

Data Management — Infrastructure 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 3 — Data Management — Infrastructure

Definition*: The level of implementation and development of your data management infrastructure: the systems and pipelines that store, check, process, and protect the data, including the checks that keep it complete, valid, and plausible.

1. What infrastructure do you use to manage data for secondary use?*

- | | |
|---|--------|
| ☐ No dedicated infrastructure. Data stays in the source systems and is extracted by hand. | 1 pt. |
| ☐ An emerging infrastructure, with some checks on data going in and out. | 2 pts. |
| ☐ A defined infrastructure that is used as our standard way of working. | 3 pts. |
| ☐ An automated and validated infrastructure for the main flows, supporting near real-time data management. | 4 pts. |
| ☐ A dependable infrastructure that manages data in real time, fully automated, with every step verified and validated. | 5 pts. |

You can support this answer with: your architecture diagram, system inventory, pipeline documentation.

2. How do you check the quality of the data in your systems?*

- | | |
|--|--------|
| ☐ There are no quality checks. | 1 pt. |
| ☐ Manual spot checks, when somebody has time. | 2 pts. |
| ☐ Documented checks on our main data flows, such as completeness, valid values, and duplicates. | 3 pts. |
| ☐ Automated checks on all flows, with errors logged and corrected. | 4 pts. |
| ☐ Automated checks on all flows, with quality results reported and independently confirmed. | 5 pts. |

You can support this answer with: your validation rules, data quality reports, error logs.

3. How automated is your data processing?*

- | | |
|--|--------|
| ☐ Not automated, with every step done by hand. | 1 pt. |
| ☐ A few scripts, run manually. | 2 pts. |
| ☐ Partly automated, with the main pipelines running on a schedule. | 3 pts. |
| ☐ Largely automated, with monitoring. | 4 pts. |
| ☐ Fully automated end-to-end, with alerting and automatic recovery. | 5 pts. |

You can support this answer with: pipeline configuration, scheduler, monitoring dashboard.

4. How current is the data you can supply?*

- | | |
|--|--------|
| ☐ We cannot supply a refreshed extract. What exists is whatever was exported at the time. | 1 pt. |
| ☐ On request only, with a lead time of several weeks. | 2 pts. |
| ☐ On a fixed refresh cycle, for example, monthly or quarterly. | 3 pts. |
| ☐ Frequent refresh, daily or weekly, for most data. | 4 pts. |
| ☐ Continuous or near-real-time availability. | 5 pts. |

You can support this answer with: refresh schedule, timestamps on the last delivered extracts.

5. How is the infrastructure secured?*

- | | |
|--|--------|
| ☐ No specific measures beyond general IT security. | 1 pt. |
| ☐ Basic access control and backups. | 2 pts. |
| ☐ Documented measures: role-based access, encryption, access logging. | 3 pts. |
| ☐ The above, tested regularly, for example, penetration tests and recovery tests. | 4 pts. |
| ☐ The above, plus an information security management system that is externally audited or certified (for example, ISO/IEC 27001). | 5 pts. |

You can support this answer with: your security concept, access matrix, last penetration test, or certificate.

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 dedicated infrastructure; manual extracts; no quality checks and no specific security.
Level 2 Emerging	An emerging setup with spot checks, a few manual scripts, basic access control, and backups.
Level 3 Defined	A defined standard infrastructure; documented checks and security; scheduled refresh.
Level 4 Managed	Automated, monitored pipelines; checks on all flows; frequent refresh; security tested.
Level 5 Optimized	Fully automated with alerting, quality independently confirmed, real time, certified security.

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: European Health Data Space / FAIR. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 2024. Question 2 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

European Commission. (2022). Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM(2022) 197 final. Adopted as Regulation (EU) 2025/327.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

International Organization for Standardization & International Electrotechnical Commission. (2022). Information security, cybersecurity and privacy protection — Information security management systems — Requirements (ISO/IEC Standard No. 27001:2022).

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.