### C.1.1 Educational goals of the modules

### C.1.2 Module subject material

We propose to develop two collections of modules, **Improving the Reproducibility of Experimental Data Recording** and **Improving the Reproducibility of Experimental Data Pre-Processing**.

The **Improving the Reproducibility of Experimental Data Recording** collection will include the following modules:

1. The principals of "tidy" data

2. Creating spreadsheet templates for experimental data collection

3. Example spreadsheet template: A template for collecting CPU data for a tuberculosis study [make more specific]

4. Choosing a spreadshet program for reproducible data collection: Excel, Google Sheets, and RStudio

5. Organizing data recording and meta-data recording through RStudio "Projects"

6. Creating "Project Templates" for consistency across projects

7. Example of creating a "Project Template": A project template for a lab group studying tuberculosis [make more specific]

8. Harnessing version control (git and GitLab) to improve transparency in data recording

9. Using git from RStudio

10. Using GitLab for version controlled collaborations

The **Improving the Reproducibility of Experimental Data Pre-Processing** collection will include the following modules:

1. An introduction to R code scripts

2. The benefits of scripts for data pre-processing

3. Getting started with RMarkdown

4. The relationship between R code scripts and RMarkdown documents

5. Creating reproducible data pre-processing protocols using Rmarkdown

6. Example of a reproducible data pre-processing protocol: Automated gating for flow cytometry data

7. Example of a reproducible data pre-processing protocol: Measuring metabolite feature intensities for metabolomics LC/MS data

8. Complex data types in R and their use in Bioconductor packages

9. Converting from complex data types to "tidy" formats for data analysis and visualization with R's "tidy" data tools