Data Quality in EHDS

- Painful reality of individual health data
- EHDS Data Quality & Utility Level
- How AIDAVA can help
- COmparison between QUANTUM Label and AIDAVA DQ score

Why do I need data quality for my health record ?

The painful reality about health data

- **Scattered** across multiple systems (±40% hospital, ±60% others)
- **Heterogeneous**: different standards (HL7, openEHR..): no interoperability
- **Not readily processable**: up to 80% in narrative format (±40% full text, ± 40% chunks of text)
- **Not well documented:** most EHRs old (90's) with limited documentation
- **Redundant & error-prone:** up to 30% redundancy, 40% of errors (10% potentially life threatening)

Data Quality enhancement of INDIVIDUAL records is not scalable as long as they are not interoperable and reusable,

Why do I need data quality for my health record ?
The consequence of suboptimal health data

Fragmentation and lower quality of care:
providers make decisions on fragmented and potentially incorrect data

- Providers do not have enough time
  - As medical problem increases, size of records increases
  - Shortage of providers
- Providers may miss important information leading to
  - Redundancies, requesting data already available *(avoidable cost)*
  - Medical errors with complications *(avoidable errors).*

Bad quality in individual records impacts the whole healthcare system, not just the patients

EHDS: Article 78. Data quality and utility label

1. Datasets made available through HDABs may have a Union **data quality and utility label** applied by the health data holders.

2. Datasets with electronic health data collected and processed with the support of Union or national public funding shall have a **data quality and utility label** covering the elements set out in paragraph 3.

3. The data quality and utility label shall cover the following elements, where applicable:

(a) for **data documentation**: metadata, support documentation, the data dictionary, the format and standards used, the source of the data and, where applicable, the data model;
(b) for **assessment of technical quality**: the completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;
(c) for **data quality management processes**: the level of maturity of the data quality management processes, including review and audit processes, and bias examination;
(d) for **assessment of coverage:** the period, population coverage and, where applicable, representativity of the population sampled, and the average timeframe in which a natural person appears in a dataset;
(e) for **information on access and provision**: the time between the collection of the electronic health data and their addition to the dataset and the time needed to provide electronic health data following the issuing of a data permit or a health data request approval;
(f) for **information on data modifications**: merging and adding data to an existing dataset, including links with other datasets.

4. Where a HDAB has reason to believe that a data quality and utility label might be inaccurate, it shall assess whether the dataset covered by the label meets the quality requirements forming part of the elements of the data quality and utility label as referred to in paragraph 3 and, in the event the dataset does not meet the quality requirements, shall revoke the label.

5. ….

Implementing data quality at individual level

How can AIDAVA Help: hypothesis

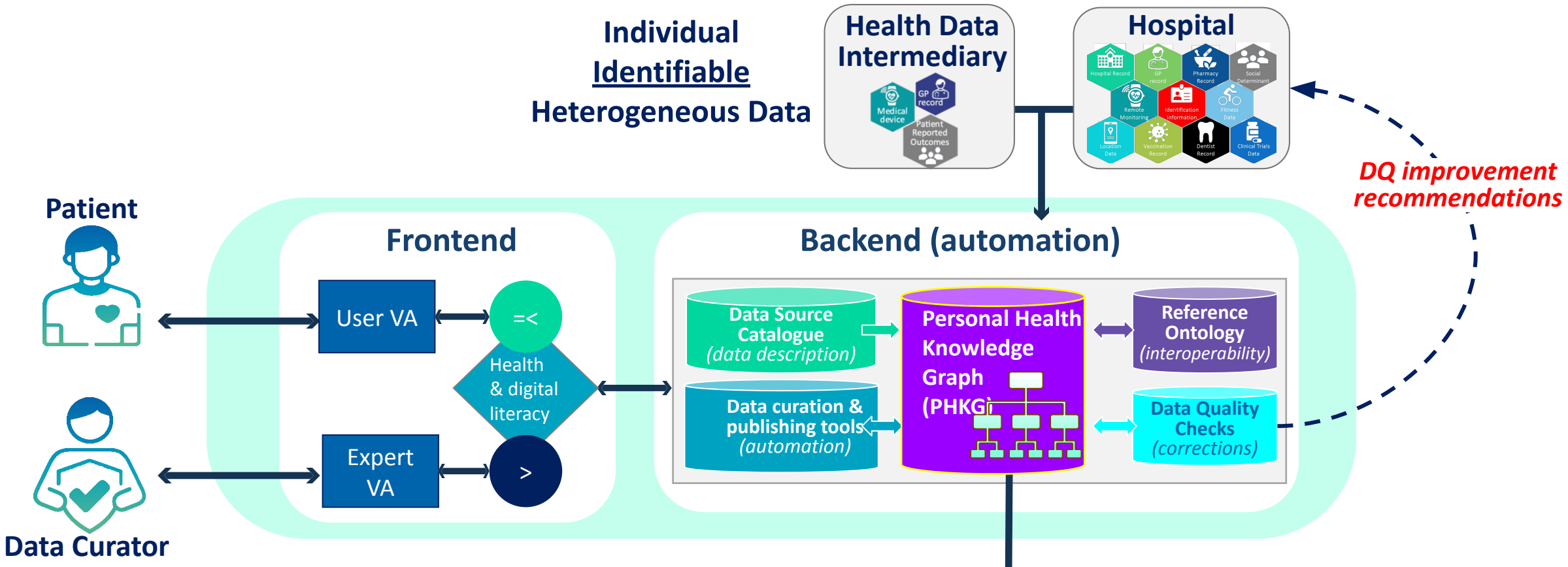
- Data quality assessment of individual records is not scalable as data sources are scattered, heterogeneous.

- An alternative approach is to assess quality of the curated record
  - integrating all available data ACROSS data sources,
  - available in a standardized semantic format, a “personal health knowledge graph”, based on an ontology aligning existing standards.

This is what AIDAVA aims to achieve !

Implementing data quality at individual level

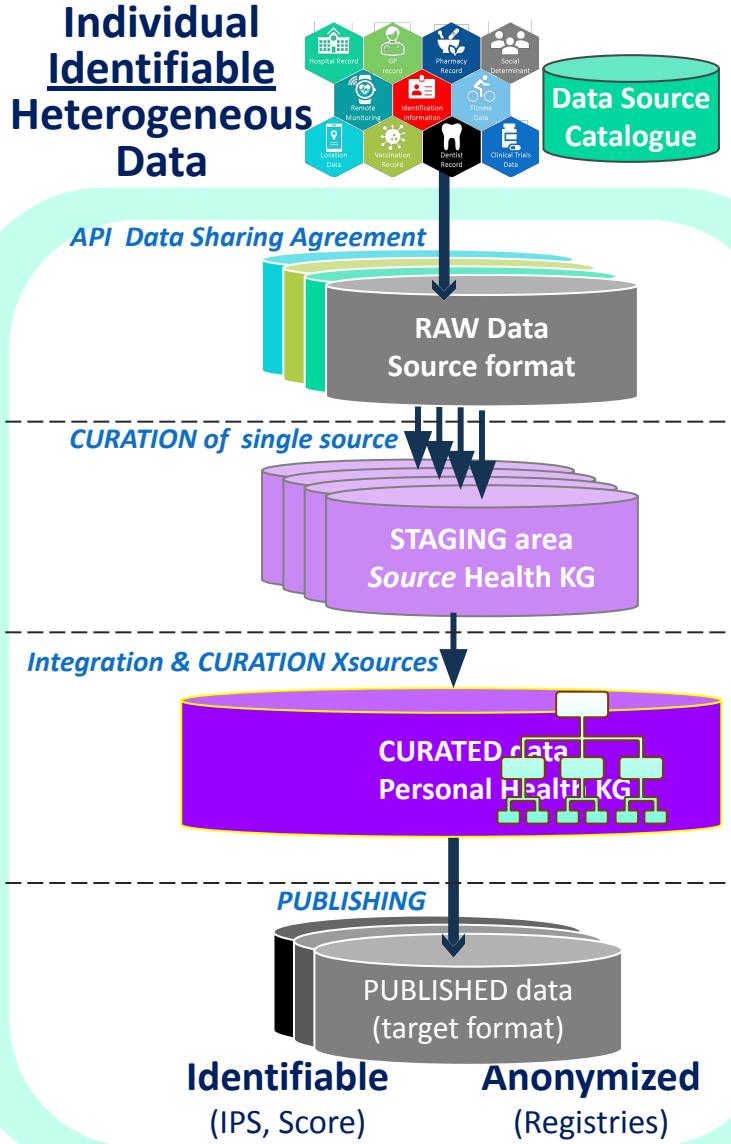
Support of AI based Data Curation Virtual Assistant (AIDAVA)



January 2026

Implementing data quality at individual level

Data quality at individual level: How can AIDAVA Help



Implementation to Data Quality differs based on the type of data

L1. Data Source: Heterogenous formats

Scalability possible for context metadata

Structural checks on context

L2. Curated Single Data Source (SHKG)

Standard format (1 source)

=> scalable checks

L3. Curated Integrated Record (PHKG)

Standard format (complete record), scalable checks

SHACL DQ Checks on patient data with Human intervention for correction

L4.Published Data - consistency to format

- if individual output: linked to L3
- If population output: Quantum label

How to measure data quality

Population (QUANTUM) versus Individual (AIDAVA) DQ

Feature	**QUANTUM** Quality Assurance (QA)	**AIDAVA** Quality Control (QC)
Scope	BROAD: Population level Involves the entire organisation and team across all patients	NARROW: Individual level Focused on inspection by dedicated staff or tool on specific data of a patient
Objective	Prevents defects	Detects defects
Orientation	Process-oriented: focuses on the process within organisation	Data Product-oriented: focuses on individual patient data
Timing	Proactive (before and during production)	Reactive (after the product - i.e. data - is collected)

QUANTUM label and AIDAVA DQ score are complementary

Conclusions:
Health Data Quality, What is My Return as A Citizen?

Health Data Quality

- at **population level**: key for research, limited value for a citizen
- at **individual level:** needed for integrated care and personalized medicine

Health Data Quality assessment at **individual** level comes with challenges that could be overcome if we work at integrated, **curated** personal health record.

Some citizens and patients are ready to contribute, if they are provided with a simple infrastructure to increase the quality of their data.

AIDAVA aims to support citizens in improving the quality of their data ensuring quality personal care