Check for updates

RESEARCH ARTICLE

Physical Activity and Healthy Ageing & Longevity Data: Interoperability, Reuse, and Innovation in a Federated European Context

[version 1; peer review: 1 not approved]

Theo Zacharis [1,2]

[1]Kinesis Innovation Center, Athens, 151 26, Greece
[2]Greek Scientists Society, Athens, 151 26, Greece

v1 **First published:** 29 Jul 2026, **6**:263
https://doi.org/10.12688/openreseurope.23637.1
Latest published: 29 Jul 2026, **6**:263
https://doi.org/10.12688/openreseurope.23637.1

Abstract

Background

Physical activity (PA) is a cornerstone of **Healthy Ageing and Longevity (HAL)** and a high-leverage target for prevention-oriented health systems. Despite major advances in sensing technologies and digital health, PA and HAL research in Europe remains constrained by weak cross-study discoverability and limited data reuse. The bottleneck is not data scarcity but the absence of consistent cataloguing, taxonomies, and machine-actionable metadata enabling interoperability, federated analytics, and reproducible evidence synthesis. European policy initiatives emphasise secondary use of health data, real-world evidence (RWE), and AI-enabled analytics as key drivers of innovation within a **federation-first governance model.**

Methods

This article presents a conceptual and policy-facing analysis structured around a six-layer interoperability stack — **cataloguing, metadata, semantics, provenance, access/governance,** and **computability** — applied to current PA/HAL data practices in Europe. A present-situation analysis, gap analysis, and alignment agenda are situated within the EU's evolving data governance landscape, including the European Health Data Space (EHDS) and the European Open Science Cloud (EOSC).

Open Peer Review

Approval Status ✗

	1
version 1 29 Jul 2026	✗ view

1. **Ivan Miguel Pires**, Universidade de Aveiro Escola Superior de Tecnologia e Gestao de Agueda (Ringgold ID: 386349), Águeda, Portugal
Instituto de Telecomunicacoes (Ringgold ID: 131676), Aveiro, Portugal

Any reports and responses or comments on the article can be found at the end of the article.

Results

PA/HAL data practices exhibit systematic gaps across all interoperability layers. Generic cataloguing fails to capture domain-relevant characteristics; metadata inconsistency undermines comparability; semantic fragmentation prevents federated analytics; and governance metadata is rarely machine-actionable. Knowledge graph-based semantic layers, combined with BBMRI-ERIC and ELIXIR infrastructures, represent underutilised bridges for domain-coherent discovery and reuse.

Conclusions

An alignment agenda for **EHDS-compatible** and **Open Science**-ready PA/HAL data is feasible using existing European components. Success depends on coherent assembly of cataloguing standards, shared taxonomies, machine-actionable provenance, and operational governance — not on centralisation. PA/HAL data represent a critical stress test for European data strategy: their effective reuse will depend on whether heterogeneous, context-sensitive data can be made meaningfully interoperable within a federated system.

Plain Language summary

Physical activity is one of the most powerful things people can do to stay healthy as they age. Researchers across Europe collect enormous amounts of data on how people move — from wearable devices, clinical studies, and surveys. But this data is rarely shared or compared across studies, not because researchers are unwilling, but because the data is described in inconsistent ways. Different devices, different definitions, and different processing methods mean that results from one study often cannot be meaningfully compared with another.

This paper examines why this happens and what can be done about it. We analyse how physical activity and healthy ageing data are currently catalogued, described, and made available for reuse across European research systems. We identify six recurring gaps — from how datasets are discovered to how analysis can be run across multiple sites without moving sensitive data. We then propose a practical agenda for making this data more comparable, reusable, and ready for artificial intelligence applications, using existing European infrastructure including the European Health Data Space, BBMRI-ERIC, and ELIXIR — without centralising data or compromising privacy.

Keywords

physical activity, healthy ageing, longevity, EHDS, EOSC, FAIR, metadata, taxonomy, ontology, data catalogues, federated analytics, interoperability, knowledge graphs, BBMRI-ERIC, ELIXIR

This article is included in the COST Actions gateway.

Corresponding author: Theo Zacharis (theo@greek-scientists-society.org)

Author roles: Zacharis T: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the European Union's Framework Programme for Research & Innovation as part of the COST Action CA20104-Physagenet, supported by the COST Association (European Cooperation in Science and Technology).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Zacharis T. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Zacharis T. **Physical Activity and Healthy Ageing & Longevity Data: Interoperability, Reuse, and Innovation in a Federated European Context [version 1; peer review: 1 not approved]** Open Research Europe 2026, **6**:263 https://doi.org/10.12688/openreseurope.23637.1

First published: 29 Jul 2026, **6**:263 https://doi.org/10.12688/openreseurope.23637.1

1. Introduction: why metadata and semantics are now the limiting factor

Physical activity is among the most evidence-supported determinants of HAL, influencing cardiometabolic health, functional ability, mental health, and the risk trajectory for multiple chronic diseases across the life course. Over the past decade, research capacity has expanded markedly: activity is increasingly measured through wearable devices and smartphones; interventions are delivered in workplaces, clinics, and communities; and longitudinal cohorts collect behavioural, clinical, and environmental variables at unprecedented resolution. However, the ability to translate this growth into cumulative knowledge is uneven. A recurring, practical barrier is that PA and HAL datasets are difficult to **find**, **interpret**, and **reuse** across studies, countries, and infrastructures. This problem is not simply "data sharing." It is a problem of *data description quality*: cataloguing practices, semantics, and metadata completeness. The FAIR Guiding Principles explicitly emphasise machine-actionable findability and reuse as a precondition for scalable science. In domains like PA, where measures are context-sensitive (intensity, modality, setting, device, sampling rate, algorithm), the absence of shared metadata and taxonomies produces non-comparable outcomes and limits both reproducibility and policy relevance.[1]

At the same time, Europe's governance direction has become clearer. The EHDS Regulation was published in the Official Journal on 5 March 2025 and entered into force on 26 March 2025, beginning the transition phase toward application. EHDS does not "centralise" health data into a single repository; it establishes an architecture and rules for enabling cross-border access, including secondary use pathways. In such a model, metadata and semantics become the infrastructure that makes federation workable.[2]

This article therefore focuses on a specific question: **How can PA and HAL research in Europe become interoperable, federated, and AI-ready under EU Open Science and EHDS-aligned principles, starting from cataloguing, taxonomies, and metadata?**

Scope and limits. This paper is a policy- and methods-facing analysis. It does not propose new legislation, nor does it prescribe a single technical standard as universally optimal. It instead identifies recurring interoperability failure modes in PA and HAL research and proposes an alignment agenda that is compatible with Europe's federation-first approach.

2. EU policy and infrastructure context: federation-first, reuse-oriented

2.1 EHDS as an interoperability-and-governance framework

The EHDS establishes a European framework for health data use, including cross-border access and secondary use conditions and structures. Crucially for PA and HAL, EHDS-era interoperability is not only about clinical records; it concerns the broader ecosystem of health-related data that may support prevention, public health, and innovation, while respecting lawful bases, safeguards, and access controls.

A federation-first model makes three requirements unavoidable:

1. **Discoverability across nodes** (can you find relevant datasets without centralising them?)
2. **Interpretability across contexts** (can you understand what was measured and how?)
3. **Computability across boundaries** (can you run analyses without relocating sensitive data?)

All three depend on structured metadata, consistent taxonomies, and compatible cataloguing.

The EHDS architecture is being operationalised through several interconnected initiatives. **MyHealth@EU** enables cross-border exchange of patient summaries, ePrescriptions, and laboratory results as the primary use infrastructure. **HealthData@EU** provides the secondary use infrastructure for cross-border research access. **TEHDAS** (Joint Action Towards the European Health Data Space) developed the governance framework and data quality recommendations that underpin EHDS implementation across Member States. Domain-specific data spaces are being built in parallel: **EUCAIM** targets cancer imaging data federation, and the **Genomic Data Infrastructure (GDI)** enables cross-border access to genomic datasets. Together these initiatives define the practical interoperability environment within which PA/HAL data reuse must be situated.[3]

2.2 Open Science and EOSC as "catalogue-of-catalogues" infrastructure

The EU's Open Science direction and European Open Science Cloud (EOSC) aim to provide researchers and innovators with an open, trusted environment to publish, find, and reuse data, tools, and services for research and innovation, scaling resources across Europe to accelerate reproducibility and trust. In practice, EOSC's value is tightly coupled to metadata

quality and catalogue interoperability: datasets and services become reusable when they are consistently described, persistently identified, and linked to protocols, software, and provenance.

For PA and HAL research, EOSC-aligned practices can reduce "post-project disappearance" by anchoring outputs (data, code, protocols) in persistent repositories and searchable catalogues, provided the domain semantics are captured adequately.

2.3 Alignment with key EU strategies and regulatory initiatives

The challenges identified in PA/HAL data practices (fragmented cataloguing, heterogeneous metadata, and weak semantic interoperability) do not exist in a policy vacuum and must be understood within the broader European policy landscape that governs data, research, and public health. PA and HAL data do not constitute a standalone technical domain; rather, they are grounded in **normative public health strategies** that define policy objectives and monitoring needs for PA and HAL across Europe; they cut across multiple EU strategies, regulatory instruments and coordination initiatives that shape **how health-related data are generated, described, accessed, and reused** across borders and sectors, and to prioritise **federation, interoperability, reuse, and trustworthiness** as core design principles.

At the highest strategic level, the **European Strategy for Data** introduces sectoral European data spaces and positions interoperability as a central driver of societal and economic value. Health is a flagship domain within this strategy, yet PA and HAL data constitute a **cross-sectoral challenge**, linking health, social, occupational, and environmental data. In such domains, the success of the European Strategy for Data depends less on additional infrastructure and more on **semantic interoperability**, shared **vocabularies**, and consistent **metadata frameworks**.

Within this strategic framework, the **European Health Data Space (EHDS)**, provides the concrete governance and infrastructure model for the primary and secondary use of health data in Europe. EHDS adopts a **federation-first architecture**, enabling cross-border access while data remain at their source. In this model, metadata, cataloguing practices, and shared semantics are not optional technical enhancements but **foundational enablers**. PA and HAL data, spanning clinical, behavioural, functional, and lifestyle dimensions, sit precisely at the intersection of data types that EHDS must make interoperable without centralisation.

Regulatory expectations concerning data quality and traceability are further reinforced by the **Artificial Intelligence Act (AI Act)**. Many AI applications in health, prevention, and ageing-related decision support are expected to fall within high-risk categories, triggering requirements for transparency, traceability, and quality documentation of training and input data. Poorly described PA datasets, lacking explicit metadata on measurement protocols, processing pipelines, or contextual variables, undermine AI Act compliance by design. From this perspective, robust metadata and provenance documentation are prerequisites for lawful, explainable, and trustworthy AI in PA and HAL-related applications.[4]

The translational relevance of PA and HAL data is further emphasised by the **Accelerating Clinical Trials in the European Union (ACT-EU)** initiative. ACT-EU seeks to modernise and accelerate clinical research through methodological innovation, improved trial efficiency, and increased use of real-world data and digital endpoints. Functional capacity, physical activity levels, and behavioural outcomes are increasingly incorporated into prevention-oriented and ageing-related trials.

However, ACT-EU ambitions cannot be fully realised without **harmonised outcome definitions, interoperable metadata, and reusable data descriptions** that allow PA-derived measures to be compared and aggregated across studies and settings.[5]

Complementing EU regulatory and infrastructure initiatives, the **WHO Regional Office for Europe Physical Activity Strategy (2016–2025)** provides the **normative public health framework** underpinning physical activity and healthy ageing policies across the European region. The strategy defines policy targets, recommended indicators, and population-level monitoring needs for PA across the life course. Its effective implementation therefore depends on the availability of comparable, well-described, and reusable PA and HAL data. Weak metadata, inconsistent taxonomies, and poor cataloguing practices directly undermine the ability to monitor progress against WHO Europe targets and to translate evidence into policy action.[6]

Finally, the EU's Open Science agenda and the **European Open Science Cloud (EOSC)** provide the research-facing layer through which PA and HAL data are expected to become discoverable and reusable. EOSC functions as a *catalogue-of-catalogues*, relying on FAIR principles to support cross-border discovery and reuse of datasets, software, and workflows. In practice, PA and HAL datasets often remain technically "open" but scientifically unreusable due to

insufficient domain-specific metadata and taxonomies. This gap limits the effectiveness of Open Science mechanisms and constrains cumulative knowledge building in healthy ageing research.

Together, these strategies and initiatives converge on a common requirement: PA and HAL data must be described in ways that are **machine-actionable, semantically consistent, and governance-aware.** Figure 1 provides a high-level conceptual overview of this convergence, while Table 1 summarises how key EU and WHO Europe strategies relate to PA and HAL data.

3. Analytical approach: the interoperability stack

To avoid a generic narrative review, the analysis is structured around a pragmatic interoperability stack reflecting how PA and HAL data are produced, described, and reused. This layered approach is adopted to enable a systematic comparison of heterogeneous PA and HAL data practices across domains and jurisdictions, while maintaining a clear link between observed gaps and EHDS-aligned reuse requirements. The stack provides a consistent analytical framework for characterising current practice, identifying recurring gaps, and formulating an alignment agenda that is compatible with the European Health Data Space and aligned with Open Science principles.

1. **Cataloguing layer.** How PA and HAL datasets are registered and discovered, including dataset records, persistent identifiers, and indexing mechanisms.

2. **Metadata layer.** Descriptive fields characterising datasets: who collected the data, what was measured, when and how data were generated, and information on devices and processing steps.

3. **Semantic layer.** Taxonomies, ontologies, and controlled vocabularies ensuring consistent interpretation of PA and HAL data across datasets and contexts.

4. **Provenance layer.** Documentation of processing pipelines, algorithms, software components, and dataset versions over time.

5. **Access and governance layer.** Conditions for data reuse: licensing, consent constraints, data protection requirements, and access control mechanisms.

6. **Computability layer.** Readiness for federated querying and analytics across distributed data sources without data movement.

This interoperability stack is used to describe the present situation in PA and HAL data practices, identify common gaps across layers, and articulate an alignment agenda that is compatible with EHDS requirements and Open Science infrastructures (Figure 2).

4. Present situation analysis: how PA and HAL data are currently described and what are the outcomes they produce

4.1 PA and HAL data outcomes and their implications for innovation

In PA and HAL, data are converted into societal value through four outcome channels: **(i) publications** (knowledge and methods), **(ii) clinical studies** (translation into interventions and evidence generation), **(iii) patents** (protectable technical inventions), and **(iv) innovative products and services** (devices, software, programmes, and care pathways). The continuity across these channels determines whether PA/HAL data ecosystems primarily generate publishable findings, or whether they also support scalable validation, protectable invention, and deployable solutions.

Publications. The PA/HAL literature is large and growing: a PubMed query using "physical activity" OR "healthy aging" returns **224,254** records overall, **198,127** over the last 20 years (2006–2025), and **144,976** in the last 10 years. The same results on the **EU 27 countries** are **68,511, 62,770** and **49,424** respectively (verified April 2026). These volumes indicate a strong evidence base, but publication output alone does not ensure comparability or downstream uptake: where data are not described with sufficient semantics and provenance, results accumulate without becoming easily interoperable, reusable assets.

Clinical studies and real-world evidence. A search of ClinicalTrials.gov for "Physical Activity OR Exercise" returns **28,760** registered studies overall and **20,861** over the 20-year window 2006–2025. The EU-27 share represents 9,457

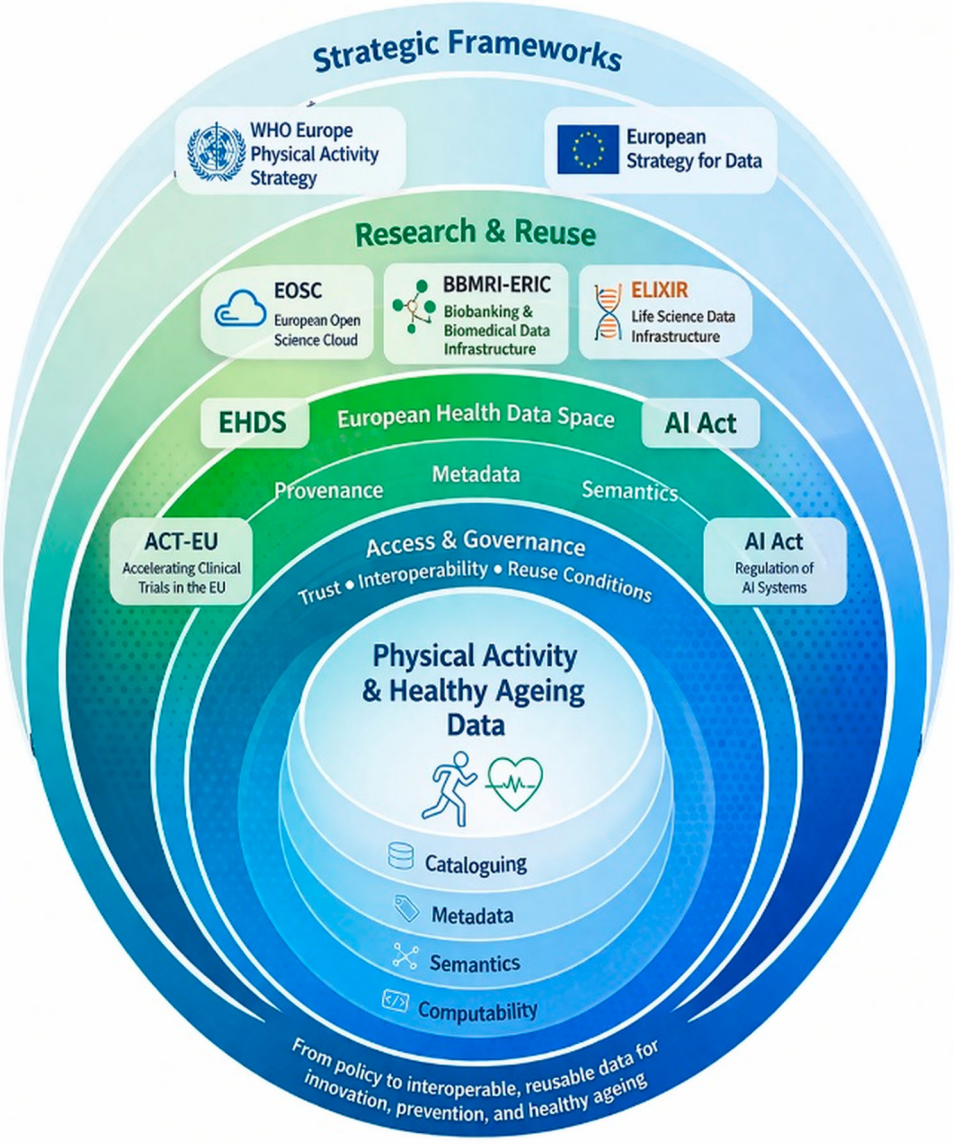


Figure 1. From policy to interoperable, reusable data for innovation, prevention, and healthy ageing. Conceptual depiction of the layered data-readiness framework for physical activity and healthy ageing data, situated within the European strategic, regulatory, and research ecosystem.

trial-site registrations across 27 member states (see Figure 3), with Spain (1,721), France (1,446), and Germany (968) accounting for 43% of EU-27 activity.

In practice, the translation value of this channel depends on whether PA/HAL variables and derived endpoints are specified consistently enough to support cross-study comparison and multi-site replication.

Patents. Patenting activity provides a proxy for conversion of PA/HAL measurement and analytics into protectable technical solutions. A Lens.org search using CPC class G16H20/30 (ICT for physical therapy/activity analytics) returns **48,874** patent records (**16,714** simple families) overall, and **45,705** records (**15,342** simple families over the 20-year window) (Figure 4).[8]

An interactive version with country-level tooltips is available as Supplementary File S1.

Table 1. EU strategies and their implications for PA & HAL data.

Strategy/ Initiative	Role in the policy stack	Relevance to PA & HAL	Implications for metadata & cataloguing
European Strategy for Data	*Overarching EU data strategy*	Establishes sectoral data spaces; prioritises interoperability	Shared semantic frameworks; cross-domain metadata alignment[7]
European Health Data Space (EHDS)	*Health data governance & infrastructure*	Primary and secondary use; federation-first access	Machine-actionable metadata; access & governance descriptors
Artificial Intelligence Act (AI Act)	*Regulation of AI systems*	Trustworthy AI for health and ageing	Data provenance; quality documentation; traceable pipelines
ACT-EU	*Clinical research coordination*	Faster trials; reuse of RWE; functional endpoints	Harmonised outcome definitions; interoperable PA metadata
European Open Science Cloud (EOSC)	*Research discovery & reuse layer*	Open Science implementation	Persistent identifiers; domain taxonomies; interoperable catalogues
BBMRI-ERIC	*Biobanking and health data infrastructure*	Provides access to large-scale, population-level and clinical datasets relevant to ageing and disease trajectories	Standardised biobank metadata (e.g. MIABIS); structured access frameworks; supports cross-cohort data discovery and harmonisation
ELIXIR	*Life science data infrastructure and interoperability framework*	Supports integration of multi-omics, clinical, and population data relevant to ageing and health	FAIR-aligned data standards, identifiers, and interoperability tools; enables scalable, cross-domain data integration
WHO Europe Physical Activity Strategy	*Normative public health framework*	PA targets, indicators, monitoring for healthy ageing	Comparable PA indicators; standardised definitions; metadata for population monitoring

Source: Authors' synthesis. PA = physical activity; HAL = healthy ageing and longevity; EHDS = European Health Data Space; EOSC = European Open Science Cloud; ACT-EU = Accelerating Clinical Trials in the EU.

However, where datasets and endpoints remain poorly comparable across contexts, incentives for standardisation and broad claims weaken, and invention-to-deployment cycles slow.

Innovative products and services. The innovation surface of PA and HAL extends well beyond consumer wearables. PA/HAL data underpin advances in sensing and measurement technologies, digital endpoints and biomarkers, clinical and public health decision support, intervention design and evaluation, and market-facing products ranging from devices to age-friendly environments. Globally, adjacent market evidence suggests that demand is shifting toward science-backed, data-driven wellness and fitness solutions, with McKinsey estimating a $1.8 trillion wellness market in 2024 and highlighting strong innovation momentum around products and services that combine personalization, efficacy, and measurable outcomes.[9]

In Europe, however, the competitive logic is more strongly mediated by regulation, interoperability, and evidence generation. European Commission analysis of digital health innovation, particularly in wearable medical devices, points to the strategic importance of standards, regulatory pathways, and industrial policy for scaling data-enabled health solutions.

At the ecosystem level, EIT Health's 2024 results (12 new healthcare innovations, support for 372 start-ups and scale-ups, and €227 million raised by supported ventures) indicate that Europe has a strong innovation base, but one that depends on translating heterogeneous data assets into reusable, trustworthy, and interoperable evidence. Across these domains, the ability to convert PA/HAL data into scalable products and services depends less on data volume alone than on semantic clarity, provenance, and reuse conditions.[10]

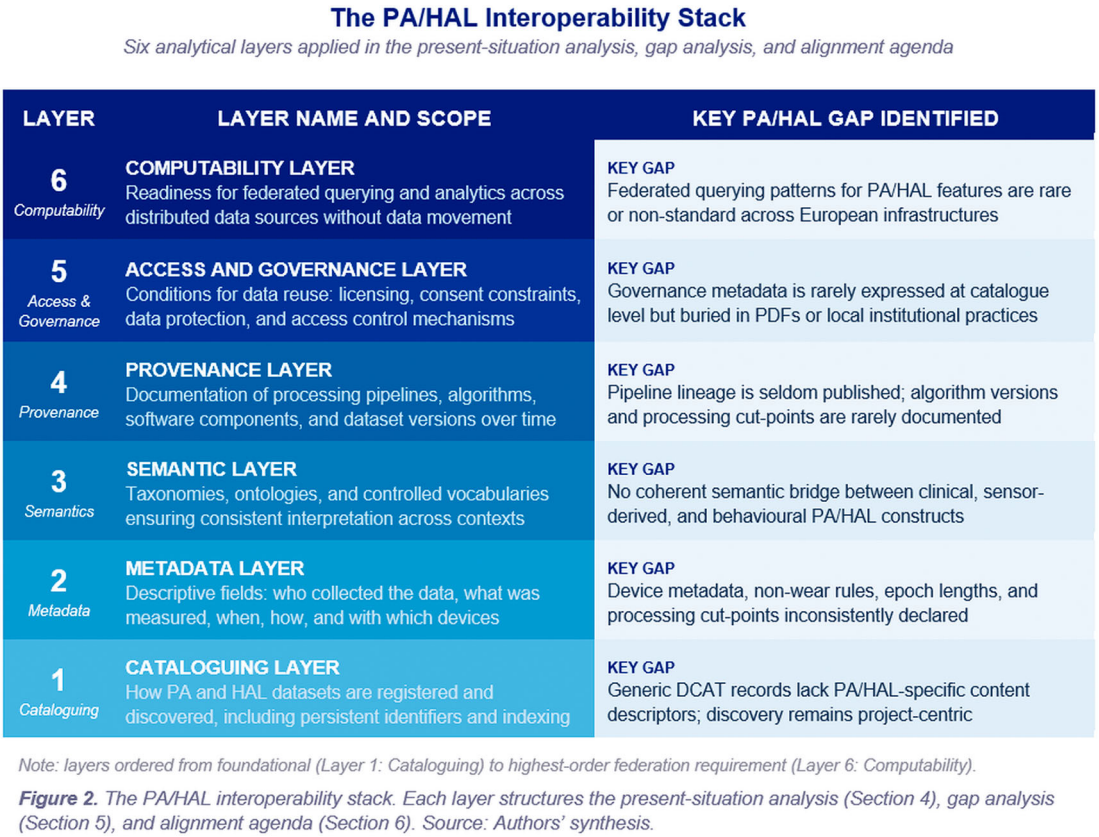


Figure 2. The PA/HAL interoperability stack. Each layer structures the present-situation analysis (Section 4), gap analysis (Section 5), and alignment agenda (Section 6). Source: Authors' synthesis.

EU position and implication. From a European perspective, the challenge is not a lack of PA/HAL research activity, but limited continuity across the four channels: data are frequently produced in ways that support publication, yet are less consistently described for reuse, cross-study synthesis, and scalable RWE. This raises the cost of replication and multi-site analytics, slows translation into protected inventions, and constrains product/service scaling, setting the context for the gap analysis (Section 5), including EU positioning relative to other major jurisdictions.

These patterns are summarised across the four innovation channels in Table 2.

Data sources and why semantics are harder than "clinical data"

PA and HAL research combines **heterogeneous data sources** spanning clinical, behavioural, and environmental domains: **(i) wearable and smartphone sensing** (e.g., steps, accelerometry-derived activity measures, heart rate, sleep proxies, mobility indicators); **(ii) clinical measurements** (diagnoses, medications, biomarkers, functional tests); **(iii) self-reported behaviour and context** (activity type, setting, barriers, wellbeing); and **(iv) environmental and social determinants** (built environment, walkability, seasonality, work patterns). This breadth is a defining strength of PA/HAL research, but it also creates a distinct semantic challenge compared with traditional clinical data. Unlike many clinical variables, where shared terminologies, care pathways, and regulatory requirements impose a baseline level of standardisation, PA/HAL variables are **highly sensitive to measurement context and analytical choices**. "Steps" varies by device, firmware, and algorithm; "moderate activity" depends on calibration methods, cut-points, population characteristics, and study protocols; "sedentary time" depends on posture inference, wear location, and non-wear detection rules; and key HAL constructs such as "frailty" correspond to multiple operational definitions across clinical and population-health traditions.

As a result, PA/HAL is a metadata- and provenance-intensive domain. Meaningful interpretation requires explicit documentation of devices and placement, sampling and epoching, wear-time processing, feature extraction pipelines, and algorithm versions. Where such descriptors are absent, reuse barriers arise earlier and more sharply than in many clinical datasets, undermining comparability, reproducibility, and the feasibility of federated RWE workflows.

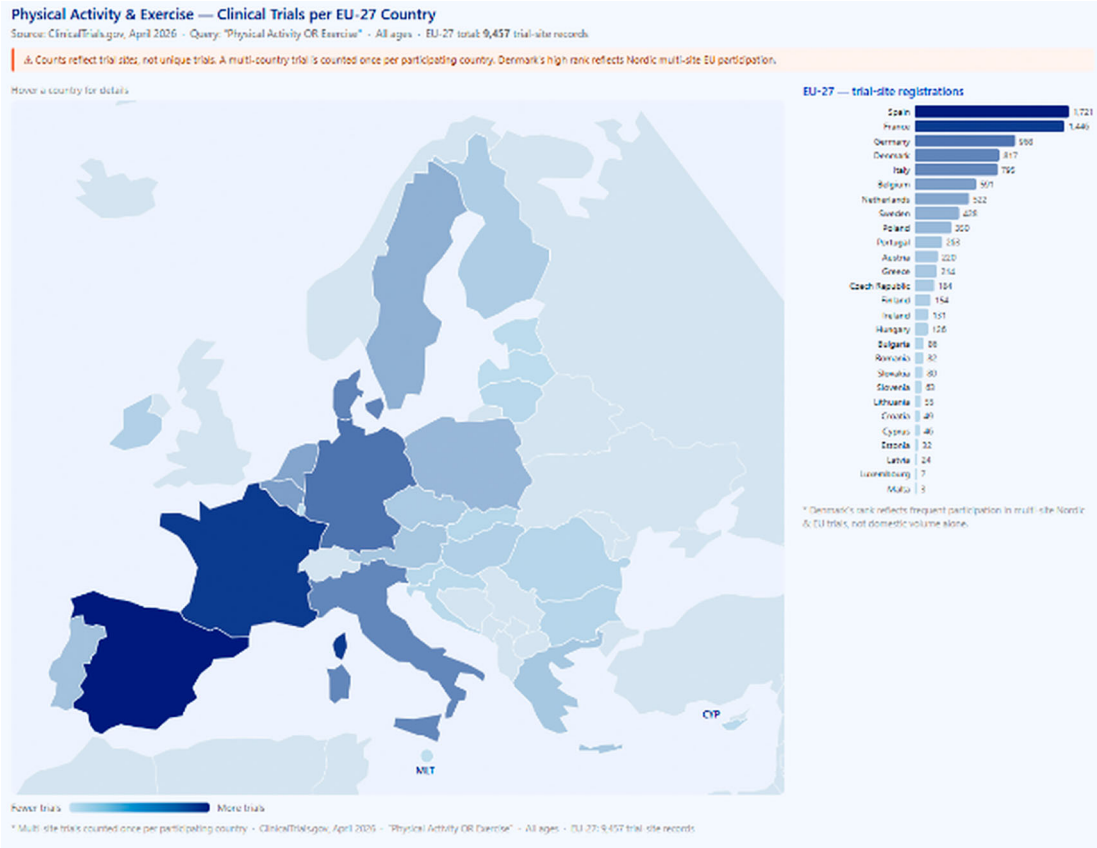


Figure 3. Clinical trial registrations by EU-27 country for interventions involving physical activity or exercise (ClinicalTrials.gov, April 2026; n = 9,457 trial-site records). Counts reflect trial-site combinations: a multi-country trial is counted once per participating country. Colour intensity reflects trial volume on a logarithmic scale.

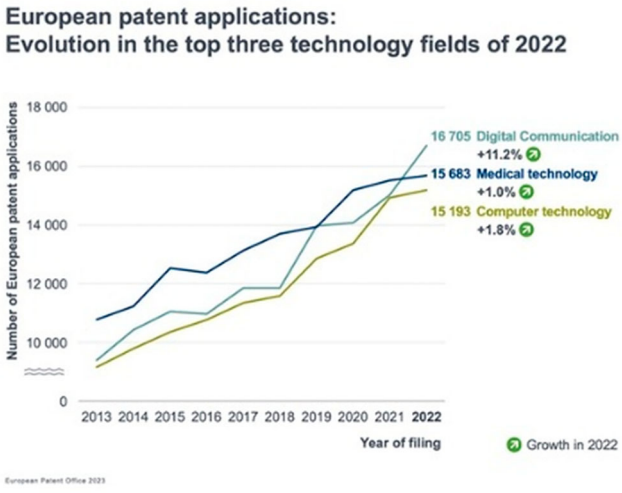


Figure 4. European patent applications: evolution in the top three technology fields of 2022. Source: EPO Patent Index 2024.[8]

This semantic variability has **direct practical consequences**. In PA/HAL, innovation depends on the ability to compare results across studies, validate digital endpoints, aggregate evidence for regulatory or payer decisions, and scale products and services across settings. When semantics and processing choices are under-specified, results may be publishable but not reusable; comparable in name but not in meaning; and credible locally but not transferable across sites or jurisdictions.

Table 2. PA/HAL innovation output across four channels: global volume and EU-27 position (verified April 2026).

Innovation channel Source & query	Global total ll-time/20-yr	EU-27 volume ll-time/20-yr	EU-27 share f global	Key implication for PA/HAL interoperability
Publications PubMed: "physical activity" OR "healthy aging"	**224,254** 198,127 (20-yr)	**68,511** 62,770 (20-yr)	**~30.5%** Verified April 2026	Strong output base but limited cross-study reusability. EU publications accumulate without becoming interoperable evidence assets in the absence of consistent metadata and semantics.
Clinical trials ClinicalTrials.gov: "Physical Activity OR Exercise"	**28,760** 20,861 (20-yr)	**9,457*** trial-site records	**~33%**	EU trial density is high, but endpoint harmonisation is weak. PA/HAL variables are rarely described consistently across multi-site studies, limiting RWE generation and meta-analysis.
Patents Lens.org: CPC G16H20/30 (digital PA/HAL analytics)	**48,874** records; 16,714 families (all-time) 45,705/15,342 (20-yr)	**~40%** EPO member states' share of global medtech filings (EPO Patent Index 2024)	**~40%** EU-specific PA/HAL families: verify via Lens.org/PATSTAT (A61B5 + G16H20/30)	Invention-to-deployment cycles slow when datasets and endpoints lack comparable descriptions. Weak semantics reduce incentives for broad standardised patent claims.
Market & innovation Wellness market + EIT Health 2024	**$1.8 trillion** global wellness market (McKinsey 2024)	**12** new innovations; 372 start-ups supported; €227 M raised (EIT Health 2024)	**Strong base** EU share of global wellness market est. ~20–25%	EU innovation is regulation- and evidence-mediated. Scaling PA/HAL products requires interoperable, reusable data assets. The federation-first approach must be operationalised in practice.

Sources: PubMed (April 2026); ClinicalTrials.gov (April 2026); Lens.org CPC G16H20/30 (April 2026); EPO Patent Index 2024; McKinsey Global Wellness Market Report 2024; EIT Health Annual Report 2024. EU-27 publication share verified via PubMed affiliation filter. * Clinical trial counts reflect trial-site combinations (a multi-country trial is counted once per participating country).

These constraints propagate directly into the innovation pipeline, limiting the transition from data to validated endpoints, protectable inventions, and scalable PA/HAL products and services. The following sections examine how these issues manifest across the remaining layers of the interoperability stack.

4.3.1 Cataloguing practices: fragmented discoverability

In Europe, PA and HAL data are catalogued using a combination of generic data cataloguing standards, clinical terminologies, and research repositories. The dominant cataloguing approach is based on **DCAT (Data Catalogue Vocabulary)**, which underpins many European data portals and is promoted for cross-domain dataset discovery within EOSC and national open data portals. However, DCAT is **intentionally domain-agnostic**, effective for high-level discovery ("a dataset exists"), but unable to capture content-level characteristics required to meaningfully discover and compare PA/HAL datasets.

For clinically anchored data sources, cataloguing relies on **SNOMED CT**, **ICD**, and **LOINC**. These provide strong semantic control for diagnoses and laboratory measurements, but PA/HAL variables, particularly those derived from wearables, functional assessments, or behavioural instruments, are only partially covered. In research and observational data ecosystems, **OMOP** improves analytical interoperability but focuses on data already mapped into the model, leaving upstream discovery of heterogeneous PA/HAL datasets underserved. Repository-based cataloguing (Zenodo, OpenAIRE) relies on generic metadata schemas with limited support for domain-specific discovery.

Example: MOBILISE-D - harmonising digital mobility outcomes across Europe.

The Mobilise-D consortium (IMI2, 2019–2024) illustrates both the scale of the PA/HAL semantic harmonisation challenge and what a structured response looks like. Spanning 34 organisations, 17 clinical sites in 10 countries, and four disease cohorts (Parkinson's disease, COPD, multiple sclerosis, proximal femoral fracture), the consortium set out to validate real-world digital mobility outcomes (DMOs) from a single lower-back-worn inertial measurement unit (IMU).

Its data standardisation protocol (Bonci et al., Scientific Data, 2023[11]) explicitly documents the heterogeneity problem encountered: different devices, firmware versions, sampling rates, wear locations, and gold-standard systems produced non-comparable raw signals even when measuring nominally identical constructs. Mobilise-D's response — a standardised data format, device-agnostic algorithm framework, and shared lexicon for mobility outcomes — required dedicated infrastructure and sustained consensus effort across the full consortium over five years.

Key lesson for PA/HAL interoperability: harmonisation is achievable, but disproportionately costly when semantics, provenance, and processing metadata are not captured by design at the point of data collection. The Mobilise-D Technical Implementation Manual (Zenodo, 2025)[12] now offers a transferable template for what a minimum viable metadata and provenance profile looks like for wearable mobility data.

As a result, current PA/HAL cataloguing practices exhibit recurring patterns:

- Dataset registration at a high, non-domain-specific level (e.g. DCAT-based records);
- Semantic description concentrated on clinical variables, with limited coverage of PA/HAL-specific constructs;
- Fragmented linkage between datasets, study protocols, code, and derived measures;
- Discovery dependent on project knowledge rather than systematic search.

Even when PA/HAL data are formally "open" or "shareable under conditions," they are often **not discoverable in a way that supports reuse, comparison, or innovation**. This fragmentation directly affects the innovation pipeline: datasets that cannot be reliably found and characterised are unlikely to be reused across studies, incorporated into federated analyses, or translated into scalable PA/HAL products and services. The limitations of current cataloguing practices therefore form a critical bottleneck between data production and innovation outcomes. PA/HAL reuse depends on an ecosystem of cataloguing standards (e.g. DCAT), clinical terminologies (e.g. SNOMED CT, LOINC, ICD), research infrastructures (e.g. EOSC, OpenAIRE, **BBMRI-ERIC**, **ELIXIR**), and federation tools (e.g. OMOP/OHDSI, DataSHIELD), but these components are typically adopted unevenly and rarely assembled into an end-to-end, domain-coherent discovery-to-compute pipeline (see Appendix).

4.3.2 Metadata: variable-level ambiguity and insufficient processing provenance

PA datasets frequently lack consistent metadata on the following elements: *device make and model, firmware version, and wear location* (e.g. wrist, hip, ankle); *sampling frequency, epoch length, and calibration protocol*; *non-wear detection and data-cleaning rules*; *feature extraction pipelines* (e.g. raw accelerometry → activity counts → intensity

Example: UK Biobank — strengths and limits of large-scale wearable metadata

The UK Biobank accelerometry sub-cohort (Doherty et al., PLOS ONE, 2017[13]) represents one of the most extensively documented wearable PA datasets in existence: over 100,000 participants, 7-day wear protocol, Axivity AX3 wrist-worn triaxial accelerometer at 100 Hz, with calibration procedures, non-wear detection rules, and feature extraction pipelines published in peer-reviewed literature and maintained in an open-source analysis tool.

This level of metadata documentation makes secondary analysis feasible at scale: studies linking MVPA to 697 incident diseases (Ye et al., npj Digital Medicine, 2022[14]) and derivation of age-referenced centiles (Rowlands et al., JSHS, 2025[15]) became possible because the processing chain was explicitly documented and traceable.

However, even UK Biobank illustrates the limits of current practice. Its accelerometry documentation explicitly notes that "data processing methods are under continual development, and so data processed with different versions of the tool and of models may not be directly comparable." This version-sensitivity is the norm across wearable PA research, and underscores why machine-actionable provenance and algorithm versioning are minimum requirements for FAIR-compliant PA/HAL data, not optional enhancements.

categories); and *versioning of algorithms, models, and derived measures.* Where metadata are mapped to healthcare standards (e.g. **HL7 FHIR Observation resources** or **LOINC** codes), this is typically limited to high-level clinical observations and does not capture the full processing context of PA/HAL variables. As a result, many datasets are technically findable and accessible, but fall short of the **FAIR Reusability (R)** requirements, which explicitly depend on rich metadata, provenance, and references to community standards.

These limitations lead to two recurring failures. First, **comparability failure**, where variables share a common label but differ in operational meaning ("same label, different meaning"). Second, **reproducibility failure**, where analytical pipelines cannot be reliably reconstructed, audited, or re-executed. Both directly constrain the reuse of PA/HAL data for cross-study synthesis, federated analysis, and innovation pipelines.

4.4.1 Taxonomies and semantics: partial reuse, limited domain alignment

In practice, semantics are assembled ad hoc: some clinical variables may align with established biomedical vocabularies; behavioural and context descriptors are often free text or local codes; and intervention descriptions vary widely (frequency, intensity, delivery mode, adherence metrics). A **"semantic gap"** emerges where PA and HAL sit between clinical medicine, behavioural science, and public health. This is precisely the gap that federated data spaces must bridge as federation depends on **shared meaning** more than shared storage.

Across PA and HAL research, datasets frequently lack **consistent, domain-relevant metadata**, even when they are deposited in repositories or registered in catalogues that nominally support FAIR principles. In practice, metadata provision is often limited to **generic schemas** (e.g. Dublin Core, DataCite) and repository-level descriptors, while **PA/HAL-specific metadata standards are rarely applied systematically**.

4.4.2 Taxonomies and semantics: partial reuse and limited domain alignment

Semantic annotation of PA/HAL data currently relies on a **fragmented combination of clinical terminologies, research vocabularies, and local conventions**. Established biomedical standards, such as **SNOMED CT** (clinical concepts), **ICD-10/11** (diagnostic classifications), and **LOINC** (laboratory tests and clinical observations), are widely used where PA/HAL variables intersect with healthcare data. However, these terminologies provide **limited and uneven coverage** of PA/HAL-specific constructs, particularly those derived from wearables, functional assessments, and behavioural measures.

As a result, behavioural and contextual descriptors are often recorded as **free text, study-specific variables, or local code lists**, while intervention characteristics (frequency, intensity, duration, delivery mode, adherence) vary substantially in structure and meaning across studies. Even when observational data models such as **OMOP** are adopted, PA/HAL variables are frequently mapped in non-standard ways, or excluded altogether, due to the lack of agreed semantic representations.

This gap has direct implications for innovation and federation. Federated data spaces and compute-to-data approaches depend less on shared storage than on **shared meaning**. Where PA/HAL semantics are weak or implicit, federation becomes technically feasible but analytically fragile: queries may execute, but results are not comparable. Consequently, semantic misalignment limits the generation of robust real-world evidence and slows the translation of PA/HAL data into validated endpoints, interoperable services, and scalable innovations.

Europe already has strong, mature components for **cataloguing (DCAT-AP), clinical semantics (SNOMED/LOINC/ICD), open science discovery (EOSC/OpenAIRE/Zenodo), biobanking infrastructure (BBMRI-ERIC), life science data management (ELIXIR), and observational federation (OMOP/OHDSI, DataSHIELD)**, and is now adding a major secondary-use governance and infrastructure layer via **EHDS/HealthData@EU**, but PA/HAL reuse remains limited because these components are rarely assembled into an end-to-end pipeline that preserves PA/HAL-specific meaning from capture to federated analysis (see Appendix).

5. Gap analysis: where current PA/HAL data practices constrain reuse and innovation

Despite the breadth of infrastructures, standards, and initiatives available in Europe, PA and HAL data practices exhibit systematic gaps across the interoperability stack. These are not isolated technical shortcomings but structural mismatches between how data are produced and how they need to function to support reuse, federation, and innovation.

Table 3. Minimum viable federation requirements for PA/HAL data under EHDS-aligned and Open Science-ready reuse.

Interoperability layer	Minimum requirement	Why it matters for PA/HAL federation
Cataloguing	Dataset records include PA/HAL-specific descriptors (device type, population, protocol, context) and persistent identifiers	Enables discovery beyond project knowledge; supports cross-site dataset identification without centralisation
Metadata (dataset level)	Machine-actionable metadata describing population, study design, data collection period, and access conditions	Allows researchers and access bodies to assess fitness-for-use before requesting or federating data
Metadata (variable/ feature level)	Explicit documentation of PA/HAL variables, including device, placement, sampling, epoching, and processing rules	Prevents "same label, different meaning" errors that undermine cross-study comparison
Semantics/ taxonomies	Use of controlled descriptors for PA modality, intensity constructs, context, and functional outcomes, with mappings where needed	Makes PA/HAL constructs queryable and comparable across datasets and jurisdictions
Provenance	Versioned documentation of processing pipelines, algorithms, and derived measures	Supports reproducibility, auditability, and AI Act-aligned traceability
Governance metadata	Structured expression of reuse conditions, purpose limitations, and access routes at catalogue level	Enables lawful discovery and secondary use under EHDS without manual negotiation at each step
Computability	Representation of PA/HAL features in forms compatible with federated querying and compute-to-data workflows	Allows analysis across distributed datasets while preserving meaning and governance constraints

Source: Authors' synthesis. PA/HAL = physical activity, healthy ageing and longevity.

5.1 Minimum metadata sets are rare or inconsistently applied

Even when teams collect rich contextual information, it is not consistently expressed as structured fields, machine-actionable descriptors, or persistent links to protocols/software. FAIR's emphasis on machine-actionability is directly relevant here: without it, AI-readiness and federated reuse remain aspirational.

5.2 Catalogue interoperability is not treated as "core infrastructure"

EOSC's mission includes enabling researchers to publish, find, and reuse data, tools, and services. If PA datasets are not consistently described and indexed, they remain invisible to cross-domain discovery. In a federation-first ecosystem, catalogue interoperability becomes a necessary condition for scaling evidence synthesis.

5.3 Weak linkage between datasets, derived features, and algorithms

AI-ready datasets are not simply "large" datasets. They require stable feature definitions, transparent pipelines, versioned transformations, and traceable provenance. Without these, AI systems risk learning dataset-specific artefacts rather than generalisable patterns.

5.4 Governance metadata is often underspecified

EHDS-era reuse requires clarity on what can be accessed, under which conditions, and with which safeguards. Operationalising this in PA research requires governance metadata to be captured and exposed at catalogue level, not buried in PDFs or local practices.[2]

Collectively, these gaps reduce the effective continuity between the four innovation outcome channels identified in Section 4.1. Data frequently support publication, but less consistently support scalable validation, protection, and deployment. Table 3 summarises the minimum viable federation requirements for PA/HAL data across all interoperability layers.

At the **cataloguing level**, PA/HAL datasets are often registered using generic dataset descriptors that enable high-level discovery but fail to capture domain-relevant characteristics. Discovery remains project-centric rather than data-centric, limiting the ability to identify datasets suitable for secondary analysis, cross-study synthesis, or innovation pipelines.

Micro-case: 'MVPA' across three datasets — same label, three different meanings

Consider a researcher attempting to pool physical activity data from three European studies, each reporting "minutes of moderate-to-vigorous physical activity (MVPA) per day":

Study A (UK Biobank): MVPA derived from wrist-worn Axivity AX3 at 100 Hz, using ENMO (Euclidean Norm Minus One) ≥40 mg, 7-day continuous wear.

Study B (Mobilise-D COPD cohort): MVPA proxied via walking speed ≥1.0 m/s from a lower-back IMU, bout-based detection using the Mobilise-D algorithm suite.

Study C (RADAR-base deployment[16]): MVPA derived from Fitbit step count >100 steps/min via vendor API; no raw signal access; no calibration record.

Each dataset reports 'MVPA' but operationalises it through a different device, placement, signal type, threshold, algorithm, and processing chain. Without explicit metadata on each of these elements, federated queries return results that appear comparable but are not. No post-hoc statistical harmonisation can fully recover comparability when the underlying measurement semantics are undocumented. This micro-case illustrates why the metadata, semantic, and provenance layers of the interoperability stack are prerequisites, not options, for meaningful PA/HAL federation.

At the **metadata level**, inconsistent documentation of devices, protocols, preprocessing steps, and derived variables undermines comparability and reproducibility. While datasets may be formally accessible, the lack of rich, standardised metadata prevents reliable interpretation and reuse, particularly across institutional and national boundaries.

At the **semantic level**, PA/HAL constructs remain weakly aligned across clinical, behavioural, and public health domains. Existing terminologies cover fragments of this space, but no coherent semantic layer bridges sensor-derived measures, functional assessments, and ageing-related outcomes. This semantic fragmentation is a critical barrier to federated analytics and cumulative evidence generation.

At the **provenance and computability levels**, analytical pipelines and algorithmic transformations are rarely described in a machine-actionable manner, limiting traceability and auditability. As a result, federated or compute-to-data approaches (central to EHDS and Open Science ambitions) remain difficult to operationalise for PA/HAL data at scale.

6. A novel framing: the PA/HAL dataset as a research object

Before proposing the alignment agenda, this section offers an argument that has not, to the author's knowledge, been made explicitly in the PA/HAL interoperability literature.

Current discussions of PA data interoperability treat datasets as objects of exchange: things to be shared, accessed, and harmonised. This framing leads naturally to questions about data governance, consent, and access control. These questions matter. But they are not the limiting factor.

The limiting factor is that PA and HAL datasets are not yet adequately constituted as *research objects* in the sense used by FAIR and EOSC. A research object is not just data; it is data plus the metadata, protocol, code, provenance, and access conditions needed to reproduce and reuse what was done. A PA dataset without its measurement protocol, its processing pipeline code, its algorithm version, and its operational definitions of key constructs is not a research object, it is an archive. Archives can be stored and shared; **research objects can be reused, federated, and built upon.**

This distinction has practical consequences. The FAIR Principles were not designed to govern archives. They were designed to govern research objects, entities that carry enough information about their own production to be interpreted independently of the team that created them. Applying FAIR to PA/HAL data without first constituting those datasets as research objects is technically correct but **practically ineffective**. The alignment agenda proposed below is therefore framed not as a metadata compliance exercise but as a ***research object assembly* agenda**: a set of steps through which PA and HAL datasets are progressively constituted as entities that can function independently, be reused without bespoke harmonisation, and be federated without losing analytical validity.

This framing has a further implication. Research objects are citable, versioned, and linked. They can be incorporated into systematic reviews and meta-analyses in the same way as publications. For PA and HAL, where evidence accumulation has outpaced evidence synthesis, the ability to cite and reuse datasets as first-class research outputs rather than as supplementary files or data-on-request, would fundamentally change the economics of secondary analysis and cross-study comparison.

7. What needs to be done: an EHDS-compatible, Open Science-ready alignment agenda for PA & HAL data

This section proposes a practical "alignment agenda" organised by the interoperability stack. An effective agenda does not require new infrastructures, but coherent use of existing ones.

7.1 Cataloguing: make PA datasets discoverable as a domain

Minimum direction:

- Use persistent identifiers for datasets, protocols, and software components.
- Ensure dataset records include domain-relevant descriptors (device, protocol, population, context).
- Ensure catalogue entries link to: *protocol documentation, processing pipelines (code), and version information.*

This directly supports the EOSC goal of enabling discovery and reuse of research objects at scale.

7.2 Metadata: define machine-actionable "minimum information" for PA measurement

A practical minimum set should cover:

- **Population:** age range, inclusion criteria, health status, key covariates.
- **Measurement:** device/sensor details, placement, sampling, epoch rules.
- **Processing:** non-wear detection, calibration, feature extraction, cut-points.
- **Outcome semantics:** exact definitions for intensity categories, sedentary time, etc.
- **Context:** setting (work, leisure), intervention delivery mode, adherence metrics.

This is not about imposing one measurement method; it is about making differences explicit and computable.

7.3 Semantics: develop a shared, extensible taxonomy for PA & HAL variables

A reusable approach:

- Start with a **core taxonomy** for: activity modality (walking, cycling, resistance training); intensity constructs and operationalisation; setting/context (occupational, transport, leisure); functional outcomes (mobility, balance, ADL proxies).
- Provide **extensions** for subdomains (workplace PA, multimorbidity cohorts, etc.).
- Ensure mappings support cross-study search and cross-study comparability.

7.4 Provenance and AI-readiness: treat feature engineering as a first-class research output

To support AI-ready and federated analysis:

- Publish feature definitions as reusable research objects.
- Version pipelines and document transformations.
- Link models to data provenance and measurement metadata.
- This aligns with FAIR's machine-actionable intent and supports reproducibility.

7.5 Governance metadata: express "as open as possible, as closed as necessary" operationally

In EHDS-era practice:

- Access conditions must be explicit at catalogue level.
- Data availability must not be confused with "open download."
- Restrictions (consent, privacy, contractual terms) should be represented as structured metadata to enable lawful discovery and access requests.

EHDS establishes the framework and transition timeline; this agenda translates that direction into domain-specific data stewardship practice.

7.6 Strategic implications: Europe's federated approach in a global PA/HAL landscape

The identified gaps have **strategic implications** for Europe's position in PA/HAL research and innovation. While Europe hosts world-class cohorts, strong public research institutions, and a dense policy framework, it faces increasing competition from regions that have prioritised data integration and industrial-scale reuse. In other jurisdictions, PA/HAL innovation increasingly benefits from vertically integrated data ecosystems where sensing, analytics, validation, and product development are closely coupled, accelerating iteration cycles and facilitating rapid translation into products and services.

Europe's strategic response differs in orientation. Rather than pursuing platform concentration, European initiatives emphasise **federation, data protection, and public trust**, as reflected in EHDS, Open Science policies, and FAIR principles. This approach offers long-term advantages in legitimacy and sustainability, but only if interoperability is realised in practice, not just in principle. For PA/HAL, Europe's competitiveness will depend less on data volume and more on its ability to align semantics, metadata, and provenance across distributed sources. Under the EHDS paradigm, "AI-ready data" does not primarily mean more data or larger repositories, but **machine-actionable data assets**: datasets that can be discovered, interpreted, and analysed with minimal bespoke harmonisation. The EU's advantage is the possibility to scale innovation while preserving public trust and legal certainty; its risk is that federation becomes technically feasible but analytically fragile when shared meaning is absent.

Relative to other major jurisdictions, Europe is building an innovation ecosystem where competitiveness derives from **interoperability and lawful reuse** rather than data centralisation. Where innovation elsewhere may be accelerated through vertically integrated ecosystems (device → platform → analytics → product), the EU's route to scale is through **standardised, governable reuse across many independent data holders**. This creates a different optimisation problem: Europe's bottleneck is less the production of PA/HAL data and more the **systematic conversion of heterogeneous datasets into reusable evidence assets** that support cross-site validation, real-world evidence generation, and scalable deployment across health systems and markets. For PA/HAL, this places an unusually high premium on semantic coherence, machine-actionable provenance, and operational governance.

Europe's approach to secondary use and innovation from health-related data is structurally distinct from jurisdictions that rely on platform concentration or centralised data aggregation. The EU's strategy is anchored in **federation by design**; data remain distributed under national and institutional governance while becoming reusable through common rules, interoperability specifications, and trusted access mechanisms. In practice, this model assumes that scalable evidence generation will be achieved not by moving PA/HAL data into a single repository, but by enabling **compute-to-data**, interoperable querying, and consistent interpretation across distributed sources. In PA/HAL, where constructs are measurement- and context-dependent, the success of the EU model depends disproportionately on community-level alignment around minimum descriptors, semantic mappings, and reusable pipeline patterns.

7.7 Knowledge graphs as a semantic integration layer for PA/HAL

The increasing scale, heterogeneity, and distribution of PA/HAL data highlight the need for mechanisms that go beyond traditional catalogues and metadata registries. While existing European initiatives provide essential building blocks for data governance, interoperability, and reuse, there remains a gap in how these elements are connected, interpreted, and made actionable across stakeholders.

A promising complementary approach is the use of **knowledge graph-based discovery and integration layers**. Knowledge graphs enable the representation of datasets, variables, protocols, organisations, publications, and policy

frameworks as interconnected entities linked through explicit semantic relationships. This allows both humans and machines to navigate complex data ecosystems, identify relevant resources, and understand their context and comparability.

Within the European research and Open Science landscape, initiatives such as **OpenAIRE** have demonstrated the feasibility of large-scale research knowledge graphs that interlink publications, datasets, software, and funding information.[17] **BBMRI-ERIC's** Directory and Negotiator tools and **ELIXIR's** RDMkit and FAIR Implementation Profiles provide concrete building blocks for such a semantic layer at the biobanking and life science infrastructure level.[18,19]

For PA/HAL, a domain-oriented semantic layer could serve as a bridge between distributed datasets and repositories; metadata and taxonomies; governance and access conditions; and policy objectives. Such a layer would not replace existing infrastructures, but enhance their usability by enabling *improved dataset discovery*, *contextual understanding*, and *traceability*, directly supporting AI Act compliance and EHDS secondary use requirements.

The main elements of the EU data agenda and their implications for PA/HAL are summarised in Table 4.

8. Implications for EU research, policy, and infrastructures

8.1 For researchers

1. Increased cross-study reuse and faster evidence synthesis.

2. Higher reproducibility through explicit processing provenance.

3. Lower duplication of data cleaning and harmonisation effort.

8.2 For policy and prevention

4. Better comparability of intervention outcomes across settings and populations.

5. More reliable evidence for prevention-oriented HAL strategies.

6. Improved capacity to monitor intervention effectiveness across regions without centralising data.

8.3 For infrastructures (EOSC-aligned and EHDS-era)

7. Domain metadata profiles become a strategic investment.

8. Catalogue interoperability becomes as important as storage or compute.

9. Federated analysis becomes feasible when semantics and metadata are treated as infrastructure.

8.4 Alignment agenda: towards EHDS-compatible and innovation-ready PA/HAL data

An effective alignment agenda for PA/HAL does not require new infrastructures, but **coherent use of existing ones**. Across the interoperability stack, alignment priorities emerge clearly.

At the **cataloguing level**, PA/HAL datasets need to be registered with domain-relevant descriptors that support meaningful discovery beyond project boundaries. Dataset records should explicitly reference access routes, reuse conditions, and links to protocols, code, and derived variables.

At the **metadata and semantic levels**, PA/HAL requires community-agreed minimum description profiles that capture device characteristics, processing pipelines, and operational definitions of key constructs. Alignment does not imply uniform measurement, but **transparent, machine-actionable documentation** of differences.

At the **provenance and computability levels**, analytical workflows and transformations should be described in ways that support traceability and federated execution. This is a precondition for reliable RWE generation and cross-site analysis under EHDS.

Crucially, alignment should be framed as an **enabler of innovation**, not a compliance exercise. When PA/HAL data are reusable by design, they can move more effectively across the publication–trial–patent–product continuum, strengthening Europe's innovation capacity without compromising decentralisation or trust.

Table 4. Main elements of the EU data agenda and their relevance to PA/HAL.

EU data agenda element	Core purpose	Relevance to PA/HAL	Practical implication for PA/HAL data
Strategic & governance frameworks			
European Strategy for Data	*Strategic framework for sectoral data spaces and interoperability*	Positions health as a flagship domain; supports reuse across sectors	PA/HAL data must be linkable across health, social, occupational, and environmental contexts
EHDS	*Governance and infrastructure for primary and secondary use of health data*	Creates the federation-first model for lawful secondary use	PA/HAL datasets need machine-actionable metadata, governance descriptors, and discoverability
HealthData@EU	*Cross-border infrastructure for EHDS secondary use*	Operationalises federated dataset discovery and access	PA/HAL data must be describable in ways compatible with federated catalogue environments
Data Governance Act	*Trusted data sharing, intermediaries, and data altruism*	Relevant for voluntary or mediated sharing of PA/HAL data	Governance and consent conditions should be more explicitly represented and reusable
Regulation & compliance			
AI Act	*Regulation of trustworthy AI, including high-risk systems*	Raises requirements for traceability, data quality, and documentation	PA/HAL data used in AI must include provenance, quality, and processing metadata
Research & open science infrastructure			
EOSC	*Research discovery and reuse ecosystem*	Supports cross-border discovery of datasets, software, and workflows	PA/HAL data need persistent identifiers, rich metadata, and stronger integration with research catalogues
OpenAIRE Graph/ knowledge graph approaches	*Interlinking research outputs and metadata at scale*	Demonstrates how distributed research assets can be connected and contextualised	PA/HAL could benefit from a semantic discovery layer linking datasets, variables, protocols, and software
ELIXIR/BBMRI-ERIC	*Life science and biobanking research infrastructure*	Provide FAIR data, metadata standards, and cross-border access components relevant to health-related data	PA/HAL should connect more explicitly to RI-supported metadata and access practices
Public health policy			
WHO Europe Physical Activity Strategy	*Normative public health strategy for PA monitoring and policy*	Defines policy demand for comparable PA indicators across populations	Metadata and semantic harmonisation are needed to make PA monitoring and reuse policy-relevant

Source: Authors' synthesis. PA/HAL = physical activity, healthy ageing and longevity. EHDS = European Health Data Space; EOSC = European Open Science Cloud; BBMRI-ERIC = Biobanking and BioMolecular Resources Research Infrastructure.

Table 5 translates the alignment agenda into concrete recommendations directed at specific actors in the European research and governance ecosystem, identifying the most relevant implementation mechanisms and indicative time-frames.

8.5 What is missing today: from "components" to an end-to-end PA/HAL pipeline

Europe already has many relevant components (cataloguing standards, clinical terminologies, research repositories, and federation tools), but PA/HAL reuse is constrained because these components are **not assembled into an end-to-end, domain-coherent discovery-to-compute pipeline**. The missing elements are largely "integration glue" rather than new institutions:

Table 5. PA/HAL interoperability alignment agenda: actor-specific recommendations. Priority ratings reflect the assessment of the authors based on institutional capacity, existing mandate, and impact on near-term PA/HAL reuse outcomes. Timeframes are indicative.

Actor	Recommended action	Mechanism/instrument	Priority	Timeframe
European Commission/ DG SANTE	Mandate a minimum PA/HAL metadata profile for EHDS-registered health data sources	EHDS secondary use implementing acts; Health Data Access Body guidance	**High**	2025–2027
EOSC Association/ National EOSC Nodes	Extend DCAT-AP with a PA/HAL domain profile; require domain metadata as indexing condition	EOSC Interoperability Framework; national open data portal guidelines	**High**	2025–2026
BBMRI-ERIC/ ELIXIR	Develop and publish FAIR Implementation Profiles (FIPs) for PA and HAL data collections	BBMRI-ERIC Directory; ELIXIR Human Data Communities; joint 2023 collaboration agreement	**High**	2025–2027
Research Funders (NWO, ANR, DFG, Horizon Europe)	Require Data Management Plans to include PA/HAL-specific provenance and semantic metadata as grant condition	DMP templates; Open Science requirements in calls; COST Action data management guidance	Medium	2025–2026
PA/HAL Research Community (COST Actions, ISA, ISBNPA)	Establish a community consensus process for PA/HAL minimum metadata standards (analogous to Mobilise-D lexicon)	COST Action networking grants; working group publications; journal data availability policies	Medium	2026–2028
Data Stewards and Repository Managers	Implement PA/HAL-specific metadata fields and version-controlled algorithm registries in institutional repositories	Zenodo community guidelines; institutional DMP review; EOSC onboarding requirements	Medium	Ongoing

Source: Authors' synthesis. EHDS = European Health Data Space; EOSC = European Open Science Cloud; BBMRI-ERIC = Biobanking and BioMolecular Resources Research Infrastructure; FIP = FAIR Implementation Profile; DMP = Data Management Plan; COST = European Cooperation in Science and Technology.

1. **PA/HAL discovery that is domain-relevant.** Generic cataloguing supports "a dataset exists," but not "this dataset contains comparable PA/HAL constructs with known processing and access conditions." Missing: a consistent way to expose PA/HAL-specific descriptors (device, protocol, derived measures, cohort context) in catalogue records and to link datasets to protocols, code, and access routes.

2. **Minimum PA/HAL metadata profiles adopted at scale.** The field lacks a widely adopted minimum set of metadata that makes PA/HAL variables interpretable across studies (device/placement/firmware; sampling/ epoching; wear-time rules; processing; derived endpoint definitions). Missing: community agreement plus incentives/requirements in repositories, calls, and access bodies.

3. **Semantic bridges between clinical, behavioural, and sensor-derived constructs.** Clinical terminologies cover diagnoses and labs well, but PA/HAL sits between domains. Missing: mappings and controlled descriptors that make PA constructs queryable and comparable across datasets, including explicit operational definitions for common constructs (e.g. MVPA, sedentary time, frailty proxies).

4. **Machine-actionable provenance as a norm.** In PA/HAL, derived measures are often the product. Missing: standard practice to publish pipeline lineage (raw→clean→features→endpoints), including algorithm and version identifiers, so results are auditable and reusable.

5. **Governance that is operational, not only legal.** Europe has strong legal guardrails; what's missing is executable governance at scale: consistent expression of reuse conditions, purpose limitations, access controls, and auditability that can be implemented across infrastructures and federated analysis contexts.

6. **Federated computability for PA/HAL features, not only clinical tables.** Federation works best where data models and semantics are mature. Missing: computability patterns for PA/HAL (how features/endpoints are represented so queries across sites produce comparable results without bespoke harmonisation each time).

These gaps explain why Europe can be strong in producing PA/HAL evidence, yet slower in converting it into scalable RWE workflows and innovation outcomes.

8.6 How to move from today's gaps to an EU-aligned PA/HAL reuse pathway

Moving from principle to practice requires aligning PA/HAL community conventions with EU federation and Open Science mechanisms. The practical route is to make PA/HAL datasets reusable by design:

1. **Define and adopt a minimum PA/HAL metadata profile** used by repositories, cohorts, and projects (as a requirement for "reusable" claims).

2. **Publish PA/HAL semantics as mappings and controlled descriptors**, ensuring that common constructs and derived endpoints are explicitly operationalised and queryable across datasets.

3. **Treat provenance as a first-class research output**, requiring that derived measures link to pipelines, parameterisation, and versioned algorithms.

4. **Enable federation by standardised representations of PA/HAL features**, so compute-to-data yields comparable outputs across sites.

5. **Align governance workflows with reuse**, so access conditions are consistently expressed, auditable, and implementable across distributed infrastructures.

This agenda is not about inventing new standards; it is about making existing European components interoperable in a PA/HAL-specific way, thereby allowing federation to deliver analytical validity and innovation velocity.

From a community perspective, the central requirement is not merely access to more PA/HAL data, but access to **interpretable and comparable PA/HAL evidence assets**; datasets whose meaning, processing history, and reuse conditions are explicit enough that multi-site analysis can be conducted without re-building harmonisation from scratch for each study.

Limitations

This analysis has several limitations that should be noted. First, the paper adopts a **conceptual and policy-facing analytical approach** rather than a systematic literature review; findings represent a structured synthesis of current evidence and policy frameworks, not a comprehensive or reproducible search of the literature. The identification of gaps and the formulation of the alignment agenda are necessarily interpretive and reflect the analytical framework applied, rather than a statistically derived consensus.

Second, the **innovation channel data** presented in Section 4.1 (publication counts, clinical trial registrations, and patent records) are proxy indicators of research activity and do not directly measure interoperability outcomes or the extent of data reuse. They are intended to contextualise the scale of PA/HAL research activity and the potential impact of improved interoperability, rather than to serve as precise measurements of innovation output.

Third, the **alignment agenda** proposed in Section 7 is prospective and has not been empirically validated. The recommendations are grounded in existing European infrastructure capabilities and policy directions, but their practical implementation will depend on institutional adoption, funding availability, and sustained community engagement that extend beyond the scope of this analysis.

Fourth, comparisons with non-EU jurisdictions, including the United States (NIH All of Us programme[20,21]) and other major data ecosystems, are illustrative rather than systematic. Jurisdictional differences in data governance, infrastructure maturity, and measurement traditions make direct comparison methodologically complex and beyond the scope of this paper.

Finally, the rapidly evolving nature of the European data governance landscape, including the ongoing EHDS transition phase and emerging EOSC governance arrangements, means that some policy and infrastructure details may change between the time of writing and publication. The analysis reflects the state of European data policy and infrastructure as of early 2026.

9. Conclusions and way forward

PA and HAL data sit at the intersection of clinical medicine, behavioural science, and public health, and therefore face **distinct interoperability challenges**. This paper has shown that Europe already possesses many of the necessary components (standards, infrastructures, and governance frameworks) but that these are rarely assembled into an end-to-end, domain-coherent pipeline.

The consequences extend beyond data management. Fragmented semantics, metadata, and provenance constrain the translation of PA/HAL data into scalable innovation, real-world evidence, and societal impact. Addressing these challenges is therefore not a technical refinement, but a strategic necessity.

Europe's PA and HAL research ecosystem is well-positioned for high-impact prevention science, but constrained by structural weaknesses in cataloguing, metadata, and semantics. EHDS strengthens the federation-first direction for health data governance, while EOSC and Open Science mechanisms aim to scale discovery and reuse of research objects across Europe. The central task for PA and HAL is therefore not to create new data silos, but to make existing and future datasets *computably comparable* through consistent metadata, shared taxonomies, and transparent provenance.

By aligning PA/HAL data practices with FAIR principles and EHDS requirements, while respecting Europe's federated, trust-based approach, Europe can strengthen its position in PA/HAL research and innovation. The opportunity lies not in centralising data, but in making distributed data meaningfully reusable, enabling PA/HAL evidence to flow more effectively from measurement to impact.

Ethics and consent

Not applicable. This study did not involve human participants, personal data, or animal subjects.

Data availability

No new datasets were generated or analysed for this study. All data sources referenced are derived from publicly available policy documents, standards documentation, and published research.

Extended data

1. PA/HAL Data Systems, Standards, Infrastructures, and Tools. A comprehensive reference overview of the data systems, standards, infrastructures, and tools referenced in this analysis, organised by seven domains: cataloguing and dataset discovery (A); metadata standards (B); semantic terminologies and ontologies (C); provenance and traceability (D); access, governance, and secondary-use mechanisms including EHDS implementation initiatives (E); computability and federation tools (F); and PA/HAL data capture platforms (G). Available at: **https://doi.org/10.5281/zenodo.19696719.**[22]

Acknowledgements

The author thanks the European research and policy communities working on physical activity, healthy ageing, Open Science, and federated health data infrastructures for ongoing dialogue that has informed the perspectives presented in this article.

Artificial intelligence assistance: The author used Claude (Anthropic) and ChatGPT (OpenAI) to assist with drafting and editing this manuscript. All content was reviewed and verified by the author.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci. Data.* 2016; **3**: 160018.
Publisher Full Text

2. European Commission: **Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space (EHDS).** *Off. J. Eur. Union.* 2025 March 5; L 2025/327.

3. European Commission: **European Health Data Space — digital strategy and implementation initiatives.**[Accessed April 2026].
Reference Source

4. European Parliament and Council: **Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).** *Off. J. Eur. Union.* 2024; L 2024/1689.

5. European Commission: *Accelerating Clinical Trials in the EU (ACT EU).* Brussels: European Commission; 2022.
Reference Source

6. World Health Organization Regional Office for Europe: *Physical Activity Strategy for the WHO European Region 2016–2025.* Copenhagen: WHO/Europe; 2016.

7. European Commission: *A European Strategy for Data. COM(2020) 66 final.* Brussels: European Commission; 2020.

8. European Patent Office: *EPO Patent Index 2024.* Munich: EPO; 2024. [Accessed April 2026].
Reference Source

9. McKinsey & Company: *The trends defining the $1.8 trillion global wellness market in 2024.* McKinsey Consumer Packaged Goods; 2024. [Accessed April 2026].
Reference Source

10. EIT Health: *EIT Health Annual Report 2024.* Brussels: EIT Health; 2024. [Accessed April 2026].
Reference Source

11. Bonci T, Keogh A, Del Din S, *et al.*: **Mobility recorded by wearable devices and gold standards: the Mobilise-D procedure for data standardization.** *Sci. Data.* 2023; **10**: 38.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Mikolaizak AS, Rochester L, Maetzler W, *et al.*: **Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement — the Mobilise-D study protocol.** *Digital Biomarkers.* 2022; **66**(3): 167–187.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Doherty A, Jackson D, Hammerla N, *et al.*: **Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank study.** *PLOS ONE.* 2017; **12**(2): e0169649.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Ye Y, Benzon HT, Sherban V, *et al.*: **Wearable accelerometer-derived physical activity and incident disease.** *npj Digital Medicine.* 2022; **5**(1): 131–140.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Rowlands AV, Harrington DM, Bodicoat DH, *et al.*: **Population-referenced centiles for wrist-worn accelerometer-measured physical activity and sedentary time in UK adults.** *J. Sport Health Sci.* 2025.
Publisher Full Text

16. Ranjan Y, Rashid Z, Stewart C, *et al.*: **RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices.** *JMIR Mhealth Uhealth.* 2019; **7**(8): e11734.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. OpenAIRE: **OpenAIRE Research Graph.**[Accessed April 2026].
Reference Source

18. BBMRI-ERIC: **BBMRI-ERIC Directory.**[Accessed April 2026].
Reference Source

19. ELIXIR: **ELIXIR Research Data Management.**[Accessed April 2026].
Reference Source

20. Denny JC, Rutter JL, Goldstein DB, *et al.*: **The "All of Us" Research Program.** *N. Engl. J. Med.* 2019; **381**(7): 668–676.
Publisher Full Text

21. Singh R, Tetrick MK, Fisher JL, *et al.*: **Analysis of physical activity using wearable health technology in US adults enrolled in the All of Us Research Program: multiyear observational study.** *J. Med. Internet Res.* 2024; **26**: e65095.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Zacharis T: **Extended Data: PA/HAL Data Systems, Standards, Infrastructures, and Tools.** *Zenodo.* 2026.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✗

Version 1

Reviewer Report 25 August 2026

https://doi.org/10.21956/openreseurope.25591.r79454

© **2026 Pires I.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✗ **Ivan Miguel Pires**
[1] Universidade de Aveiro Escola Superior de Tecnologia e Gestao de Agueda (Ringgold ID: 386349), Águeda, Aveiro District, Portugal
[2] Instituto de Telecomunicacoes (Ringgold ID: 131676), Aveiro, Aveiro District, Portugal

The manuscript addresses a timely and important topic: how physical activity and healthy ageing data can be made more interoperable, reusable, and suitable for federated analysis within the European research environment. The proposed six-layer framework covering cataloguing, metadata, semantics, provenance, governance, and computability is useful and provides a clear structure for the discussion. The examples from wearable data, Mobilise-D, and UK Biobank also help illustrate why differences in devices, processing methods, and variable definitions can create serious problems for data reuse.

Overall, I think the paper has academic merit, but several important issues should be addressed before indexing. My main concern is that the manuscript presents a “gap analysis” without giving enough detail about how the literature, policy documents, databases, and examples were selected. The authors themselves acknowledge that this was not a systematic or reproducible literature review. This is acceptable for a conceptual or perspective paper, but the manuscript should be clearer about this and avoid wording that suggests the identified gaps have been established systematically across the whole field.

There are also problems with some of the quantitative comparisons. In particular, the ClinicalTrials.gov section compares the global number of studies with EU “trial-site registrations,” which are not equivalent measures. The reported EU percentage should therefore be recalculated using the same unit for the numerator and denominator. The patent analysis also needs clarification, as the approximately 40% figure appears to relate to broader EPO medtech activity rather than specifically to EU-27 PA/HAL patents. Table 2 even contains wording indicating that EU-specific PA/HAL patent families still need to be verified. These points should be corrected before indexing.

Another important issue is the coverage of current European metadata developments. For a

manuscript centred on EHDS-compatible cataloguing, the discussion should engage much more directly with HealthDCAT-AP and explain whether the proposed PA/HAL metadata requirements are already represented there or whether a PA/HAL-specific extension is genuinely needed. This would make the recommendations more practical and better connected to the current EHDS implementation environment.

The literature is relevant but could also be strengthened. More recent work on physical-activity ontologies, wearable measurement standards, and reproducible digital mobility analysis would provide stronger support for the manuscript's central claims. There is also at least one citation problem: the manuscript refers to a 2025 Mobilise-D Technical Implementation Manual as reference 12, but reference 12 is actually a 2022 study protocol. I would therefore recommend checking the full reference list carefully.

The data-availability statement should also be reconsidered. Although no new participant-level dataset was collected, the manuscript does analyse public databases such as PubMed, ClinicalTrials.gov, and Lens.org. To make these results reproducible, the authors should provide the exact search terms, dates, filters, exported results, and any calculations or scripts used to produce the figures and percentages.

In summary, I see value in the paper and would encourage revisions. The conceptual framework is useful and the central message—that PA/HAL data require better metadata, semantics, and provenance to support meaningful reuse—is convincing. However, the quantitative analyses need correction, the methodology needs to be described more transparently, the current EHDS metadata landscape should be incorporated, and some of the broader claims should be moderated unless additional evidence is provided.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Digital health, physical activity, healthy ageing, health data science, and

research data interoperability.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 04 Sep 2026

Theo Zacharis

I thank Dr Pires for his careful and constructive review. I appreciate the time he spent engaging with the manuscript. The comments are clear and useful. I agree with the feedback on all the points. A revised version (v2) is in preparation. My planned response to each point is set out below.
1. Methodology and scope. I agree that the manuscript needs to clearly say it is a conceptual and policy-focused analysis, not a systematic review. It should also explain how the literature, policy documents, databases, and examples were chosen. V2 will clearly state the analytical approach and selection basis. The wording will change. The identified gaps will show as results of the applied framework, not as systematic findings in the field.
2. ClinicalTrials.gov figures. I recognise that the reported EU share compares different types of units. The global figure counts studies, but the EU figure adds up trial-site records from member states. v2 will report this comparison on a matched unit basis, with the query, filters, execution date, and raw export documented in the Extended Data. Figure 3 and Table 2 will be revised accordingly, and the counting rule will be stated explicitly in the caption.
3. Patent analysis. I understand that the 40% figure comes from EPO member states' share of global medtech filings and is not a specific EU-27 figure for PA/HAL. The placeholder wording in Table 2 should not have remained in the published version. v2 will either report a jurisdiction-resolved figure for the relevant CPC scope, with the query and attribution rule documented, or, if a defensible domain-specific figure cannot be established, attribute the medtech figure clearly to its actual scope and remove the inference from it to PA/HAL patenting. The placeholder wording will be removed in either case.
4. HealthDCAT-AP. I agree that a manuscript focused on EHDS-compatible cataloguing should connect with HealthDCAT-AP. V2 will include a treatment to check which PA/HAL descriptors are in HealthDCAT-AP. It will also explore if a domain-specific extension is needed. I agree this will make the recommendations clearer and better fit the current EHDS setup.
5. Reference [12]. I see the citation mistake: the text mentions the Mobilise-D Technical Implementation Manual, but reference 12 is the 2022 study protocol. I'll fix this and audit the entire reference list against each in-text citation, not just this one.
6. Reproducibility of data-source analyses. I agree the database analyses in Section 4.1 need full query provenance. The data availability statement must reflect this. A new version of the Extended Data record will be uploaded to Zenodo. It will include the verbatim query for each source, field tags, all filters and their settings, the execution date, counts returned, raw exports, and calculations for reported percentages. The rule for assigning records to EU-27 will be clearly stated, as this definition significantly impacts the figures.
7. Literature. I agree with the idea to enhance coverage of physical activity ontologies, wearable measurement standards, and reproducible digital mobility analysis. I will add

recent work in these areas.

The Section 4.1 analysis aims to place EU-27 PA/HAL research and innovation capacity in a global context. This supports the idea that interoperability is key to turning that capacity into competitive strength. v2 will ensure the underlying figures are constructed on a basis that properly supports that framing. I appreciate the review for pinpointing concrete issues and suggesting solutions. A revised version and a detailed point-by-point response will follow.

Competing Interests: No competing interests were disclosed.