Real World Evidence Community Forum

Real World Evidence Meets AI and Privacy

phuse.global

Hosts & Presenters

connect·share·advance

Ashwin Rai
(Thermo Fisher Scientific)

Rethinking Real-World Evidence in the Age of GenAI

Rafik Ghazariyan (BioBrain PRO LLC)
Tuhin James Paul (ISF College of Pharmacy)

Unlocking Real-World Data:
Compliance, De-Identification, Trust & Privacy-Preserving Innovation

Hosts:

- Elizabeth Merrall, *Johnson & Johnson;* Mary Anne Rutkowski, *Merck;* Nicky Newton, *PHUSE*

Disclaimer: The contents, opinions and view expressed in this presentation are of authors only and not of the organisations they represent.

Content

connect·share·advance

- Introduction – 5’
- Ashwin – 20’
- Q&A – 5’
- Rafik and Tuhin – 20’
- Q&A – 10’

Please enter any questions you have for the presenters in the Q+A box within Zoom

RWE Working Group

connect·share·advance

https://phuse.global/Working_Groups

The Real-World Evidence Working Group aims to support, address and answer pertinent questions around real-world evidence. The Working Group is dedicated to sharing across the PHUSE Community (through Community Forums) and aligning on the best industry practices. Some of the questions we intend to address are:

- What are the requirements, technologies and processes/best practices needed to use RWD as a source for data analysis?
- incorporate real-world data into clinical trials?
- to support the use of real-world evidence as part of regulatory submissions?

Several Projects

- Applying Advanced Data Privacy Methods to Real World Data (RWD)
- Using OMOP and Other Real World Data Standards to Support Regulatory Submissions
- Estimands For RWD/RWE
- Quality & Reusability of RWD
- RWD Guideline for Programming & Analysis Processes
- Submitting RWD
- Missing Data Imputation in RWD Exploration of Multiple Techniques on Open-Source Data

Blog Articles & White Papers
Community forums
Conference activities

US Connect 2027 – Call for Papers now open

EU Connect 2026 – Register now

If you would like to get involved, please email us:
workinggroups@phuse.global

Rethinking Real-World Evidence in the Age of GenAI

Ashwin Rai

Director of Data Science & Analytics, ThermoFisher Scientific

July 2026

The world leader in serving science

Before we begin: What You'll Take Away from This Talk

Frame where GenAI can improve RWE workflows today

Accelerate evidence review, data-source assessment, and early study concept development.

Focus on review-ready evidence inputs

Convert fragmented sources into structured summaries, assumptions, and questions for experts.

Keep scientific judgment visible

Use GenAI to reduce manual friction, not to replace validation, governance, or decisions.

Key message

GenAI is not a replacement for scientific judgment. It is a way to reduce manual friction, structure evidence, and focus expert time where it matters most.

Disclaimer: The contents, opinions and views expressed in this presentation are those of the author only and not necessarily of the organizations represented.

Agenda

1 **RWE project landscape**

Where GenAI fits across evidence planning, execution, and communication

2 **Opportunity map**

Uses across the RWE lifecycle, with a focus on productivity and viability

3 **High-value use cases**

Evidence synthesis, study design, data source fit, unstructured data, analysis

4 **GenAI-assisted analysis**

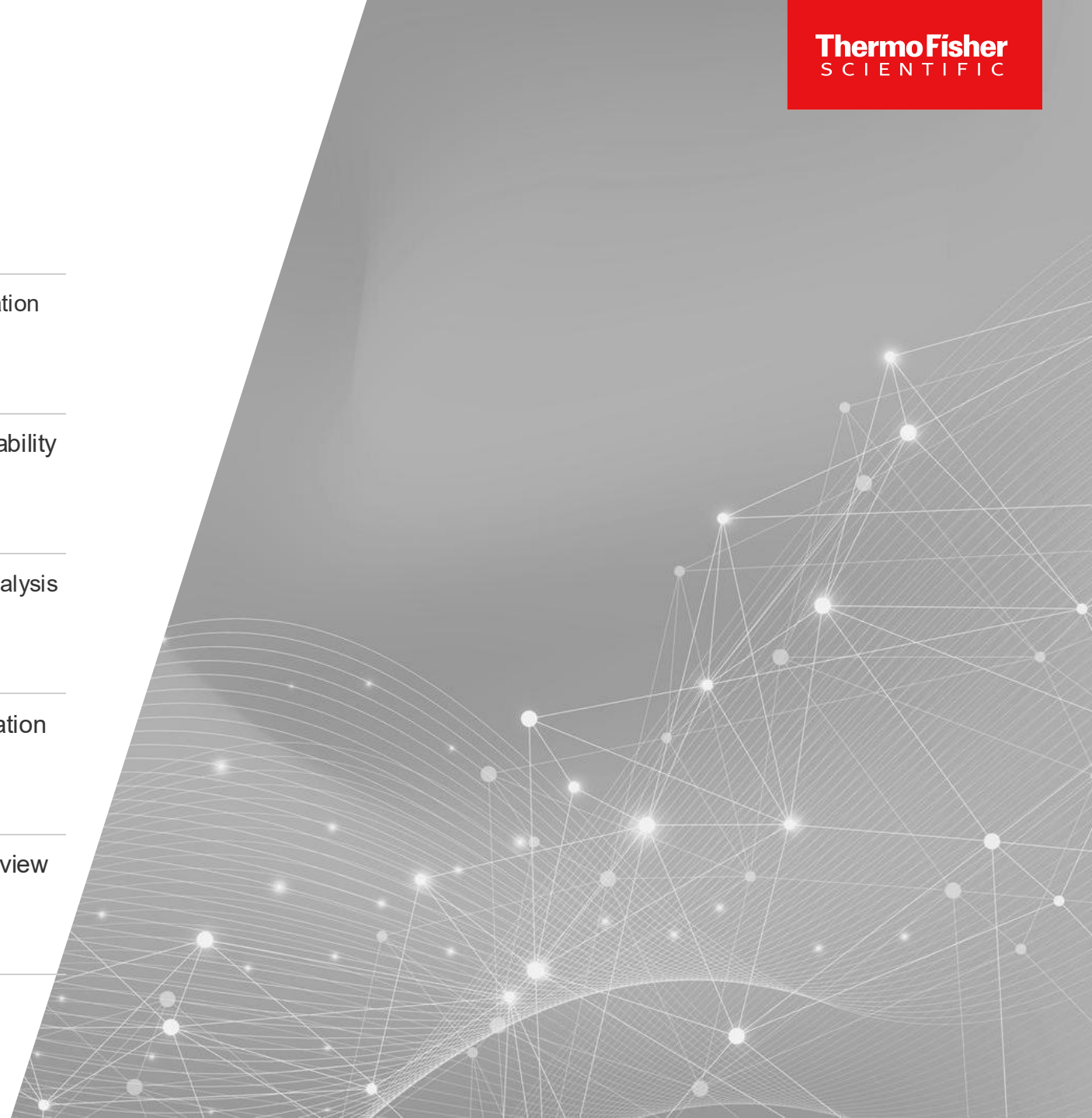
Code, QC, TFL shells, exploratory summaries — with expert validation

5 **Responsible adoption**

Scientific validity, auditability, privacy, bias, governance, human review

6 **Roadmap and takeaways**

How to pilot, validate, and scale trusted RWE workflows



Why RWE Needs a Rethink Now

RWE teams are asked to transform larger, messier, and more fragmented information into reliable evidence — faster

Expanding RWD sources

EHR, claims, registries, patient-generated data, digital health, and unstructured narratives

Fragmented information

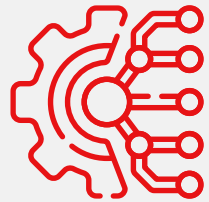
Evidence, protocols, data dictionaries, codebooks, prior studies, and notes live in different places

Manual interpretation load

Teams spend significant effort reading, extracting, checking, and synthesizing

Need for traceable decisions

Every output must remain scientifically valid, auditable, and fit for purpose

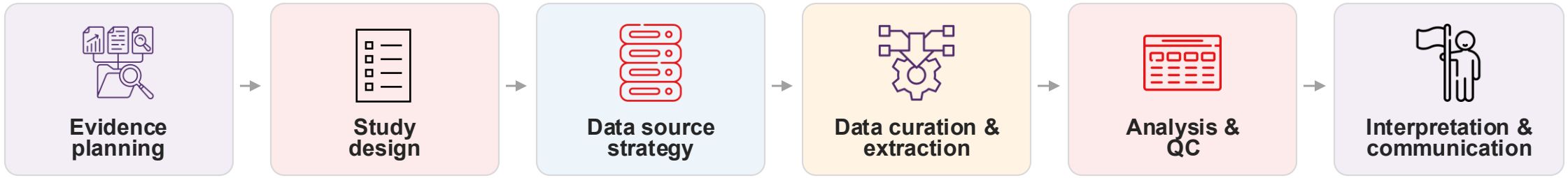


GenAI opportunity

Compress the repetitive work so experts can spend more time on judgment, strategy, and evidence quality.

RWE project landscape: where GenAI can assist

GenAI is relevant across RWE projects, but the practical value today is highest in expert-assisted workflows.



launch planning | post-market evidence | medical strategy | HEOR / access evidence | regulatory questions

What GenAI can do well

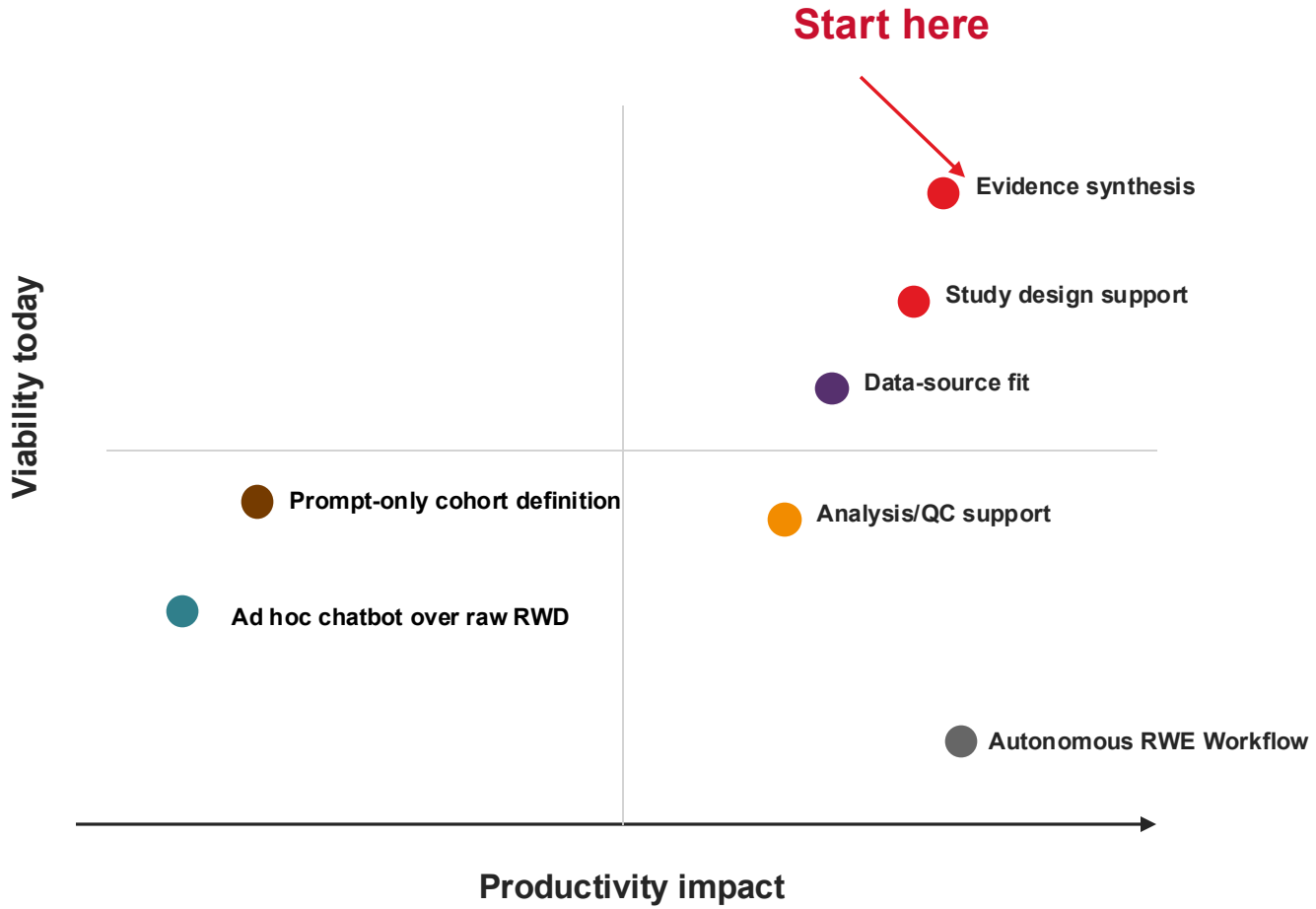
Read, extract, classify, compare, summarize, draft, and structure information for expert review.

What experts must still own

Scientific validity, protocol/SAP decisions, causal inference, interpretation, and accountability.

Prioritize use cases by productivity and viability

For RWE projects, the best starting points are high-manual-effort tasks where outputs can be reviewed, traced, and validated.



High-value RWE starting points

- Evidence synthesis
- Study concept support
- Data-source fit assessment
- Unstructured extraction
- Analysis code / QC support
- Draft reports & summaries

High-productivity, viable use cases for RWE teams

Evidence landscape review

Find, compare, and summarize prior evidence for expert review.

Study concept & protocol support

Draft research questions, cohorts, endpoints, and feasibility prompts.

Data-source fit assessment

Compare EHR, claims, registries, narratives, and patient-generated data.

Unstructured-to-structured inputs

Extract variables from notes, reports, clinical narratives, and patient text.

Analysis productivity support

Draft code, TFL shells, QC checks, missingness summaries, and output narratives.

Documentation & communication

Create traceable first drafts, study summaries, and decision-ready materials.

Decision rule: use GenAI to prepare expert-ready inputs — not to replace scientific judgment or accountability.

Use case 1: Evidence landscape & study concept acceleration

RWE study teams often spend a large share of early project time on search, synthesis, framing, and feasibility thinking.

Inputs

PICO, indication, prior studies, clinical context, target decision

GenAI assistance

Synthesize evidence, draft research questions, highlight feasibility considerations

Structured concept

Draft study objectives, population, endpoints, data needs, assumptions

Expert review

Clinical, epidemiology, biostats, HEOR, regulatory alignment

What improves

Faster landscape review, cleaner framing, fewer blank-page starts, better reuse of prior evidence.

Guardrails

Ground outputs in source documents, keep citations/traceability, and require SME review.

RWE value

Moves teams faster from question framing to a feasible, fit-for-purpose study concept.

Use case 2: Data source fit and feasibility assessment

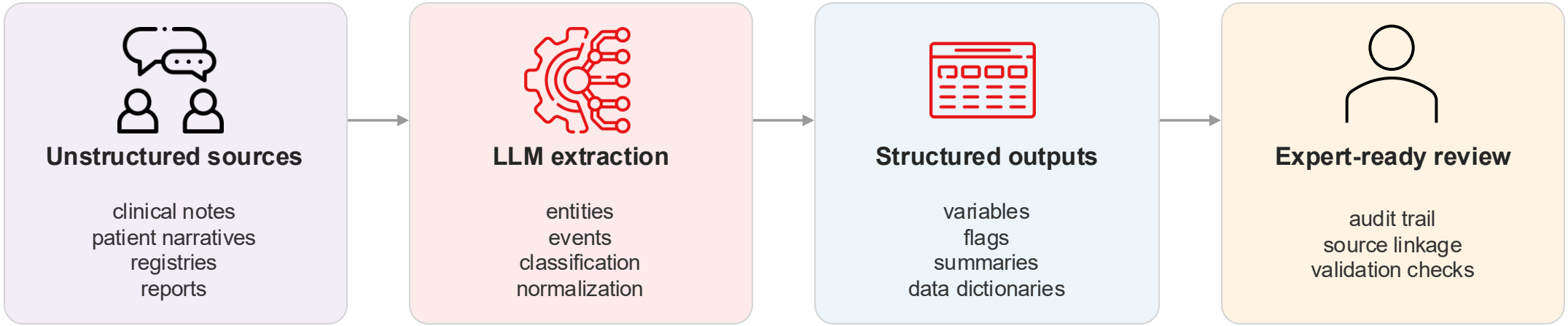
GenAI can help organize the evidence needed for data-source decisions, while experts judge whether the source is scientifically fit for the study question.

RWD source	What GenAI can prepare	Human decision criteria
 **EHR / clinical notes**	Availability of encounters, labs, procedures, narratives, and timing	Completeness, clinical specificity, missingness, and capture of outcomes
 **Claims**	Coverage windows, utilization, treatments, costs, and enrollment logic	Coding validity, confounding, lag, and lack of clinical detail
 **Registries / devices**	Relevant variables, follow-up, patient populations, and endpoint definitions	Selection bias, standardization, and endpoint comparability
 **Patient-generated / narratives**	Patient language, symptoms, treatment burden, and lived experience signals	Representativeness, privacy, consent, and validation strategy

Output: a structured, transparent data-source recommendation package that can be reviewed by RWE, clinical, regulatory, and HEOR stakeholders.

Use case 3: Turning unstructured information into structured RWE inputs

LLMs are well suited to organize fragmented text and prepare structured variables for expert review and downstream analysis.



Potential value

Less manual abstraction and faster creation of structured evidence inputs.

Validation need

Sampling, adjudication, inter-reviewer checks, and error characterization.

RWE use

Prepare structured variables for feasibility, cohorts, outcomes, and patient experience.

Use case 4: GenAI-assisted analysis — productivity, not autopilot

GenAI can accelerate analytic preparation and review, but RWE findings must remain protocol-driven, statistically validated, auditable, and expert-reviewed.

Draft analysis code
SAS / R / Python scaffolding and reusable snippets

Cohort attrition summaries
Logic checks, flow descriptions, and study population tracing

Missingness & data quality exploration
Flags, patterns, and first-pass QC narratives

TFL shells & output summaries
Table, figure, listing shells and interpretive first drafts

Sensitivity / subgroup suggestions
Hypotheses to consider — not protocol changes without review

Guardrail principle

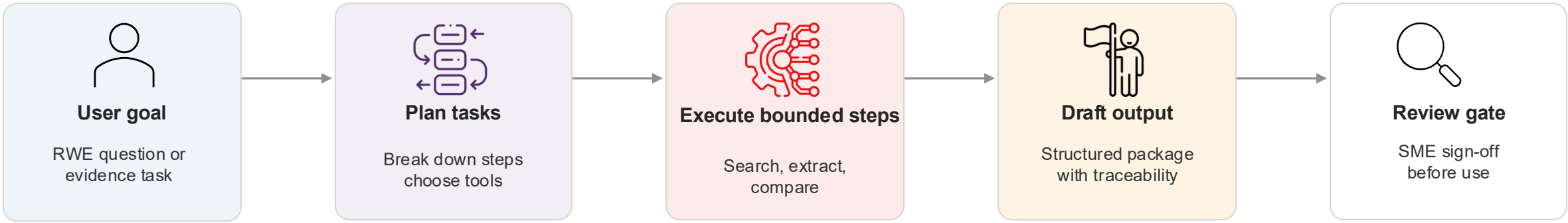
GenAI can assist the analyst, but it does not become the analyst of record.

- Pre-specified protocol and SAP remain authoritative
- Generated code must be peer-reviewed and tested
- All outputs reconcile to source data and logs
- No unsupported causal or regulatory claims
- Interpretation owned by epidemiology, statistics, and clinical SMEs

Best fit: repeatable programming, QC, documentation, and exploratory synthesis tasks that can be checked against source outputs.

Use case 5: Agentic workflow support for RWE projects

Agentic AI can coordinate multi-step RWE work — search, extract, compare, draft, and route — while maintaining checkpoints for expert review.



Good fit

Routine evidence and documentation tasks with clear inputs, schema, and review criteria.

Design requirement

Use approved sources, bounded tools, version control, and audit logs.

Review requirement

Human checkpoints before any analysis, publication, or decision-facing output.

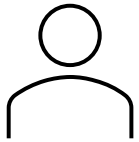
What GenAI should not own in RWE projects

The more the task affects validity, interpretation, or regulatory confidence, the stronger the human control must be.

AI-assisted

Draft, extract, structure, compare, summarize, and flag issues.

- Retrieve approved sources
- Create structured extraction tables
- Compare variable definitions
- Route low-confidence outputs to experts



Human-led decisions

- Final research question, estimand, and causal framework
- Protocol Assumptions, SAP, endpoint definitions, and cohort logic
- Fit-for-purpose assessment and validation thresholds
- Final statistical conclusions and medical interpretation
- Regulatory-facing claims and risk-benefit implications

Position GenAI as an expert-assist layer — not a replacement for epidemiology, biostatistics, clinical judgment, or governance.

Responsible adoption: controls that make GenAI usable for RWE

ThermoFisher SCIENTIFIC

Scientific validity

Define intended use, performance targets, review process, and validation sample.

Transparency & auditability

Keep prompts, sources, versions, outputs, decisions, and reviewer approvals.

Privacy & security

Use approved environments, de-identification, access controls, and PHI safeguards.

Bias & representativeness

Assess source bias, model error patterns, subgroup performance, and data gaps.

Governance & change control

Define owners, approval gates, monitoring, escalation, and retirement criteria.

Human-in-the-loop review

Require expert adjudication before analytic use, publication, or decision support.

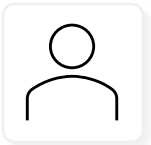
Principle: higher-impact outputs require stronger validation, traceability, and expert accountability.

RWE GenAI adoption is a cross-functional operating model, not a tool rollout

The work succeeds when teams agree on roles, review checkpoints, and shared artifacts before the pilot starts.

Data science

- Model design and prompt testing
- Monitoring and evaluation
- Workflow automation



Clinical / HEOR / Medical

- Study question and design
- Outcome relevance
- Scientific interpretation

Data operations

- Source inventory
- Data access and refresh
- Metadata management

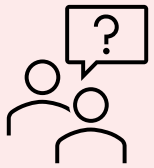
Regulatory / privacy

- Intended-use guardrails
- Data boundaries
- Compliance review

Shared artifacts: use-case brief, source inventory, validation plan, reviewer checklist, audit log, and change-control process.

Practical implementation roadmap: pilot, validate, scale

1



Select RWE use case

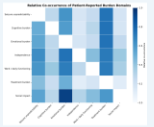
Pick a reviewable, high-manual-effort task such as evidence synthesis, feasibility, or QC narration.

2

Build bounded workflow

Define approved sources, prompts, output schema, human checkpoints, and audit logging.

3



Validate and quantify

Measure accuracy, time saved, reviewer burden, error modes, and reproducibility.

4

Scale with governance

Add SOPs, training, monitoring, change control, and reusable assets for similar RWE projects.

Success metric: not “AI generated a result” — but “RWE teams produced higher-quality, traceable inputs faster with controlled risk.”

Key takeaways

1

Think lifecycle, focus selectively

GenAI can support many RWE steps, but start where productivity and viability are both high.

2

Use GenAI as an expert-assist layer

The strongest value is structuring, summarizing, extracting, drafting, and checking information.

3

Include analysis support carefully

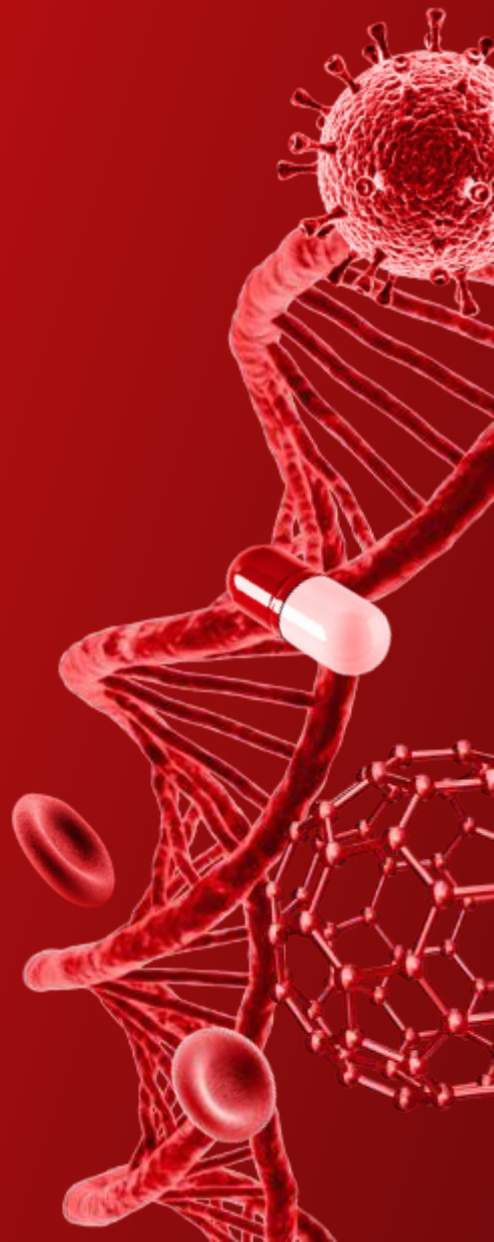
Use GenAI for code scaffolding, QC, missingness, TFL shells, and output narratives — not final findings.

4

Build trust before scale

Human review, validation, audit trails, privacy controls, and governance determine whether adoption is credible.

Thank you



Q&A

Applying Advanced Data Privacy Methods to RWD Project

connect·share·advance

- **Our focus areas:**

✓ Real-World Data (RWD): Structured and Unstructured/EHR and Hospital RWD

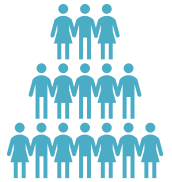
✓ Technology: existing traditional data privacy methods vs more advance methods

✓ AI-Driven Processes: usage of AI/ML methods

- **Working Streams:**

✓ WS1: How to incorporate the advanced data privacy methods in the industry with the current technologies, and possible usage of ML.

✓ WS2:Review of the existing data regulations –too broad and lack of specific process flow.



Join the RWE Working Group via PHUSE Advance Hub.

Volunteer for subprojects (e.g., privacy, estimands, OMOP standards).

Attend webinars/community forums; contribute case studies and reviews.

PHUSE Applying Advanced Data Privacy Methods to Real World Data Project

connect·share·advance

phuse.global

connect·share·advance

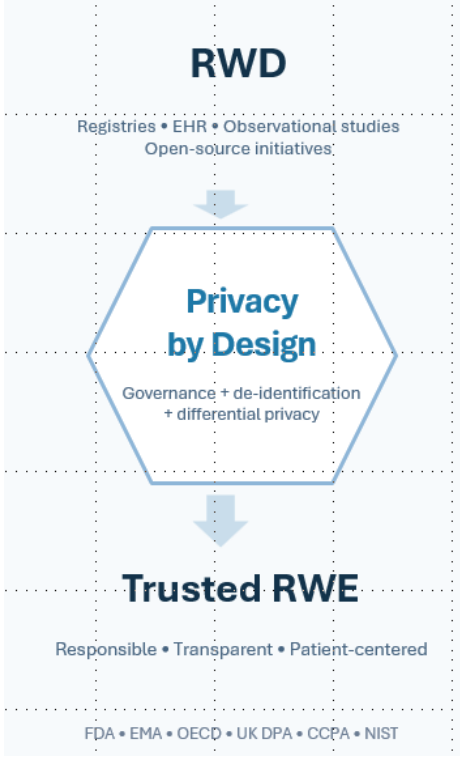
Unlocking Real-World Data: Compliance, De-Identification, Trust, and Privacy-Preserving Innovation

Speakers and Contributors:
Rafik Ghazariyan, BioBrain PRO LLC
Tuhin James Paul, ISF College of Pharmacy
Aditee Dani, Arcus Biosciences
Elena Valkonava, Independent Consultant

Turning Real-World Data into Trusted Real-World Evidence

Turning RWD into RWE

Real-world data can accelerate research, but trust must be built into every step — from access and governance to analysis and evidence generation.



1. **Privacy & governance essentials** — DUAs, HIPAA de-identification, GDPR principles, DPIAs, auditability, and 21 CFR Part 11
2. **Practical Alzheimer's disease case study** — Differential privacy applied to protect patient data while enabling research
3. **Future of trusted RWE** — Scalable RWD infrastructure, multi-site collaboration, trusted research environments, and privacy-enhancing technologies

Roadmap

Roadmap

1. Why Trust Comes First — From compliance language to operational reality

2. Accessing Real Hospital Data — What actually happens before data are in hand

3. Data Use Agreements and Governance — Rules, responsibilities, and institutional safeguards

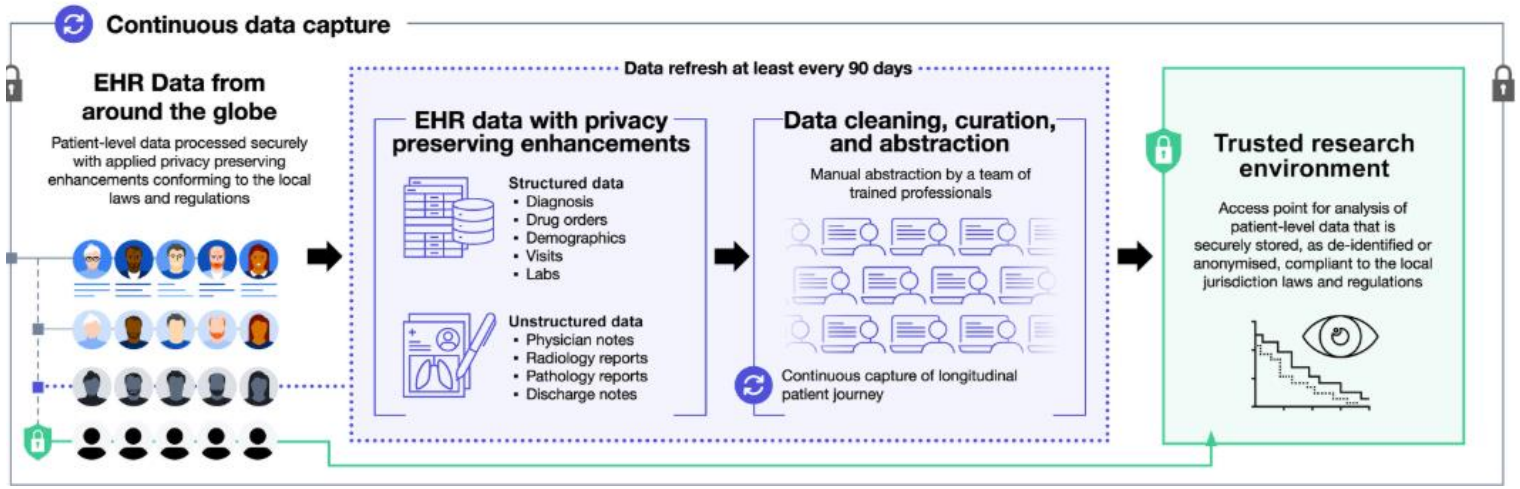
4. HIPAA Framework and Compliance Controls — Privacy, security, and governance requirements for PHI

5. De-identification in Practice — Turning sensitive records into research-ready data

6. Q&A and Closing Discussion — Open questions for the RWD/RWE community

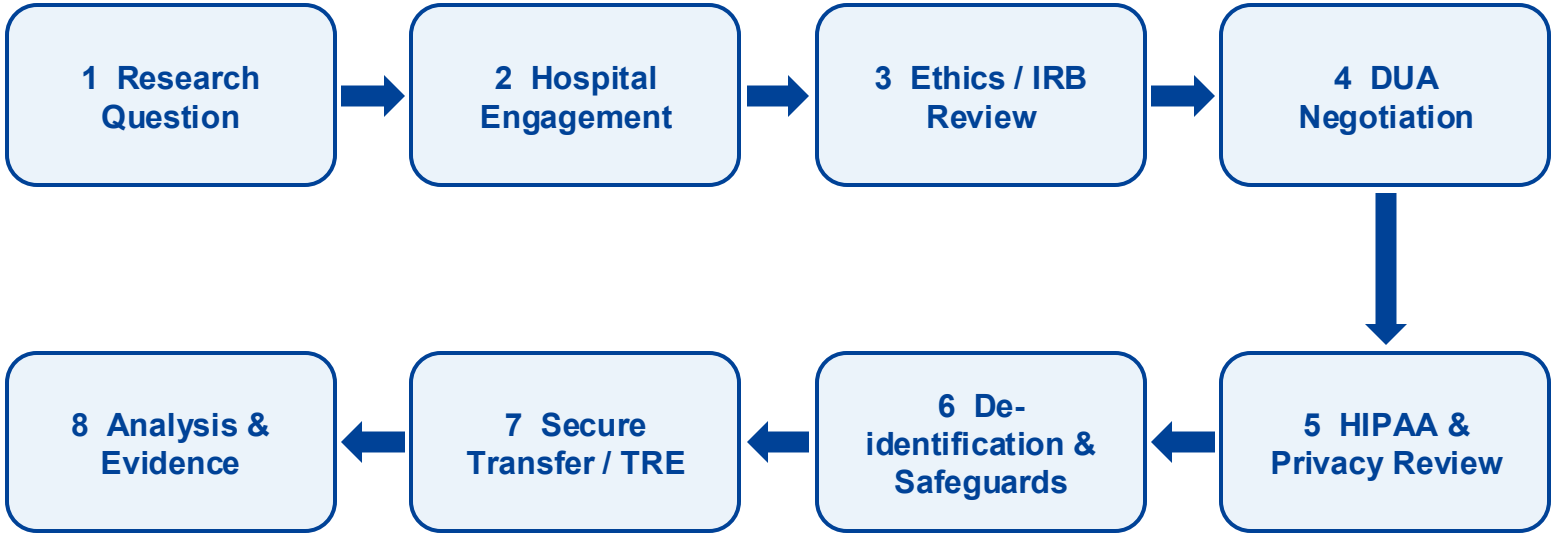
What It Actually Took to Access Real Hospital Data?

Accessing Hospital Data



Source:
Adamson, B., Horne, E., Xu, C., et al. *Characterisation of oncology EHR-derived real-world data in the UK, Germany, and Japan.* **ESMO Real-World Data and Digital Oncology**, 2025.

Process Followed: From Data Request to Trusted RWE



Key message: Data access is not one approval - it is a controlled workflow of trust-building, safeguards, and accountable use.

The Journey - From Cold Email to Actual Data

Each phase came with its own surprises. The hospitals didn't reply to my first email, so I followed up with a second explaining how I'd de-identify their data and share the final results with them - that got their attention. But government ethics approval wasn't enough; they wanted their own committee reviews too. More documentation, more discussions, but eventually - data in hand.

Key takeaway: Persistence, transparency, and a willingness to give back (analysis help, shared results) made the difference.

DATA USE AGREEMENT

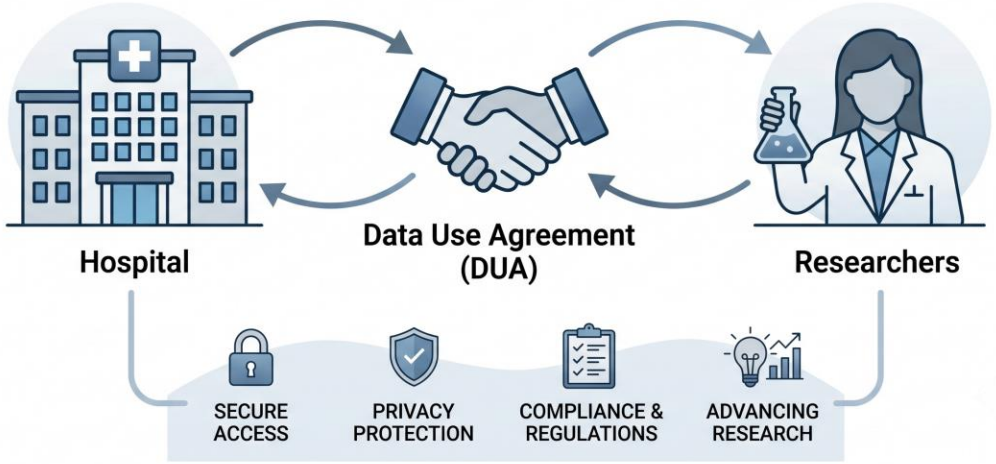
Data Use Agreements: The Contract Behind Data Access

phuse.global

DATA USE AGREEMENT (DUA)

DUA

- A **DUA** is a legally binding agreement governing data sharing. It defines what data can be shared and why. It protects both the data provider and the data recipient.



Note: In practice, the DUA is often the first major milestone in obtaining real-world data (RWD).

- "No data sharing begins without clearly defined rules."

References: Chen JJ. Multicenter observational studies: understanding the basics of data sharing and data user agreements. Journal of Neuro-Ophthalmology. 2024 Mar 1;44(1):1-4.

Core Components of a DUA

DUA

Purpose	Access	Protection
Research objectives	Who can access data	Confidentiality obligations
Scope of data use	User roles and responsibilities	Security requirements
Duration of use	Permitted disclosures	Regulatory compliance

- DUAs answer three critical questions: Why are we using the data? Who can use it? How will it be protected?"
- A DUA establishes the boundaries for responsible data use.

References: Girman CJ, Ritchey ME, Re III VL. Real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. Pharmacoepidemiology and drug safety. 2022 May 3;31(7):717.

Where the DUA Fits in RWD Access

DUA

- The DUA bridges trust and operational execution.

- Negotiating a DUA can take weeks or months, but it transforms institutional trust into actionable collaboration.

References: Girman CJ, Ritchey ME, Re III VL. Real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. Pharmacoepidemiology and drug safety. 2022 May 3;31(7):717.

HIPAA Framework and Compliance

phuse.global

HIPAA Framework : The Foundation for RWD Access

HIPAA

- HIPAA's Role in Real-World Data Research
  1. Establishes national standards for protecting patient health information.
  2. Enables healthcare organizations to share data while safeguarding patient privacy.
  3. HIPAA establishes the legal framework governing the use and disclosure of protected health information for research.

Why HIPAA matters for RWD/RWE

1. Hospitals cannot simply provide patient data upon request.
2. Every data-sharing activity must comply with HIPAA requirements.
3. Compliance is essential for maintaining patient trust and institutional participation.

References: Pillai CS, Panico EE, Zou KH, Filipowska E. Patient data privacy, protected health information, and ethics of real-world evidence. InReal-world evidence in a patient-centric digital era 2022 Aug 3 (pp. 51-72). Chapman and Hall/CRC.

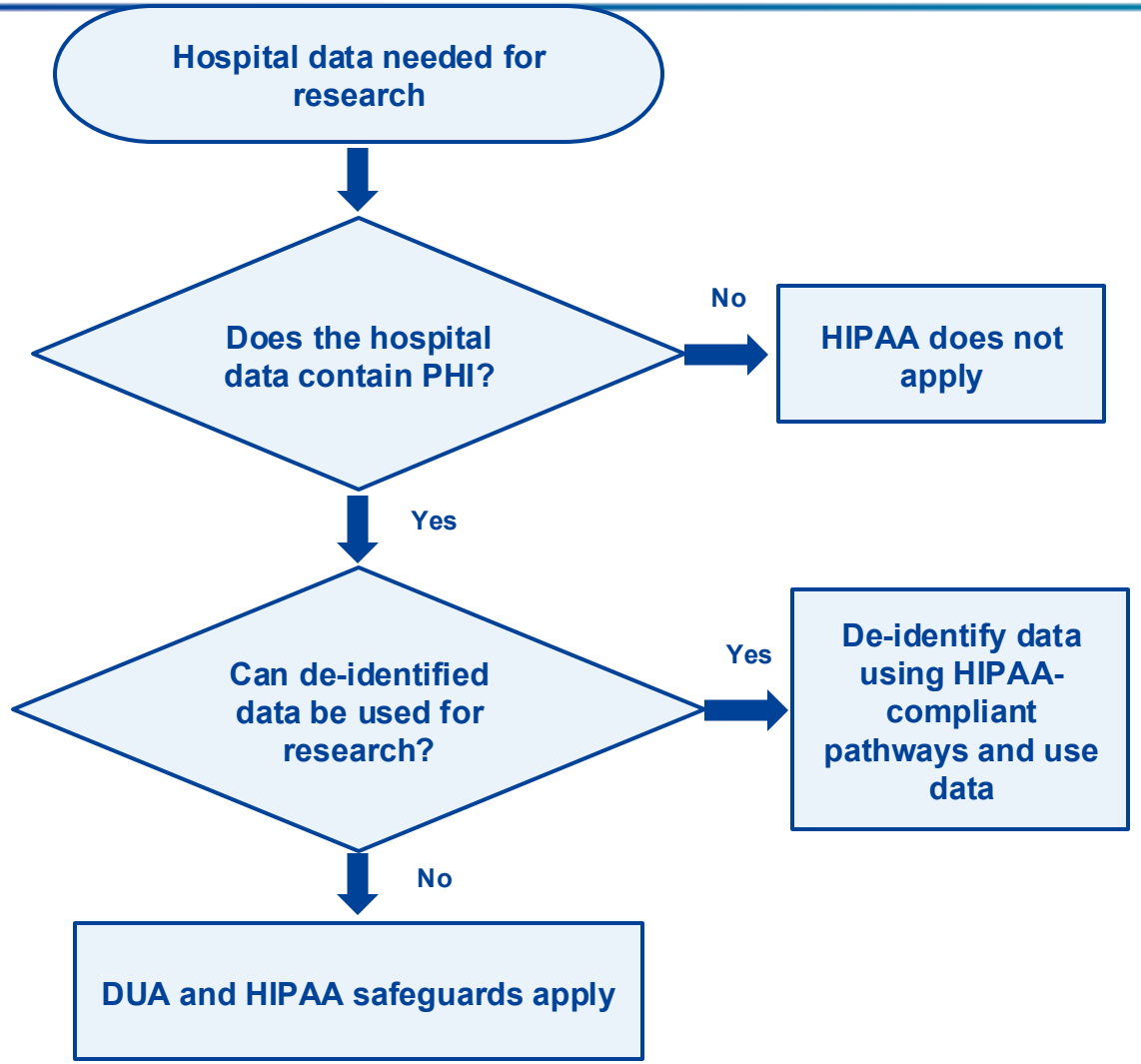
HIPAA Components

HIPAA

Rule	Primary Focus
Privacy Rule	Governs when PHI can be used or disclosed
Security Rule	Protects electronic PHI (ePHI) through technical and administrative safeguards
Breach Notification Rule	Requires reporting of unauthorized disclosures
Enforcement Rule	Defines penalties and regulatory oversight

References: Barbaria S, Jemai A, Ceylan Hİ, Muntean RI, Dergaa I, Boussi Rahmouni H. Advancing compliance with HIPAA and GDPR in healthcare: a blockchain-based strategy for secure data exchange in clinical research involving private health information. InHealthcare 2025 Oct 15 (Vol. 13, No. 20, p. 2594). MDPI.

HIPAA decision pathway for RWE



HIPAA

References: Chen JJ. Multicenter observational studies: understanding the basics of data sharing and data user agreements. Journal of Neuro-Ophthalmology. 2024 Mar 1;44(1):1-4.

HIPAA Compliance Controls in RWD Projects

HIPAA

- Privacy Controls
  - Minimum Necessary Principle
  - Role-based access
  - Data access approvals
  - Purpose limitation
- Security Controls
  - Encryption at rest and in transit
  - Secure file transfer mechanisms
  - Audit logging and monitoring
  - User authentication and access management
- Governance Controls
  - Data Use Agreements
  - Institutional Review Board (IRB) oversight
  - Vendor and researcher accountability
  - Periodic compliance reviews

References: Kahn SD, Terry SF. Who owns (or controls) health data?. Scientific Data. 2024 Feb 1;11(1):156.

HIPAA

HIPAA expected outcomes

- Patients
  - Confidence that their information is **protected**
- Hospitals
  - Confidence that data sharing is **legally compliant**
- Researchers
  - Confidence data can be accessed and **used responsibly**

References:

45 CFR Part 160—General Administrative Requirements: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-160
45 CFR Part 164—Security and Privacy: https://www.hhs.gov/hipaa/for-professionals/security/index.html
HHS OCR — The HIPAA Privacy Rule: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html
HHS OCR — HIPAA De-identification Guidance: https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

Data Privacy: Protecting Patients While Enabling Research

phuse.global

Data Privacy Guidelines

Guideline	Region	Description
HIPAA (Health Insurance Portability and Accountability Act)	USA	Protects personal health information (PHI); defines de-identification methods; sets rules for data use and disclosure.
GDPR (General Data Protection Regulation)	EU	Enforces data minimization, consent, purpose limitation, and data subject rights; covers pseudonymization, anonymization, and DPIAs.
21 CFR Part 11	USA (FDA)	Governs electronic records and signatures, including privacy and audit trails for clinical systems.
Data Protection Act (2018)	UK	Implements GDPR standards with UK-specific modifications.
California Consumer Privacy Act (CCPA)	California	Requires transparency and opt-out options for data collection/sharing.
EMA Reflection Paper on Use of RWD in Regulatory Decision Making	EU	Emphasizes ethical use, transparency, and data protection when using RWD for regulatory submissions.
OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data	Global	Promotes harmonization of international privacy principles.
U.S. FDA RWE Framework (2023)	USA	Highlights data privacy considerations in use of RWD for regulatory submissions.
Evaluating Differential Privacy Guarantees (NIST SP 800-226) (2025)	USA	The primary goal of this publication is to help practitioners of all backgrounds better understand how to think about differentially private software solutions.

Data Privacy and De-identification

References: Al Zaabi M, Alhashmi SM. Big data security and privacy in healthcare: A systematic review and future research directions. Information Development. 2026 Jun;42(2):907-21.

De-identification: Making Data Research-Ready

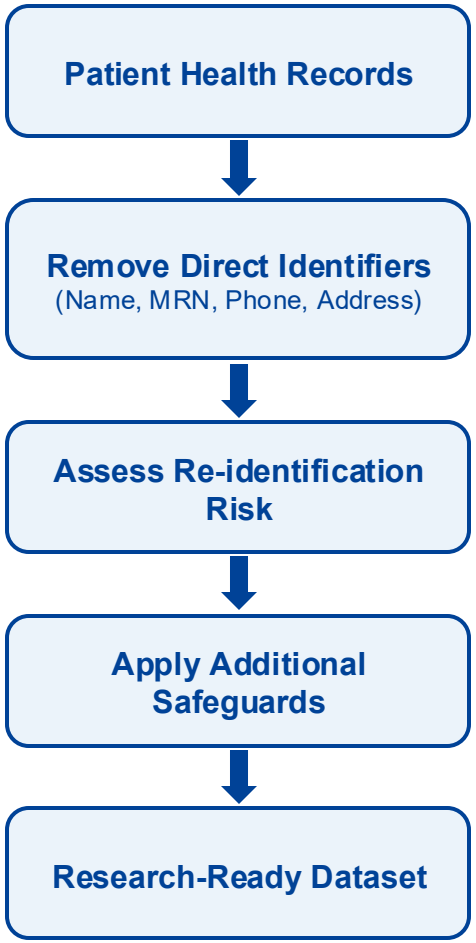
Data Privacy and De-identification

- De-identification transforms sensitive information into a research asset.

- Removes direct personal identifiers from healthcare data.
- Reduces privacy risks while preserving research value.
- Enables the responsible use of RWD for evidence generation.

References: Zambare N, Aghakasiri K, Lin C, Ye C, Mitchell JR, Abdalla M. Towards Fair and Efficient De-identification: Quantifying the Efficiency and Generalizability of De-identification Approaches. InFindings of the Association for Computational Linguistics: EACL 2026 2026 Mar (pp. 4242-4257).

The De-identification Process



- Direct identifiers are removed or masked.
- Potential re-identification risks are evaluated.
- Additional safeguards are applied when necessary.
- Privacy protection is a process, not a single step.

- **Note:** Traditional privacy methods, such as removing names, dates of birth, or direct identifiers, may reduce disclosure risk but do not fully guarantee individual privacy. When datasets are linked with other available information, individuals may still be re-identified.

References: Shahid A, Bazargani MH, Banahan P, Mac Namee B, Kechadi T, Treacy C, Regan G, MacMahon P. A two-stage de-identification process for privacy-preserving medical image analysis. InHealthcare 2022 Apr 19 (Vol. 10, No. 5, p. 755). MDPI.

Protecting Patients Beyond De-identification

Data Privacy and De-identification

Remove	Protect	Enable
Personal identifiers	Against re-identification	Responsible research
Sensitive data	Through governance controls	Evidence generation
Direct patient links	Through ongoing oversight and applying advanced data privacy methods	Public trust

Protecting patients requires more than removing direct identifiers. As real-world healthcare data becomes richer and easier to link, stronger governance and advanced privacy-preserving methods, such as differential privacy, are needed to reduce re-identification risk while enabling trustworthy research.

References: Stukes E. Verifiable Healthcare Data Governance, Provenance, and Accountability in the Age of Artificial Intelligence. Provenance, and Accountability in the Age of Artificial Intelligence (March 31, 2026). 2026 Mar 31.

Question and Discussions

What evidence is needed to show that patient privacy is protected enough before real-world data can be shared or analyzed?

Keep in Touch!

connect·share·advance

If you would like to get involved, email: workinggroups@phuse.global

https://phuse.global/Working_Groups

Slides & recording will shortly be available on our event page

Welcome to reach out to any of the presenters directly:

- Ashwin: ashwin.rai@thermofisher.com
- Rafik and Tuhin: rafik.ghazaryan@biobrainpro.net
  tuhinjamespaul@gmail.com

Real-World Evidence Virtual Event

- 30 September & 1 October
- 8 Selected High Quality Presentations
- Q & A sessions
- Working group updates

- Organizing Committee:
  - Working group leads
  - Community Forum leads

connect·share·advance

THANK YOU!