RegenMed

WHITE PAPER

CIRCLE DATASETS MEET THE CHALLENGES OF FEDERATED HEALTHCARE DATA CAPTURE

November 2025

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The First Challenge: Healthcare Data Quality

Artificial intelligence (AI) has entered medicine with extraordinary promise—and deep fragility. From diagnostic imaging to population health analytics, thousands of algorithms have been published, yet only a fraction deliver reliable results in clinical practice. The reason is fundamental: *AI in healthcare has outgrown its data foundation.*

AI and other "big data" approaches to healthcare data solutions are trained on unstructured or opportunistic data extracted from electronic health records, claims systems, or research silos. These sources were never designed for scientific reproducibility, regulatory compliance, or continuous learning. They are riddled with bias, missingness, and unverifiable provenance. The result is a widening gap between technological potential and clinical trust.

If are to achieve clinical reliability, its foundation must change—from uncurated data to **structured, validated, longitudinal, and interoperable real-world evidence**. That is the core insight behind **RegenMed's Circle Datasets**. A deep look into these issues, and how Circle Datasets solve them, can be found in our White Paper, ***Circle Datasets As Ground Truth for AI in Healthcare***.

The Second Challenge: Healthcare Data Ownership and Control

The era of AI-augmented medicine alters fundamentally who *owns*, *controls*, and *is accountable for* health-data assets. Traditionally, healthcare data has been generated within provider settings, coded for operations or billing, and held under institutional stewardship. With advanced analytics and real-world evidence platforms, new actors emerge: AI vendors, platform providers, federated data networks, and third-party aggregators. This creates dual pressure: an ethics mandate (patient autonomy, fairness) and a business mandate (data as asset).

Recent analysis of individual-level health data finds that neither pure private ownership nor pure public ownership models satisfactorily balance autonomy, justice and utility. Instead, scholars argue for governance frameworks that emphasize *procedural control* (e.g., data access committees, managed access) over simple notions of "ownership".

In short: data control is no longer limited to the institution that captured the record — it spans networks, pipelines and models. For AI to deliver safe, interoperable, longitudinal insight, ownership and governance must be re-defined in the context of provenance, auditing and accountability.

Each Circle contributes to a federated network of participating institutions and independent providers —a **Circle Ecosystem** where data quality improves as participation grows.

- Clinicians receive feedback through benchmarking dashboards.
- Researchers gain access to de-identified longitudinal cohorts.
- Regulators obtain continuous, auditable real-world evidence.
- Patients benefit from safer, explainable, and continuously improving AI tools.

This creates a self-reinforcing feedback loop: better data → better models → better care → broader participation → still better data.

The Future

Healthcare is entering a phase where *data integrity* and *data ownership/control* will matter more than *model novelty* and *data quantity*. RegenMed's Circles Platform establishes the infrastructure for that era—a verifiable, federated, and continuously learning evidence system that bridges human expertise and machine intelligence.

Circle Datasets are not another data repository; they are the **scientific substrate** for trustworthy AI in medicine. They turn artificial intelligence into accountable intelligence.

DATA OWNERSHIP AND ACCOUNTABILITY IN THE AGE OF AI

Accountability (Liability, Traceability, and Governance)

When AI and other big data models make or support clinical decisions, the question of *who is accountable* becomes critical. Is liability with the vendor, the institution, the clinician, the data-owner or some combination? A policy brief from Stanford HAI underscores that many courts lack precedents for AI-driven harm, making it difficult to assign liability under traditional tort frameworks.

Specific issues in accountability include:

- **Traceability of data lineage**: Without a documented chain of custody from capture to model output, it becomes difficult to audit or assign liability. For example, the concept of data provenance has been defined as the "attributes about the origin of health information at the time it is first created and tracks the uses and permutations of the health information over its lifecycle".
- **Auditability of AI models**: Governance frameworks emphasize the need for metadata standards, transparency "nutrition-labels", and ongoing monitoring of real-world performance.
- **Governance gap**: Even as institutions adopt AI, many find they lack internal governance structures for AI-enabled digital health tools — for instance, institutional review boards (IRBs) may not be configured to assess data pipeline risks, model drift or federated architecture.

Ownership Implications For Data Capture, Sharing and Monetization

Ownership and control of data underpin several commercial and ethical dynamics:

- **Access and sharing protocols**: If data is held by one institution but integrated into multi-site networks or pooled for AI modelling, the original "owner" may lose operational control over derivatives, models, and insights. Governance mechanisms must therefore define rights for derivative creation, secondary use, and downstream model outputs.
- **Monetization and value capture**: As RWE platforms monetize datasets, clarity on ownership and rights (e.g., commercial licensing, reuse, downstream models) is imperative. Without clear rights, legal risk escalates and value may diminish.
- **Patient vs institutional rights**: Patients contribute data, but often have limited visibility into how it is used, shared or monetized. Ethical frameworks emphasize autonomy, transparency and informed consent; a narrow legal notion of "ownership" alone cannot guarantee these. (See our White Paper, ***Circle Datasets As The Basis For Circle Health Coins***.)

Liability And Organizational Risk In The AI-Enabled Environment

Healthcare provider organizations, AI vendors and platform operators must address evolving liability risks associated with AI. Key considerations include:

- **Standard of care redefinition**: AI tools may shift or expand what is deemed the acceptable standard of care. If AI is integrated into decision workflows without

appropriate proof-points, institutions may face liability for failing to monitor or validate performance.

- **Vendor contracts and indemnification**: With unclear legal precedents, contracts must clarify responsibilities, licensing terms and indemnification agreements for misuse or unintended outcomes. A failure to allocate risk explicitly may leave institutions exposed.
- **Data quality and adverse outcomes**: If AI models feed off data lacking provenance, the downstream risk of model error rises. Institutions must ensure data pipelines are auditable and trackable — otherwise, liability may attach to data provider/model chain failures rather than just model design.
- **Regulatory breach and transparency**: Non-compliance with privacy laws (HIPAA, GDPR), lack of transparency, or unexplained model outcomes increase legal risk. Legal reviews state that the regulatory frameworks (FDA, FTC, EU AI Act) remain fragmented.

A Governance-Centric Asset View For High-Integrity Data

Given these complexities, the emerging consensus is that high-integrity medical data must be treated as a **governance-centric asset** — not merely a database. Key features of such an asset include:

- **Documented provenance**: Every data element should carry metadata about origin, capture protocol, versioning, transformation and access history. This enables auditability, liability assignment and trust.
- **Access controls and stewardship**: Data stewardship frameworks (rather than simply "ownership") define who may access, process or derive insights from datasets — including patient rights, institutional responsibilities and third-party obligations.
- **Interoperability and standards alignment**: Data mapped to standard terminologies (ICD, CPT, LOINC, SNOMED) and structured for interoperability strengthen accountability and reuse-rights.
- **Refresh and lifecycle governance**: High-integrity datasets have defined refresh cycles, version control, archival methods, and decommissioning protocols — all part of the accountability chain.
- **Liability alignment in contracts and processes**: Legal contracts and institutional processes must mirror data governance decisions — embedding audit logs, performance monitoring, and oversight into data-model-deployment pathways.

Implications For The Circles Approach

This governance-centric asset view underpins the value-proposition of its Circle Datasets. Because the dataset is captured under predefined Observational Protocols, it carries full provenance, is longitudinal, and is mapped to standard terminologies. It reduces ownership ambiguity, strengthens accountability (both data and model side), and enables downstream model-centric liabilities to be managed more proactively.

By framing dataset ownership and accountability within a structured governance posture, RegenMed addresses not only the technical *dataset integrity* problem but also the broader *institutional, legal and ethical* risk environment that many AI and other big data projects neglect.

FEDERATION, PRIVACY, AND LOCAL CONTROL

The Decentralization Imperative

Modern healthcare data landscapes are fragmented across hospitals, specialty practices, and regional networks. Historically, this fragmentation hindered research, yet the opposite extreme—centralized "data lakes"—has proven equally problematic: they invite privacy risk, loss of institutional control, governance complexity, and inability to establish data provenance.

Regulators and privacy experts now recognize that **federated or distributed models** offer a path between these poles: enabling computation across datasets while preserving local ownership and compliance. Both the **U.S. Office of the National Coordinator for Health Information Technology (ONC)** and the **European Medicines Agency (EMA)** have issued guidance emphasizing local control with standardized data interfaces.

What "Federated Capture" Means In The Circle Context

In conventional federated learning, models are trained locally at each data-holding institution, and only model parameters—not raw data—are shared centrally. RegenMed's approach extends this concept to **federated data capture**, meaning that:

- Each participating site maintains its **own operational dataset** under an Observational Protocol (OP) "fit-for-purpose" for that particular clinical site.
- Data remain **physically resident** at the site, stored within a secure institutional environment.

- The Circles Platform provides **standardized schemas, APIs, and validation routines** that allow data to be aggregated or queried *virtually*—without ever relinquishing ownership.

This structure preserves the trust dynamics of traditional research networks (e.g., PCORnet, OHDSI, Sentinel) while introducing modern interoperability and real-time auditability.

Advantages Of Federated Capture

Regulatory And Privacy Alignment

Federated capture respects both U.S. and European privacy regimes by maintaining local custodianship:

- In the United States, it aligns with **HIPAA's "minimum necessary" standard** and allows covered entities to maintain "designated record sets" within their own HIPAA environment.
- In the European Union, it satisfies **GDPR's data minimization and purpose-limitation principles**, as personal data never leave the controller's domain.
- For cross-border collaborations, metadata—not identifiers—can be exchanged under federated governance, meeting both jurisdictions' adequacy requirements.

Scalability And Data Integrity

Federation scales naturally: each new site activates a pre-defined OP, runs local validation, and joins the virtual network without massive data migrations. Because validation and provenance are enforced locally, Circle Datasets maintain consistent quality across sites. Federation also guards against single-point breaches and ensures continuity if one institution pauses participation.

Trust and Institutional Participation

Clinicians and institutions are more willing to contribute data when they retain control. By allowing each site to govern its dataset under its own IRB and/or other data-use policies, RegenMed reduces friction and accelerates multi-center adoption. This approach mirrors the trust-enhancing dynamics observed in **FDA Sentinel** and **OHDSI** networks.

Provenance And Governance

Each federated node records its own provenance ledger—metadata describing capture events, transformations, and exports. These ledgers can be queried centrally to verify

compliance and reproducibility without revealing patient-level data. Such an approach is aligned with the **W3C PROV** and **GA4GH Data Connect** standards.

Competitive Differentiation

While many RWE platforms are being designed to consolidate data into single silos, RegenMed's federated capture model scales ethically and defensibly across hundreds of sites and thousands of clinicians. It will transform data collection from a one-time extraction into a **continuous, auditable, multi-tenant collaboration**—a distinguishing capability for future health-AI governance.

Policy-As-Code: Governing Without Meetings

Federation can create complexity if governance depends on manual coordination. The Circles model will apply a **policy-as-code framework**, in which access rules, validation thresholds, and audit triggers are encoded into software. This reduces administrative overhead and ensures **consistent enforcement** of rules across sites. Examples include:

- Automated checks that data leaving a node are fully de-identified.
- Time-boxed queries with expiration policies.
- Logging of every inter-site request for audit purposes.

This approach follows best practices outlined in **NIST SP 1800-36: Implementing a Zero Trust Architecture for Healthcare** (and **GAIA-X Federated Trust Model 2024**.

Privacy With Performance

Federation within the Circle ecosystem achieves a rare balance—**data privacy, institutional autonomy, and analytic performance**. By retaining data locally while enabling standardized, auditable aggregation, RegenMed provides:

- Regulatory confidence (HIPAA, GDPR, IRB alignment).
- Scalability across diverse sites.
- Trust among clinicians and patients.
- Sustainable, privacy-preserving infrastructure for longitudinal AI development.

As healthcare AI faces increasing scrutiny over data provenance and consent, federated capture and local control will define the operational gold standard for trustworthy real-world evidence generation.

From Isolated Datasets To A Federated Network

Most real-world evidence (RWE) initiatives remain fragmented: registries operate independently, EHR exports lack standardization, and AI projects often use private datasets that cannot be reproduced or verified.

The **Circle Ecosystem** replaces this fragmentation with an **integrated, continuously learning evidence network**. Each **Circle Dataset** contributes to a federated, interoperable graph of validated, longitudinal clinical data linked by shared observational protocols.

This ecosystem is not merely a collection of databases — it is an evolving infrastructure for evidence generation. Every participating clinician, patient, and institution simultaneously contributes to and benefits from the collective learning process. This approach reflects the **Learning Health System (LHS)** model championed by the U.S. National Academy of Medicine, where continuous feedback from practice data informs research and quality improvement.

COLLABORATION: CLINICIANS, RESEARCHERS, AND REGULATORS

In the Circle Ecosystem, clinicians, academic researchers, and regulators engage through defined roles:

- **Clinicians** contribute high-quality data at the point of care through inCytes™ and receive outcome benchmarks in return.
- **Researchers** access de-identified, longitudinal data for hypothesis generation, validation studies, and AI benchmarking.
- **Regulators and policymakers** can reference Circle Datasets as "regulatory-grade" RWE for post-market surveillance, device safety monitoring, and comparative effectiveness studies.

This collaborative model echoes frameworks like the **FDA Sentinel Initiative** and the **EMA DARWIN EU network**, which both rely on federated, real-world data to inform regulatory decision-making. However, Circles extend this paradigm by integrating **real-time**

clinician participation and **patient-reported outcomes** — capabilities that conventional regulatory networks lack.

Real-World Evidence Supporting Clinical Innovation

Because Circle Datasets capture longitudinal, validated outcomes across diverse sites, they provide a continuously refreshed foundation for:

- **Comparative effectiveness research (CER):** Evaluating treatment pathways and outcomes using real-world cohorts consistent with **PCORI Methodology Standards**.
- **Post-market safety monitoring:** Feeding ongoing regulatory assessment under frameworks such as **FDA REALM** and **EMA Registry Guidelines**.
- **Precision-medicine discovery:** Linking phenotypic, procedural, and outcome data across populations to identify predictors of response.
- **Reimbursement and value-based care:** Supporting outcomes-based contracting by providing transparent, auditable evidence streams that payers and providers both trust.

These functions transform RWE from a static reporting mechanism into a **dynamic innovation substrate** — one that shortens the "lab-to-bedside" feedback cycle.

Benchmarks And Publication Networks

Every Circle Dataset contributes to **network-level benchmarking**, generating de-identified, cross-site performance dashboards. Participating clinicians can compare their results against aggregated network data — similar to but more rigorous than registry benchmarking. This design promotes data completeness and fosters professional collaboration while protecting privacy.

Results generated from Circle Datasets feed a **publication pipeline** supporting peer-reviewed manuscripts, abstracts, and conference presentations. Each publication is linked to dataset DOIs, protocol identifiers, and validation metrics. This continuous publication ecosystem ensures transparency, reproducibility, and recognition for contributors — accelerating the adoption of validated RWE practices.

Interoperability As A Network Effect

Interoperability is often treated as a technical requirement, but in the Circle Ecosystem it becomes a **source of acceleration**. Each new participating site strengthens the network by:

- Expanding the diversity of patient populations and clinical contexts.
- Increasing statistical power for rare outcomes.
- Enabling meta-analysis and benchmarking across geographies and subspecialties.

Because every dataset is FHIR-compatible and mapped to shared terminologies (ICD-10, CPT, LOINC, SNOMED CT), new nodes can connect with minimal friction. The resulting network achieves what traditional registries cannot: **scalable, harmonized data capture with provable provenance**. This aligns with interoperability goals defined in **ONC's United States Core Data for Interoperability (USCDI)** and international standards such as **HL7 FHIR R4**.

Future-Proofing AI with Continuously Refreshed Data

AI models degrade over time as clinical practice, patient demographics, and disease patterns evolve — a phenomenon known as **model drift**. Because Circle Datasets are longitudinal and refreshed in near real-time, they can be used for **continuous performance monitoring and re-training**.

- Drift detection algorithms can be run against updated Circle datasets to flag loss of calibration.
- Federated learning workflows can retrain models locally without exporting raw data, maintaining privacy and compliance.
- Regulators can rely on these refresh intervals as part of adaptive-AI oversight under **FDA GMLP** and **EU AI Act** provisions.

In effect, Circle Datasets provide the *ground-truth substrate* necessary for trustworthy model lifecycle management — ensuring that healthcare AI evolves safely alongside medicine itself.

Toward A Global Standard For Trustworthy Medical AI

The broader goal of the Circle Ecosystem is to establish a **de facto global standard** for medical AI data integrity:

- Structured and longitudinal rather than opportunistic.
- Verified and auditable rather than opaque.
- Federated and privacy-preserving rather than centralized.
- Continuously refreshed and peer-reviewed rather than static.

By demonstrating that this model works across institutions and borders, RegenMed helps define a new category of infrastructure: *real-world evidence as the ground truth for healthcare AI.*

GOVERNANCE AND REGULATORY MAPPING

Overview

Healthcare data governance now operates under converging but heterogeneous global frameworks. In the United States, the **Food and Drug Administration (FDA)**, the **Department of Health and Human Services (HHS)**, and the **Office for Civil Rights (OCR)** set expectations for privacy, real-world evidence (RWE), and trustworthy AI.

In Europe, the **European Medicines Agency (EMA)**, the **European Data Protection Board (EDPB)**, and the **European Commission** collectively define rules through the **General Data Protection Regulation (GDPR)** and the **Artificial Intelligence Act (AI Act 2024)**.

The **RegenMed Circles Platform** is designed to operate at the intersection of these frameworks. Its federated architecture and provenance-centric governance model are being built to satisfy the common objectives shared across jurisdictions: transparency, accountability, privacy, and auditability.

FDA Alignment: Real-World Evidence And Trustworthy AI

The FDA formally recognizes that real-world data (RWD) and RWE can support regulatory decision-making when they meet standards for reliability and relevance. Key alignment points include:

FDA Framework	Requirement	Circle Implementation
RWE Framework (2021)	Data reliability and traceability for regulatory submissions	Immutable provenance metadata, protocol-locked data dictionaries
21 CFR Part 11	Audit trails for electronic records	Automated event logging at each capture and export

FDA Framework	Requirement	Circle Implementation
Good Machine Learning Practice (GMLP 2021)	Continuous learning, dataset documentation, version control	Policy-as-code governance and data-card documentation
HIPAA Privacy Rule (45 CFR 164)	Minimum-necessary use and patient authorization	Federated capture—no raw data transfer; dynamic consent
HIPAA Security Rule (45 CFR 164.302-318)	Technical and administrative safeguards	AES-256 encryption, zero-trust architecture, audit controls
FDA AI/ML Action Plan (2021)	Lifecycle transparency for adaptive models	Continuous validation using refreshed Circle Datasets

Together, these alignments ensure that Circle Datasets can be used for **regulatory-grade submissions**, post-market surveillance, and algorithmic validation without violating U.S. privacy statutes.

EMA Alignment: Registry-Based Studies And DARWIN EU

The **European Medicines Agency (EMA)** has likewise prioritized federated RWE generation through its **Data Analysis and Real-World Interrogation Network (DARWIN EU)** and its **Guideline on Registry-Based Studies.** Core congruences include:

EMA Directive	Requirement	Circle Implementation
Guideline on Registry-Based Studies (2024)	Standardized data capture, auditability, traceability	Observational Protocols (OPs) + provenance logs
EMA Data Standardisation Strategy (2023)	Common data models and terminologies	FHIR R4, ICD-10, CPT, LOINC, SNOMED CT

EMA Directive	Requirement	Circle Implementation
DARWIN EU Model	Federated data network for regulatory use	Circle federated nodes with policy-as-code governance
Good Pharmacovigilance Practice (GVP Module VIII)	Timely and complete post-authorization safety evidence	Longitudinal follow-up and real-time refresh cadence

The Circle model thus mirrors EMA's own shift from centralized registries to **distributed, auditable RWE networks** that protect patient privacy while enabling cross-member-state analytics.

GDPR Compliance: Data Protection And Subject Rights

The **General Data Protection Regulation (2016/679)** establishes foundational principles for processing personal data in the EU. Circle Datasets will adhere to these through architectural and procedural controls:

- **Article 5 – Lawfulness, fairness, transparency** → Patient consent recorded and versioned within Benchmarc™; transparency dashboards available for participants.
- **Article 25 – Data protection by design and by default** → Privacy preserved through federated capture, de-identification, and role-based access.
- **Articles 15–22 – Data subject rights** → Mechanisms for access, rectification, and erasure requests via site-level controllers.
- **Article 30 – Records of processing activities** → Immutable provenance logs maintained at each site.
- **Articles 44–49 – Cross-border transfers** → Metadata exchange only; no raw data movement; Standard Contractual Clauses (SCCs) used when required.

Independent audits confirm compliance equivalence under **EDPB Guidelines 05/2021** on controller–processor roles.

The AI Act 2024: Risk-Based Governance

The **European AI Act (2024)** introduces a tiered regulatory system distinguishing *high-risk* medical AI from lower-risk applications. For high-risk systems, obligations include dataset documentation, risk management, human oversight, and post-market monitoring. Circle Datasets directly enable compliance by providing:

- **Dataset documentation** – Provenance and quality metadata ("data cards") for each model input.
- **Risk management** – Traceable lineage for model-training and validation datasets.
- **Human oversight** – Audit trails allowing clinician review of AI-assisted outputs.
- **Post-market monitoring** – Continuous validation through refreshed longitudinal data.

This architecture anticipates the AI Act's implementation phase, positioning RegenMed's infrastructure as a *ready-compliant environment* for European deployments.

Inter-Jurisdictional Harmonization

Federated design allows simultaneous adherence to U.S. and EU frameworks without duplicative storage or conflicting jurisdictional claims. Key bridging mechanisms include:

- **Controller-to-controller governance** → Each site acts as its own data controller while adhering to shared protocol governance.
- **Metadata federation** → Cross-border data discovery through GA4GH Data Connect without personal-data transfer.
- **Audit reciprocity** → Standard audit schema compatible with both FDA 21 CFR Part 11 and GDPR Article 30 records.
- **Regulatory reporting integration** → Datasets can be referenced in both FDA and EMA submissions under harmonized metadata identifiers (UUID + DOI).

This dual-compliance model ensures that multi-national studies can operate seamlessly under a single operational and ethical framework.

Oversight And Continuous Assurance

Governance does not end at compliance; it is maintained through **continuous assurance**:

- **Internal oversight** – Dedicated Data Protection Officers (DPOs) at each institution monitor adherence to Observational Protocols.
- **External audits** – Annual reviews by independent auditors under ISO 27001 and SOC 2 Type II standards.
- **Incident response** – Zero-trust network segmentation limits breach scope; incident protocols comply with GDPR Articles 33–34 and HHS Breach Notification Rule.
- **Governance review cycle** – Annual harmonization with evolving FDA and EMA guidelines, recorded in version-controlled governance registry.

This "living compliance" model replaces static certification with demonstrable, measurable governance performance.

Summary

Circle Datasets can meet -- and in several respects exceed -- the core governance expectations of the world's leading regulatory bodies. Their federated architecture ensures that data protection, auditability, and reproducibility are built into the system's design—not added post hoc. By aligning operationally with **FDA RWE and GMLP**, **EMA DARWIN EU**, **GDPR**, and the **AI Act 2024**, RegenMed has created a platform that can sustain global, regulator-grade trust in AI-enabled healthcare.

CIRCLES FEDERATED DATA CAPTURE: TECHNICAL OVERVIEW

Overview

Traditional real-world evidence (RWE) systems have relied on centralized data aggregation, where raw patient data are exported from local systems into a single repository for analysis. While convenient for analytics, this model creates intractable problems of **privacy, governance, and provenance**. Once data leave the originating institution, the original audit trail fragments, compliance responsibility blurs, and clear ownership/control is lost.

To address these issues, the **RegenMed Circles Platform** is implementing a **federated data-capture architecture**, in which all data remain physically resident under local institutional control, but share a common structure, schema, and validation logic.

This model enables **standardized, privacy-preserving real-world evidence generation** at scale—an approach increasingly endorsed by regulators and data-governance consortia such

as **FDA Sentinel**, **EMA DARWIN EU**, and the **Global Alliance for Genomics and Health (GA4GH).**

Core Architectural Principles

The design of RegenMed's federated architecture will rest on five interlocking principles:

- **Local custodianship** — Each participating site (hospital, clinic, or research center) retains physical custody of its data and remains the data controller under applicable law (HIPAA, GDPR).
- **Shared schema and validation logic** — All nodes implement the same Observational Protocols (OPs) and data-quality routines, guaranteeing structural consistency even though data are not centralized.
- **Federated query orchestration** — Aggregate analyses are performed through secure query orchestration: queries are sent to local nodes, computed locally, and only de-identified or aggregated results are returned.
- **Immutable provenance metadata** — Every capture event, transformation, or query execution writes a metadata record compliant with the **W3C PROV** and **GA4GH Data Connect** standards.
- **Policy-as-code enforcement** — Governance rules (access control, de-identification, expiration, audit triggers) are encoded directly into the platform, ensuring uniform compliance without manual intervention.

This architecture achieves **logical centralization with physical decentralization**—an essential property for privacy-preserving analytics in healthcare.

Data-Flow Summary

A typical Circle Dataset federation event will proceed through the following stages:

- **Capture** – Clinicians and patients enter data into inCytes™ and Benchmarc™ interfaces following an approved OP.
- **Local validation** – Automated schema checks verify completeness, type conformity, and temporal plausibility before data are committed to the local node.
- **Provenance stamping** – Each record receives a unique event identifier (UUID + timestamp) linking it to the OP version, site, and user role.
- **Federated registration** – Metadata describing the dataset (not the data themselves) are published to a central registry enabling discoverability.

- **Distributed query execution** – When authorized studies run, the federated query engine dispatches analytic scripts (e.g., SQL, Python) to each node. Local execution yields aggregated statistics (counts, means, risk ratios) returned via encrypted channel.
- **Aggregation and reporting** – The Circles Platform consolidates the anonymized aggregates, ensuring that no patient-level data ever cross institutional boundaries.

This process will conform to emerging **federated-analytics frameworks** used by OHDSI, PCORnet, and Sentinel, while adding real-time validation and provenance capture at every stage.

Security and Privacy Controls

Security controls operate at both the node and orchestration layers:

- **Data isolation:** Each node stores patient data in its own secure environment; no raw data are replicated externally.
- **Encryption:** AES-256 encryption at rest and TLS 1.3 in transit.
- **Access management:** OAuth 2.0 and federated identity via OpenID Connect; role-based permissions limit query scope.
- **Differential privacy:** Optional noise addition for small-cell suppression in aggregate results.
- **Zero-trust networking:** All cross-node communication follows the **NIST SP 1800-36 Zero Trust Architecture**.
- **Continuous monitoring:** All API calls and data exports logged to an immutable audit ledger (block-hash chained, non-financial).

Collectively, these controls will satisfy -- and in several respects exceed -- the technical safeguards required under **HIPAA Security Rule** (45 CFR §164.302–318) and the **GDPR Articles 5–32** on data protection.

Governance And Consent Framework

Every federated node can operate under a local IRB or MEC approval referencing RegenMed's master governance framework. Key governance elements include:

- **Dynamic consent:** Patients opt in via Benchmarc™ and can withdraw or limit use at any time.

- **Data-use agreements (DUAs):** Standardized templates specify rights, obligations, and audit provisions across all sites.
- **Cross-border harmonization:** Where data span jurisdictions, federated participation occurs under controller-to-controller clauses compatible with GDPR Article 46 and HIPAA's Business Associate Agreements.
- **Audit frequency:** Quarterly internal audits and annual third-party validation ensure continuing compliance.

This governance structure will reflect best practices articulated in **OECD Data Governance for Health 2021**.

Technical Performance and Scalability

Benchmark testing within pilot networks will seek to demonstrate that federated query execution across ≥10 institutions introduce minimal latency (< 10 %) compared with centralized execution while eliminating cross-site data transfers.

Data-quality metrics (completeness, conformance, plausibility, timeliness) will be monitored automatically using a derivative of the **OHDSI Data Quality Dashboard.**

Because all sites share identical OPs and schemas, adding new nodes will require configuration, not redesign—enabling exponential scaling with linear administrative overhead.

Regulatory Alignment

Federated capture directly supports compliance with major regulatory expectations for trustworthy AI and RWE:

Framework	Relevant Principle	Circle Alignment
FDA Real-World Evidence Framework (2021)	Data reliability, transparency, traceability	Local validation + provenance metadata
EU AI Act (2024)	Risk-based AI oversight, dataset documentation	Immutable audit trail, data cards

Framework	Relevant Principle	Circle Alignment
NIST AI RMF (2023)	Governance, accountability, privacy	Policy-as-code + zero-trust design
EMA DARWIN EU Network (2023)	Federated data access for regulatory use	Node-level custodianship
GA4GH Data Connect (2023)	Cross-network discoverability	Metadata registry + query federation

This alignment will ensure that Circle Datasets can serve not only as research resources but also as **regulatory-grade evidence sources** for model validation and approval submissions.

Summary

Federated data capture represents the technological and ethical foundation of the Circle model. It will unite privacy preservation, institutional autonomy, and analytic scalability within a single architecture that is simultaneously **auditable**, **standards-based**, and **regulatorily aligned**.

By embedding validation and provenance at the point of origin rather than at the point of aggregation, RegenMed will create a living network of trustworthy data—an infrastructure capable of supporting the next generation of transparent, reproducible, and continuously learning medical AI.

REFERENCES

https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf

https://assets.ctfassets.net/opszt4tga0mx/14I8maINZnuuNMvzcnCU2l/5521bcd412ab71f205681d6e2e1bf641/TEFCA-v1-Overview-Final.pdf

https://www.nejm.org/doi/full/10.1056/NEJMp1809643

https://hai.stanford.edu/assets/files/2024-02/Liability-Risk-Healthcare-AI.pdf

https://www.physicianleaders.org/articles/doi/10.55834/plj.5448947832

https://datafoundation.org/news/reports/697/697-Data-Provenance-in-AI

https://www.bmj.com/content/369/bmj.m1328

https://gaia-x.eu/

https://www.fda.gov/safety/fdas-sentinel-initiative

https://gdpr-info.eu/art-5-gdpr/

https://www.tandfonline.com/doi/pdf/10.1080/20502877.2025.2482282

https://www.oecd.org/en/topics/health.html

https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

https://www.mdpi.com/1424-8220/23/14/6495

https://academic.oup.com/jamia/advance-article/doi/10.1093/jamia/ocaf133/8264332

https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-022-00848-y

https://www.mcpdigitalhealth.org/article/S2949-7612%2824%2900083-X/fulltext

https://www.fda.gov/media/145022/download

https://www.ga4gh.org/

https://csrc.nist.gov/publications/detail/sp/1800-36/final

https://oecd.ai/en/ai-principles

https://www.nejm.org/doi/full/10.1056/NEJMp1809643

https://nam.edu/resources/publications/

https://www.talkinghealthlaw.com/post/navigating-ai-liability-in-healthcare-key-considerations-for-health-system-leaders

https://www.pcori.org/research-related-projects/about-our-research/research-methodology

https://hl7.org/fhir/R4/

https://www.pcori.org/research-related-projects/about-our-research/research-methodology

https://ohdsi.github.io/DataQualityDashboard/index.html

https://www.ga4gh.org/

https://www.nejm.org/doi/full/10.1056/NEJMp1809643