Lessons learnt

Overall strategy

- Very clear and quantifiable **objectives** allow the development of an effective sampling plan, including the selection of appropriate measurement techniques and the up-front definition of criteria for the measurements (e.g. detection limits, uncertainties). On the contrary, the absence of explicit objectives complicates all aspects of designing and implementing the characterization strategy and the planning down to the selection of appropriate measurement techniques, determination of minimal detection limits, etc. This is illustrated in UC1. If feasible, data sets containing large amounts of below **detection limit** data should be avoided and whenever possible tackled during the strategy development (e.g. UC2). However, this is not always possible; for example due to low threshold values, the limitations of measurement techniques, but also due to unclear initial objectives or potentially changing objectives or thresholds during the characterisation process. Depending on the case, it might be very wise to use advanced statistical methods for dealing with samples below detection limit (Kim, et al., 2020). In UC1 and UC3a, data analysis methods were tested with and without uncertainties and data below detection limit were applied. Especially UC3a shows that the impact on certain parameter estimates can be significant.
- Performing a radiological characterization program in two or more **stages/phases** as implemented in UC2 can be efficient and effective to tackle areas with higher uncertainties. Unfortunately, this might not always possible due to planning constraints. In addition, historical data, sometimes gathered with other objectives (e.g. analysis liner samples in UC2) could optimize the characterization process.
- In a first attempt to organize and select methods for sampling design and data analysis, we produced streamlined flowcharts including yes/no answer questions and decision trees. However, this approach was abandoned rather soon and replaced by the organization of a non-limited list of methods in Venn diagrams, guiding the end-user to the selection of appropriate methods. Indeed, there is not a unique and absolute solution. This **flexibility for the selection of suitable methods** is vital, since the same problem can be tackled using various methods as mainly demonstrated in UC1, UC3a and UC3b.
- In UC1, the mean value of the activity for several nuclides was calculated using various **robust methods**. The methods included empirical mean, Wilks median, bootstrap estimation of mean, Bayesian estimation and mean estimated from a fitted theoretical distribution law. The Wilks method provided the most conservative estimates, particularly so for data sets with outliers.
- Decommissioning is a multi-disciplinary operation and the **involvement** of specialized staff performing the next stages of the decommissioning project has proved to be beneficial in UC2. Technical feasibilities/constraints in the next stages might strongly influence the initial characterization program. Effective communication and common understanding is essential. Extensive compartmentalizing might result in misinterpretation and wrong decisions.

Sampling design approach

- In case of tanks filled with liquids as in UC1, there is always an amount of sludge present. **Representative sampling** is obviously a challenge in this case. In theory, one should sample the population as a fully homogeneous entity. According to the historical information, this might have been more or less the case for the historical sampling campaign of UC1. An alternative is to apply stratified or targeted sampling of solid and liquid fraction after settling. The latter was however not feasible due to accessibility constraints. Instead, additional samples were taken during the INSIDER project after trying to homogenize the contents of the tanks for a few hours. Results from sample analysis showed a large level of heterogeneity, resulting in the cancellation of the INSIDER benchmarking exercise.
- The sampling design approach is not uniquely defined in all of the real use cases. Often, a combination of approaches is being implemented. In any case, the **sample locations** should be selected so that subsequent extrapolation during data analysis is avoided. As shown in UC2 (stage 2) and UC3a, this does not only concern the expected activity concentrations (lowest and highest), but also the physical location. It is obvious that in a certain stage in the dismantling process, it might be difficult or impossible to access locations with expected extreme values for performing in-situ measurements or for taking samples. In this case, it is necessary to foresee

the measurements at a later stage in the project and to allow for updating the existing data analysis and post processing. Just ignoring the information might result in unacceptable uncertainties. Many characterisation projects have the tendency to focus their sampling efforts on the highest affected areas, neglecting areas with lower activity concentration levels. Nonetheless, it is necessary to sample the supposedly least impacted zones as well as the most impacted zones to achieve a realistic understanding of the statistical distribution of the activity concentration. Confirming some non-impacted areas is often as important as (or even more important than) confirming historically impacted areas. From the point of view of waste volume management, transition zones are more critical, since it is difficult to categorize them with respect to the reference thresholds. Uncertainty being the most important in these areas for proper delineation (and limiting misclassification errors), the sampling distribution should favour them over other areas that only require confirmation of impacted or non-impacted.

- Defining the **sample size** or density often results in lively discussions, due to the impact on planning and overall costs. The STRATEGIST web tool provides for example a computation of the minimal sample size to estimate a quantile (or a tolerance interval) with a given confidence level according to Wilks formula. However, implementing it in day-to-day practice might be challenging when several physical parameters need to be assessed (i.e. total activity, activity concentration, thresholds) based on various data sources that might not always be representative. Moreover, a data set might contain considerable amounts of values below detection limit and confidence levels required might not always be unequivocally defined. Sensitivity analysis on the UC2 data showed that reducing the size of the primary dataset with about 75%, could result in extreme under/over estimation of the volume exceeding the threshold. The sensitivity analysis also shows that this deviation can be strongly reduced by combining the limited higher quality and costly primary dataset (in-lab sample measurements) with a large cheap secondary data set (in-situ measurements). Also in the UC3a case the effect of sample reduction was examined. Due to the already limited amount of cores, the total number of samples was decreased by reducing the number of samples in each core in a systematic way, attempting to keep spatial sample distribution. Sample reduction generally leads to an increasing uncertainty and estimated remediation volume. The effect is still limited for a 50% sample reduction but can strongly increase when sample density is further reduced. Also in this case, the additional uncertainty can be mitigated by integrating secondary data from gamma scanning.

Data analysis methods

- The STRATEGIST web tool foresaw some additional quality assurance/control checks on the existing data during the pre-processing step (e.g. errors, outliers, sample representativeness). However, the first version did not mention the use of **validation techniques for assessing the results** obtained. Both UC2 and UC3a show methods to assess the reliability of the data analysis process.
- There is no unique solution to the data analysis method selection problem and in many cases various methods might be combined. In the case relevant physical processes can be simulated, the use of a sound process model may serve as additional data or trend model as shown for activation in UC3b.

Uncertainties

Parameters influencing the uncertainty might depend on the physical quantity to be estimated (e.g. volume, activity concentration, total activity) and the required confidence level, but might as well strongly depend on objectives and sampling design. Some parameters can be more important than others. The return on experience from the three use cases regarding uncertainties is summarised as follows:

- The impact of the **sampling design** is most fundamental (see section on sampling design).
- Using advanced statistical methods for dealing with **samples below detection limit** can have an important (positive) impact on estimated values and should be considered in case the data set contains a considerable amount of results below detection limit (see section on overall strategy). Obviously, the potential impact depends on the objective (estimated quantity).

- Apart from the general aspects mentioned above, we also observe effects which could be occasionally generalised as well, but in other instances might be specifically related to the sampling design and data analysis implemented in the following use cases:
  - UC1: In the case of **small data sets**, the presence of **outliers** could clearly increase the uncertainty. Additional verification (process driven or error) might be necessary when estimations are becoming close to a decision threshold.
  - UC2: Sensitivity analysis shows that the most important uncertainties for the physical parameters estimated (e.g. volume categorisation, total activity), are due to **spatial uncertainty** (geostatistical simulations) and **heteroscedasticity**. Heteroscedasticity occurs more often in datasets that have a large range between the largest and smallest observed values, which is typically the case for radiological data. The contribution to the global uncertainty, resulting from measurement uncertainties on primary and secondary data, the relation between secondary and primary data, and the trend model are relatively limited. Effects of heteroscedasticity on the uncertainty could be reduced using rescaling techniques or, in case sufficient data is available, sub-dividing the data in groups and applying data analysis on each group separately.
  - UC3a: Sensitivity analysis shows that the **impact of sample measurement uncertainty** on the volume categorisation is **quite limited** (as it is in the case of UC2). The analysis shows as well that the effect on the estimated quantity (e.g. total activity or threshold) is not always the same as on the uncertainty. In addition, the integration of non-destructive measurements in a multivariate approach significantly reduces uncertainties when sampling is reduced.
  - UC3b: The variance decomposition was fully possible and highlights that most of the uncertainty budget comes from the spatial distribution of values in comparison to the uncertainty related to sample duplicates and even more to measurement replicates. For one third of the radionuclides (Co-60, Ba-133, Eu-154 and Pu-239 + Pu-240), estimations without measurement uncertainty treatment are up to 95% less than estimations with measurement uncertainty treatment and for one third of the radionuclides (C-14, Cl-36, Eu-155 and Ni-63), estimations with measurement uncertainty treatment are up to 28% less than estimations without. That shows the impact of uncertainty is not negligible on final activity estimation in particular in the case of small datasets (less than 20 measurement data for 75% of the radionuclides).