

## About this Guide

The EHDS Guide supports you in implementing the European Health Data Space. It helps you understand the necessary steps, manage your data transparently and securely, and assess your digital maturity in handling health data.

## — Who is this for?

Whether you want to provide data, use it, or meet regulatory requirements, this guide offers concrete guidance and the right tools for your needs. The maturity assessment is designed mainly for organizations holding or providing data and preparing to participate in the European Health Data Space (EHDS).

## — What you get

After downloading the 10 questionnaires and overall maturity assessment document, you answer questions across ten well-defined dimensions of data maturity and immediately receive a personalized assessment regarding your maturity toward EHDS readiness.

## — How it works — at a glance

1

### Answer

Work through 10 short questionnaires, one per dimension of data maturity.

2

### Score

Each dimension yields a maturity level from 1 to 5 based on your answers.

3

### Reflect

Combine all dimension levels to get your overall maturity rating.

### Tip

Answer honestly based on your current situation, not where you aim to be. The assessment is most useful when it accurately reflects where you are today — it then points to where you can grow.

## — The framework

The maturity assessment is built on the QUANTUM framework (Deliverable 1.2 by Jones & Bernal-Delgado, 2025), which structures data holder maturity into ten complementary dimensions. Each dimension is assessed independently with a short questionnaire of 4 to 5 questions.

|          |                                                                                                                                        |           |                                                                                                                                 |
|----------|----------------------------------------------------------------------------------------------------------------------------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------|
| <b>1</b> | <b>Data Collection Process</b><br>How well defined are your processes for capturing data from multiple sources for secondary purposes. | <b>2</b>  | <b>Data Management Governance</b><br>Maturity of your data management plans, their implementation, and compliance.              |
| <b>3</b> | <b>Data Management Infrastructure</b><br>Implementation and development of your data management infrastructure.                        | <b>4</b>  | <b>Data Provenance</b><br>Clarity and documentation of the lineage of your dataset and data pipeline.                           |
| <b>5</b> | <b>Data Access</b><br>How well defined are data access processes legally, ethically, and technically.                                  | <b>6</b>  | <b>Data Analytics Environment</b><br>Analytical services, tooling, and access to secure data environments.                      |
| <b>7</b> | <b>Data Enhancement — Augmentation</b><br>Application of techniques to make data more usable for specific purposes.                    | <b>8</b>  | <b>Data Enhancement — Enrichment</b><br>Enrichment of data with annotations, image labels, derivations, and NLP-derived labels. |
| <b>9</b> | <b>Data Model</b><br>Availability of a clear, documented data model providing structure and standardisation.                           | <b>10</b> | <b>Data Dictionary</b><br>Provision of a documented data dictionary and terminologies.                                          |

## — How scoring works

The assessment uses a three-tier scoring system: each answer earns points, those points combine into a level per dimension, and the dimension levels combine into an overall maturity level. All scores are on a 1-to-5 scale, where 1 means "no maturity" and 5 means "comprehensive maturity".

### Step 1 — Question level

Each answer is worth between 1 and 5 points, depending on how mature the described situation is. Some questions offer the full 1–5 range; others offer fewer options (e.g., 1, 3, 5) where intermediate levels do not apply.

### Step 2 — Dimension level

For each dimension, sum the points of all your answers and divide by the number of questions in that dimension. This gives you a decimal score between 1 and 5. Round this number down to the nearest whole number to determine your dimension maturity level.

#### Example

A dimension with 4 questions and answers worth 3, 4, 5, and 3 points gives a sum of 15. Divided by 4 questions = 3.75. Rounded down: Dimension Level 3.

### Step 3 — Overall maturity level

In the Overall Score Calculation document, select the maturity level you reached for each of the ten dimensions. Sum all ten dimension levels and divide by 10. Round this number down to the nearest whole number to get your overall maturity level (1–5).

## — The maturity levels

|                                                            |                                                            |                                                               |                                                           |                                                             |
|------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>No or ad-hoc processes | <b>Level 2</b><br><b>Emerging</b><br>Basic processes exist | <b>Level 3</b><br><b>Defined</b><br>Documented and consistent | <b>Level 4</b><br><b>Managed</b><br>Measured and verified | <b>Level 5</b><br><b>Optimized</b><br>Continuously improved |
|------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------------|

#### A note on interpretation

Rounding down is intentional: you only reach a level when you fully meet its criteria. A score of 3.8 means you are close to Level 4 but have not fully achieved it yet — your current level is 3, with clear room to grow.

# 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?\*

- |                                                                                                                              |        |
|------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We have no defined process for collecting or capturing data for secondary purposes.                 | 1 pt.  |
| <input type="checkbox"/> We have an informal, ad hoc process. It works in practice, but it is not written down.              | 2 pts. |
| <input type="checkbox"/> We have a written process for some of our data collections.                                         | 3 pts. |
| <input type="checkbox"/> We have a written process for every data collection we make available for secondary use.            | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                             |        |
|---------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> None of it.                                                        | 1 pt.  |
| <input type="checkbox"/> One or two collections, described informally.                      | 2 pts. |
| <input type="checkbox"/> Our main collections, roughly half of what we hold.                | 3 pts. |
| <input type="checkbox"/> All collections we produce ourselves.                              | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                                                  |        |
|------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No. It has never been used beyond its original purpose.                                 | 1 pt.  |
| <input type="checkbox"/> Once or twice, as a one-off project.                                                    | 2 pts. |
| <input type="checkbox"/> Some of it is used regularly, for example, internal research or quality reporting.      | 3 pts. |
| <input type="checkbox"/> Most of it is used regularly, including by parties outside our organization.            | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                               |        |
|-----------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Fully manual, with data exported, copied, and retyped by hand.       | 1 pt.  |
| <input type="checkbox"/> Mostly manual, with a few helper scripts.                            | 2 pts. |
| <input type="checkbox"/> Partly automated, with recurring extracts running on a schedule.     | 3 pts. |
| <input type="checkbox"/> Largely automated, with monitoring and error alerts.                 | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                                                                                                 |        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We do not look at data quality.                                                                                                        | 1 pt.  |
| <input type="checkbox"/> We notice quality problems when somebody complains about them.                                                                         | 2 pts. |
| <input type="checkbox"/> We check completeness and plausibility when something changes, and correct what we find.                                               | 3 pts. |
| <input type="checkbox"/> We measure defined quality indicators, such as completeness, plausibility, and timeliness, on a fixed schedule and record the results. | 4 pts. |
| <input type="checkbox"/> 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

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

# Data Management — Governance 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 2 — Data Management — Governance

**Definition\*:** The level of maturity of your data management processes: the plan that says how data is handled, whether it is actually followed, and who is accountable for it.

### 1. What is the status of your data management plan?\*

- |                                                                                                                          |        |
|--------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We have nothing documented.                                                                     | 1 pt.  |
| <input type="checkbox"/> We have a written data management plan, approved and dated, but no evidence that it is applied. | 2 pts. |
| <input type="checkbox"/> We can show evidence that the plan has been implemented, such as records, logs, and minutes.    | 3 pts. |
| <input type="checkbox"/> We measure compliance with the plan and can demonstrate it internally.                          | 4 pts. |
| <input type="checkbox"/> Compliance with the plan has been verified by an external party (audit or certification).       | 5 pts. |

*You can support this answer with: the data management plan itself, with its version and approval date.*

### 2. What does your data management plan cover?\*

- |                                                                                                 |        |
|-------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We have no plan.                                                       | 1 pt.  |
| <input type="checkbox"/> Planning only, what we intend to do.                                   | 2 pts. |
| <input type="checkbox"/> Collection, quality checking, and day-to-day management.               | 3 pts. |
| <input type="checkbox"/> The above, plus retention, deletion, access, and incident handling.    | 4 pts. |
| <input type="checkbox"/> The full data lifecycle, including secondary use and EHDS obligations. | 5 pts. |

*You can support this answer with: the table of contents of your plan.*

### 3. How far has the plan been implemented?\*

- |                                                                                                                                  |        |
|----------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> It has not been implemented.                                                                            | 1 pt.  |
| <input type="checkbox"/> Implemented in one department or for one dataset.                                                       | 2 pts. |
| <input type="checkbox"/> Implemented for our main datasets.                                                                      | 3 pts. |
| <input type="checkbox"/> Implemented across the whole organization.                                                              | 4 pts. |
| <input type="checkbox"/> Implemented across the organization and reviewed on a fixed cycle, with changes fed back into the plan. | 5 pts. |

*You can support this answer with: implementation records, roll-out plan, system configuration.*

### 4. How do you know the plan is being followed?\*

- |                                                                                                                  |        |
|------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We do not check.                                                                        | 1 pt.  |
| <input type="checkbox"/> We only hear about it when somebody raises a problem.                                   | 2 pts. |
| <input type="checkbox"/> Somebody reviews compliance from time to time, without a schedule and without a record. | 3 pts. |
| <input type="checkbox"/> We check on a fixed schedule and record the result (internal audit).                    | 4 pts. |
| <input type="checkbox"/> An external party verifies compliance (external audit or certification).                | 5 pts. |

*You can support this answer with: your last internal or external audit report.*

### 5. Who is responsible for data governance?\*

- |                                                                                                                            |        |
|----------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nobody in particular.                                                                             | 1 pt.  |
| <input type="checkbox"/> Somebody does it informally, alongside their main job.                                            | 2 pts. |
| <input type="checkbox"/> A named person has it in their job description.                                                   | 3 pts. |
| <input type="checkbox"/> A named person plus a standing group (for example, a data governance board) that meets regularly. | 4 pts. |
| <input type="checkbox"/> As above, with a documented mandate and a reporting line to senior management.                    | 5 pts. |

*You can support this answer with: the job description, terms of reference of the governance board, organization chart.*

#### 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

|                             |                                                                                                |
|-----------------------------|------------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | Nothing documented, nobody accountable, and no check that anything is followed.                |
| <b>Level 2</b><br>Emerging  | A written plan and somebody informally responsible, but no evidence it is being applied.       |
| <b>Level 3</b><br>Defined   | The plan is followed for the main datasets, with a named owner and occasional checks.          |
| <b>Level 4</b><br>Managed   | Implemented organization-wide, compliance measured on a schedule, governance board in place.   |
| <b>Level 5</b><br>Optimized | Externally audited or certified, with a documented mandate and a reporting line to management. |

**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: Adapted HDRUK. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

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.

# 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?\*

- |                                                                                                                                               |        |
|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No dedicated infrastructure. Data stays in the source systems and is extracted by hand.                              | 1 pt.  |
| <input type="checkbox"/> An emerging infrastructure, with some checks on data going in and out.                                               | 2 pts. |
| <input type="checkbox"/> A defined infrastructure that is used as our standard way of working.                                                | 3 pts. |
| <input type="checkbox"/> An automated and validated infrastructure for the main flows, supporting near real-time data management.             | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                                                        |        |
|------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> There are no quality checks.                                                                  | 1 pt.  |
| <input type="checkbox"/> Manual spot checks, when somebody has time.                                                   | 2 pts. |
| <input type="checkbox"/> Documented checks on our main data flows, such as completeness, valid values, and duplicates. | 3 pts. |
| <input type="checkbox"/> Automated checks on all flows, with errors logged and corrected.                              | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                            |        |
|--------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Not automated, with every step done by hand.                      | 1 pt.  |
| <input type="checkbox"/> A few scripts, run manually.                                      | 2 pts. |
| <input type="checkbox"/> Partly automated, with the main pipelines running on a schedule.  | 3 pts. |
| <input type="checkbox"/> Largely automated, with monitoring.                               | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                                                  |        |
|------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We cannot supply a refreshed extract. What exists is whatever was exported at the time. | 1 pt.  |
| <input type="checkbox"/> On request only, with a lead time of several weeks.                                     | 2 pts. |
| <input type="checkbox"/> On a fixed refresh cycle, for example, monthly or quarterly.                            | 3 pts. |
| <input type="checkbox"/> Frequent refresh, daily or weekly, for most data.                                       | 4 pts. |
| <input type="checkbox"/> 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?\*

- |                                                                                                                                                          |        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No specific measures beyond general IT security.                                                                                | 1 pt.  |
| <input type="checkbox"/> Basic access control and backups.                                                                                               | 2 pts. |
| <input type="checkbox"/> Documented measures: role-based access, encryption, access logging.                                                             | 3 pts. |
| <input type="checkbox"/> The above, tested regularly, for example, penetration tests and recovery tests.                                                 | 4 pts. |
| <input type="checkbox"/> 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

|                             |                                                                                                |
|-----------------------------|------------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | No dedicated infrastructure; manual extracts; no quality checks and no specific security.      |
| <b>Level 2</b><br>Emerging  | An emerging setup with spot checks, a few manual scripts, basic access control, and backups.   |
| <b>Level 3</b><br>Defined   | A defined standard infrastructure; documented checks and security; scheduled refresh.          |
| <b>Level 4</b><br>Managed   | Automated, monitored pipelines; checks on all flows; frequent refresh; security tested.        |
| <b>Level 5</b><br>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.

# Data Provenance 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 4 — Data Provenance

**Definition\*:** Clear description of the lineage of the dataset, providing transparent and comprehensive documentation of the data pipeline used to produce it, covering where the data came from, what was done to it, and what that did to its quality.

### 1. What do you document about where your dataset comes from?\*

- |                                                                                                                                             |        |
|---------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nothing is documented.                                                                                             | 1 pt.  |
| <input type="checkbox"/> We document the source of the dataset.                                                                             | 2 pts. |
| <input type="checkbox"/> We document the source plus the transformations, rules, and exclusions applied.                                    | 3 pts. |
| <input type="checkbox"/> We list all original data items and all transformations, rules, and exclusions, and document what effect they had. | 4 pts. |
| <input type="checkbox"/> We can show earlier versions, including the raw source data, and review the effect of each step.                   | 5 pts. |

*You can support this answer with: your provenance or lineage document for one representative dataset.*

### 2. How complete is that documentation?\*

- |                                                                                                                     |        |
|---------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nothing is written down.                                                                   | 1 pt.  |
| <input type="checkbox"/> The source system is named.                                                                | 2 pts. |
| <input type="checkbox"/> Source, plus the main transformation rules.                                                | 3 pts. |
| <input type="checkbox"/> Source, all transformations, rules, and exclusions, with record counts before and after.   | 4 pts. |
| <input type="checkbox"/> As above, maintained automatically by the pipeline itself, so it can never be out of date. | 5 pts. |

*You can support this answer with: the lineage document, or the pipeline metadata it is generated from.*

### 3. Can you retrieve and compare earlier versions of the dataset?\*

- |                                                                                                                     |        |
|---------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No version control. There is only "the current file."                                      | 1 pt.  |
| <input type="checkbox"/> Old versions are kept somewhere, informally.                                               | 2 pts. |
| <input type="checkbox"/> Dataset versions are numbered and dated.                                                   | 3 pts. |
| <input type="checkbox"/> Versions are numbered and dated, and the differences between them are documented.          | 4 pts. |
| <input type="checkbox"/> Any earlier version, including the raw source, can be retrieved and compared step by step. | 5 pts. |

*You can support this answer with: your version register, repository, or archive of previous releases.*

### 4. Do you document what transformations and exclusions do to the quality of the dataset?\*

- |                                                                                                                                           |        |
|-------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Not at all.                                                                                                      | 1 pt.  |
| <input type="checkbox"/> We know exclusions happen, but not how large they are.                                                           | 2 pts. |
| <input type="checkbox"/> We record which records were excluded and why.                                                                   | 3 pts. |
| <input type="checkbox"/> We record how many records each rule affected.                                                                   | 4 pts. |
| <input type="checkbox"/> We report the effect on data quality, such as completeness, plausibility, and representativeness, for each step. | 5 pts. |

*You can support this answer with: your exclusion log, record counts per pipeline step, data quality report.*

### 5. Can the people who use your data see this provenance information?\*

- |                                                                                                                    |        |
|--------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> It is not shared outside the team.                                                        | 1 pt.  |
| <input type="checkbox"/> Available on request, if somebody asks the right person.                                  | 2 pts. |
| <input type="checkbox"/> A provenance document is supplied together with the dataset.                              | 3 pts. |
| <input type="checkbox"/> Provenance is part of the standard machine-readable metadata delivered with the dataset.  | 4 pts. |
| <input type="checkbox"/> As above, and published in a dataset catalog so users can read it before requesting data. | 5 pts. |

*You can support this answer with: your dataset catalog entry or the metadata file shipped with a release.*

## 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.

|                                                     |                                                       |                                                     |                                                  |                                                     |
|-----------------------------------------------------|-------------------------------------------------------|-----------------------------------------------------|--------------------------------------------------|-----------------------------------------------------|
| <b>Level 1</b><br>Initial<br>Starting out · 1 Point | <b>Level 2</b><br>Emerging<br>Written down · 2 Points | <b>Level 3</b><br>Defined<br>In practice · 3 Points | <b>Level 4</b><br>Managed<br>Measured · 4 Points | <b>Level 5</b><br>Optimized<br>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.

|                             |                                                                                                |
|-----------------------------|------------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | Nothing documented; no versions; no view of what transformations do; nothing shared.           |
| <b>Level 2</b><br>Emerging  | The source system is named; old versions kept informally; exclusions known but not quantified. |
| <b>Level 3</b><br>Defined   | Source, transformations, and exclusions documented and versioned; supplied with the dataset.   |
| <b>Level 4</b><br>Managed   | All items, rules, and record counts documented; version differences tracked; machine-readable. |
| <b>Level 5</b><br>Optimized | Lineage maintained by the pipeline itself; any version retrievable; quality effects published. |

**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: Adapted HDRUK. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 2024. Question 4 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

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.

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.

# Data Access 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 5 — Data Access

**Definition\*:** How well defined and implemented your data access processes are, from a legal, ethical, and technical perspective. Under the EHDS, a Health Data Access Body (HDAB) will forward data requests to you; this dimension asks how ready you are to handle them.

### 1. When somebody asks for your data, what happens?\*

- ☐ There is no process. Requests are handled case by case, by whoever is asked. 1 pt.
- ☐ There is a written process, but responses are neither timely nor consistent, with no agreed response time and varying outcomes. 2 pts.
- ☐ Responses are timely and consistent: an agreed response time, a named responsible person, and the same route every time. 3 pts.
- ☐ The process covers the technical and the legal side, with tracked service levels, for example, "90% answered within 30 days." 4 pts.
- ☐ A complete access system covering technical, legal, and ethical review, aligned with EU rules and reviewed regularly. 5 pts.

You can support this answer with: your data access policy, the request log for the last 12 months.

### 2. How timely and consistent are your responses?\*

- ☐ No defined response time. The outcome depends on who handles the request. 1 pt.
- ☐ We aim to respond promptly, but nothing is agreed or written down. 2 pts.
- ☐ A target response time is agreed and written down, but we do not measure whether we meet it. 3 pts.
- ☐ We measure our response times against the target and meet it in most cases. 4 pts.
- ☐ Response times are measured, published, and consistently met; decisions follow documented criteria. 5 pts.

You can support this answer with: your request log with dates received and answered, service-level report.

### 3. What does your access process cover?\*

- ☐ Nothing is in place. 1 pt.
- ☐ The technical route only, so we can extract and send data. 2 pts.
- ☐ Technical route, plus a legal basis and a data sharing agreement. 3 pts.
- ☐ The above, plus a documented decision procedure with defined approval criteria. 4 pts.
- ☐ The above, plus an ethical review step. 5 pts.

You can support this answer with: your access procedure, template data sharing agreement, ethics review route.

### 4. How does your access process relate to EU rules?\*

- ☐ We have not assessed it against EU rules. 1 pt.
- ☐ We are aware of GDPR obligations and apply them informally. 2 pts.
- ☐ GDPR compliance is documented for data sharing. 3 pts.
- ☐ GDPR compliance is documented, and we have mapped our process against the EHDS Regulation. 4 pts.
- ☐ Our process is aligned with the EHDS Regulation and is reviewed as implementing acts appear. 5 pts.

You can support this answer with: your records of processing, DPIA, EHDS gap analysis.

### 5. Can potential users find out what you hold and how to ask for it?\*

- ☐ No. They would have to know somebody here. 1 pt.
- ☐ They can ask us, and we will tell them. 2 pts.
- ☐ Our datasets and access conditions are described in a standard document, supplied on request. 3 pts.
- ☐ Datasets and access conditions are published on our website or in a catalog. 4 pts.
- ☐ Published in a machine-readable catalog, ready for the national EHDS dataset catalog. 5 pts.

You can support this answer with: your public dataset list or catalog entry.

### 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

|                             |                                                                                                 |
|-----------------------------|-------------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | No process, no response time, no legal assessment, and nobody can find out what you hold.       |
| <b>Level 2</b><br>Emerging  | A written route exists, but responses are unpredictable and only GDPR is applied informally.    |
| <b>Level 3</b><br>Defined   | Agreed response time and named owner; legal basis and sharing agreement; described on request.  |
| <b>Level 4</b><br>Managed   | Response times measured and met; documented approval criteria; EHDS mapped; datasets published. |
| <b>Level 5</b><br>Optimized | Ethics review included, aligned with EU rules and reviewed; machine-readable dataset catalog.   |

**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. 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.

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.

# Data Analytics Environment 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 6 — Data Analytics Environment

**Definition\*:** Analytical services, tooling, and access to secure data environments. A secure data environment (SDE) is a controlled workspace in which approved users analyze data without being able to copy it out; a federated setup runs the analysis at each organization so that data never leaves the site.

### 1. Where can your data be analyzed?\*

- |                                                                                                                                                               |        |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nowhere on our side. We can only send files.                                                                                         | 1 pt.  |
| <input type="checkbox"/> Analyses are run by our own staff, and results are returned in aggregated or anonymized form.                                        | 2 pts. |
| <input type="checkbox"/> Approved users can work with the data inside a secure data environment (SDE).                                                        | 3 pts. |
| <input type="checkbox"/> The dataset can be used in an SDE, and users can bring in additional data and their own tools.                                       | 4 pts. |
| <input type="checkbox"/> The dataset can be used in a federated environment, where analysis runs across several organizations without data leaving each site. | 5 pts. |

You can support this answer with: a description of your analysis environment, user access records.

### 2. Who can run an analysis on your data?\*

- |                                                                                                  |        |
|--------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nobody, as we cannot support analysis at all.                           | 1 pt.  |
| <input type="checkbox"/> Only our own staff, running simple predefined queries.                  | 2 pts. |
| <input type="checkbox"/> Our own staff, running custom analyses on request.                      | 3 pts. |
| <input type="checkbox"/> External users, running their own analyses in a controlled environment. | 4 pts. |
| <input type="checkbox"/> External users, running their own analyses across federated datasets.   | 5 pts. |

You can support this answer with: your list of external analysis projects, environment user list.

### 3. How is the analysis environment secured?\*

- |                                                                                                  |        |
|--------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> There is no controlled environment, so data is sent out as files.       | 1 pt.  |
| <input type="checkbox"/> Access is controlled by a password and an agreement only.               | 2 pts. |
| <input type="checkbox"/> Named user accounts, access logging, and defined roles.                 | 3 pts. |
| <input type="checkbox"/> The above, in an environment from which data cannot be downloaded.      | 4 pts. |
| <input type="checkbox"/> The above, independently assessed or certified, and reviewed regularly. | 5 pts. |

You can support this answer with: your access control policy, environment security assessment.

### 4. How flexible is the environment for its users?\*

- |                                                                                                    |        |
|----------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Only fixed outputs. Users cannot choose what is produced.                 | 1 pt.  |
| <input type="checkbox"/> A single tool that we provide, with no way to add anything.               | 2 pts. |
| <input type="checkbox"/> The common analysis tools are available, for example, R, Python, and SQL. | 3 pts. |
| <input type="checkbox"/> Users may install additional tools themselves within the environment.     | 4 pts. |
| <input type="checkbox"/> Users may install their own tools and link in additional datasets.        | 5 pts. |

You can support this answer with: the tool catalog of your environment, user guidance.

### 5. How do you control what leaves the environment?\*

- |                                                                                                                 |        |
|-----------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> There are no checks on results leaving.                                                | 1 pt.  |
| <input type="checkbox"/> Somebody glances at results before they go out.                                        | 2 pts. |
| <input type="checkbox"/> A documented disclosure rule is applied, for example, a minimum cell size.             | 3 pts. |
| <input type="checkbox"/> Disclosure control is applied by a trained reviewer and recorded.                      | 4 pts. |
| <input type="checkbox"/> Disclosure control is documented, applied consistently, audited, and partly automated. | 5 pts. |

You can support this answer with: your output checking policy, disclosure control log.

### 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

|                             |                                                                                             |
|-----------------------------|---------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | No environment; files are sent out; no checks on results and no choice of output.           |
| <b>Level 2</b><br>Emerging  | Your staff run predefined queries; access by password and agreement; results glanced at.    |
| <b>Level 3</b><br>Defined   | Approved users work in a secure environment with named accounts and a disclosure rule.      |
| <b>Level 4</b><br>Managed   | External users analyze in a no-download environment with their own tools; output reviewed.  |
| <b>Level 5</b><br>Optimized | Federated analysis, independently assessed; users bring tools and data; disclosure audited. |

**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: Adapted HDRUK. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

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.

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.

# Data Enhancement — Augmentation 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 7 — Data Enhancement — Augmentation

**Definition\*:** The application of various techniques to make data more usable for specific purposes, for example, cleaning, deduplication, unit harmonization, or mapping the data to a recognized common data model.

### 1. Do you apply techniques to make your data more usable?\*

- |                                                                                                                                                                     |        |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No techniques are applied.                                                                                                                 | 1 pt.  |
| <input type="checkbox"/> Some techniques are applied case by case, for example, cleaning, deduplication, unit conversion.                                           | 2 pts. |
| <input type="checkbox"/> Defined, documented techniques are applied.                                                                                                | 3 pts. |
| <input type="checkbox"/> The techniques are managed: versioned, tested, and applied consistently.                                                                   | 4 pts. |
| <input type="checkbox"/> A comprehensive set of techniques is applied, and the data is mapped to a recognized data model (for example, OMOP, openEHR, or HL7 FHIR). | 5 pts. |

You can support this answer with: your data preparation documentation, transformation scripts.

### 2. How much of your data is prepared in this way?\*

- |                                                                                        |        |
|----------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> None of it.                                                   | 1 pt.  |
| <input type="checkbox"/> A single dataset or one project.                              | 2 pts. |
| <input type="checkbox"/> Our main datasets.                                            | 3 pts. |
| <input type="checkbox"/> All datasets offered for secondary use.                       | 4 pts. |
| <input type="checkbox"/> All datasets, prepared automatically as part of the pipeline. | 5 pts. |

You can support this answer with: your data inventory, showing which datasets are prepared.

### 3. How are these techniques managed?\*

- |                                                                                                                 |        |
|-----------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Not managed. Each person does it their own way, and nothing is kept.                   | 1 pt.  |
| <input type="checkbox"/> One person's script or routine is kept but not documented, so only they can repeat it. | 2 pts. |
| <input type="checkbox"/> Documented and repeatable by a colleague.                                              | 3 pts. |
| <input type="checkbox"/> Version-controlled, tested, and reviewed.                                              | 4 pts. |
| <input type="checkbox"/> Version-controlled, tested, reviewed, and monitored in production.                     | 5 pts. |

You can support this answer with: your code repository, review records, test results.

### 4. Is your data mapped to a recognized data model?\*

- |                                                                                                                                   |        |
|-----------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No mapping, and none is planned.                                                                         | 1 pt.  |
| <input type="checkbox"/> A target model has been chosen, and the mapping is planned, but nothing is in use yet.                   | 2 pts. |
| <input type="checkbox"/> Some key fields are mapped to a recognized model.                                                        | 3 pts. |
| <input type="checkbox"/> Most of the dataset is mapped, with a documented mapping specification.                                  | 4 pts. |
| <input type="checkbox"/> Fully mapped to a recognized model (OMOP, openEHR, HL7 FHIR), with the mapping maintained and validated. | 5 pts. |

You can support this answer with: your mapping specification, mapping coverage report.

### 5. Do you check the quality of the result?\*

- |                                                                                                      |        |
|------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We do not check.                                                            | 1 pt.  |
| <input type="checkbox"/> We look at the output informally.                                           | 2 pts. |
| <input type="checkbox"/> We compare key figures before and after.                                    | 3 pts. |
| <input type="checkbox"/> We apply documented quality checks and record the outcome.                  | 4 pts. |
| <input type="checkbox"/> Quality checks are automated, reported, and used to improve the techniques. | 5 pts. |

You can support this answer with: your before/after quality report, automated check results.

### 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.

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

Round the dimension score down to the nearest whole number.

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.

**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.

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: OMOP / OpenEHR / HL7 FHIR / ISO/IEC 22989:2022. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Observational Health Data Sciences and Informatics (OHDSI). (n.d.). OMOP Common Data Model.

openEHR International. (n.d.). openEHR Architecture Specifications.

Health Level Seven International. (n.d.). FHIR Release 4 — Fast Healthcare Interoperability Resources.

International Organization for Standardization & International Electrotechnical Commission. (2022). Information technology — Artificial intelligence — Artificial intelligence concepts and terminology (ISO/IEC Standard No. 22989:2022).

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.

# Data Enhancement — Enrichment 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 8 — Data Enhancement — Enrichment

**Definition\*:** Data sources enriched, for example, with annotations, image labels, phenotype definitions, derived variables, and labels produced by natural language processing. A derived field is a value your organization calculates or assigns; it is not recorded directly at the source.

### 1. Does your data contain derived fields or added annotations?\*

- |                                                                                                                                                   |        |
|---------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No. The data contains only what was recorded at source.                                                                  | 1 pt.  |
| <input type="checkbox"/> Yes, some derived fields or annotations that we created ourselves.                                                       | 2 pts. |
| <input type="checkbox"/> Yes, and they are also used by other data sources or organizations.                                                      | 3 pts. |
| <input type="checkbox"/> Yes, and they are produced by, or used in, a peer-reviewed algorithm, one published and assessed by independent experts. | 4 pts. |
| <input type="checkbox"/> Yes, and they come from a national or international report.                                                              | 5 pts. |

You can support this answer with: your list of derived fields with their definitions.

### 2. How widely are derived fields used across your data?\*

- |                                                                                      |        |
|--------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nowhere.                                                    | 1 pt.  |
| <input type="checkbox"/> A few fields in one dataset.                                | 2 pts. |
| <input type="checkbox"/> Across our main datasets.                                   | 3 pts. |
| <input type="checkbox"/> Across all datasets offered for secondary use.              | 4 pts. |
| <input type="checkbox"/> Across all datasets, generated and refreshed automatically. | 5 pts. |

You can support this answer with: your data dictionary, marking which fields are derived.

### 3. Who uses the enriched data?\*

- |                                                                                                                |        |
|----------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nobody uses it.                                                                       | 1 pt.  |
| <input type="checkbox"/> Our own teams only.                                                                   | 2 pts. |
| <input type="checkbox"/> Our own teams and named partners.                                                     | 3 pts. |
| <input type="checkbox"/> External researchers, routinely.                                                      | 4 pts. |
| <input type="checkbox"/> External researchers, routinely, and it is cited in publications or official reports. | 5 pts. |

You can support this answer with: project list, publication list referencing your derived variables.

### 4. What are the derived fields based on?\*

- |                                                                                                                    |        |
|--------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No defined basis, so they were created ad hoc.                                            | 1 pt.  |
| <input type="checkbox"/> An internal rule that is written down.                                                    | 2 pts. |
| <input type="checkbox"/> A national guideline, or an algorithm validated locally.                                  | 3 pts. |
| <input type="checkbox"/> A peer-reviewed published algorithm.                                                      | 4 pts. |
| <input type="checkbox"/> A national or international standard, or a definition taken from an international report. | 5 pts. |

You can support this answer with: the reference or citation for each derivation rule.

### 5. Can users tell derived fields apart from source fields?\*

- |                                                                                                     |        |
|-----------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No. Derived and source fields are indistinguishable.                       | 1 pt.  |
| <input type="checkbox"/> It is mentioned informally when we hand over data.                         | 2 pts. |
| <input type="checkbox"/> Derived fields are marked and described in the dataset documentation.      | 3 pts. |
| <input type="checkbox"/> The derivation rule and its source are documented for every derived field. | 4 pts. |
| <input type="checkbox"/> As above, machine-readable and published together with the dataset.        | 5 pts. |

You can support this answer with: your data dictionary entry for a derived field.

### 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.

|                                         |                                                                                                   |
|-----------------------------------------|---------------------------------------------------------------------------------------------------|
| <div>Level 1</div> <div>Initial</div>   | No derived fields or annotations, so nothing to use, document, or distinguish.                    |
| <div>Level 2</div> <div>Emerging</div>  | A few self-made derived fields in one dataset, used internally, mentioned informally.             |
| <div>Level 3</div> <div>Defined</div>   | Derived fields across the main datasets, shared with partners, described in the documentation.    |
| <div>Level 4</div> <div>Managed</div>   | Peer-reviewed derivations across all shared datasets, used by external researchers, documented.   |
| <div>Level 5</div> <div>Optimized</div> | Standard or report-based definitions, generated automatically, cited, published machine-readable. |

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: Adapted HDRUK. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

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.

Data Enhancement — Enrichment Maturity Assessment | Please answer all questions

Page 2 of 2

# Data Model 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 9 — Data Model

**Definition\*:** Availability of a clear, documented data model that provides structure and standardization. A local standard is a code list agreed inside your own organization; a national or international standard is one maintained outside it, such as ICD-10, SNOMED CT, or LOINC.

### 1. What data model do you use?\*

- |                                                                                                                                                  |        |
|--------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We have no data model; the structure is whatever the source system produces.                                            | 1 pt.  |
| <input type="checkbox"/> A data model exists and is accepted internally, but important fields are free text or uncoded.                          | 2 pts. |
| <input type="checkbox"/> Key fields are coded using a local standard and kept up to date.                                                        | 3 pts. |
| <input type="checkbox"/> Important fields use a national or international code system, and we keep them current.                                 | 4 pts. |
| <input type="checkbox"/> The data model itself conforms to an international standard, and key fields use national or international code systems. | 5 pts. |

You can support this answer with: your data model documentation, entity-relationship diagram.

### 2. How are your key fields coded?\*

- |                                                                                                                                |        |
|--------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Not coded, free text.                                                                                 | 1 pt.  |
| <input type="checkbox"/> Some fields coded with an in-house list.                                                              | 2 pts. |
| <input type="checkbox"/> Key fields coded with a local standard.                                                               | 3 pts. |
| <input type="checkbox"/> Key fields coded with a national or international standard, for example, ICD-10, SNOMED CT, LOINC.    | 4 pts. |
| <input type="checkbox"/> All fields relevant for secondary use coded with an international standard, with mappings maintained. | 5 pts. |

You can support this answer with: your code list register, field-to-terminology mapping table.

### 3. How is the data model maintained?\*

- |                                                                                                            |        |
|------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> No updates. The model is whatever it was when it was created.                     | 1 pt.  |
| <input type="checkbox"/> Updated only when a source system changes.                                        | 2 pts. |
| <input type="checkbox"/> Reviewed and updated on a fixed schedule.                                         | 3 pts. |
| <input type="checkbox"/> Reviewed on a fixed schedule and aligned with new releases of the standards used. | 4 pts. |
| <input type="checkbox"/> As above, with a version history and a documented change procedure.               | 5 pts. |

You can support this answer with: the version history of your data model, change log.

### 4. How easily can your data be combined with data from other organizations?\*

- |                                                                                                                                                       |        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> There is no route. Every combination is a one-off project, negotiated and mapped from scratch.                               | 1 pt.  |
| <input type="checkbox"/> A reusable export route exists, but our data model is not standardized, so every new partner needs its own mapping.          | 2 pts. |
| <input type="checkbox"/> We exchange data with local partners in a format agreed between us, and the mapping is documented and reused.                | 3 pts. |
| <input type="checkbox"/> We exchange data in a national or international format, for example, HL7 FHIR, so partners need no bespoke mapping.          | 4 pts. |
| <input type="checkbox"/> Our data conforms to an international standard and can be combined across organizations and borders without further mapping. | 5 pts. |

You can support this answer with: your export formats, interface specifications, examples of past exchanges.

### 5. Is the data model documented for the people who use your data?\*

- |                                                                                                         |        |
|---------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> It is not documented.                                                          | 1 pt.  |
| <input type="checkbox"/> It exists as a diagram or spreadsheet somewhere.                               | 2 pts. |
| <input type="checkbox"/> A documented model covering entities, fields, and relationships is maintained. | 3 pts. |
| <input type="checkbox"/> Documented and published to data users with each version.                      | 4 pts. |
| <input type="checkbox"/> Documented, versioned, published, and machine-readable.                        | 5 pts. |

You can support this answer with: the published model documentation, catalog entry.

### 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.

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

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.

|                                         |                                                                                                    |
|-----------------------------------------|----------------------------------------------------------------------------------------------------|
| <div>Level 1</div> <div>Initial</div>   | No data model, no coding, no maintenance, nothing documented, and no way to combine data.          |
| <div>Level 2</div> <div>Emerging</div>  | A model exists internally with in-house lists; updated only when a source system changes.          |
| <div>Level 3</div> <div>Defined</div>   | Key fields coded to a local standard, reviewed on a schedule, documented, exchanged locally.       |
| <div>Level 4</div> <div>Managed</div>   | National or international coding, aligned with standard releases, published to users each version. |
| <div>Level 5</div> <div>Optimized</div> | Conforms to an international standard; versioned, machine-readable, combinable across borders.     |

**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: Adapted HDRUK / OMOP / OpenEHR / HL7 FHIR. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

Observational Health Data Sciences and Informatics (OHDSI). (n.d.). OMOP Common Data Model.

openEHR International. (n.d.). openEHR Architecture Specifications.

Health Level Seven International. (n.d.). FHIR Release 4 — Fast Healthcare Interoperability Resources.

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.

Data Model Maturity Assessment | Please answer all questions

Page 2 of 2

# Data Dictionary 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 10 — Data Dictionary

**Definition\*:** A documented data dictionary and terminologies. The data dictionary explains, field by field, what each item means, what values it may take, and how it should be interpreted, so that somebody outside your organization can use the data correctly.

### 1. Do you have a data dictionary?\*

- |                                                                                                                 |        |
|-----------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> We have nothing that documents what our fields mean.                                   | 1 pt.  |
| <input checked="" type="checkbox"/> Definitions exist for some fields, scattered across documents.              | 2 pts. |
| <input type="checkbox"/> All definitions are collected in one dictionary of our own, published online.          | 3 pts. |
| <input type="checkbox"/> The dictionary uses national definitions.                                              | 4 pts. |
| <input type="checkbox"/> Our definitions follow international standards, and the mapping to them is documented. | 5 pts. |

*You can support this answer with: the data dictionary file or page itself.*

### 2. How complete are the definitions?\*

- |                                                                                                              |        |
|--------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> There are no definitions.                                                           | 1 pt.  |
| <input type="checkbox"/> Definitions for some fields.                                                        | 2 pts. |
| <input type="checkbox"/> Definitions for all fields used for secondary purposes.                             | 3 pts. |
| <input type="checkbox"/> As above, including value sets, units, and permitted ranges.                        | 4 pts. |
| <input type="checkbox"/> As above, including missing-value conventions and known limitations for each field. | 5 pts. |

*You can support this answer with: a sample of dictionary entries for your most-used dataset.*

### 3. Who can read the data dictionary?\*

- |                                                                                                                         |        |
|-------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Only our own team.                                                                             | 1 pt.  |
| <input type="checkbox"/> Anyone, on request.                                                                            | 2 pts. |
| <input type="checkbox"/> It is published online, so users can read it before requesting data.                           | 3 pts. |
| <input type="checkbox"/> It is published online and supplied with every dataset release, versioned in step with it.     | 4 pts. |
| <input type="checkbox"/> It is published online in machine-readable form, for example, through an API or catalog entry. | 5 pts. |

*You can support this answer with: the public link to your dictionary, or the API endpoint.*

### 4. What is the dictionary standardized against?\*

- |                                                                                                                                |        |
|--------------------------------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> Nothing. Each field is described in whatever wording the author chose.                                | 1 pt.  |
| <input type="checkbox"/> An in-house term list is used for part of the dictionary.                                             | 2 pts. |
| <input type="checkbox"/> Standardized against a local or organizational standard.                                              | 3 pts. |
| <input type="checkbox"/> Standardized against national definitions.                                                            | 4 pts. |
| <input type="checkbox"/> Based on international standards, with the mapping documented, for example, SNOMED CT, LOINC, or ICD. | 5 pts. |

*You can support this answer with: your mapping table from local fields to the standard used.*

### 5. How is the dictionary kept up to date?\*

- |                                                                                                         |        |
|---------------------------------------------------------------------------------------------------------|--------|
| <input type="checkbox"/> It is not maintained.                                                          | 1 pt.  |
| <input type="checkbox"/> Updated when somebody notices an error.                                        | 2 pts. |
| <input type="checkbox"/> Updated whenever the data model changes.                                       | 3 pts. |
| <input type="checkbox"/> Reviewed on a fixed schedule, with a version number and date.                  | 4 pts. |
| <input type="checkbox"/> Reviewed on a fixed schedule, versioned, with a change log published to users. | 5 pts. |

*You can support this answer with: the version and date on your current dictionary, its change log.*

### 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.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>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.

|                             |                                                                                                   |
|-----------------------------|---------------------------------------------------------------------------------------------------|
| <b>Level 1</b><br>Initial   | No dictionary, no definitions, visible only internally, in your own wording, not maintained.      |
| <b>Level 2</b><br>Emerging  | Scattered definitions for some fields, available on request, in-house terms, fixed when noticed.  |
| <b>Level 3</b><br>Defined   | One dictionary covering all fields for secondary use, published online, updated with the model.   |
| <b>Level 4</b><br>Managed   | Value sets, units, and ranges included; national definitions; versioned and reviewed on schedule. |
| <b>Level 5</b><br>Optimized | International standards with documented mapping, machine-readable, with a published change log.   |

**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: Adapted HDRUK / EHDS. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 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

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

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.

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.

# Maturity Overall Score Calculation

Please mark the level you reached in each of the ten dimensions. Take each level from the corresponding dimension questionnaire.

| Dimension                                             | Level 1<br>Initial       | Level 2<br>Emerging      | Level 3<br>Defined       | Level 4<br>Managed       | Level 5<br>Optimized     | N/A                      |
|-------------------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Dimension 1: Data Collection Process                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 2: Data Management — Governance <b>CORE</b> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 3: Data Management — Infrastructure         | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 4: Data Provenance <b>CORE</b>              | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 5: Data Access <b>CORE</b>                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 6: Data Analytics Environment               | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 7: Data Enhancement — Augmentation          | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Dimension 8: Data Enhancement — Enrichment            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Dimension 9: Data Model <b>CORE</b>                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |
| Dimension 10: Data Dictionary <b>CORE</b>             | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | —                        |

**TOTAL MATURITY SCORE**

Add up your ten dimension levels. The total runs from 10 to 50. It is the figure to quote when your result is compared with other assessments.

**AVERAGE MATURITY**

Divide the total by the number of dimensions you answered (10, or 8 if you marked Dimensions 7 and 8 as not applicable). Keep one decimal, which is what shows progress year on year.

**YOUR EHDS BASELINE**

Write down the **lowest** level you reached across the five core dimensions marked **CORE**, namely 2, 4, 5, 9, and 10. These are what a Health Data Access Body needs from you, so the weakest of them is what limits you today.

**YOUR NARRATIVE**

A short supporting narrative is expected for each dimension: the practices and tools behind the score, recent improvements, and the gaps that explain the lower scores. The evidence lines on each questionnaire give you the material.

|                                                            |                                                              |                                                            |                                                         |                                                            |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|
| <b>Level 1</b><br><b>Initial</b><br>Starting out · 1 Point | <b>Level 2</b><br><b>Emerging</b><br>Written down · 2 Points | <b>Level 3</b><br><b>Defined</b><br>In practice · 3 Points | <b>Level 4</b><br><b>Managed</b><br>Measured · 4 Points | <b>Level 5</b><br><b>Optimized</b><br>Improving · 5 Points |
|------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------|

**How to read your result.** Your ten dimension levels together form your maturity profile, and that profile, not a single number, is the useful result. The two dimensions with the lowest levels are where the next improvement will pay off most. **Dimensions 7 and 8 (Data Enhancement)** describe added value rather than an entry requirement for the EHDS; if they do not apply to your organization, mark N/A.

**Why the core dimensions matter.** Governance (2), Provenance (4), Access (5), Data Model (9), and Data Dictionary (10) are the areas a Health Data Access Body relies on to accept, describe, and pass on your data. The obligation for data holders to make data available under Chapter IV of the EHDS Regulation (EU) 2025/327 applies from 26 March 2029, with genomic, molecular and clinical trial data following from 26 March 2031.

**This is a self-assessment.** The model is applied at organizational level, so one assessment per data holder is expected; a second is only needed where maturity differs markedly between data domains, for example, clinical versus genomic data. It is not a certificate and not the EHDS data quality and utility label.

## References\*

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.

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.

Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space. OJ L, 2025/327, 5.3.2025.

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.