PET Decision Tree – Checklist

Version: 05 October 2021

Before starting your path through the decision tree, we advise you to look at this checklist to ensure that you have the adequate information to be able to answer the questions in the decision tree. We also advise you to read the Guide.

More elaborate documentation of the decision tree can be found in the Guide.

Please provide as much detail as possible and take care to describe multiple answers / situations / scenarios. Considering scenario's that are not realistic (at first thought) can still help in clarifying the challenge at hand!

- **What challenge are seeking to tackle?** What is currently seen as an unsolved problem, where data from multiple parties needs to be combined?
- In an ideal world (from your perspective), please **describe various solutions** to the challenge.
    - **Also** consider solutions that seem **infeasible** (e.g. legal or practical constraints) at this stage; these solutions can pinpoint important considerations or serve as inspiration for feasible solutions.
    - **Don't limit** yourself to PET-based solutions. A solution could also be: organisation X performs computation Y and sends my organization the final result.
    - For every solution, please **indicate why** you think that the solution may or may not be feasible (e.g.: GDPR prevents me to share microdata, may have legal ground to see aggregated result, …).
    - **Be specific!** Visualize the real-world problem that you encounter, walk through the steps that you would like to take, the insights that you need to have, the data that is required, the organizations or persons that are involved.
    - Be critical and **think outside the have-all-data box**: do you really need to know all cash flows from a person, or do you just want to know whether their net yearly income is above some threshold?
- All data providing parties should have a **list of information elements** they have available. Preferably using FAIR descriptions:
    - Findable: know what elements are available (and you can search them)
    - Accessible: know how all information elements can be accessed internally
    - Interoperable: know in what format (syntax) the information is, and in which terminologies (semantics) are used
    - Reusable: know what the license connected to this data is, and who is (legally) allowed to access the data.
- Based on the previous step, identify **which data elements are used**, and which parties do have these data elements?
    - **Again: be as specific** as possible! 'Social or bank data' is much too wide a scope – 'decade of birth or annual cash inflow' is already much better.
- How **sensitive** are the **information elements** (covered partly in Reusable), and how sensitive is the **result of the analysis**?
    - Is the shared analysis allowed from **ethical, legal and societal viewpoints** (seen from multiple stakeholders; both beneficiaries and subjects)?

- Keep in mind that the term information elements is used here to denote not only raw data, but any other type of element that is part of the intended analysis (e.g. the resulting model in the case of Machine Learning).

❖ Which **legal grounds** may justify or prohibit the data exchange, and what *kind* of data exchange are we talking about (directly vs. indirectly identifiable data, or encryption on identifiable data)?
- Some of the privacy-preserving techniques considered allow for computations on encrypted data. That is, a special form of encryption (or similar forms) can be used such that data is encrypted, sent to another organization and then meaningfully processed by the other party *without decrypting the data*. As such, depending on legal-ethical supports, it can be important to distinguish raw data access, encrypted data exchange (i.e. transfer of encrypted data) and access to decrypted *or still-encrypted* data at another location.

❖ Are there any **other regulations** (e.g. compliancy) that may be relevant in the exchange of the data?

❖ **Who can have access to which data**, within and outside of the organization and how does this access materialize? *Consider this question from every stakeholder's perspective; some data can be processed by one organisation but not by another.*
- Is it allowed that a person from outside your organization triggers the execution of an analysis on your internal data?