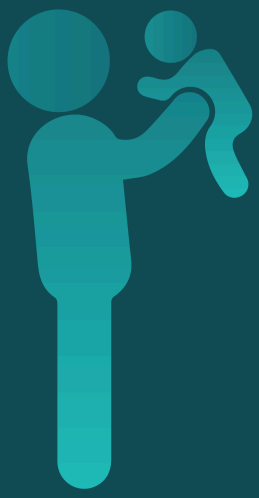
DATAcc by DiMe

INDUSTRY BRIEF

Structuring digital measurements for pediatric rare disease:

Data standards, platforms, and common data elements for registries



Introduction

Pediatric rare diseases (PRD) are characterized by small, heterogeneous patient populations, complex phenotypes, and long, uncertain diagnostic and treatment journeys. Currently, our understanding of these conditions relies heavily on qualitative measures, anecdotal evidence, and phenotype documentation. However, looking to the future, **digital health technologies** (DHTs) - such as wearables, smartphone apps, in-home sensors - offer a way to evolve this approach by capturing clinically meaningful aspects of children's lives where they actually happen: at home, at school, and in the community.

The Digital Medicine Society's (DiMe) DATAcc initiative defined a core set of digital measures for pediatric rare disease to accelerate clinical research and transform care. These measures are grounded in a conceptual model that links:

- **Meaningful aspects of health** for children and families (e.g., walking to school, playing, communicating, sleeping well),
- **Concepts of interest** (e.g., motor function and mobility, cognition, adaptive function, seizure control, sleep, pain), and
- **Outcomes to be measured** (e.g., stability and balance, gait, ambulation volume, upper limb function, feeding independence, seizure frequency, nocturnal seizure events, core measures of sleep).

Digital measures can quantify these outcomes with higher resolution than traditional clinic-based assessments, but they only deliver their full value if they are **structured**, **standardized**, and **interoperable**. Pediatric rare disease programmes are beginning to rely on complex clinical data from multiple sources, and without shared semantic standards, curated common data elements (CDEs), common data models (CDMs), and agreed exchange formats, those data remain siloed, hard to compare, and difficult to reuse for regulatory and health technology assessment (HTA) decisions. A standards-based ecosystem that links ontologies, PRD focused CDEs, CDMs, data exchange standards, platforms, and networks make digital measures interoperable by design, reduces harmonization effort, and allows evidence to accumulate across trials, registries, and real-world datasets over time.

To support the development of the data ecosystem components, this brief answers three questions:

- What standards, platforms, and networks already exist that PRD stakeholders can build on?
- How can we structure the PRD core digital measures using common data elements (CDEs), so that registries and trials collect compatible data?
- How can these structured data support drug development and regulatory pathways such as qualification of digital measures as Medical Device Development Tools (MDDT),evidence generation for health technology assessment (HTA) or payor decision-making?

Who is this brief for?

This industry brief is designed as a foundational guide for stakeholders navigating the intersection of pediatric rare disease (PRD) and digital health.

It is particularly valuable for **patient advocacy leaders** and **registry custodians** seeking to understand the critical role of data standards without needing deep technical expertise. It also serves **biopharma sponsors** and **clinical researchers** new to the digital measures space, providing a strategic overview of how to build interoperable, regulatory-ready evidence. Finally, **data architects** and **technology developers** looking for a high-level conceptual framework will find this a useful roadmap for integrating novel digital data streams into the established rare disease ecosystem. Use this resource to align your team on the "big picture" of interoperability before diving into technical implementation.

Landscape overview: data standards, platforms, and networks

The rare disease ecosystem has invested heavily in a comprehensive suite of **data standards and infrastructures**, including semantic standards, common data elements, common data models, data exchange standards, registry platforms, and multi-site research networks. PRD digital measures should plug into this ecosystem rather than create yet another silo.

Taken together, these components form a **complete interoperability stack** that stakeholders can use end-to-end:

- **Semantic standards** provide the *language*: controlled vocabularies and ontologies that ensure diseases, phenotypes, and meaningful aspects of health are described unambiguously.
- **Common Data Elements** provide the *minimum dataset*: standardized variables and definitions collected consistently across registries and studies.
- **Common Data Models (CDM) and data exchange standards** provide the *structure and movement*: how data are organised in databases and how they move between systems,
- **Data platforms and registry infrastructures** provide the *technical environment*: where these data standards are implemented in practice.
- **Networks, consortia, and registry collaborations** provide the *community alignment and scale*: multi-country and multi-institution groups that harmonize data collection, governance, and shared analytics.

Together, these components enable rare disease stakeholders (registries, sponsors, digital health companies, clinicians, and regulators) to **collect, structure, exchange, and analyze** PRD digital measures in a consistent, interoperable way.

Semantic standards: Standardized terminologies and ontologies

- They are controlled vocabularies and hierarchies (codes, terms, relationships) that ensure your measurement and context are described unambiguously. This semantic layer is essential if we want digital measures (such as gait features or sleep metrics) to be discoverable and analyzable across diseases and datasets.
- **Examples:** Orphanet, The Human Phenotype Ontology (HPO), Mondo Disease Ontology (Mondo), Disease Ontology (DO), MAxO, SNOMED-CT, COEL Model

- **PRD-specific use:** These resources allow PRD digital datasets to use universally recognized codes and terms to:
    - Encode **diagnoses**
    - Encode **activities of daily living**
    - Encode **clinical features and meaningful aspects of health** (e.g., abnormal gait, dystonia, ataxia, seizures, sleep disturbance, feeding difficulties, pain)

  Map DiMe's PRD core measures to applicable ontologies to standardize data coding.

Common Data Elements (CDEs) for rare disease registration

- Standardized data fields (variables, definitions, permissible values) that specify what information should be collected in a registry or study and how it should be recorded. CDEs create a consistent "minimum dataset" across projects so that data can be pooled, compared, and reused without extensive post-hoc harmonization. For digital measures, they ensure that each recorded value (e.g., step count, seizure frequency, sleep efficiency) is anchored in a well-defined context: who the patient is, what condition they have, when and how the data were collected.
- **Examples:** EU Rare Disease Platform: Common Data Elements (CDEs), NIH/NCATS/GRDR® Common Data Elements
- **PRD-specific use:** For pediatric rare diseases, CDEs provide the backbone for registries that aim to incorporate core digital measures. They:
    - Ensure that every digital outcome (e.g., gait variability, upper-limb activity, nocturnal seizure counts, sleep continuity metrics) is systematically linked to shared core fields such as demographics, diagnosis, disease stage, and comorbidities.
    - Enable registry datasets to directly support drug development, MDDT qualification, and HTA/payor analyses, since the underlying data structure is consistent, well-defined, and interoperable with broader rare disease standards.

Common Data Models (CDMs) & data exchange standards

- They are structural and technical standards that specify how data are organised, exchanged, and integrated across systems. Common Data Models (CDMs) define the schema - tables, fields, and relationships - that allow different datasets to be harmonized and queried in the same way. Data

exchange standards define the format and transport - how data move between systems (e.g., registries, EHRs, digital platforms) in a consistent, machine-readable manner.

- **Examples:**
  - CDMs: OMOP CDM (OHDSI), PCORnet CDM, PEDSnet CDM
  - Data Object standards: COEL Behavioural Atom
  - Data Exchange standards: HL7 FHIR, CDISC Standards, Open mHealth schemas, Phenopackets
- **PRD-specific use:**
  - Seamless integration of digital measures with clinical and registry data.
  - Preservation of event-level context and consent using a standardized privacy-preserving data object (e.g. COEL Behavioural Atom) that can feed directly into CDMs and data exchange standards.
  - Consistent structuring of wearable-derived features (e.g., gait metrics, sleep efficiency, seizure counts).
  - Rapid data movement between devices/apps and registries using FHIR APIs and Open mHealth schemas.
  - Multi-site PRD research in learning health systems where analytic code runs identically across hospitals and registries.
  - Regulatory readiness, since CDISC, OMOP, and FHIR standards are accepted or preferred by regulatory agencies for digital health and real-world evidence submissions.
  - HTA/payor-grade analyses, because consistently structured data can be pooled to demonstrate functional benefit, natural history characteristics, and real-world change over time.

Data platforms, repositories, and registry platforms

- They are technical infrastructures that store, curate, and manage rare disease data. They provide mechanisms for registry creation, metadata management, secure data hosting, and controlled-access sharing. These platforms act as the operational layer supporting CDEs, semantic standards, and CDMs, making them the practical environment where real datasets live and evolve.
- **Examples:** ERDRI (European Rare Disease Registry Infrastructure), RD-Connect Registry & Biobank Finder, PEDSnet data platform, RDCA-DAP, IAMRARE®, Phenome Central, disease-specific registry platforms (e.g., TREAT-NMD's DMD registry infrastructure, metabolic disease registry platforms, oncology registries etc.)

- **PRD-specific use:** These platforms serve as the anchor points for PRD digital measures. They provide:
    - A place to store digital measurement datasets with consistent CDEs and semantic annotations.
    - Infrastructure to integrate home-based digital data with clinical, genetic, and developmental data.
    - Governance frameworks for controlled data sharing, enabling regulators, sponsors, and networks to reuse data for endpoint validation, natural history characterisation, and external control arms.
    - Incorporating the curated PRD CDE list directly into platform metadata (e.g., ERDRI.mdr-style dictionaries) ensures rapid adoption by registry developers.

Networks, consortia & registry networks

- They are multi-site collaborations, research consortia, and clinical networks that coordinate rare disease research, data sharing, natural history studies, and multi-country registry governance. They provide community-wide alignment, making standards and CDEs actionable across institutions.
- **Examples:** Solve-RD, Matchmaker Exchange (MME), ERNs (European Reference Networks, e.g., ERN-RND, ERN-NMD, ERN-ITHACA), PEDSnet network, CBTN (Children's Brain Tumor Network), disease-focused consortia (e.g. TREAT-NMD for neuromuscular diseases), international multi-registry initiatives linking natural history datasets and biobanks.
- **PRD-specific use:** These networks will be essential for scaling PRD digital measures because they enable:
    - Access to harmonized, deeply phenotyped datasets and federated analysis workflows for discovery and benchmarking.
    - Diagnostic acceleration and gene discovery through global matchmaking platforms.
    - Reusable governance and federated data-sharing models that inform new PRD digital and PGHD platforms.
    - Harmonized cohort building and rapid trial readiness across countries, supporting natural history research and real-world evidence.

PRO TIP: Connecting core digital measures for PRD to this ecosystem

Based on the described components of the ecosystem, from a top-down perspective, the workflow to integrate PRD digital measures into existing standards would look as follows:

1. **Define the condition and cohort** (e.g. using ORPHAcode/MONDO).
2. **Describe clinical features and meaningful aspects of health** using HPO (Human Phenotype Ontology).
3. **Define the digital measure concept** (e.g., ambulatory mobility, hand function, seizure burden, sleep continuity) and align it with appropriate ontologies and vocabularies.
4. **Map the resulting variables to a CDM** with clear links to visits, devices, and clinical context.
5. **Capture CDEs** for device metadata, protocol, and environment to ensure interpretability.

Structuring core digital measures of PRD with common data elements

Rare disease research is often constrained by fragmented data, inconsistent definitions, and poor comparability across studies. Our work aligns digital measures with the meaningful aspects of health most relevant to patients with rare diseases and their families. By embedding these measures in a harmonized, implementation-ready data structure, it enables more consistent data collection and greater interoperability across registries and clinical trials. This approach is set to improve data quality, reduce burden for both patients and researchers, and helps accelerate the development of therapies that reflect outcomes that truly matter to the rare disease community.

DiMe DATAcc project defined a core set of digital measures rooted in the meaningful aspects of health of the patients with rare diseases and their families. To make these measures usable in registries and trials, we propose a bottom-up data structure expanding on the available CDEs. This structure was created based on two currently available CDE structures for rare diseases: EU Rare Disease Platform: Common Data Elements (CDEs) and NIH/NCATS/GRDR® Common Data Elements, together with critical input from the pre-competitive expert group.

A proposed CDE structure for core measures of PRD:

Core registration layer	Patient ID, date of birth, sex, country, etc. (directly taken from EU RD CDEs). Rare disease diagnosis (ORPHAcode, MONDO ID [assess the need based on context & geography, e.g. MONDO for interfacing with EPIC, or ORPHA if used in the EU]) Age at onset, diagnostic confirmation, genetic variant (where available) Consent metadata (consent, purpose of processing, retention)
Phenotype & meaningful-aspect layer	HPO terms reflecting the main measured **concepts of interest**. **Example: Motor function and mobility** - Abnormal gait, ataxia, muscle weakness, dystonia
Digital measurement context layer	Device identifier (vendor, model, type (accelerometer, IMU garment, EEG, actigraphy, camera)). Device placement (e.g., wrist, ankle, chest pendant, full-body suit). Sampling properties (sampling rate, epochs, feature extraction window). Protocol vs free-living (in-clinic tests, at-home continuous monitoring). Wear time and compliance measures (valid days, hours/day).
Event & behavioral episode layer	**Event-level objects** defining the singular "what happened" records (meals, falls, exercise bouts, sleep episodes, med intake, symptom attacks, context changes) that (a) have their own timestamps/durations and (b) often serve as inputs, stratifiers, anchors, or covariates for multiple outcomes.
Outcome-specific digital feature layer	The final, analysis-ready measures derived from digital health technologies. This layer summarizes raw sensor data into clinically meaningful metrics aligned with specific outcomes, while preserving links to context, events, and source data. **Example: Mobility and gait outcome measures**

Example use of the CDE structure for core measures of PRD:

Core registration layer	
Field	*Example Value*
Patient ID	PRD-00123
Date of Birth	2015-04-12
Sex	Male
Country	Germany
Diagnosis (ORPHAcode)	ORPHA:98896 (DMD)
Diagnosis (MONDO ID)	MONDO:0007178
Age at Onset	3 years

Diagnostic Confirmation	Genetic test - confirmed
Genetic Variant	DMD gene, exon 45 deletion
Phenotype and meaningful-aspect layer	
Field	*Example Value*
Device Vendor	ActiWear Labs
Device Model	ActiWear Pro v2
Device Type	Accelerometer
Device Placement	Ankle (right)
Sampling Rate	100 Hz
Epoch Length	60 seconds
Feature Extraction Window	30 seconds sliding, 50% overlap
Protocol Type	7-day at-home continuous monitoring
Recording Dates	2025-02-01 to 2025-02-07
Valid Wear Days	6/7
Average Daily Wear Time	13.8 hours/day
Digital measurement context layer	
Field	*Example Value*
Device Vendor	ActiWear Labs
Device Model	ActiWear Pro v2
Device Type	Accelerometer
Device Placement	Ankle (right)
Sampling Rate	100 Hz
Epoch Length	60 seconds
Feature Extraction Window	30 seconds sliding, 50% overlap
Protocol Type	7-day at-home continuous monitoring
Recording Dates	2025-02-01 to 2025-02-07
Valid Wear Days	6/7
Average Daily Wear Time	13.8 hours/day
Event and behavioral episode layer	
Field	*Example Value*
Event type	Daily walking activity (aggregated)

Source	Derived from device-detected walk bouts (ActiWear Pro v2 algorithm v1.3)
Time bounds	Start Time: 2025-02-03T00:00:00+01:00 End Time: 2025-02-03T23:59:59+01:00 Duration: 24 hours Timezone: Europe/Berlin (CET)
Location/context	Location: Predominantly home / local community Context: Mixed indoor and outdoor free-living walking
Links	Linked Patient ID: PRD-00123 Linked Recording Window: 2025-02-01 to 2025-02-07 Linked Device: ActiWear Labs | ActiWear Pro v2 | Accelerometer | Ankle (right) Linked Raw Data Segment: prd00123_actiwearv2_20250203_full_day.bin Linked Outcome Features: Daily Step Count; Walking Time; Number of Walking Bouts; Maximum Walking Bout Duration; Cadence; Gait Variability Index; Stability Index
Confidence score	Confidence Score: 0.95 Confidence Method: Aggregate confidence

Outcome-specific digital feature layer (incl. raw data)	
Feature Name	*Example Value*
Daily Step Count (mean)	5,842 steps/day
Walking Time (mean)	92 minutes/day
Cadence (median)	103 steps/min
Gait Variability Index	0.42 (unitless)
Stride Regularity	0.67
Stability Index (lower = better)	0.31
Sit-to-Stand Transitions	28 transitions/day
Transition Duration (mean)	1.8 seconds
Workspace of Lower Limb Movement	0.83 m^2
Maximum Walking Bout Duration	11.4 minutes
Number of Walking Bouts/Day	62 bouts/day

Example from published research

The KineDMD study in Duchenne muscular dystrophy used a **full-body IMU suit with 17 sensors** during real-world activities of daily living (source). Investigators recorded ADLs (walking, climbing stairs, playing) alongside standard clinical assessments (6MWD, NSAA, PUL, grip strength). Data were organised in conceptual layers such as:

1. **Raw sensor layer** (high-frequency orientation and acceleration).
2. **Feature layer** for each subject × visit, ~100+ summary features.

The feature layer is essentially a **CDE-ready table**: subject, visit, feature columns, and clinical outcomes. As the next step, standardizing names, units, and definitions of these features using the **CDEs** would make it far easier to compare across cohorts and re-use them in future rare disease registries.

Implementation recommendations

These recommendations reflect consensus input from a cross-sector, pre-competitive expert group and are anchored in established rare disease data standards. They are designed to be practical, scalable, and aligned with regulatory and evidence-generation needs, helping stakeholders avoid fragmentation and accelerate real-world impact.

For registry developers and networks:

1. **Adopt the PRD curated CDE list as a default data structure**
   - Incorporate the core demographic, diagnosis, and disease history CDEs directly from the EU RD Platform and RD-Connect frameworks.
   - Add the PRD digital measurement context and outcome modules that are most relevant to your disease area.
2. **Use standard terminologies for key fields**
   - Encode diagnoses with ORPHAcode and MONDO where possible.
   - Use HPO terms to record phenotypes and meaningful aspects of health.
   - For behavioural and activity-related events, consider using a classification taxonomy such as the COEL Model to label everyday living events consistently across devices and registries.

3. **Plan for CDM mapping from the outset**

   Design registry databases so CDEs map cleanly to OMOP or PEDSnet CDM tables, enabling future integration with hospital EHR data and multi-site analyses.

4. **Use a standard event data object for behavioural data**

   Where registries ingest behavioural data from multiple measurement systems, represent individual events or observations using a standardized data object standard such as the COEL Behavioural Atom to extend behavioural health coverage and ontology alignment so that consent and measurement context are preserved while mapping into CDMs.

5. **Expose metadata and schemas openly**

   Publish registry variable dictionaries and CDE mappings in accessible formats (e.g., ERDRI.mdr-style metadata repositories), encouraging reuse and alignment (for example, as described in *An ontology-based rare disease common data model harmonising international registries, FHIR, and Phenopackets*).

For sponsors, CROs, and digital health companies:

1. **Design digital endpoints around the curated CDEs**
   - Use the same variable names, units, and derivation rules as the CDE list whenever possible.
   - Where novel features are needed, define them clearly and propose them back into the shared CDE resource.
   - When deriving endpoints from behavioural data, structure event-level data using a data object standard such as the COEL Behavioural Atom.

2. **Align trial datasets with registry structures**

   Plan clinical trial CRFs so the core PRD CDEs are collected in both trial and registry contexts, enabling linkage and external control arm development.

3. **Build evidence packages with reuse in mind**

   When planning validation studies and MDDT submissions, target designs that can be replicated across registries or CDM-mapped datasets, leveraging the shared data structure (Read more in: *Unleashing the full potential of digital outcome measures in clinical trials: eight questions that need attention*)

For regulators and HTA/payor bodies:

1. **Signal the value of standardized digital measures**
   - Encourage submissions that use well-defined, harmonized CDEs and CDMs, noting that they facilitate comparability and cumulative evidence building.

- Encourage submissions that use data object representations alongside harmonized CDEs and CDMs, since these support audit trails from raw data to aggregate endpoints.

2. **Leverage registries in assessments**

   When evaluating PRD therapies, look for digital measures that can be contextualised using registry-based natural history data collected with the curated PRD CDE list.

For patient and advocacy organizations:

1. **Champion the use of open, shared CDEs**

   Ask registry and trial partners how they are aligning with PRD CDEs and digital measurement standards.

2. **Ensure meaningful aspects of health remain central**

   Work with researchers to ensure that the curated CDE list continues to reflect outcomes that matter most to children and families (e.g., independence, participation, sleep, pain, caregiver burden).

Conclusion

Digital measures have enormous potential to make pediatric rare disease research more sensitive, less burdensome, and more reflective of children's daily lives. But this potential will only be realised if **data are collected, structured, and shared in consistent ways**.

By curating a pragmatic **PRD-focused common data elements** for digital measures, the community can create registries and trials that are **interoperable by design** and ready to support **drug development, MDDT qualification, and HTA/payor decisions**.

The examples from published research studies show that the core technical and analytic building blocks are already in place. The next step is to define and implement consistent data structures. The field has reached an inflection point: the technical and analytic foundations for digital measures in rare disease are already in place. The next step is a collective action to define and implement consistent data structures.

This industry brief, unified common data elements, as well as DiMe's open-access resources provide powerful tools to support stakeholders in taking these steps and start translating the best practices into practice to ultimately change lives of child rare disease patients worldwide.