

pISA-tree: standard directory tree as a support for reproducible research

Andrej Blejec, Kristina Gruden

National Institute of Biology, Vecna pot 111, SI-1000 Ljubljana, Slovenia

Corresponding author: Andrej Blejec <andrej.blejec@nib.si>

Basic idea of science is reproducibility of phenomena and experiments. Reproducibility of data analyses and reports is becoming more and more important. It requires structured organization of data, augmented with enough metadata for future re-use of data. Our aim is to provide a system to store data in a way that can be used for small and moderate size projects and fulfill minimal requirements of ISA-tab and FAIR paradigm.

Standard directory trees are applicable to research data storage. The main condition is that information is organized in files and we are not interested in the system that enables access to individual line/record or column/variable in a tabular data structure in the file. The tree structure is generated on the fly by use of batch files (on Windows platform) that generate necessary folders and meta-data template files.

We implemented the system of standard directory trees for support of the research in our research unit. Most often, our research projects can be hierarchically structured into, what can be called, Investigation which is composed of several Studies. Each individual study can have one or more Assays. To reflect this hierarchy, we named such directory tree the pISA-tree. To make new levels we provide three batch files: makeInvestigation, makeStudy, and makeAssay. Special attention is given to Description files that contain meta-information about research, protocols, samples, features et cetera. They are in line with the standards accepted for particular assays (e.g. MIQE, MIAMI, MIRIAM,...) which allow exchange of data with other data management services. In particular, we have in mind FAIRdom platform and some digital notebooks (e.g. SciNote).

pISA-tree structure rely on directories that are readily available on any computer platform and familiar to use by the researchers. Since the translation of meta-data into ISA-tab standard format is not too complex, pISA-tree is a step towards the FAIR paradigm.