

Data Collection Process Maturity Assessment

Please select exactly one answer per question. Select the option that best describes your current situation today, not the situation you are working toward.

Dimension 1 — Data Collection Process

Definition*: The processes relating to the capture of data from multiple sources for secondary purposes. “Secondary purposes” means any use beyond the reason the data was originally recorded, such as research, policymaking, statistics, innovation, or regulatory activities. It also covers how well you know the quality of what you collect.

1. How defined are your processes for collecting data that can be reused for secondary purposes?*

- ☐ We have no defined process for collecting or capturing data for secondary purposes. 1 pt.
- ☐ We have an informal, ad hoc process. It works in practice, but it is not written down. 2 pts.
- ☐ We have a written process for some of our data collections. 3 pts.
- ☐ We have a written process for every data collection we make available for secondary use. 4 pts.
- ☐ Our collection processes are automated end-to-end, and we review and improve them on a fixed cycle. 5 pts.

You can support this answer with: your process description, work instructions, the register of data collections.

2. How much of your data is covered by a written collection process?*

- ☐ None of it. 1 pt.
- ☐ One or two collections, described informally. 2 pts.
- ☐ Our main collections, roughly half of what we hold. 3 pts.
- ☐ All collections we produce ourselves. 4 pts.
- ☐ All collections, including data we receive from external partners. 5 pts.

You can support this answer with: your data inventory, showing which entries have a documented process.

3. Is the data you collect actually being used for secondary purposes today?*

- ☐ No. It has never been used beyond its original purpose. 1 pt.
- ☐ Once or twice, as a one-off project. 2 pts.
- ☐ Some of it is used regularly, for example, internal research or quality reporting. 3 pts.
- ☐ Most of it is used regularly, including by parties outside our organization. 4 pts.
- ☐ It is routinely used by external researchers or authorities through a documented route. 5 pts.

You can support this answer with: your list of data releases or research projects over the last 24 months.

4. How automated is your data collection?*

- ☐ Fully manual, with data exported, copied, and retyped by hand. 1 pt.
- ☐ Mostly manual, with a few helper scripts. 2 pts.
- ☐ Partly automated, with recurring extracts running on a schedule. 3 pts.
- ☐ Largely automated, with monitoring and error alerts. 4 pts.
- ☐ Fully automated and monitored, with failures logged and followed up. 5 pts.

You can support this answer with: your job scheduler configuration, pipeline documentation, error logs.

5. Do you measure and improve the quality of the data you collect?*

- ☐ We do not look at data quality. 1 pt.
- ☐ We notice quality problems when somebody complains about them. 2 pts.
- ☐ We check completeness and plausibility when something changes, and correct what we find. 3 pts.
- ☐ We measure defined quality indicators, such as completeness, plausibility, and timeliness, on a fixed schedule and record the results. 4 pts.
- ☐ Quality indicators are measured continuously, reported, and drive documented improvements to the collection process. 5 pts.

You can support this answer with: your data quality report, the indicators you track, your improvement backlog.

YOUR NOTES FOR THIS DIMENSION

What sits behind these answers: the practices and tools you rely on, recent improvements, and the gaps that explain the lower scores. This is the material for the supporting narrative on the Overall Score sheet.

DIMENSION SCORE

Add up the points of all 5 answers and divide by 5. Keep one decimal.

YOUR MATURITY LEVEL

Round the dimension score down to the nearest whole number.

Level 1 Initial Starting out · 1 Point	Level 2 Emerging Written down · 2 Points	Level 3 Defined In practice · 3 Points	Level 4 Managed Measured · 4 Points	Level 5 Optimized Improving · 5 Points
--	--	--	---	--

The five maturity levels in this dimension

Your five answers average to one level. Each row describes what that level typically looks like across all five questions, so you may be ahead of it on some points and behind on others.

Level 1 Initial	No defined process; collection is fully manual; data is never reused; quality is not looked at.
Level 2 Emerging	An informal routine, not written down; a few scripts; one-off reuse; problems noticed late.
Level 3 Defined	Written process and scheduled extracts for main collections; regular internal reuse; quality checked.
Level 4 Managed	Every own collection documented and largely automated; external reuse; quality indicators measured.
Level 5 Optimized	All collections automated and monitored; routine external reuse; quality drives ongoing improvement.

Your next step: find the row for the level you reached, then read the row directly below it. That row is what you need to achieve to move up one level in this dimension.

Note: Every question in this dimension uses the same 1–5 scale, and the same real-world situation is worth the same number of points in every question, so both the lowest and the highest level are reachable. Transfer your maturity level for this dimension to the *Maturity Overall Score Calculation* sheet, which brings all ten dimensions together.

References*

Question 1 follows the maturity levels of this dimension in the QUANTUM data holders' maturity model (Deliverable 1.2 v3.5); the wording is EHDS4ALL's own. Underlying source given there for this dimension: ISO 8000-62 data maturity model. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 2024. Question 5 assesses data quality, identified as a necessary part of the model in the EHDS4ALL workshop of 19 August 2024. The five level statements below combine all five questions, because the dimension score is the average of all five answers.

Jones, M., & Bernal-Delgado, E. (2025). **Deliverable 1.2 — Specification for the assessment of data holders' maturity** (Version 3.5.0). QUANTUM project. Zenodo. doi: 10.5281/zenodo.14944767

International Organization for Standardization. (2018). Data quality — Part 62: Data quality management: Organizational process maturity.

EHDS4ALL (2024). Protocol, Workshop Kick-off WP3, Cologne, 19 August 2024. Work Package 3, dimension validation and maturity levels.

EHDS4ALL (2024). Expert interviews with health data holders and maturity model developers, October 2024. Conducted by inav and the University of Cologne; participants anonymized.