ÚOCHB AV ČR
IOCB PRAGUE

Directive S2023-01

Addendum II

Detailed Specifications for Archiving Primary Data and Laboratory Logbooks for the Various Scientific Disciplines Pursued at IOCB

Research Integrity and Data Management

Research at IOCB must be conducted to the highest standards of integrity, and correct data management is crucial.

An integral component of the data management environment is the use of laboratory logbooks to record every experiment and an inventory system of all new compounds and their characterization data. A laboratory logbook may be either electronic (preferentially) or a bound paper book in which records are written in unerasable pen. Every experiment must be dated.

In adherence with FAIR data management principles, primary data must be kept for all published works (including peer-reviewed works and archive manuscripts and theses) in order to document their authenticity and ensure their reproducibility. All primary or raw data should be archived and safely stored for at least 10 years after the first publication on at least two independent storage media (hard-disks, CD/DVD etc.).

The aforesaid data should be made available upon a reasonable request. If the requested data is under embargo or has restricted access, basic metadata should at least be provided.

CHEM cluster

The CHEM cluster at IOCB broadly includes organic synthesis, bioorganic and medicinal chemistry, and material chemistry. All new compounds must be fully characterized by analytical methods and spectroscopy (NMR, MS, IR, etc.) and should be included in an electronic inventory system. All raw NMR data must be archived in the original FID format together with acquisition parameters; non-standard processing parameters should also be archived. Raw MS data should be stored in the format and software provided by the supplier of the spectrometer. Other raw spectra should also be archived in the format of the providers of the instruments. For X-ray crystallography, crystallographic information files (CIFs), structure factor tables, and CheckCIF should be submitted to the Cambridge Crystallographic Data Centre (CCDC). Scientists are encouraged to submit all stable new compounds (even if not intended for medicinal chemistry) to the IOCB Compound Library for biological activity testing.

BIO cluster

The BIO cluster at IOCB broadly includes molecular cell biology, structural biology, and chemical biology. Proteomics, lipidomics, and transcriptomics data must be deposited in the respective databases (PRIDE, etc.) as usually required by the journals, or be archived at IOCB. Other primary data that is usually not included in the extended or supplemental data sections may represent e.g. primary uncropped immunoblots, immunofluorescence (light microscopy) images, enzyme progress curves

used for IC50 measurements, and other primary enzymatic and binding data, etc. For all gels without any cutting, modifications, or alterations should be archived in a raw format (such as RAW). A serious and distinct issue is raw data from cryoEM experiments, which can easily amount to tens of TB from a single high-resolution experiment. The regime of backing up and accessing this data type will most likely have to be different from the rest of the data (e.g. using magnetic tapes).

PHYS cluster

Preserving data from computational works ensures the reproducibility and integrity of scientific results and further research based on them. To ensure reproducibility, researchers should keep at least all input files/data, codes, and the instructions for executing them. These files and information details include initial structures and simulation conditions, databases, and similar information that may directly and substantially influence the obtained results. Preserved information should also include the complete identification of codes/software used (e.g. version, compilation options, parameters); in-house codes or scripts should be kept under an open-source license. The usage of fully annotated protocols that contain all information that ensures procedural reproducibility (e.g. Jupyter notebooks) should be encouraged. Although not always feasible (primarily due to size issues), preserving raw data should also be encouraged, especially when used in published works.