The School of Biological Sciences at Edinburgh University is committed to improving scientific reproducibility through open science, of which data management is a core component. Our data management strategy includes:

1.    Short term and long-term storage of primary and processed data internally on University-managed servers that are subjected to routine backup at multiple locations. The Q-Exactive HFX mass spectrometer generates data in the form of .raw files (‘raw files’). These raw files are then processed using software platforms (MaxQuant, Progenesis, Skyline) to produce datasets in form of large tables in .txt format.

We estimate we will generate approx. 2 TB of data per month from the Q-Exactive HFX. Short term access to data for University of Edinburgh researchers will be made available through DataStore, a file storage system available to all research staff. Extramural users of the instrumentation will receive data via online sharing tools, including Office365 University of Edinburgh SharePoint, a secure online platform for sharing documents and data. Long term storage will be available through the University of Edinburgh DataVault. Data will be archived in according with BBSRC policies for at least ten years since last access.

2.    Sharing of primary and processed data in formats that enable frictionless re-use to maximise scientific impact to potential end users. To this end, we will systematically collect experimental metadata to improve the reusability and reinterpretation of the datasets. Our strategy is to share data in a tiered fashion, providing both raw data that can be used by experienced mass spectrometrists, large tables that can interrogated bioinformatically, and also where appropriate, data will be shared via customised graphical interfaces for biologist end users (‘bench’ researchers’) created using the R/Shiny scripting platform.

3.    Uploading primary and processed data onto the relevant public data warehouses that have secure, long-term funding. For mass spectrometry proteomics and phosphoproteomics data, the ‘industry standard’ is the ProteomeXchange  consortium, which was established to provide international coordination in standard data submission and dissemination pipelines between the main proteomics data repositories.

Where appropriate we will also upload datasets to the University of Edinburgh Data Share, a service hosted by the University Information Services. Each DataShare dataset will have a unique DOI that can be linked to publications. It is the University of Edinburgh’s policy that datasets on DataShare will be permanently made available to the public.