

Data lifecycle questionnaire

Thoughts for data-centric IT projects, with a special focus on (1) Blockchain criteria and (2) requirements for trustworthy, reproducible, or traceable IT processes.

Data

- What is size, format, and structure of the data in question?
- What are the dependencies between fields (within and among) data records?
- For which parts or aspects do you require (1) integrity, (2) traceability, (3) immutability, (4) confidentiality, (5) availability, or (6) global consensus?
- For which of them should the current state be public?
- For which of them should their history be public (traceability)?
- Are there further records, fields, or relations requiring special treatment?

Input

- What are the data sources?
- Where do the data come from?
- How do they enter the system?
- What about existing (legacy) data?
- How is quality assured?
(correctness, uniformity, consistency, authenticity, completeness, ...)
- What are the licenses and other conditions attached to the data?

Processing

- Which processing steps will the data be processed through while in the system?
- How is correctness of these steps and results asserted? Can this be performed automatically? Always?

Output

- What should happen to these data afterward?
- Which actions should be triggered by the results from these data and their processing? Automatically?
- How is the execution of these resulting actions verified and/or enforced?

Cooperation

- Which groups may read, write, perform administrative operations on the data? Who can change program code?
- Which risks might arise from authorized people abusing their access permissions (or, by unauthorized programs or third parties abusing the permissions of authorized people)?
- Is prolonged or persistent mistrust to be expected between actors? Can this mistrust be managed through hierarchies, assignment of responsibilities or (work) contracts?

Futureproofing

- Will format, structure, or dependencies between the data eventually change?
- When parts of the data itself changes: Should historic data continue to refer to the historic values referenced, or should they be updated (i.e., what should happen to recorded names/addresses on name/address change)?
- How should these changes be implemented (for both legacy and future data)?
- Do data have an expiry/deletion date? Is that date determined statically or dynamically? What should happen then?
- What should be done to wrong entries (whether accidentally or malevolent)?
Which effect does this have on traceability?
- How is the right for correction and deletion of personal data to be implemented?
- Even if no personal identifiable data is processed: How will court orders for correction/deletion be implemented?