

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?*

- | | |
|---|--------|
| ☐ No techniques are applied. | 1 pt. |
| ☐ Some techniques are applied case by case, for example, cleaning, deduplication, unit conversion. | 2 pts. |
| ☐ Defined, documented techniques are applied. | 3 pts. |
| ☐ The techniques are managed: versioned, tested, and applied consistently. | 4 pts. |
| ☐ 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?*

- | | |
|--|--------|
| ☐ None of it. | 1 pt. |
| ☐ A single dataset or one project. | 2 pts. |
| ☐ Our main datasets. | 3 pts. |
| ☐ All datasets offered for secondary use. | 4 pts. |
| ☐ 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?*

- | | |
|---|--------|
| ☐ Not managed. Each person does it their own way, and nothing is kept. | 1 pt. |
| ☐ One person's script or routine is kept but not documented, so only they can repeat it. | 2 pts. |
| ☐ Documented and repeatable by a colleague. | 3 pts. |
| ☐ Version-controlled, tested, and reviewed. | 4 pts. |
| ☐ 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?*

- | | |
|---|--------|
| ☐ No mapping, and none is planned. | 1 pt. |
| ☐ A target model has been chosen, and the mapping is planned, but nothing is in use yet. | 2 pts. |
| ☐ Some key fields are mapped to a recognized model. | 3 pts. |
| ☐ Most of the dataset is mapped, with a documented mapping specification. | 4 pts. |
| ☐ 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?*

- | | |
|--|--------|
| ☐ We do not check. | 1 pt. |
| ☐ We look at the output informally. | 2 pts. |
| ☐ We compare key figures before and after. | 3 pts. |
| ☐ We apply documented quality checks and record the outcome. | 4 pts. |
| ☐ 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.

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.

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