Scientific Working

Before diving in into the various work packages that eventually comprise a scientific project, let's briefly think about what *Scientific Working* actually means[1]. Most likely, there is more than one way to do science. However isn't it common to all these ways that you start from something that is known and established and then you want to proceed into an uncharted area to find something new? So, what are good or at least promising approaches to do science?

The hypothesis

Most of the scientific projects are *hypothesis-driven*, right? But what exactly is a *hypothesis*? Wikipedia has a nice take on defining the term 🔗 hypothesis. In brief, it is written that "A hypothesis (plural hypotheses) is a proposed explanation for a phenomenon. For a hypothesis to be a scientific hypothesis, the scientific method requires that one can test it."[2]. So, to take home:

1. you have to observe something that cannot be explained with current 🔗 scientific theories
2. you have to formulate a possible explanation for the phenomenon, i.e. a hypothesis
3. you have to make sure that your hypothesis is actually testable with current methods, only then it is a scientific hypothesis

Hypothesis testing

As a rule of thumb, hypotheses are only then scientific hypotheses if they are falsifiable. In essence, you need to be able to make an observation that cannot be explained by the hypothesis. 🟡 What does it now mean when your attempts failed to falsify/reject your hypothesis? Well, to be honest, not much. This is because you then always have a flank open to the critique that you did not try hard enough[3]. As a take home message, 🟡 it is very hard to support a hypothesis, and you are always better off to think of a hypothesis as something that is most likely wrong, but thus far explains the data best. In scientific practice, we often face the problem to decide between two competing hypotheses. Typically, we then assign one –probably the simpler, probably older and better accepted hypothesis– the term *NULL hypothesis* (H_0), whereas the newer, competing hypothesis is called the alternative hypothesis (H_A). Statistical tests, such as likelihood ratio tests if the two hypotheses are nested, or Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) together with their derivates if they are not, help then to decide whether H_A explains the data 🟡 significantly better than H_0. Only then H_0 is considered rejected. But remember, this does by no means indicate that H_A is true!

Scientific reasoning

Let's start with a simple statement: Science is arguing, because nobody knows the truth. Arguing requires...right... **Arguments**. But what are these arguments? In the first approximation, they are

statements that are backed up by ⚠ either experimental data, or ⚠ a reference to a study where somebody else collected the data in support of the statement. It might be easy to think of scientific reasoning as a building that you create. A solid foundation is essential, but a house of cards built on concrete foundations will also easily collapse, if you get the line. To cut a long story short, your reasoning is exactly as solid and stable as it is the weakest supported statement that you recruit. So, stay away from *hearsay*, *anecdotal evidences*, or *handwaving argumentation*.

Scientific documentation should be FAIR

Scientific projects have to be [documented](#) very carefully. The requirements are easily specified: ⚠

You need to document to an extent that any third person is capable of understanding ⚠ why you have done ⚠ what and ⚠ how, what the ⚠ results are. You also have to make clear to what extent your ⚠ conclusions are supported by the data. When it comes to your data, make sure to follow the [FAIR principles](#). This means that your data has to be

1. **F**indable
2. **A**ccessible
3. **I**nteroperable
4. **R**eusable

In particular, if the **F** and the **R** of the [FAIR principles](#) are not met, then the data generated in your project is useless for follow up analyses. This is a huge draw back, because the costs of the data masses that are produced nowadays–see for example the [Earth Biogenome Project](#)–are only justified if data reuse is possible and encouraged.

Keep your workspace clean

Throughout this course, we will use numerous different programs with different datasets and different parameters. ⚠ Already after a few days, this will result in a large number of files with the potential for a lot of confusion. To avoid long searches and possible mistakes by using wrong or outdated data, we highly recommend to develop a systematic scheme for naming your files and your directories. In the following, there are few points you may want to consider:

- Generate a new directory for each program, or alternatively for each step in the analysis workflow.
- If a program is used multiple times, name the output files in a way that they unambiguously show from which run they stem. This can be done, for example, by including the input files and/or parameters in the name.
- For each script you write yourself, make the script name a short description of its function. Moreover, right from the start, comment your scripts exhaustively. This takes a bit of time while writing, but it helps you to understand what your script is doing even when the course is long

over.

- **Avoid the use of whitespaces** and language specific characters such as "ä", "ö", "ü" in both file and folder names, as this can cause problems when working with the linux terminal.
- Keep input data separate from results, as you may do different analyses with the same input data. If you insist in having the input together with the results, consider the use of soft links. These are pointers to a file or a directory that can be placed anywhere in the file system without the need to duplicate the often large input files.

- [Back to MBW course](#)
- [Back to PBioC course](#)
- [Back to BEH course](#)
- [Back to DigiKomp course](#)

[1)](#)

rest assure, that books about proper ways of doing science fill large shelves. The few and crude sentences here are just meant to trigger your attention

[2)](#)

https://en.wikipedia.org/wiki/Hypothesis - accessed on Oct. 13 2021

[3)](#)

in statistical terms, the power of your test was not sufficient

From:
https://applbio.biologie.uni-frankfurt.de/teaching/wiki/ - **Teaching**

Permanent link:
https://applbio.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=general:scientificworking

Last update: **2025/10/15 11:23**