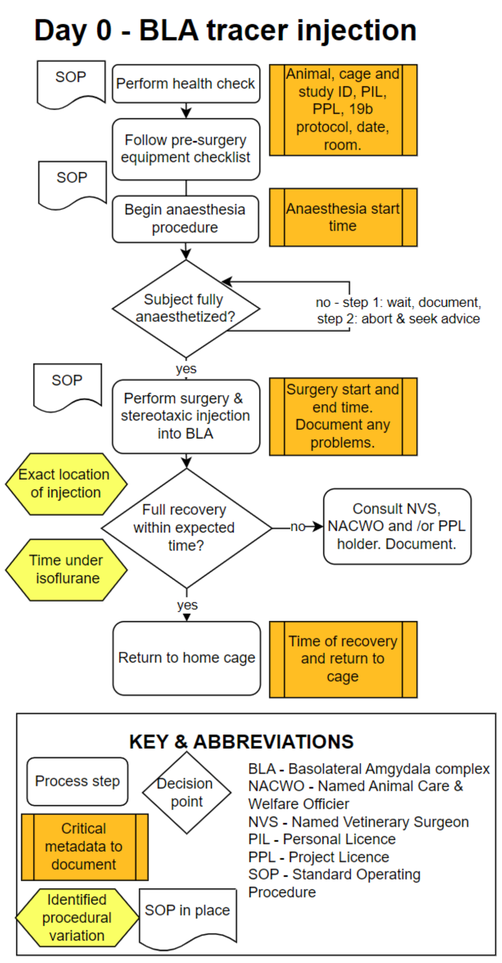
Rodent Research Case Study

Stereotaxic surgery and injections into the rodent brain enable precise delivery of therapeutic and research agents. Precise injection into the basolateral amygdala allows delivery of anterograde and retrograde tracers which reveal the intricate axonal and dendritic connections stemming from this key area of the limbic system implicated with fear and emotion.

As part of a project to create a Standard Operating Procedure (SOP) for data management, laboratory personnel mapped out all processes from the initial prestudy briefing document, receipt, health checking and habituation of animals, *in vivo* procedures, up to sacrifice and collection of brain tissue for slice electrophysiological recordings and histology.

The high level process map displayed here is a subset of a larger study process map and provides a visual guide to the fundamental steps involved in administering a licensed stereotaxic injection into the rodent brain. Mapping out this process enables researchers to first formalize the key steps involved in the procedure, and identify and agree on sources of procedural variability, and other sources of risks that could affect the quality of the output. The map also highlights parts of the process where a SOP is in place to control procedural variability.

This map identifies two common sources of variability, shown in yellow, which vary from procedure to procedure: time spent under anaesthesia with isoflurane, and the exact location of the injection.



Mapping out a research process also helps to identify contextual information required for rigorous documentation to maintain the integrity of the research data and research processes, and ultimately aid the reproducibility of the research. This is metadata, shown above in orange. Examples of other risks to study quality which process mapping can help to identify are performance bias, and data storage and security vulnerabilities. Periodic review and revision of these process maps can help to refine SOPs further to ensure that variation is minimized where possible and that all relevant metadata is captured.

Discussion

Quality by Design (QbD) was a term first coined by Joseph M Juran, an American Professor of Industrial Engineering at New York University. Juran’s quality management principles contributed to quality improvements which revolutionised the Japanese automotive industry. In his 1992 publication Juran on Quality by Design[2], he introduced the concept of process capability; this is the “inherent reproducibility of a process, the ability to repeat it’s results during multiple cycles of operation”. Whilst Juran’s expertise and experience related primarily to industrial manufacturing, there are clear benefits to understanding process capability and the extent to which it introduces inherent variability within the context of academic research. Following the adoption of QbD principles in pharmaceutical manufacturing in 2005[3], and for the conduct of clinical trials in 2021[4], we have designed a framework which utilises key elements of QbD to optimise the quality of research conducted in the academic life sciences. The two case studies above demonstrate that the principles of Reproducibility by Design such as examining the entire process of conducting research, analysing critical processes and identifying risks to quality by monitoring variation can have tangible benefits to researchers and for others that seeking to make use of the research in the future.

Like QbD, the foundation for Reproducibility by Design is process mapping, which is an easily adopted tool to gain a deep understanding of the complete process. Process maps provide a visual means to identify critical research processes and associated raw data and metadata. These are elements of research which have been identified as factors which contribute to irreproducible research[5]. Process mapping sessions should include all parties involved in the process to gain diverse perspectives and create consensus. The outputs can be used to improve the content of SOPs and as an adjunct to text to provide visual easy reference guides. They also help to inform the content of Data Management Plans by identifying the metadata which is required to support research outputs and to define areas of the research process which are critical to overall quality and therefore require robust control measures.

References:
1. Robust characterization of forest structure from airborne laser scanning – a systematic assessment and sample workflow for ecologists, Fischer et al, https://doi.org/10.1101/2024.03.27.586702 [1]).
2. Juran J. M., Juran on Quality by Design: the new steps for planning quality into goods and services, 1992, ISBN 0-02-916683-7
3. ICH Guideline Q8(R2) on Pharmaceutical Development
4. ICH guideline E8 (R1) on general considerations for clinical studies
5. Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). https://doi.org/10.1038/533452a

DOI 10.17605/OSF.IO/E853C