

AI powered Data Curation & Publishing Virtual Assistant

*Optimize interoperability & quality of health data to increase data sharing and reuse
across Clinical Registries & Personal Data Intermediaries*

EHDS Implementation **Value of a “digital twin” of patient records**

*How can AI driven data curation help to make
EHDS affordable*

As part of the AIDAVA Business Plan we assessed the impact on AIDAVA-like tools on EHDS

- **Financial aspect/Business case)**

- 2022 EHDS Impact Assessment: Recovery and Resilience Fund (RRF) will solve health data interoperability across MS
=> As of december 2025: less than €9B (64%) of the budgeted €14B has been used
- No solid figure on total cost of EHDS (complexity due to multifactorial aspects)
 - More realistic to compare relative cost of different implementation scenarios
 - **Findings: a disproportionate part of the cost is to be borne by Data Holders without tangible rewards.**

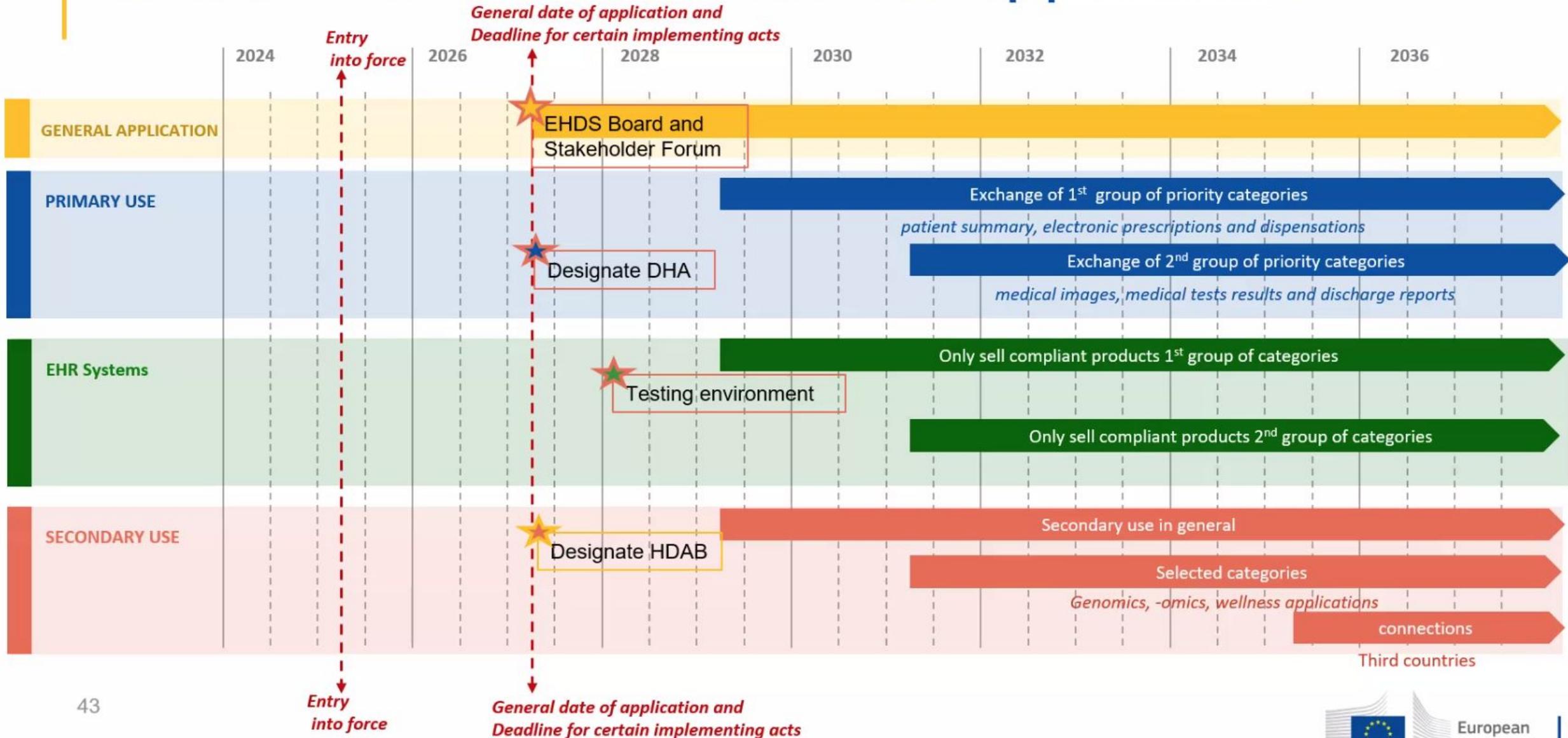
- **EU Market Maturity and Digital Health Equity**

- Digital Health Maturity - and level of interoperability - differs widely across EU
- Cost of EHDS compliance will be higher for low maturity countries
- If EHDS compliance is too expensive for Data Holders
 - Certain countries and rural hospitals with financial difficulties will struggle
 - MS may not be ready to provide additional financial support
 - **Risk to generate a digital health divide across EU**

Our results indicate that AIDAVA could decrease EHDS compliance costs, increase benefits for Data Holders and ensure EU Digital Health equity by maintaining a high quality, interoperable and reusable digital twin of patient records

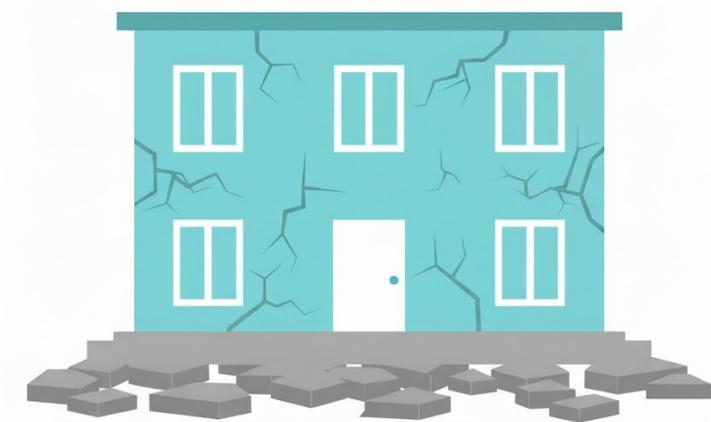
Overview of EHDS

EHDS – Overall timeline for application



Cost of EHDS: is it affordable ?

EHDS in current reality



- Data non interoperable & reusable
- Costly transformation of data
- No guarantee of individual record quality (up to 40% of errors)
- Suboptimal personalised (AI) care
- Suboptimal medical decision making

EHDS as we build it



- Unclear semantic Data foundation (maintained manually with ETLs ?)
- Unclear responsibility for data transformation for secondary use (SPE ? or Data Holders ?)
- Costly recurrent data curation for
 - XBorders EEHRxF (primary use)
 - Permit/query (secondary use)

M. Loughton, "Europe's EHDS Readiness: Progress and Persistent Hurdles," The Financial Analyst - June 2025

270 European healthcare IT leaders
1,305 EHR/EPR users across 17 EU countries

Maturity

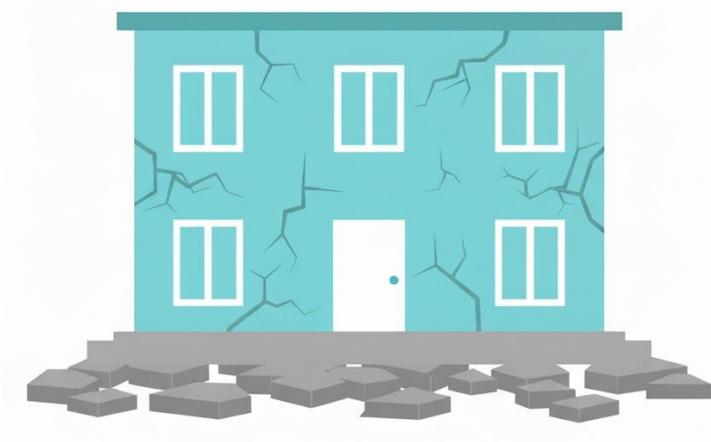
- **88%** familiar / involved in EHDS
- **76%** piloting / using HL7 FHIR,

Key barriers

- **52% lack of interoperability.**
- **47% legacy systems.**
- **43% data quality issues.**

Cost of EHDS: is it affordable ?

EHDS in current reality



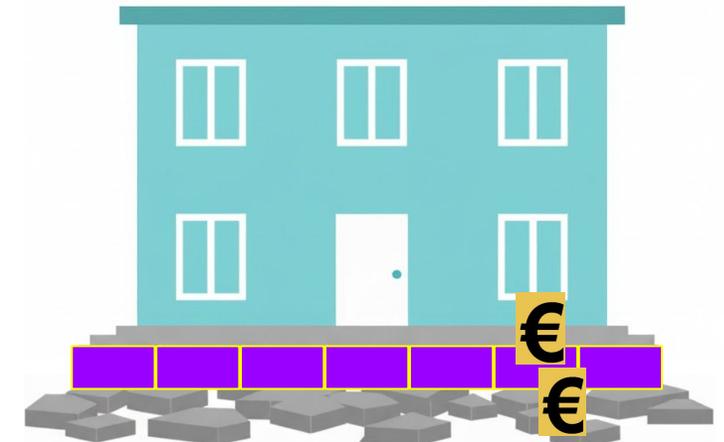
- Data non interoperable & reusable
- Costly transformation of data
- No guarantee of individual record quality (up to 40% of errors)
- Suboptimal personalised (AI) care
- Suboptimal medical decision making

EHDS as we build it



- Unclear semantic Data foundation (maintained manually with ETLs ?)
- Unclear responsibility for data transformation for secondary use (SPE ? or Data Holders ?)
- Costly recurrent data curation for
 - XBorders EEHRxF (primary use)
 - Permit/query (secondary use)

EHDS as we could build it



- Solid semantic Data foundation (maintained by AI): *“digital twin” of patient record in knowledge graph*
- Data Holders transform ONCE patient data for primary & secondary use
- Just-in-time
 - generation of EEHRxF
 - extraction of interoperable datasets

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 ($\pm 40\%$ hospital, $\pm 60\%$ others)
- **Heterogeneous**: different standards (HL7, openEHR..): no interoperability
- **Not readily processable**: up to 80% in narrative format ($\pm 40\%$ full text, $\pm 40\%$ 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

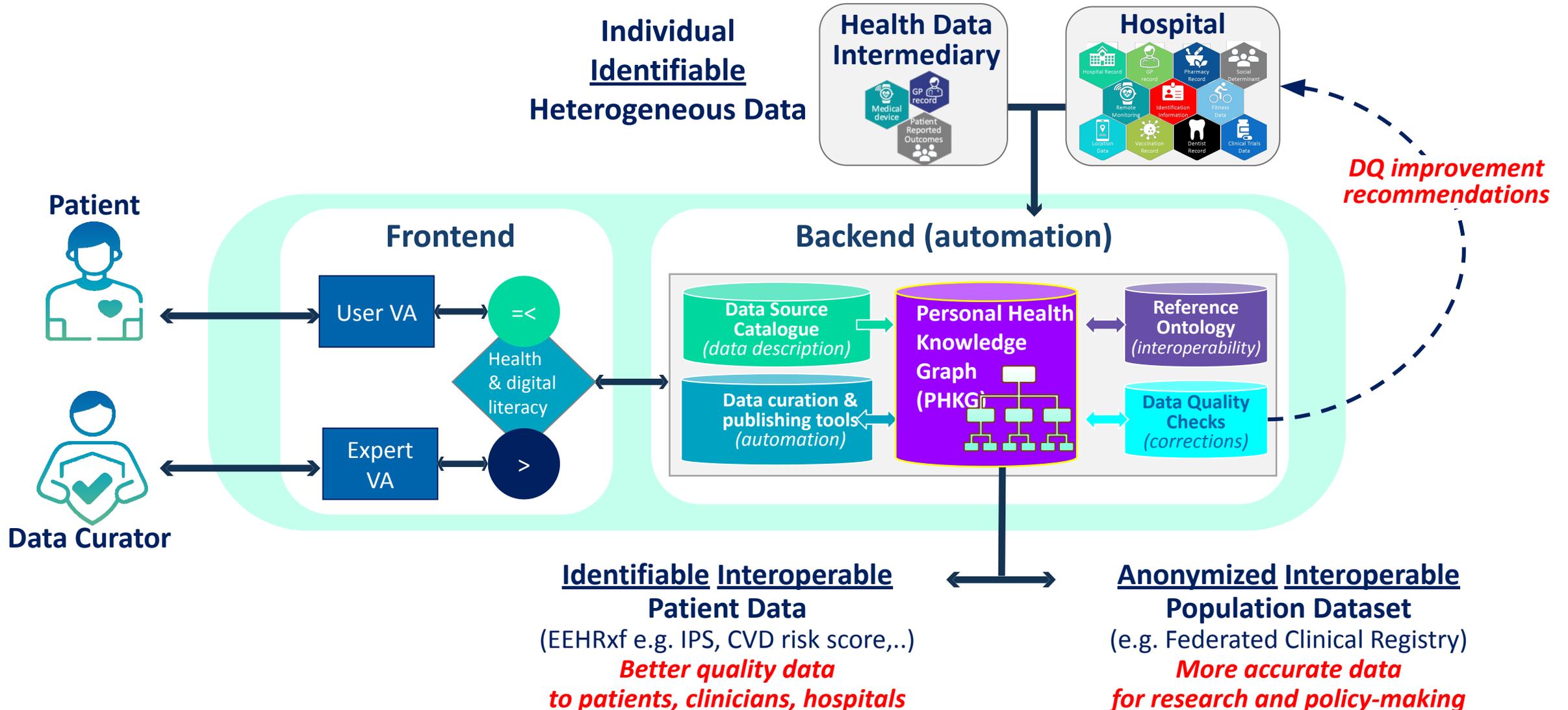


- Data quality assessment of individual records is not scalable as data sources are scattered, heterogeneous.
- An alternative approach is to assess quality of the curated record
 - integrating all available data ACROSS data sources,
 - available in a standardized semantic format, a “personal health 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)



Implementing data quality at individual level

Data quality at individual level: How can AIDAVA Help



**Individual
Identifiable
Heterogeneous
Data**



Implementation to Data Quality differs based on the type of data

API Data Sharing Agreement



L1. Data Source: Heterogeneous formats
Scalability possible for context metadata

Structural checks on context

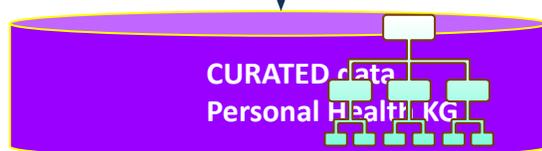
CURATION of single source



L2. Curated Single Data Source (SHKG)
Standard format (1 source)
=> scalable checks

SHACL DQ Checks on patient data with Human intervention for correction

Integration & CURATION Xsources



L3. Curated Integrated Record (PHKG)
Standard format (complete record),
scalable checks

PUBLISHING



L4. Published Data - consistency to format

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

Identifiable (IPS, Score) **Anonymized** (Registries)

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.

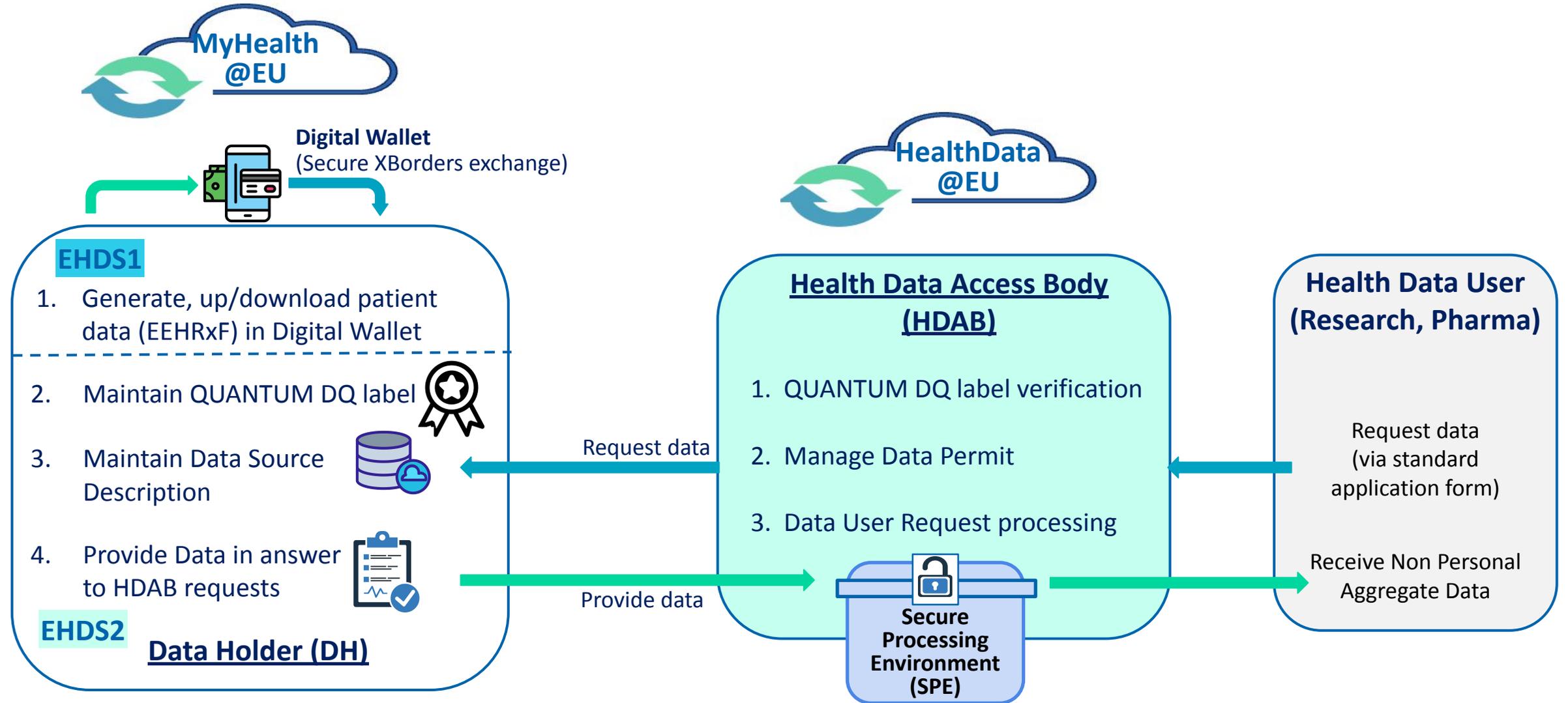
**AIDAVA aims to support citizens in improving the quality of their data
ensuring quality personal care**

Is EHDS affordable

- Current data flows
- Requirements for data holders and authorities
- Cost estimates

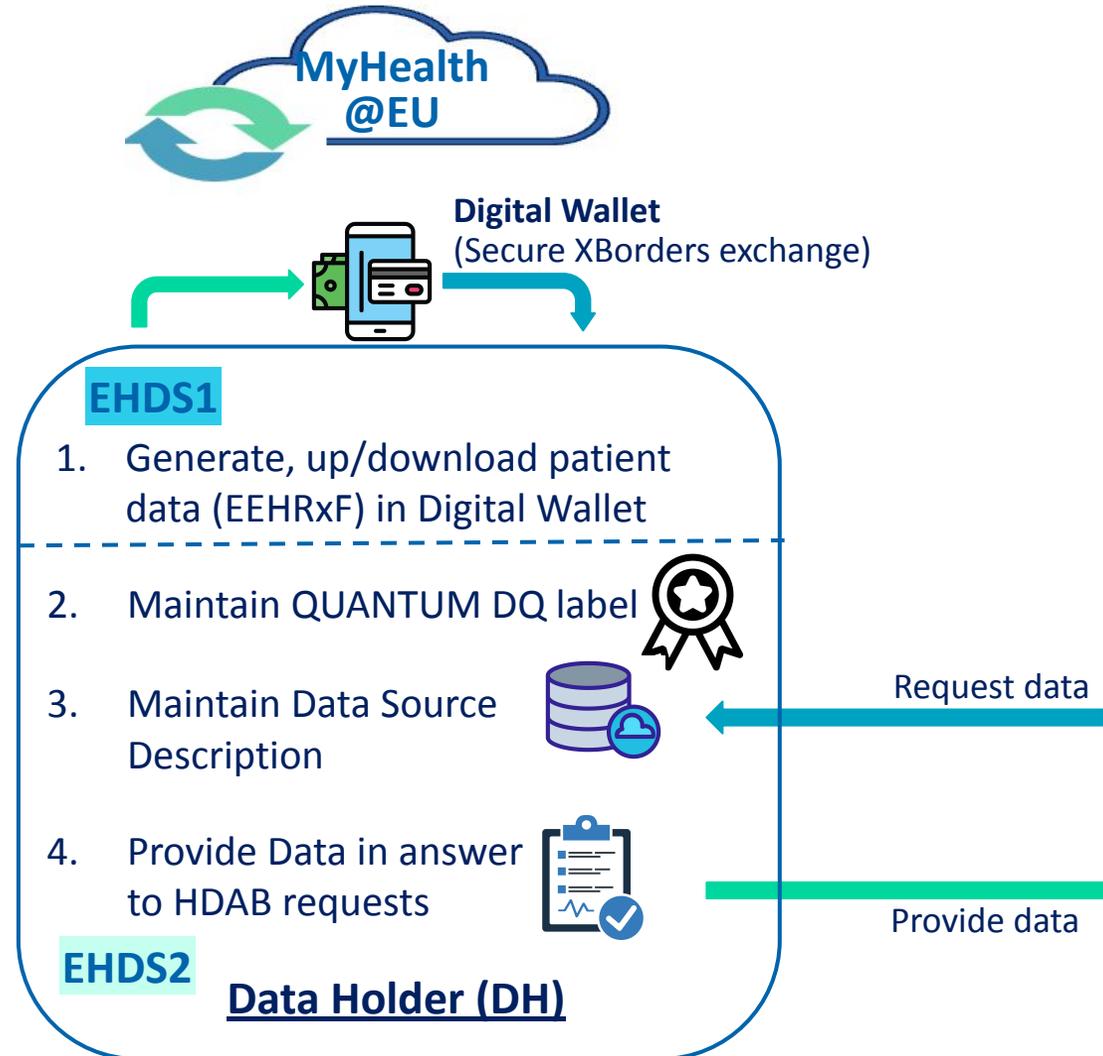
Deep dive in EHDS Data Flows

Overview



Deep dive in EHDS Data Flows

Requirements for Data Holders



Needed by Data Holders

- In depth understanding of data (attribute level) across all hospital's IT subsystems
- **Provide EEHRxF compliant data**
Develop/buy tools for
 - data extraction of a patient relevant data
 - **(manual ?) data quality checks**
 - transformation into HL7 FHIR (EEHRxF) format
- **Provide data in answer to HDABs request**
Develop/Buy Tools to support
 - extraction of data on demand
 - **transformation/curation in required format**
- **Needed skills:** HL7 FHIR, data scientists

In case of vendors managed systems,

- Pay for additional services required by EHDS
- May require integration across vendors (& managing inconsistencies)

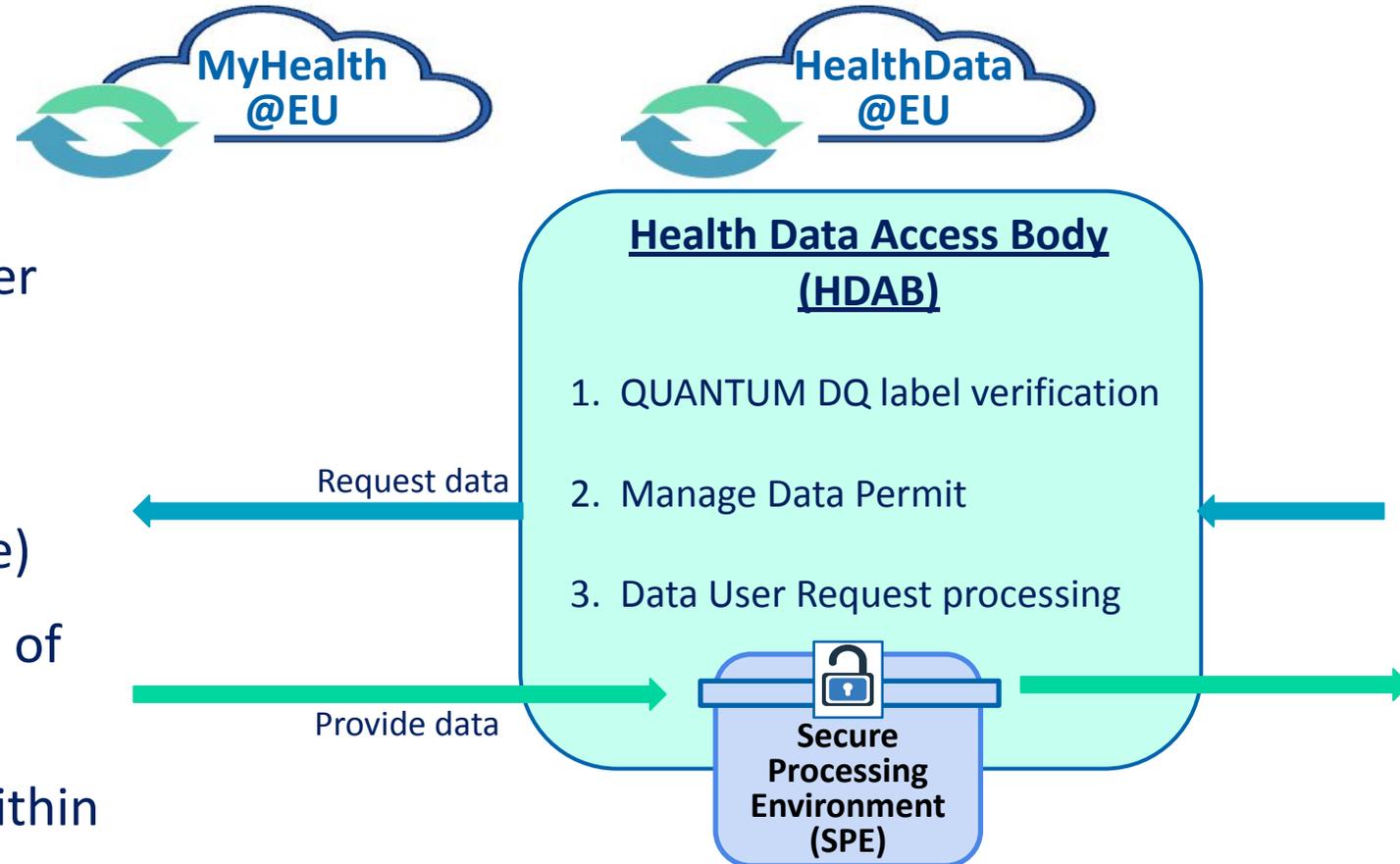
Deep dive in EHDS Data Flows

Overview - Primary & Secondary Use



Needed by Authorities

- Set up organisations (HDABs & SPEs)
- Set up EU infrastructure for cross border exchanges
 - Digital Wallet (EUDI/ eIDAS)
 - MyHealth@EU (for primary use)
 - HealthData@EU (for secondary use)
- Support development and deployment of EEHRxF implementation guides
- Support / supervise implementation within Data Holders (training, funding)



Cost ESTIMATES:

10 hospitals, 5 years, per year: 1250 records, 1250 EEHRxF, 100 HDAB queries



		Cost estimates (KEUR)				
	Steps	One of	Per unit	Yearly	Total Period	%
Data Holders	Curation	€0	€0	€0	€0	0.0%
	EEHRxf	€300	€0.10	€425	€2,125	43.5%
	Dataset Description	€50	€0.25	€52	€258	5.3%
	Answer to queries	€0	€5	€500	€2,500	51.2%
	Total/Data Holder				€4,883	
TOTAL ALL DH					€48,825	
<hr/>						
HDAB	Curation	€0	€50	€5,000	€25,000	
<hr/>						
TOTAL (KEUR)	10 hospital/ 5 years				€73,825	
<hr/>						
TOTAL (Billion EUR)	10.000 hospital/ 5 years				€738	
TOTAL (Trillion EUR)	10.000 hospital/ 5 years				€0.7	

Numbers in red: assumptions being verified

Cost ESTIMATES:

10 hospitals, 5 years, per year: 1250 records, 1250 EEHRxF, **200** HDAB queries



		Cost estimates (KEUR)				
	Steps	One of	Per unit	Yearly	Total Period	%
Data Holders	Curation	€0	€0	€0	€0	0.0%
	EEHRxf	€300	€0.10	€425	€2,125	28.8%
	Dataset Description	€50	€0.25	€52	€258	3.5%
	Answer to queries	€0	€5	€1,000	€5,000	67.7%
	Total/Data Holder				€7,383	
	TOTAL ALL DH				€73,825	
<hr/>						
HDAB	Curation	€0	€50	€10,000	€50,000	
<hr/>						
TOTAL (KEUR)	10 hospital/ 5 years				€123,825	
<hr/>						
TOTAL (Billion EUR)	10.000 hospital/ 5 years				€1,238	
TOTAL (Trillion EUR)	10.000 hospital/ 5 years				€1.2	

Numbers in red: assumptions being verified

EHDS implementing acts push a lot of requirements on 'data holders' (and more specifically hospitals). This makes sense as hospitals are the 'holders' of the patient data, at the core of health systems.

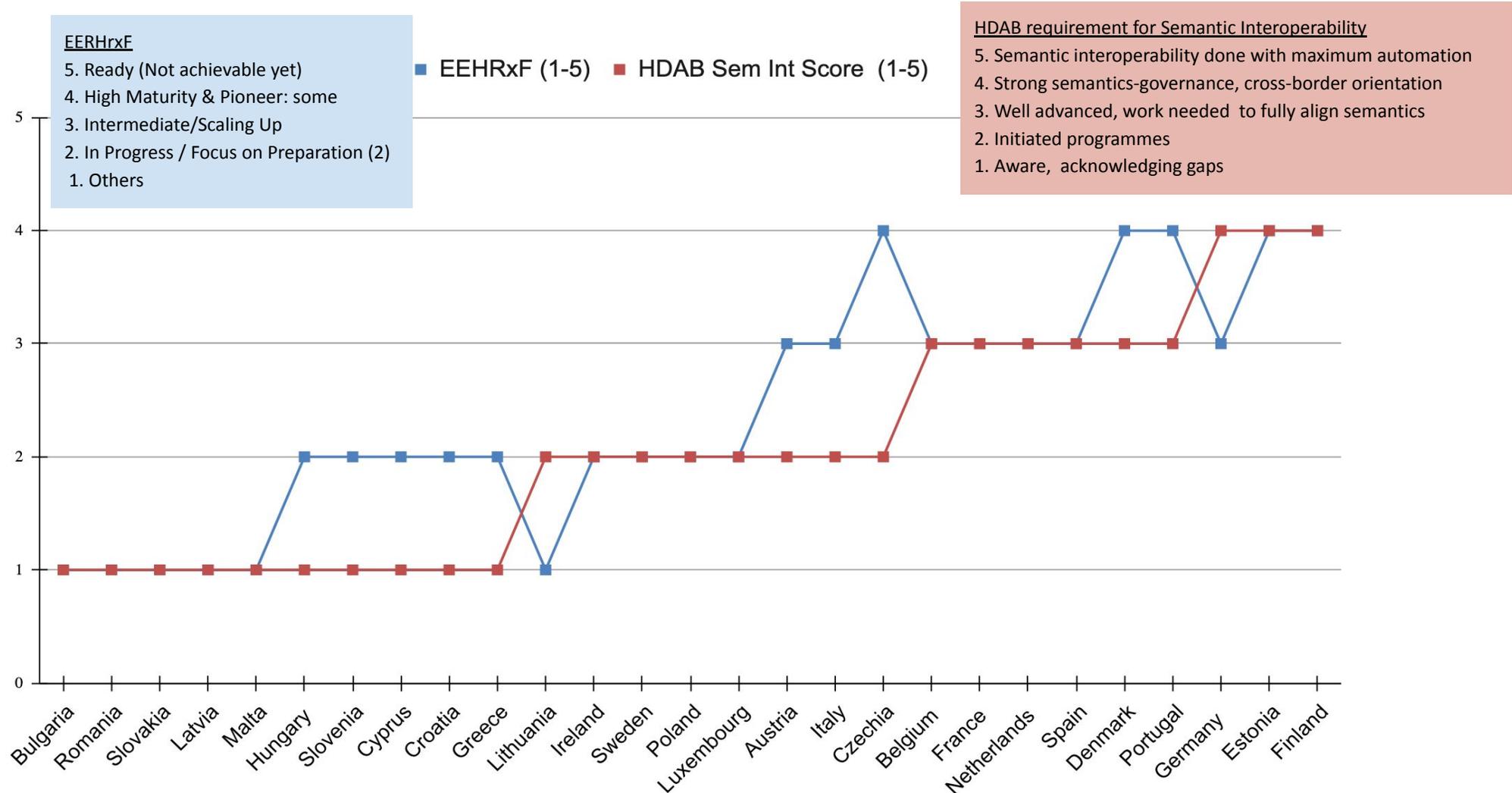
However, without supporting tools and automation, the additional burden (cost, needed skills) to hospitals might be unbearable for most and lead to EU digital divide in health. In addition, the cost rises proportionally with the number of requests.

EHDS may increase existing digital health inequity across EU

- Current data flows
- Requirements for data holders and authorities
- Cost estimates

Risk of digital health divide in EU ?

EHDS Maturity: EEHRxF generation & answer to HDAB requests



(per chatGPT/ Gemini - not validated by MS)

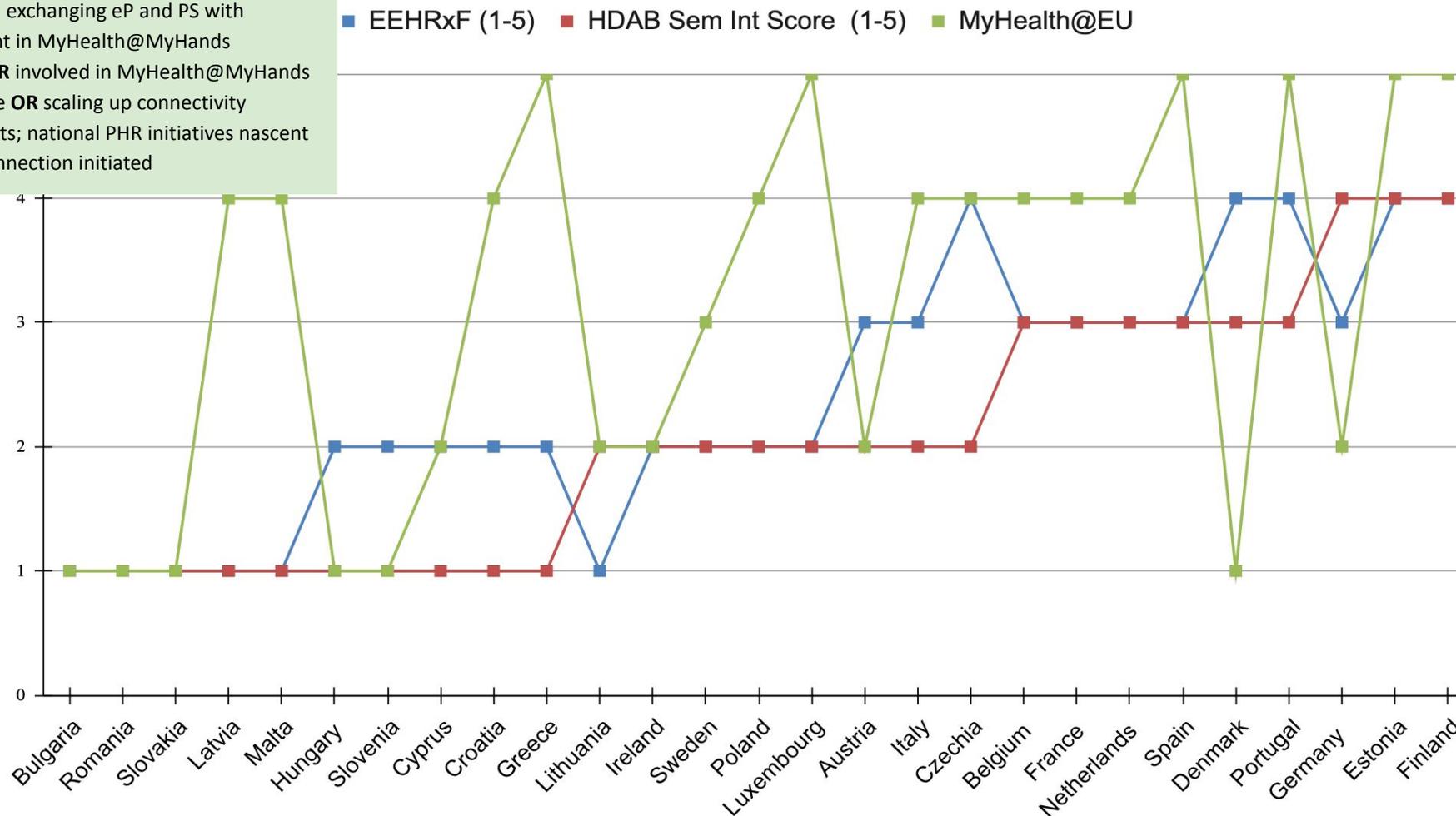
Risk of digital health divide in EU ?

EHDS Maturity: MyHealth@EU



MyHealth@EU -

5. Full Pioneer & Citizen Focus: exchanging eP and PS with multiple countries, involvement in MyHealth@MyHands
4. Scaling: exchanging eP/PS **OR** involved in MyHealth@MyHands
3. Single Service Live : 1 service **OR** scaling up connectivity
2. Audit/Pilot: Connected/ pilots; national PHR initiatives nascent
1. In Preparation: Technical connection initiated



(per chatGPT/ Gemini - not validated by MS)

Risk of digital health divide in EU ?

Conclusion



There is a disparity across Europe in Digital Health maturity,

- High maturity countries like Estonia and Finland should be able to meet EHDS requirements
- Low maturity countries may struggle for a long time

EHDS will add additional financial burden to Data Holders, with no well defined benefits

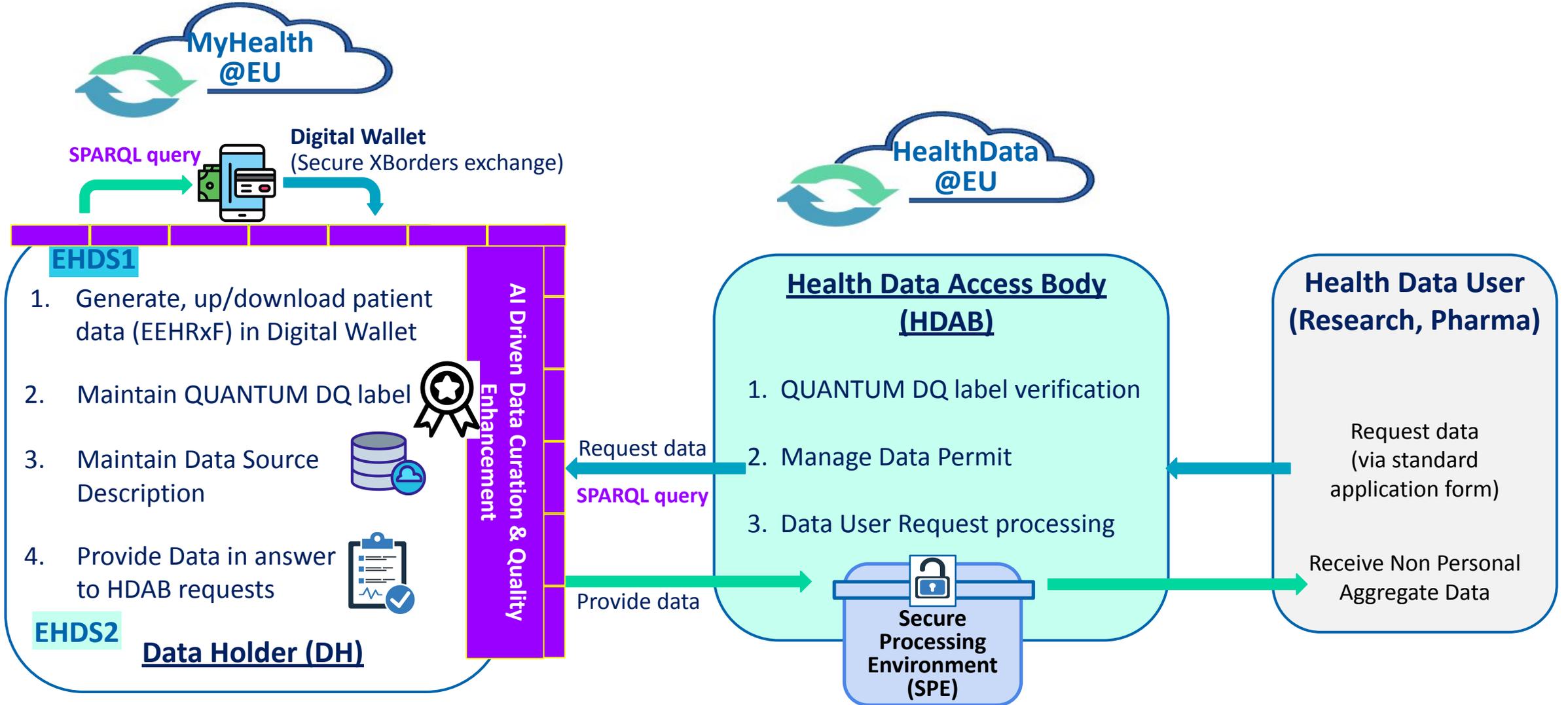
- Low maturity countries may see EHDS as an **painfull, unfunded administrative mandate**
- And potentially even decide not to participate
- This increases the risk of **Digital Health Divide** (against the core values of Europe).

EHDS Data Flows with a 'Digital Twin' of patient health records

- EHDS data flows with a Digital Twin
- How to generate a Digital Twin with AIDAVA
- Benefits of Digital Twin
- Critical Success Factors to AIDAVA Scalability

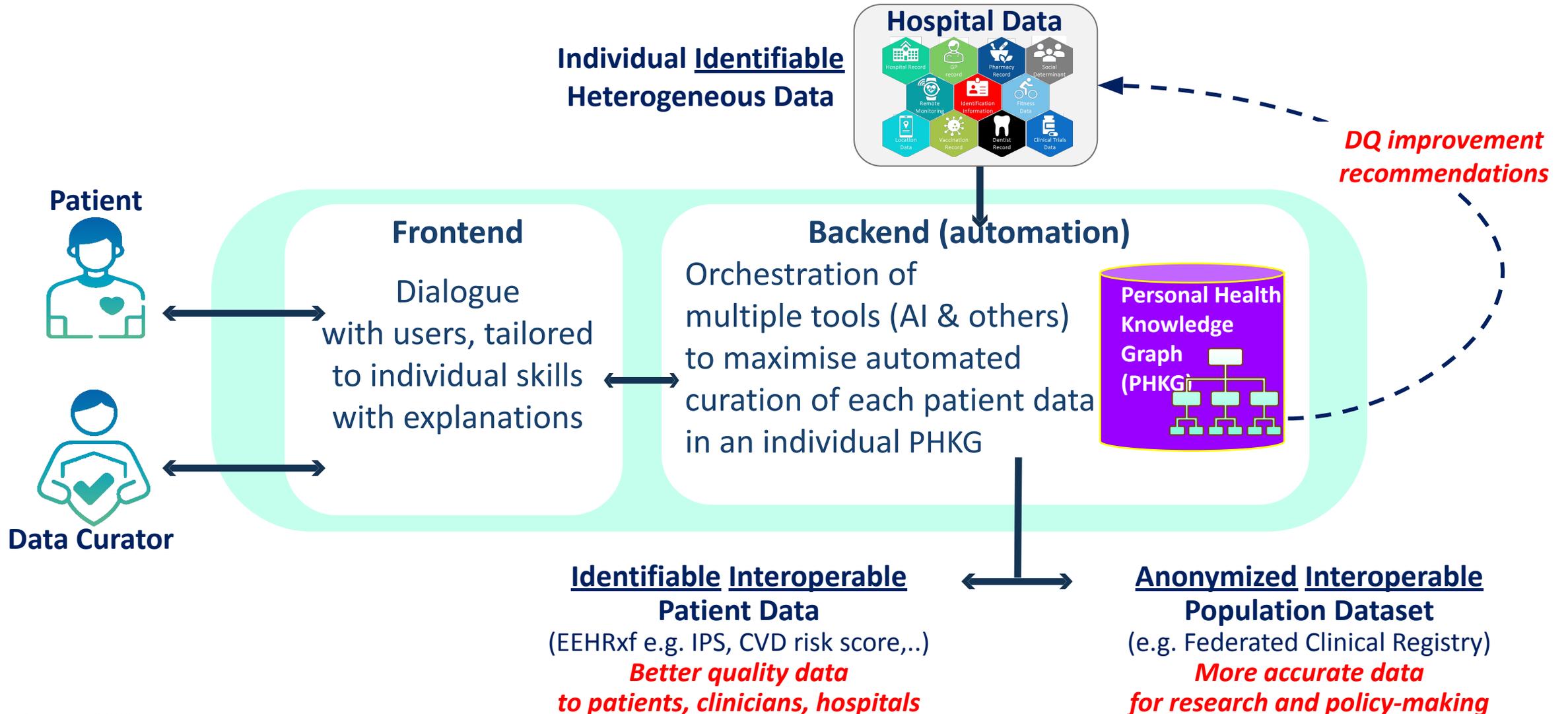
Deep dive in EHDS Data Flows

Adding AI maintained high-quality “digital twin” of patient records



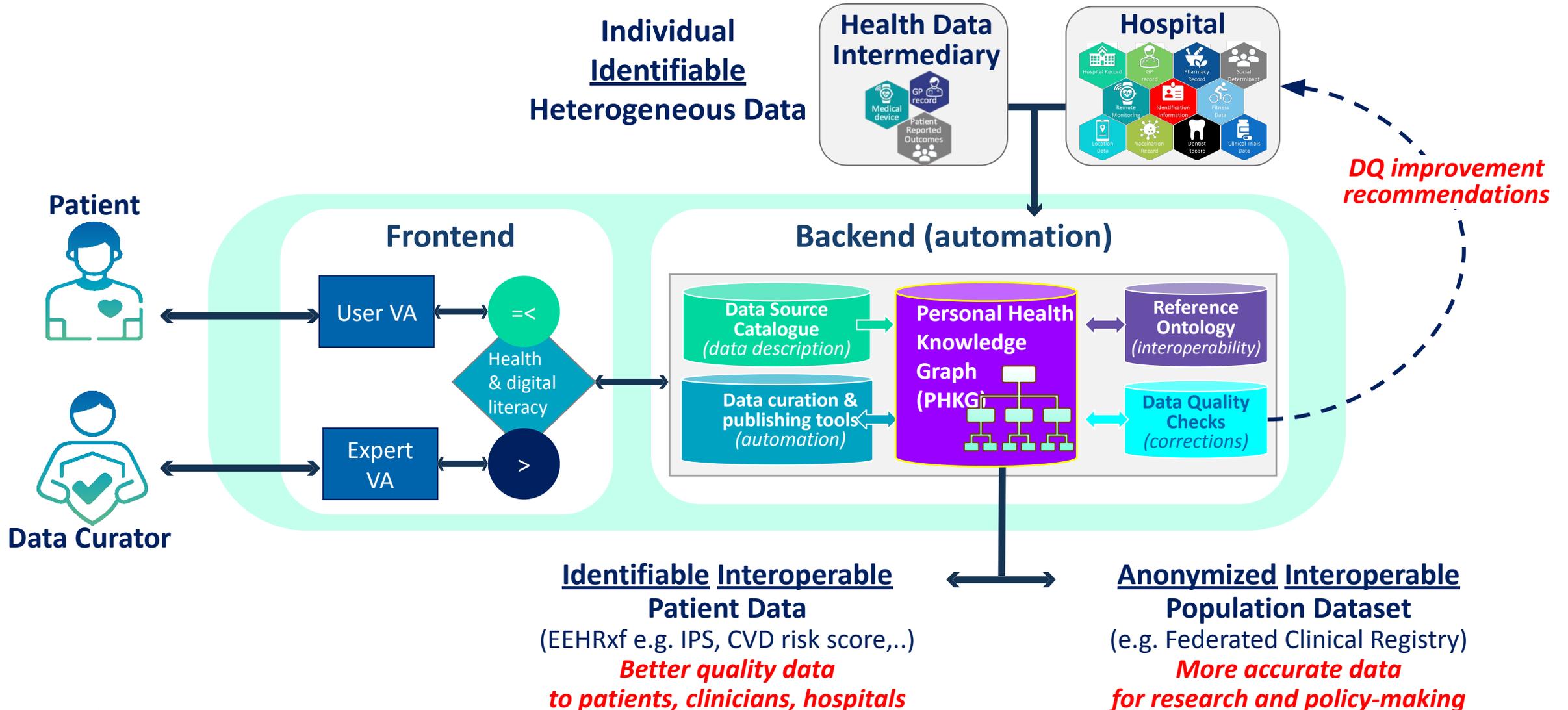
How to deliver a “digital twin” (PHKG)

Need support of AI based Data Curation Virtual Assistant (AIDAVA)



How to deliver a “digital twin” (PHKG)

Need support of AI based Data Curation Virtual Assistant (AIDAVA)



How to deliver a “digital twin” (PHKG)

Readiness for production use of AIDAVA



- AIDAVA prototype is a TRL 5-TRL 6 prototype
 - based on **21 AI and non AI tools** orchestrated through workflow to support automation in mapping, curation and data quality enhancement
 - available in **3 languages** (Estonian, German, Dutch)
- First generation (G1 - June 2024) tested with 78 patients in 4 hospitals
 - **documentation of available data** within an organisation is hard (knowledge & skills)
 - **automation in curation** is possible
 - dramatically decrease workload of current curation
 - increase interoperability of data (no risk of difference due to human mapping)
 - **major enhancements** required
 - AI tools (NLP, medical coding, mapping) across all languages
 - Explainability (context, language) in case of issues
- G2 (December 2025) being tested with 90 patients in 4 hospitals
 - Serious improvements to AI tools (more is needed)

Benefits of high quality data “digital twin”

Quality = interoperable, correct/accurate, complete AND reusable



EHDS compliance at lower cost for Data Holders

- Smoother answer to queries: HDAB data requests (“permit”) transformed in SPARQL query executable everywhere without additional transformation
- Easier generation of EEHRxF : implementation guide for each critical data category translated into SPARQL query executable everywhere
- Limited need for new skills/staff: data scientists to support AI driven curation

Additional benefits for Data Holders: Data quality at INDIVIDUAL record level

- Primary use: Improved patient outcome and personalised (AI) care: individual data in AI model ready format (extracted from digital twin)
- Secondary use: Automatic maintenance of local (or federated) clinical registries and smoother participation to Clinical Trials for all hospitals (including rural ones)

Additional benefits for Authorities: Faster & more reliable evidence

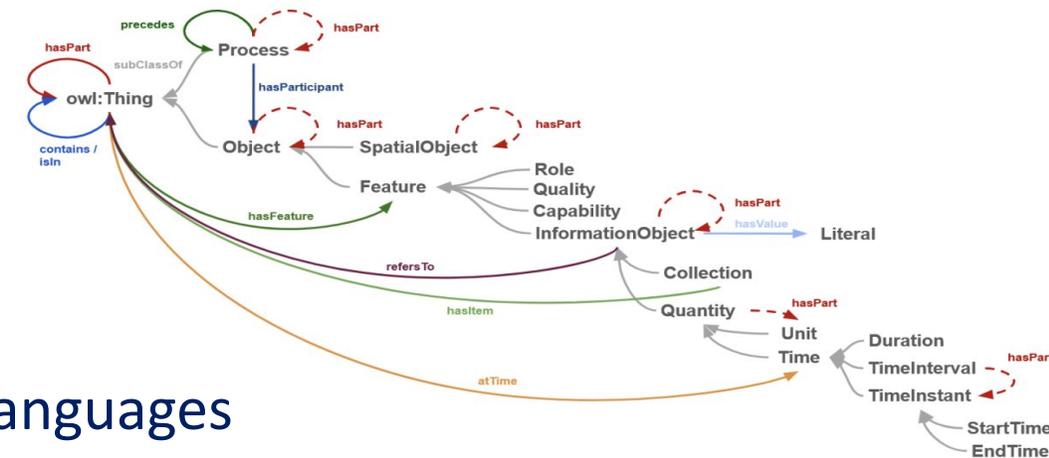
- Higher quality health systems and public health strategies,
- Faster and more impactful innovations adapted to all EU population

How to deliver “digital twin” (PHKG) at EU level

Critical Success Factors (CSF)



- AIDAVA demonstrated that (mainly) automated curation is possible by orchestrating multiple AI and non-AI tools
- AIDAVA-like systems could be deployed across EU with 2 CSF
 - Agreement & enforcement on **SULO** (Simplified Upper Level Ontology)
 - SULO supports alignment NOT harmonization across standards
 - by making their semantic explicit
 - and facilitate (AI driven) mapping
 - All “Digital Twins” should be compliant with SULO (no impact on local HIMS/EHR systems)
 - Speed of improvement of **AI tools** across EU languages (mainly NLP and Entity Linking/medical coding)
 - Availability of detailed **documentation on data sources** (extension of EHDS Data Description to each attribute)



Comparative cost estimate of EHDS - 2 scenarios [ref]

10 hospitals, 5 years, per year: 1250 records, 1250 EEHRxF, 100 HDAB queries



		CURRENT SCENARIO (K€)				Curation of data sources - with AI (K€)			
	Steps	One of	Per unit	Yearly	Total Period	One of	Per unit	Yearly	Total Period
Data Holders/ EHR Vendors	Curation	€0	€0	€0	€0	€300	€0.10	€425.0	€2,125
	EEHRxf	€300	€0.10	€425	€2,125	€20	€0.05	€82.5	€413
	Data Description	€50	€0.25	€52	€258	€5	€0.05	€5.3	€27
	Answer to permit Total/Data Holder	€0	€5	€500	€2,500	€10	€0.05	€15.0	€75
	TOTAL ALL DH				€48,825				€26,390
HDAB	Curation	€0	€50	€5,000	€25,000	€0	€5	€500	€2,500
TOTAL					€73,825				€28,890 (39%)

Numbers in red: assumptions to be verified. Their value is in comparative cost rather than total cost

Comparative cost estimate of EHDS - 2 scenarios [ref]

10 hospitals, 5 years, per year: 1250 records, 1250 EEHRxF, 200 HDAB queries



		CURRENT SCENARIO (K€)				Curation of data sources - with AI (K€)			
	Steps	One of	Per unit	Yearly	Total Period	One of	Per unit	Yearly	Total Period
Data Holders/ EHR Vendors	Curation	€0	€0	€0	€0	€300	€0.10	€425.0	€2,125
	EEHRxf	€300	€0.10	€425	€2,125	€20	€0.05	€82.5	€413
	Data Description	€50	€0.25	€52	€258	€5	€0.05	€5.3	€27
	Answer to permit Total/Data Holder	€0	€5	€1,000	€5,000	€10	€0.05	€15.0	€100
	TOTAL ALL DH				€7,383				€2,664
					€73,825				€26,640
HDAB	Curation	€0	€50	€10,000	€50,000	€0	€5	€1,000	€5,000
TOTAL					€123,825				€31,640 (25%)

Numbers in red: assumptions to be verified. Their value is in comparative cost rather than total cost

PHKG = high-quality, interoperable 'Digital Twin' of patient health records

- Curation process today
- Curation workflows in AIDAVA
- EHDS and PHKG: implementation of EEHRxF and Requests

Problem we need to solve

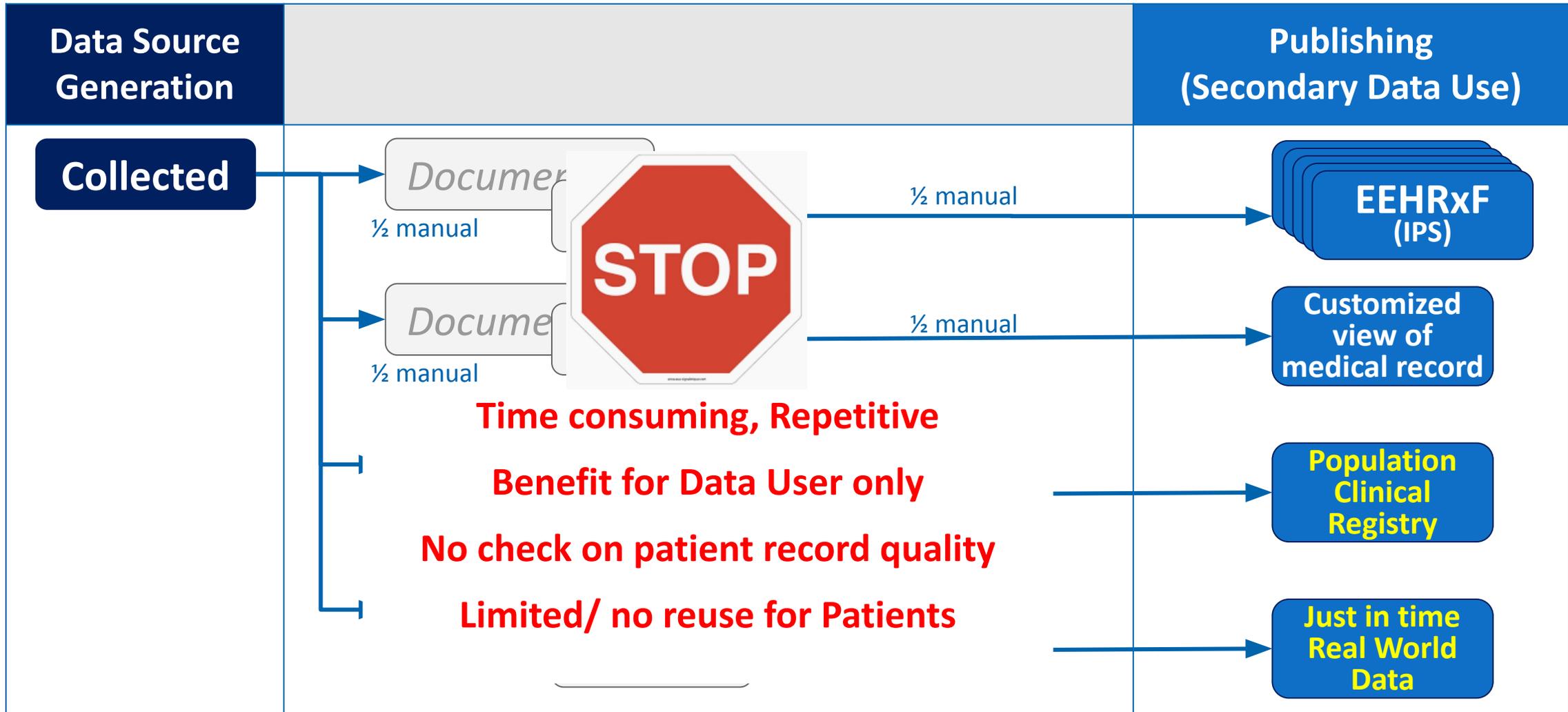
The painful reality about health data



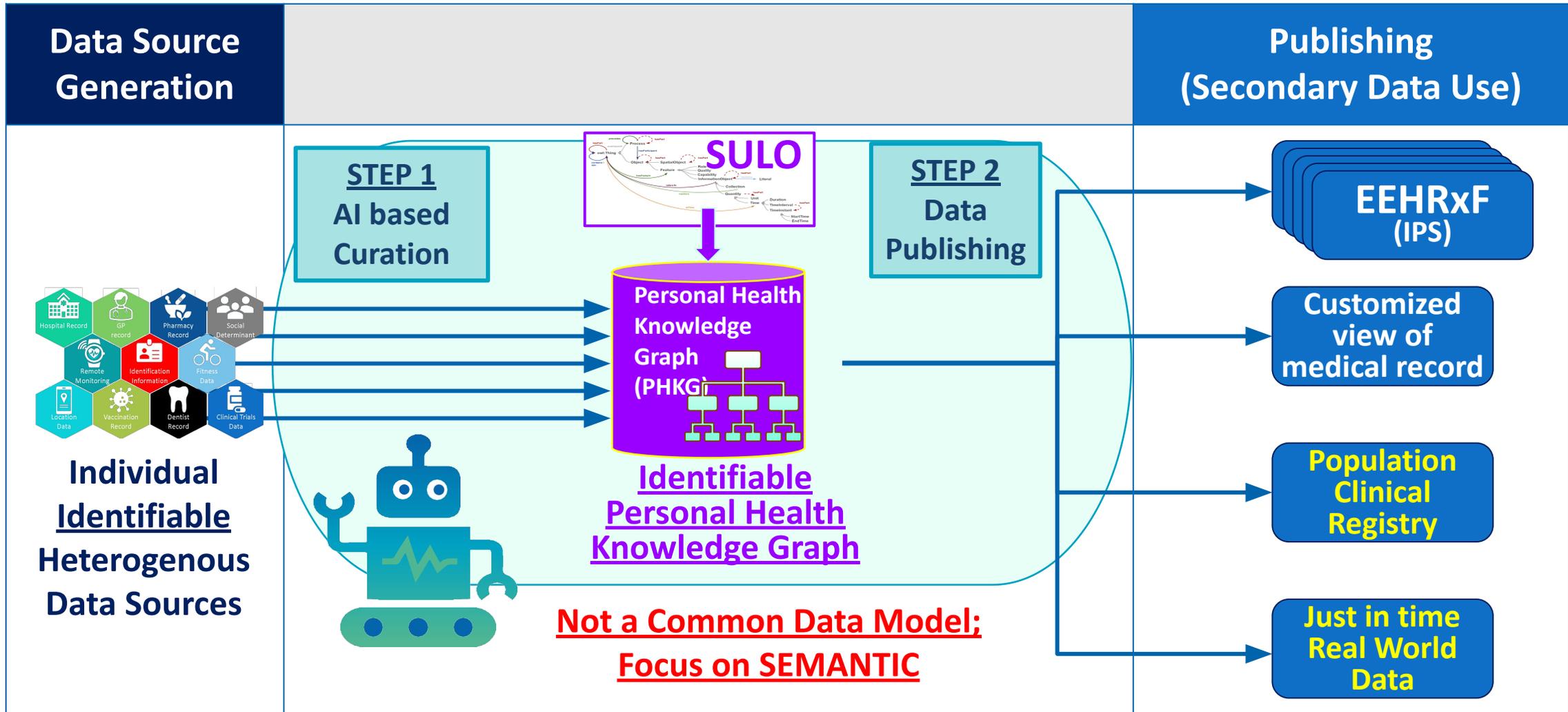
- **Scattered** across multiple systems ($\pm 40\%$ hospital, $\pm 60\%$ others)
- **Heterogeneous**: different standards (HL7, openEHR..): no interoperability
- **Not readily processable**: up to 80% in narrative format ($\pm 40\%$ full text, $\pm 40\%$ chunks of text)
- **Not well documented**: most 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,

Impact: time consuming access to high-quality, reusable data for clinical care and research

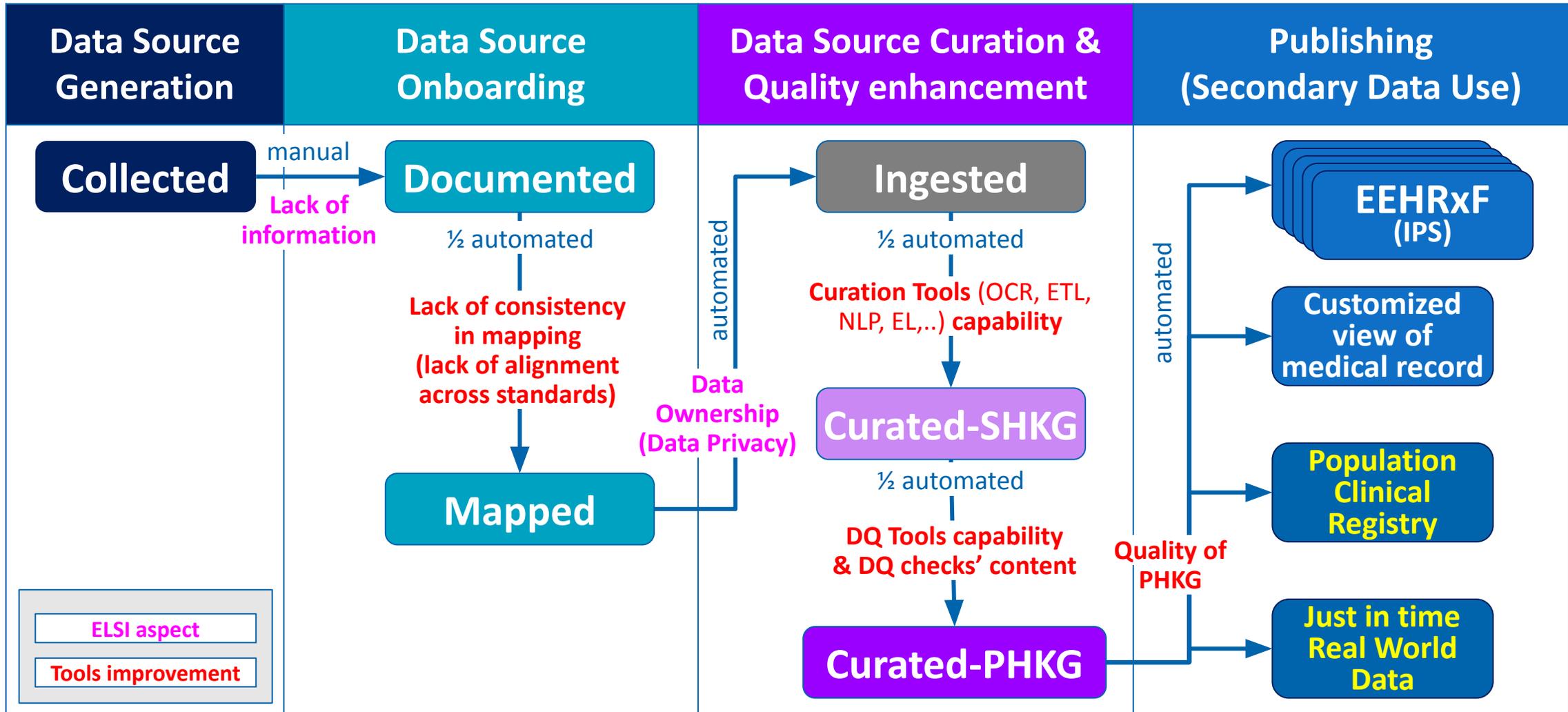


Introducing AIDAVA: automate data curation & quality enhancement of heterogeneous data with AI



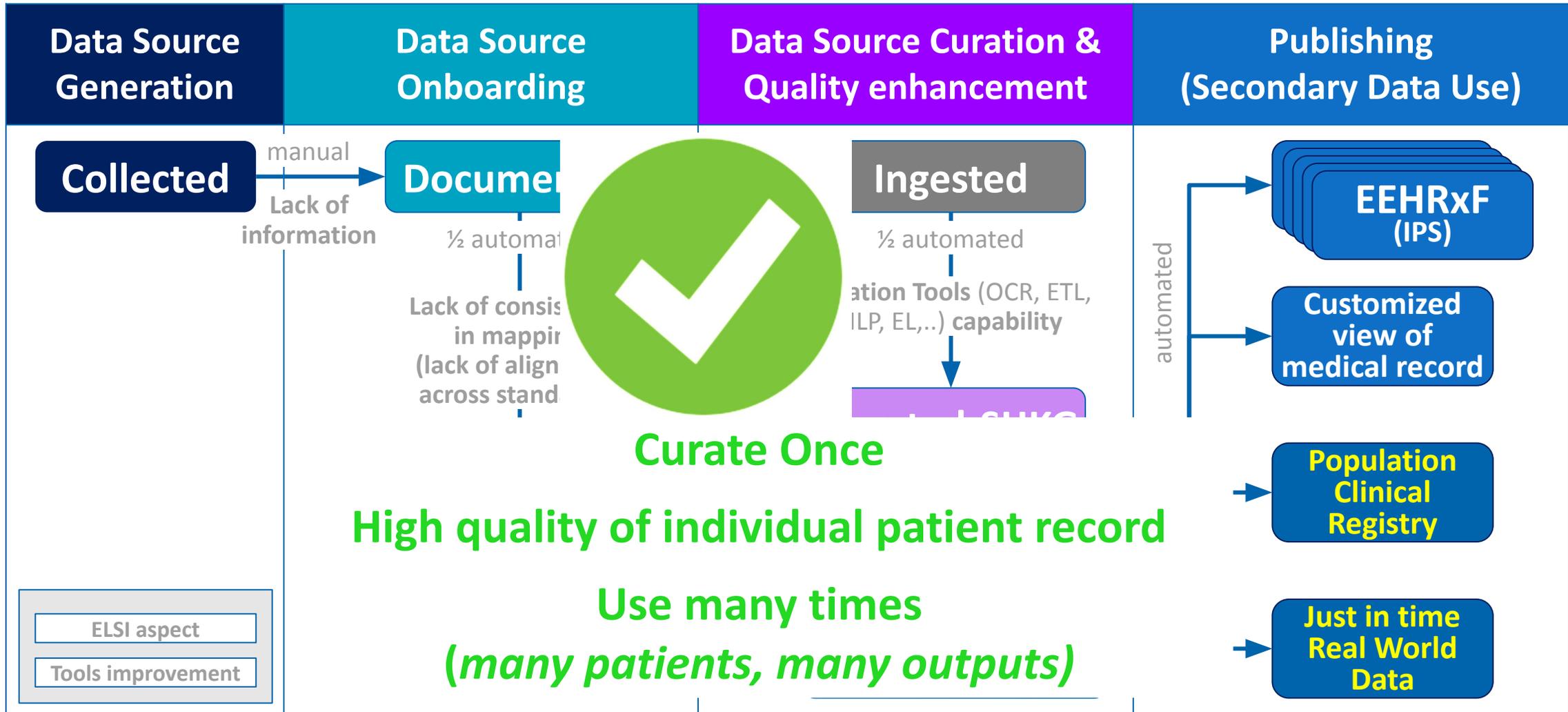
Proposed solution: a set of steps with automation potential

Journey of a data element in AIDAVA



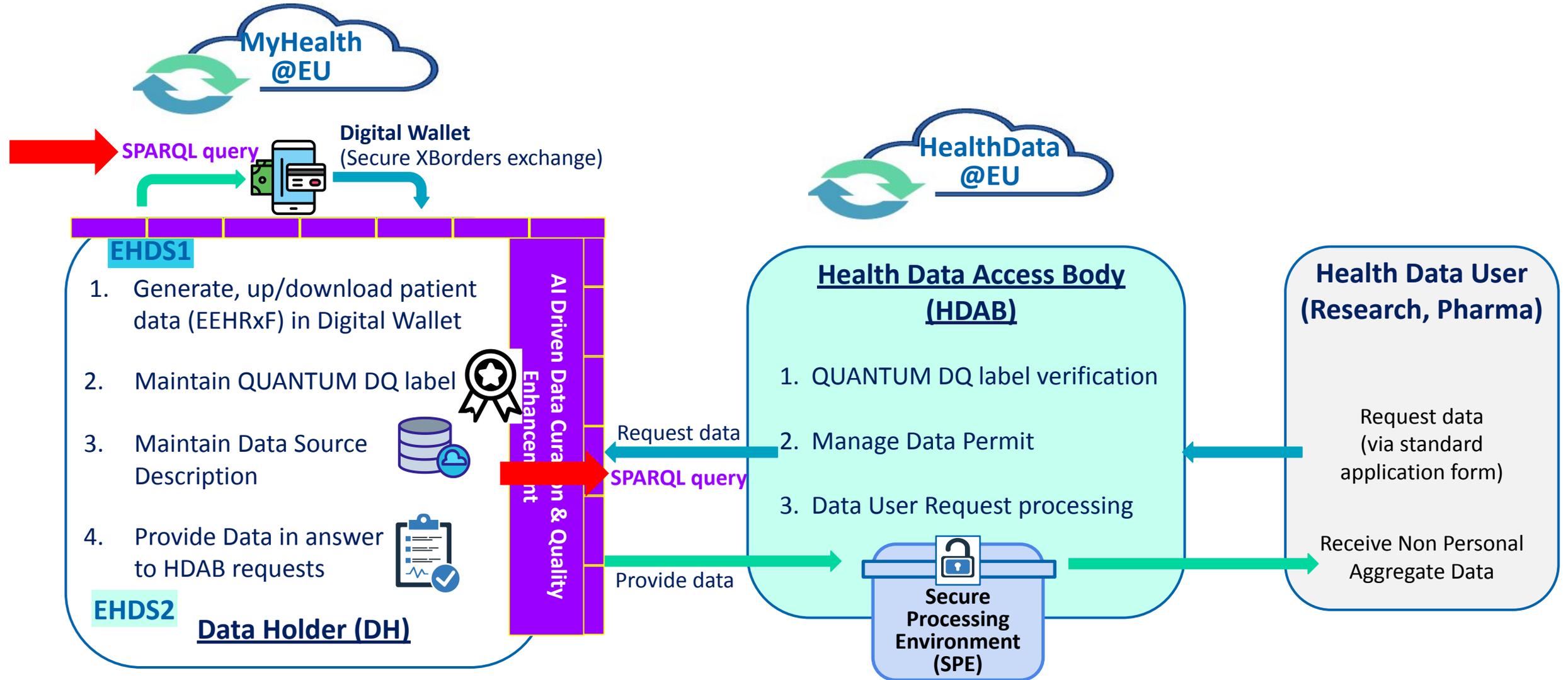
Proposed solution: a set of steps with automation potential

Journey of a data element in AIDAVA



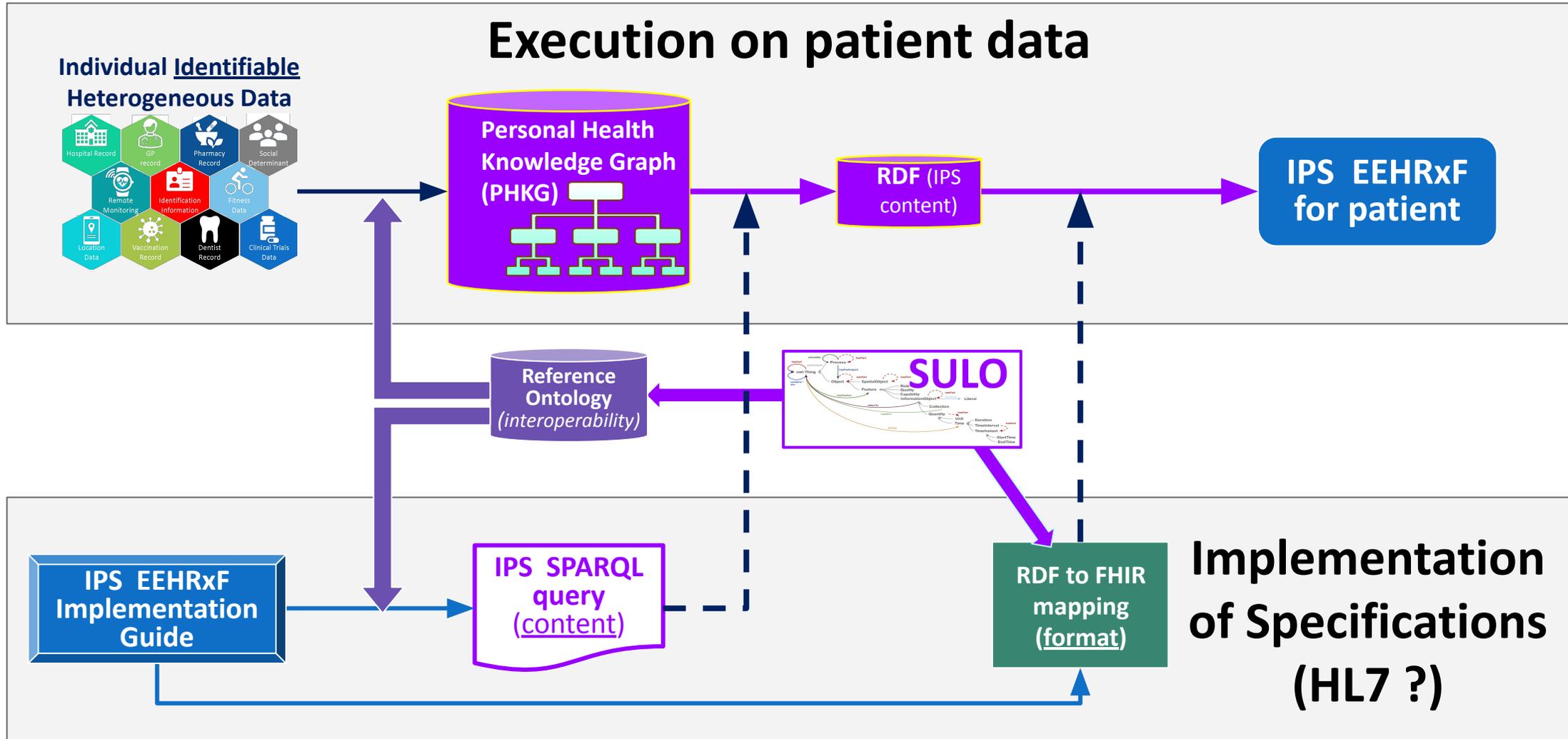
EHDS Data Flows and PHKG

PHKG as an AI maintained high-quality “digital twin” of patient records



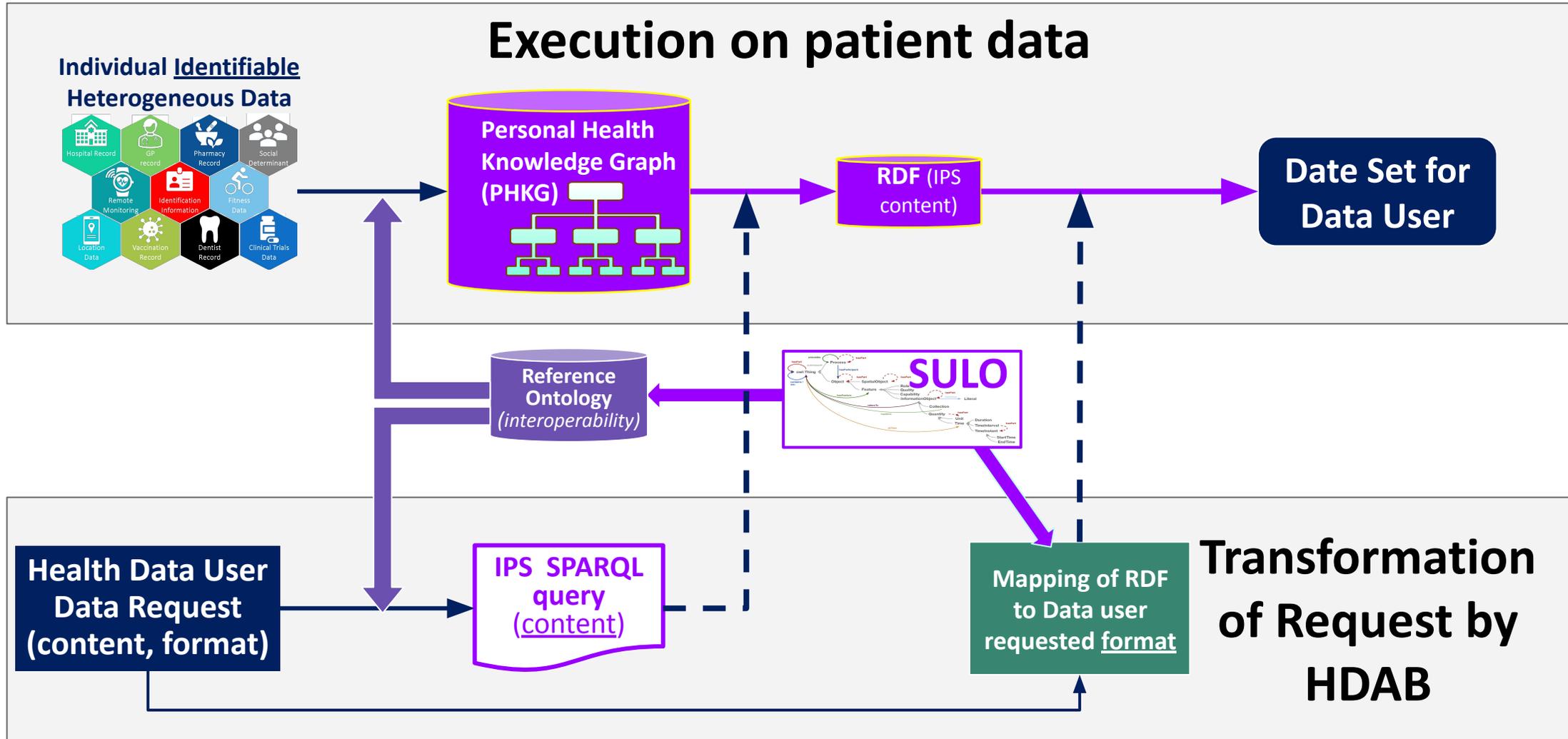
EHDS Data Flows and PHKG

Implementation of EEHRxF IG on AIDAVA 'digital twin' (PHKG)



EHDS Data Flows and PHKG

Execution of Data User request on AIDAVA 'digital twin' (PHKG)



EHDS implementing acts push a lot of requirements on 'data holders' (and more specifically hospitals). This makes sense as hospitals are the 'holders' of the patient data, at the core of health systems.

However, without supporting tools and automation, the additional burden (cost, needed skills) to hospitals might be unbearable for most and lead to EU digital divide in health.

AIDAVA - properly extended & implemented - could help without requiring changes into the regulations & implementing acts while bringing benefits to patients, providers and the whole EU health system.



Funded by
the European Union

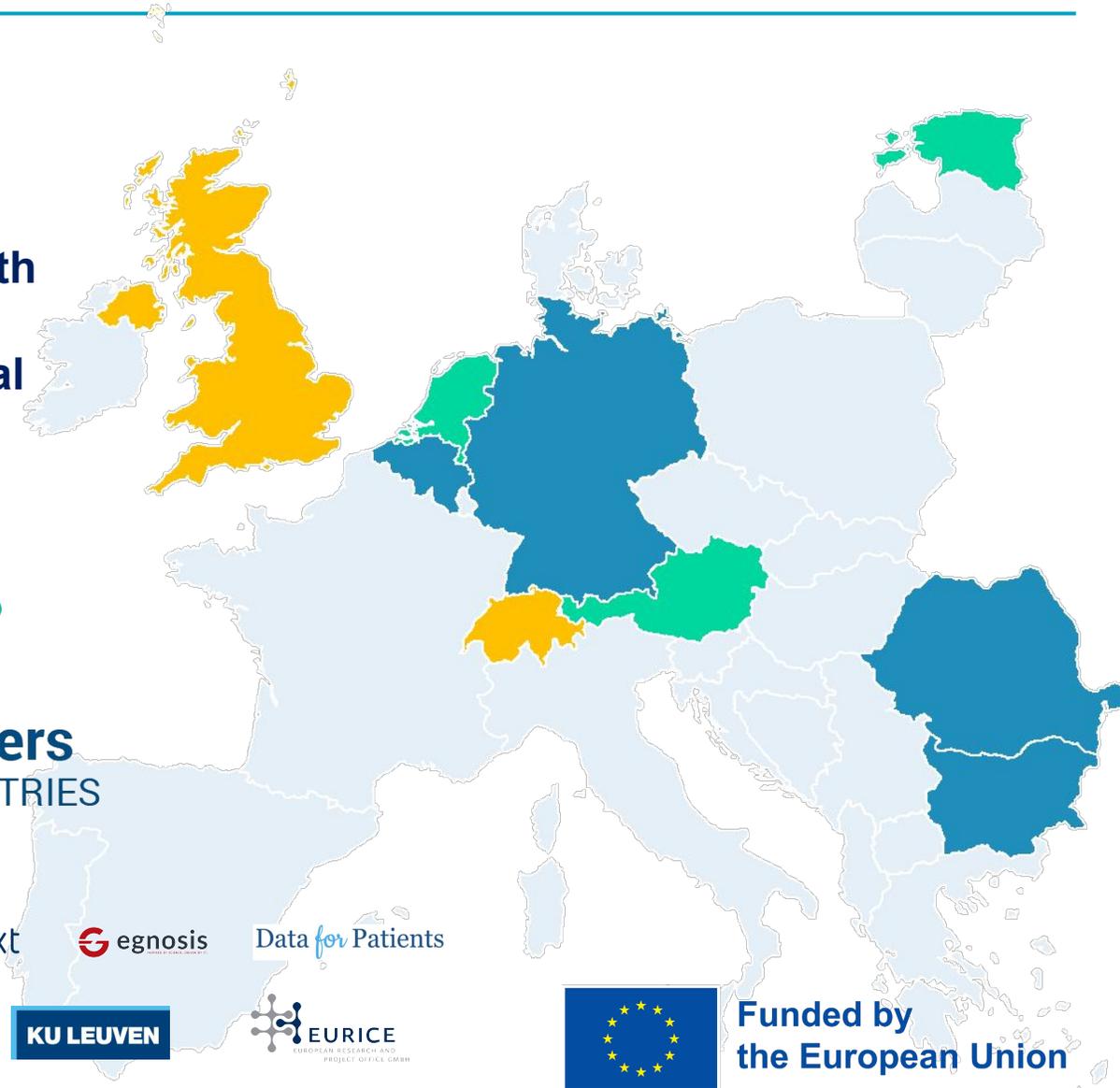


THANK YOU !

AIDAVA: Facts and Figures



The interdisciplinary AIDAVA consortium consisting of 14 European partners will develop a **digital solution**, orchestrating diverse artificial intelligence technologies, for **more efficient curation and publishing of personal health data**, delivering interoperable and reusable personal health records for the benefit of patients and physicians/clinical researchers.



01 SEP 2022
START DATE



31 AUG 2026
END DATE



14 Partners
FROM 9 COUNTRIES



**Funded by
the European Union**

Coordination Team



Michel Dumontier

Principal Investigator

Distinguished Professor

Maastricht University

*FAIR Data and Services, Knowledge
Graphs, Personalized Medicine*



Remzi Celebi

Coordinator (technical)

Researcher

Maastricht University

*FAIR Data, Knowledge Graphs
& Embeddings, Personalized
Medicine*



Isabelle de Zegher

Co-coordinator (clinical)

MD, MSc

Founder b!loba

*Digital Health, Data Standards,
Personal Health Data Sharing*



Gabrijela Radic

Project Management Office

Eurice GmbH

*Day-to-day administrative,
contractual, organisational
management*





Funded by
the European Union



BACKUP

Deep dive in EHDS Data Flows

Overview of requirements (EHDS compliance and others)

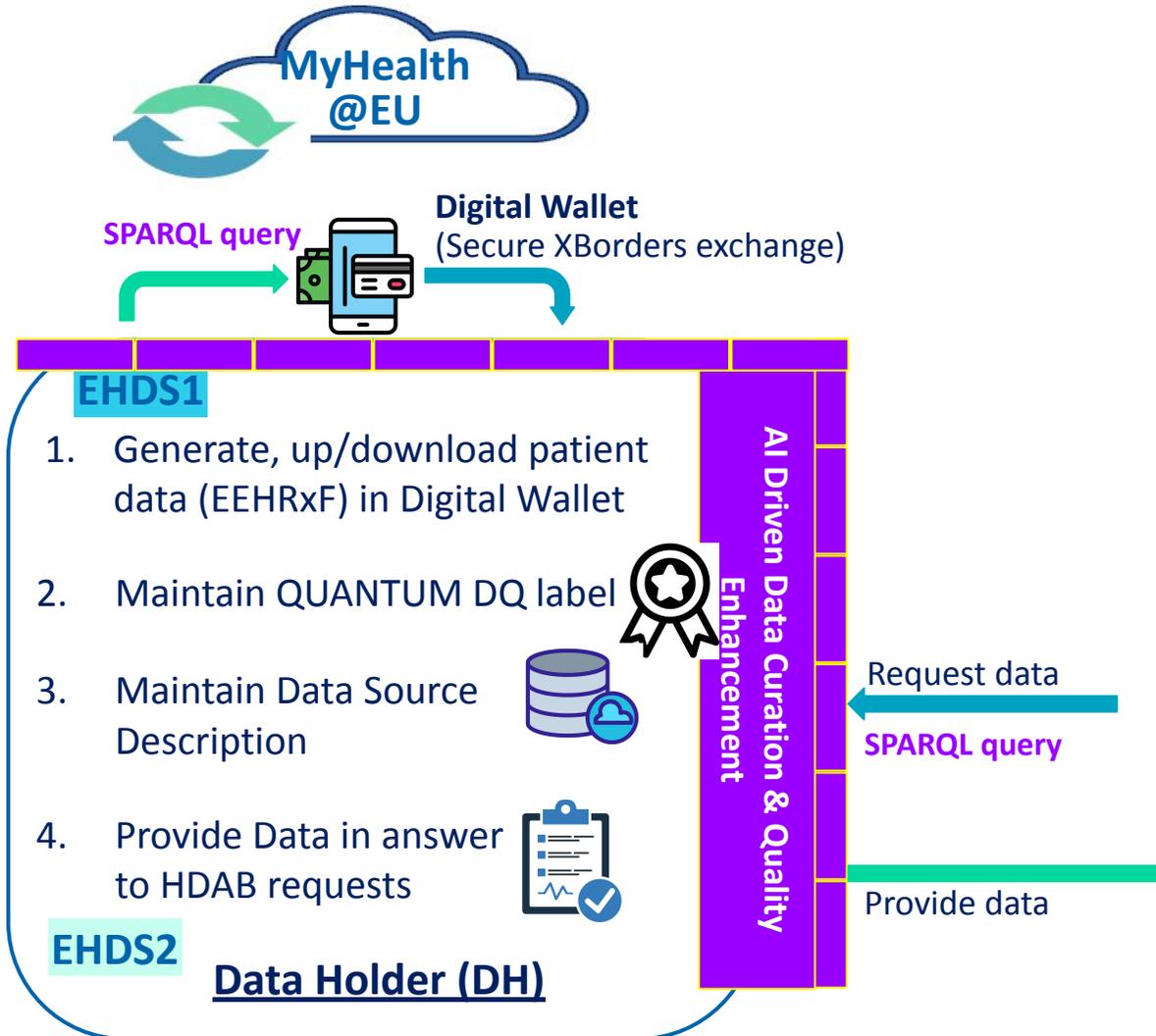


Category of requirement	Details on the requirement	As Is	MANDATORY EHDS compliance	DISCRETIONAL usable & safe
Regulatory compliance - EHDS1 and EHDS2	Maintain & self-certify data quality (QUANTUM label)	N/A	YES	??
	Describe their data sources using standardized metadata catalogues (Data Description)			
Regulatory compliance - EHDS1 ((EEHRxF)	Import and display an incoming EEHRxF message	N/A	YES	YES
	Store audit logs, supporting provenance and traceability			
	Ensure security/ data privacy			
Clinical trust in exchanged data	Put something correct into the message,	LOW	MEDIUM	HIGH
	Put something safe & reasonably complete in EEHRxF message			
	Explain how data inconsistencies are safely handled			
Usability for clinicians	Match incoming data elements to corresponding EHR data so the clinician can see it (e.g a unified medication list)	LOW	MEDIUM	HIGH
Continuity-of-care value	Store the imported data in the EHR as integrated data,	LOW	MEDIUM	HIGH
Risk of reputational backlash	Avoid liability and negative image due to mistakes/errors in the data	HIGH	MEDIUM	LOW
Operational needs	AI readiness: make patient data AI ready to maximise benefit of AI	LOW	MEDIUM	HIGH
	High-quality personalised care based on high-quality record			
	Automatic maintenance of local registries			
	Smoother answer to queries across organisations (research)			
Regulatory Compliance - EHDS2 (data prep for secondary use)	Serve a data request properly i.e. prepare the data and make it available for SPEs.	LOW	MEDIUM	HIGH

ing their

Deep dive in EHDS Data Flows

Adding AI maintained high-quality “digital twin” of patient records

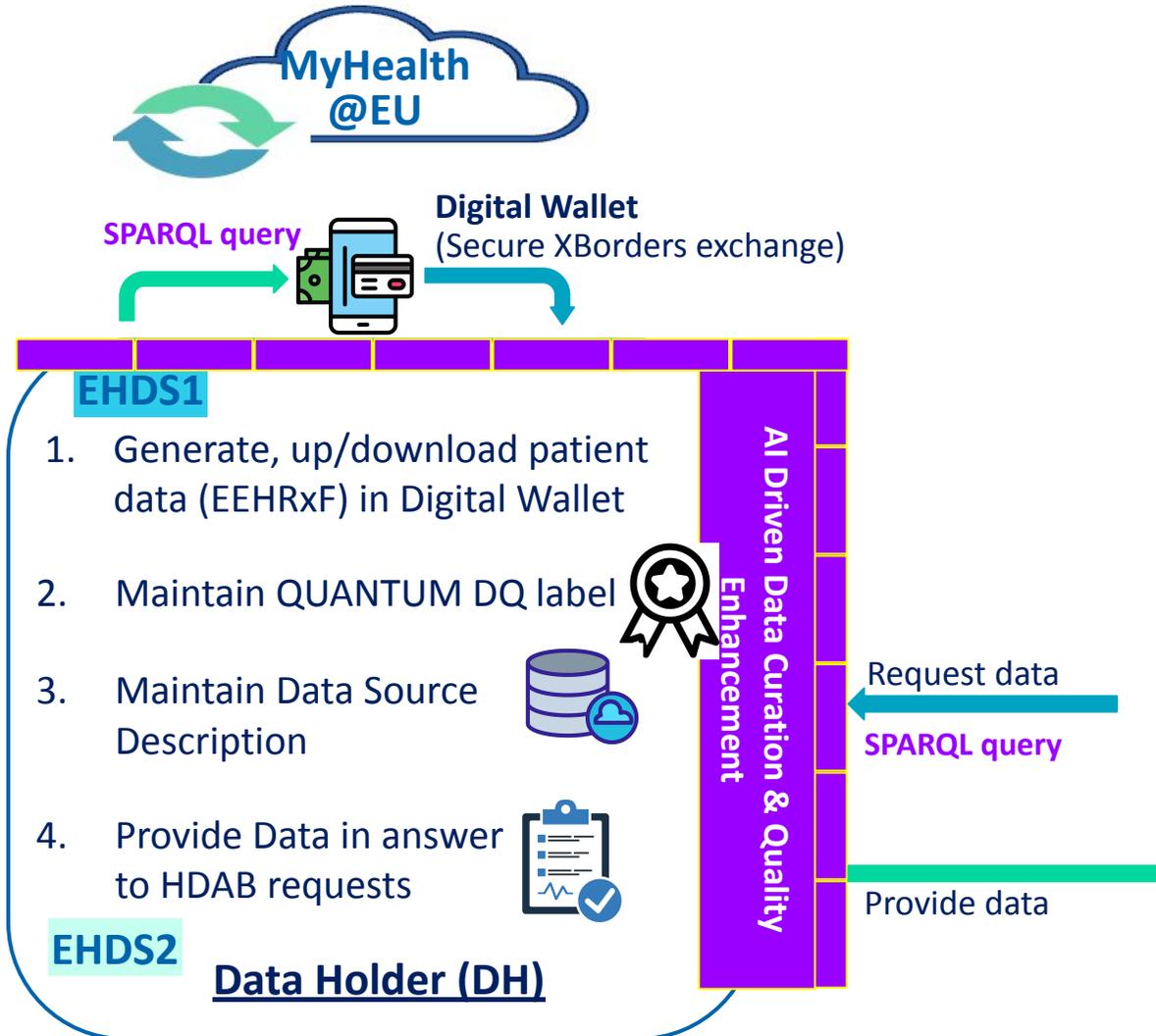


Needed by Data Holders

- Deploy AIDAVA-like virtual assistant
 - Secure local infrastructure
 - Documentation of data sources to support mapping
 - Develop tools to extract individual patient data query
- **To develop “digital-twin”** : execute AIDAVA-like virtual assistant
 - Extract data for each patient (maintain regularly)
 - Run AIDAVA and answer queries (missing information, inconsistency,..) when relevant
- **To provide EEHRxF compliant patient data:** execute SPARQL query related to category (IPS, lab, ePS, eD, discharge,...)
- **To provide data in answer to HDABs request:** execute SPARQL query received from HDAB with request

Deep dive in EHDS Data Flows

Adding AI maintained high-quality “digital twin” of patient records



Benefits by Data Holders: data quality

- **Additional benefits : data quality (within digital twin)**
 - ability to put something correct into the message,
 - ability to put something safe and reasonably complete into the message,
 - ability to match the incoming data elements to corresponding EHR data so the clinician can see, for example, a unified medication list,
 - ability to store the imported data in the EHR as integrated data,
 - ability to manage and explain how data inconsistencies are safely handled.
- **Compliance to EHDS1 and EHDS2 at low cost**
- **Additional benefits : operational efficiency**
 - reuse digital twin to make patient data AI ready and maximise benefit of AI
 - enable high-quality personalised care/decision making based on high-quality record
 - enable automatic maintenance of local registries
 - smoother answer to any queries related to patient across the organisations (research benefits)

How to deliver “digital twin” (PHKG) at EU level

Needed improvements (cost estimate to be defined)



Complexity	Improvement
Medium/High	Build consensus on Ontology (Simplified Upper Level Ontology)
High	Data Source onboarding <ul style="list-style-type: none">● Expand EHDS Data source catalogue (Health DCAT-AP)● Develop AI driven derivation of hospital data store into Data Source Catalogue● Develop AI driven mapping to Reference Ontology (SULO compliant)
High	Improve AI curation tools <ul style="list-style-type: none">● NLP, EL/medical coding in all EU languages● Entity Deduplication● KG embedding (Entity Alignment and Data Imputation)● LLM for HTIL
Medium	Use LLM to transform medical consensus/guidelines in Data Quality checks
Low	Transform HDAB permit/queries & EEHRxF standards into SPARQL
Medium	MDR and AI certification (including regular monitoring & training)

EHDS affordability: Need to support Data Holders



Data holders will bear most of the burden of EHDS with a significant 'compliance gap' (high upfront costs, lack of ROI).

Suggested solutions so far

- **Health Intermediation Entities (HIEs)** for smaller hospital (TEHDAS suggestion)
 - Requires delegating data control to third parties: governance concerns and systemic complexity
 - Funding unclear
- **Defer responsibility to Electronic Health Record (EHR) vendors.**
 - Fragmented vendor landscape across & within hospitals with different subsystems (LIS, PIS, RIS/PACS, ...)
 - Numbers of hospitals still use legacy proprietary 'home-grown' systems.
 - Funding needed for additional cost (implement EEHRxF , extraction and curation of data for HDAB permits,)
- **Deployment of Secure Processing Environments (SPE) through a decentralized network of Data Integration Centers**
 - Support secondary use in university hospitals, smaller/medium-size hospitals will struggle with IT infrastructure
 - MII Core Data Set contains extract of the patient dossier, not sufficient for EHDS in primary use

AIDAVA proposed “AI driven digital twin of health record” (to be further piloted with stepwise improvements)

- One of cost: estimated to be 500 and 150 KEUR per hospital (based on number of subsystems to be integrated)

AI readiness: Can we compete with the US (and China)?



Executive predictions for healthcare AI in 2026 in US (Dec 2025)

- 2026 will be the year AI is put to the test in healthcare
- The industry will shift
 - from pilots and hype to **accountable, integrated systems**
 - orchestrated AI workflows and multi-tasks agents tightly integrated into care plans
 - multi-modal AI (imaging, laboratory, genomic) data to provide a comprehensive patient view.
 - governed with clearer **guardrails**
 - that prove to have impact with **measurable outcomes**, making care proactive and highly personalized without increasing the time burden on the physician.

This requires AI ready data (in the format expected by the AI model)