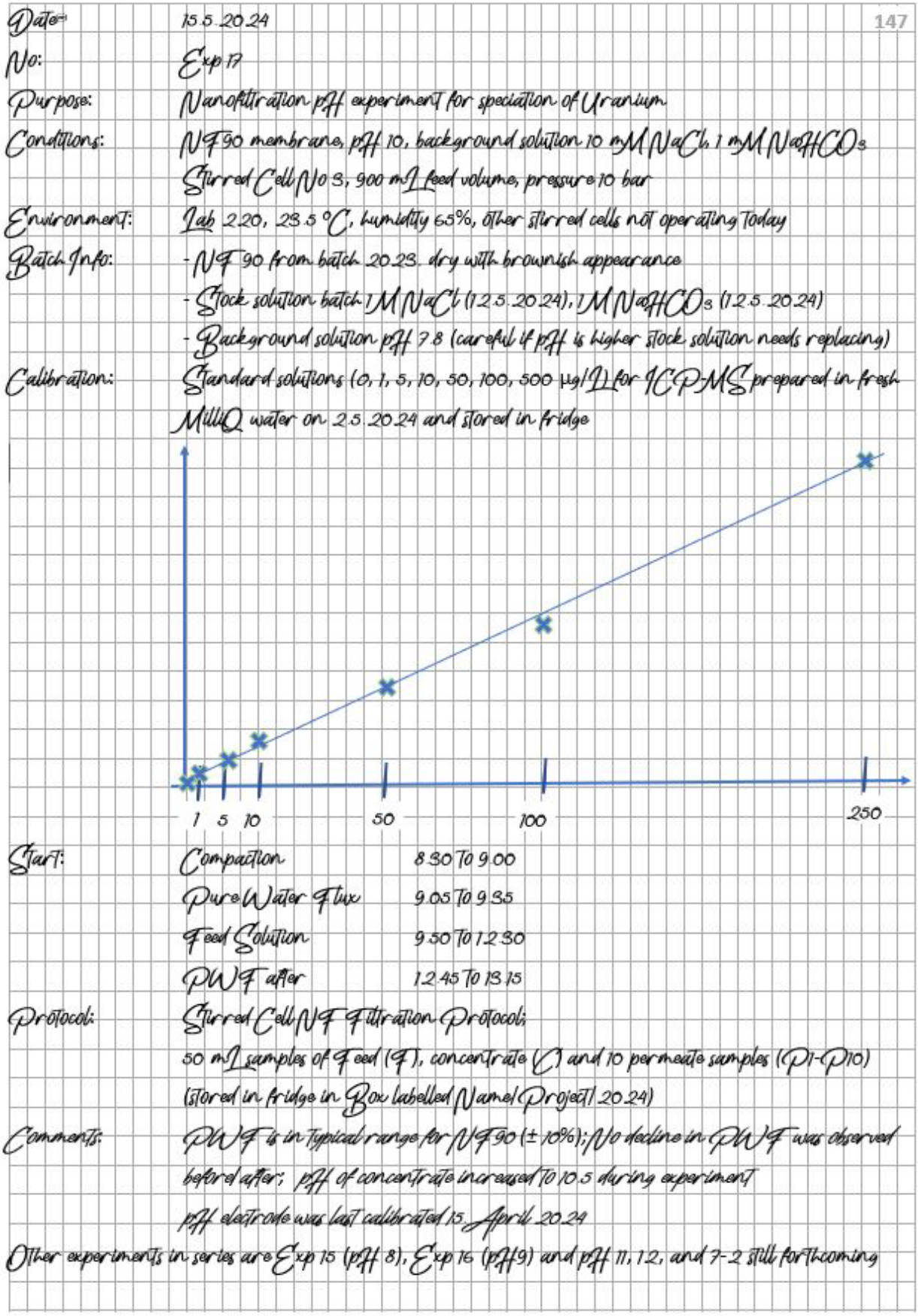
RSC 11: LAB BOOKS, ORGANIZING EXPERIMENTS AND DATA MANAGEMENT

Research data must be stored for at least 10 years and this can be a challenge when researchers are inherently mobile. Lab books are a record of research carried out in a laboratory and access needs to be restricted to the person in charge of conducting this research, it is a legal document. The entries are permanent and need to contain a certain type of information, as usual in the RSCs this is written for the field of water process engineering, at the example of membrane materials and processes. The nature of a lab book is that it has numbered pages and no page can be removed, while numbered pages allow cross-referencing. At IAMT lab books must be kept in the English language and the person conducting a particular experiment records it in their book. The entry must be made with a permanent pen, neither water soluble pens (not compatible with water search!), nor pencils (that can easily be erased) can be used. The book contains facts and observations that help with the data that is typically registered electronically (e.g. Labview and data acquisition) and the research environment (e.g. temperature, humidity, other observations such as large equipment used which may affect electric circuits or works carried out that may release contamination). It is the diary of a researcher that describes what is done in the lab. This may include also preparation and standards with the masses weighing in, this allows mistakes to be traced from when the new standard was used. In some labs a signature is required, at a university laboratory, supervisors normally assume that a researcher knows how to use/keep a lab book, which is often not the case. So, what is a lab book really?

- **PERSONAL LAB BOOKS**: At IAMT every researcher has their own personal lab book that they use for their research. This means that if different researchers collaborate on a project, entries from several lab books may contain some of the experiments (it is a good idea to cross-reference each other such that this is obvious). This book is provided and is owned by IAMT, at the end of the research stay at IAMT the book is given to the director for safe storage. The legal requirement for research data storage is 10 years. The reason is that research must be real and true, while data fabrication, modification and misinterpretation can cause a lot of damage. The responsibility for data integrity lies with every individual lab user, not the supervisor. Usually, the keeper of the lab book will be able to take a copy of their book when leaving the lab, but permission should be sought when lab books are handed over before departure.
- **EXPERIMENT ID**: It is important to have a clear experiment ID and the simple is best: name your experiments with a number from 1 to ∞. It is important that you also give a number to aborted experiments. This is so effective because it is the easiest code to have a chronological order and a clear identifier that you can then add to any lab book entry, data file or sample label. If an experiment is repeated, this gets a new number, but one can indicate that experiment Z is a repeat of experiment Y.
- **SAMPLE LABELS**: We are doing water research and water, combined with vials rubbing against each other when transported, can dissolve even permanent markers. We wash sample vials and this means stickers are a pain. It is a good idea to choose a permanent market that works and then label the vial and lid. This can be a lot of work unless the label system is a clear code. Usually, it is sufficient to put the experiment number and then a clear identifier, such as F for feed, P for permeate and C for concentrate, maybe followed by a number if there is more than one sample each per experiment. Again, a short but clearly identifiable descriptor can then be used in the analytical instrument and this reduces the risk of errors.
- **SAMPLE STORAGE:** At IAMT we store samples in Euro boxes owned by users in our walk-in fridge until manuscripts are published. This enables re-analysis of samples if problems pop up in validation. It is much easier to re-analyse samples than re-run experiments, although this may also be required at times. Once samples are discarded, we wash a lot of the vials with a very good protocol suitable for trace organic work. A small contribution to a circuilar economy and it helps the budget as well as the effort of ordering.
- **INFORMATION IN A LAB BOOK**: What should be entered in a lab book? As shown in the example below, the essentials are i) date, ii) experiment number, iii) purpose of the experiment (and if a series (e.g. pH 2-12 it is useful that other experiments that belong to the series), iv) experiment conditions (e.g. membrane type, solution chemistry (composition, conductivity, pH), pressure, flowrate), v) environmental conditions (e.g. temperature, humidity, other lab users, other equipment operating), vi) batch information (e.g. particular batch of membranes, stock solutions, standards), vii) calibration information (results of standard analysis, was equipment recalibrated recently? Note that recalibration

can introduce step change artefacts in your results!), viii) start and finish time of the experiments, ix) any odd things such as a sample vial being spilt or mistakes being made, x) key data (most data are usually recorded through data acquisition but a curious researcher keeps an eye as the experiment proceeds): are the results as expected, is everything running as it should or did the experiment have to be aborted. Examples of key data are pure water flux for example, that must be in a certain range – or data variation as a function of the key parameter investigated (the purpose of the experiment).



In addition, one may make reference to key literature that inspired the experiment or characterization data (which sample or membrane was sent for additional characterization elsewhere and where can these results be found?). Remember, the lab book is a record, of the data and what it means. It is a map to retrace the data and where it can be found. Usually, full data analysis happens later for publication or for thesis preparation, most of us will not remember details of the day and often a problem occurs and it helps to identify this problem when a super detailed record exists. Examples may be i) a lab member using a perfume that interferes with trace organic analysis, ii) power-intensive equipment causing spikes in the electric system that interfere with an analytical instrument or iii) a mistake in the stock solution or standard preparation causing all sorts of errors. Never very pleasant, but it is always really good if the actual problem can be identified. Naturally, the lab book entry depends on the nature of the research. If one is taking samples in the environment, the number still applies (1 to ∞) but the lab book entry may include GPS data, reference to photographs, volume and vast descriptions of weather, environment and information from people (e.g. water users in a community).

- **OTHER TYPES OF LAB BOOKS**: Instrument books are another type of 'lab book' at IAMT. The numbering of pages is not required as these instrument books are used to enter the users, the duration of usage, the type of samples and any problems or maintenance tasks (including calibrations). The purpose is to keep a record of usage, costs (per time and per sample), problems and any changes that may have contributed to altering results. If one runs an experimental series and someone in between changes the calibration (for tools where you do not run your own standards) this will affect results and may make them meaningless. Documented problems (and their solutions) will help future team members to reach the solution more quickly. Some samples risk contaminating an instrument and it is important to record what samples were analysed such that this can be dealt with, or if need be, prevented. Filling instrument books diligently leaves your legacy as a responsible team member and helps future researchers, especially when no technicians are available to take care of everything.
- **ELECTRONIC LAB BOOKS**: Electronic lab books are a relatively recent invention. While they appear to be popular in the same disciplines, at IAMT I have decided that they are not appropriate. Making sure that software is available through multiple updates for a decade and protected from every increasing hacking attacks or sabotage is beyond my skill. In addition, the control of lab books is a further step removed from having a quick look at someone's hard copy in the lab. Physical lab books disappearing or being 'lost' is virtually impossible (without considerable ill will) and I have decided that if we manage to keep good lab books we are doing well. Naturally, most of our data is electronic and managing this data well, downloading it from lab computers to personal research data files and carrying our regular backups is a sufficient challenge to manage well in a team.
- **DATA BACKUPS**: Having witnessed PhDs losing all data and writing weeks before completion is the most unimaginable heartbreak. This is real, no one is awarded a PhD for lost data, it is a huge waste of resources (IAMT research is extremely expensive) and hence safeguards need to be put into place. At IAMT we operate a server and request a monthly data back-up to the personal 'Research Data' folder of a team member. This data can be accessed only by the team member or myself, if others need access this can be arranged as shared access (for supervisors) or data transfer (if someone else writes up the work later). The transfer is carried out to prevent data from being tampered with by anyone other than the person who secured the data and this allows for tracking mistakes later, as the original dataset was protected. Maybe more common in some countries than others: regular power cuts. Check which instruments are vulnerable to data loss if this occurs, which instruments restart automatically and how you can make your experiments more resilient such that you do not need to repeat everything. Uninterrupted power supplies (UPS) are great assets for such cases and some instruments are damaged when power breaks suddenly. A small investment of time and money may save a lot of time and money later.
- **DATA REPOSITORIES**: The point of data repositories is to store data and to make it available to others. This is a noble intention, yet one ought to be careful. In water process engineering such data would require extensive documentation (metadata) to be useful, in the absence of all required documentation the risk is a misinterpretation of data. Documenting data well is a task that requires a lot of time and verification, which is probably not what most research teams can manage. At IAMT this has resulted in an approach of 'reflected compromise' with a priority of making sure data is well documented in manual lab books, and personal research data folders on the intranet (not accessible to anyone else) and prepared such that the validity of this data can be verified. A core challenge has been over the years that

graphs are prepared for publications, while supervisor(s) may never see the original data and this may not be able to identify challenges. As a result, it is now common practice at IAMT that all original data (pressure, flux, temperature, concentration, pH, etc.) from experiments are prepared and added to the supporting information (SI) and presented as a function of time (or volume).

- **DATA IN MANUSCRIPT SUPPORTING INFORMATION**: The data in the supporting information serves as a public repository. Given that the documentation is extensive in the manuscript, this approach was adapted by IAMT to make the data accessible, even though this is not electronic or directly usable. As indicated above there are two purposes to doing so, being i) the data being visible to supervisors in manuscript preparation (which often enables the identification of unstable experiments and the need for repeats or other systematic issues such as feed concentrations being out of range) and ii) data being available to other researchers. Other data reported in the SI are calibration data and detection limits, compared with the concentration range during the experiments.
- **DATA AVAILABLE ON REQUEST**: At IAMT the publishing option 'data available on request' is chosen. Enquiries are not frequent and will be dealt with if there is a clear purpose, as again, making data available is a tremendous effort and with some communication of purpose this can be better targeted. Ideally, all that is required is the electronic file of data in the manuscript and the SI, making it possible to direct requests to the SI. Careful documentation in the manuscript is thus a very good idea, especially seeing that such requests often come years after manuscripts are published, the main researchers have long moved on and no one can really remember details.
- **DATA MANAGEMENT PLANS**: These are required for funding applications and typically range from bizarre standard paragraphs to genuine text. As with most things, an understanding of purpose and a good level of common sense can reduce meaningless efforts for the sake of compliance. The 'Horizon Europe data management plan template' highlights some of the questions as well as the required effort to document data for a public repository. It is worthwhile to consider which data warrants this effort, that may be better invested in data validation and ensuring reproducibility.

References

1. DFG Guidelines on the Handling of Research Data, https://www.dfg.de/resource/blob/172098/b08fcad16f1ff5ddca967f1ebde3a8c3/guidelines-research-data-data.pdf (accessed 16/12/2024)
2. Horizon Europe data management plan template, https://cohortcoordinationboard.eu/toolkit/horizon-europe-data-management-plan-template/XX (accessed 16/12/2024)