Workshop #4

Innovative Models and Data Management

Session 2

Workshop #4 Session 2

Facilitators

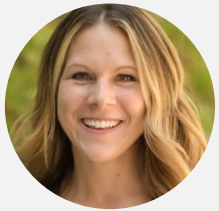
Jaime GUIDRY AUVIL - US National Cancer Institute

Gijs GELEIJNSE - Netherlands Comprehensive Cancer Organisation

Subthemes

- Are there options to consolidate data across a limited set of countries that could then provide a “grouped node”?
- Are there specific considerations for federated queries that relate to interoperability questions?

Discussants



Dr Heidi HANSON
USA
Statistics Canada
Oak Ridge National Laboratory
Group Lead Biostatistics
and Multilevel Systems Modeling,
Senior scientist

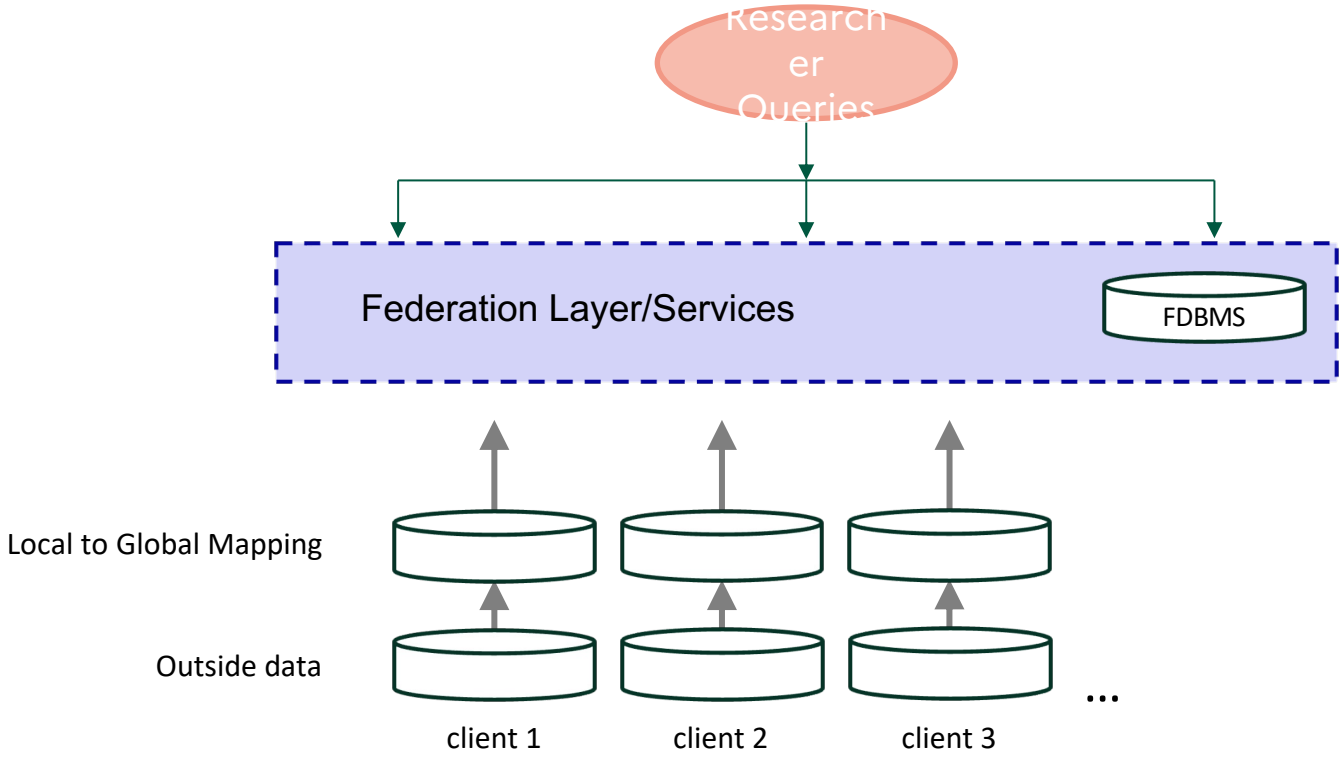
Robert MILLER
USA
Minderoo Foundation
Principal, Unlocking Patient
Data, Cancer Mission,
Minderoo Foundation

Smashing Data Silos: Opportunities and challenges for federated technologies in childhood cancer research

Dr. Heidi Hanson
Advanced Computing for Health Sciences | Oak Ridge National Laboratory

Modeling Outcomes Using Surveillance Data & Scalable Artificial Intelligence for Cancer

Federated Data



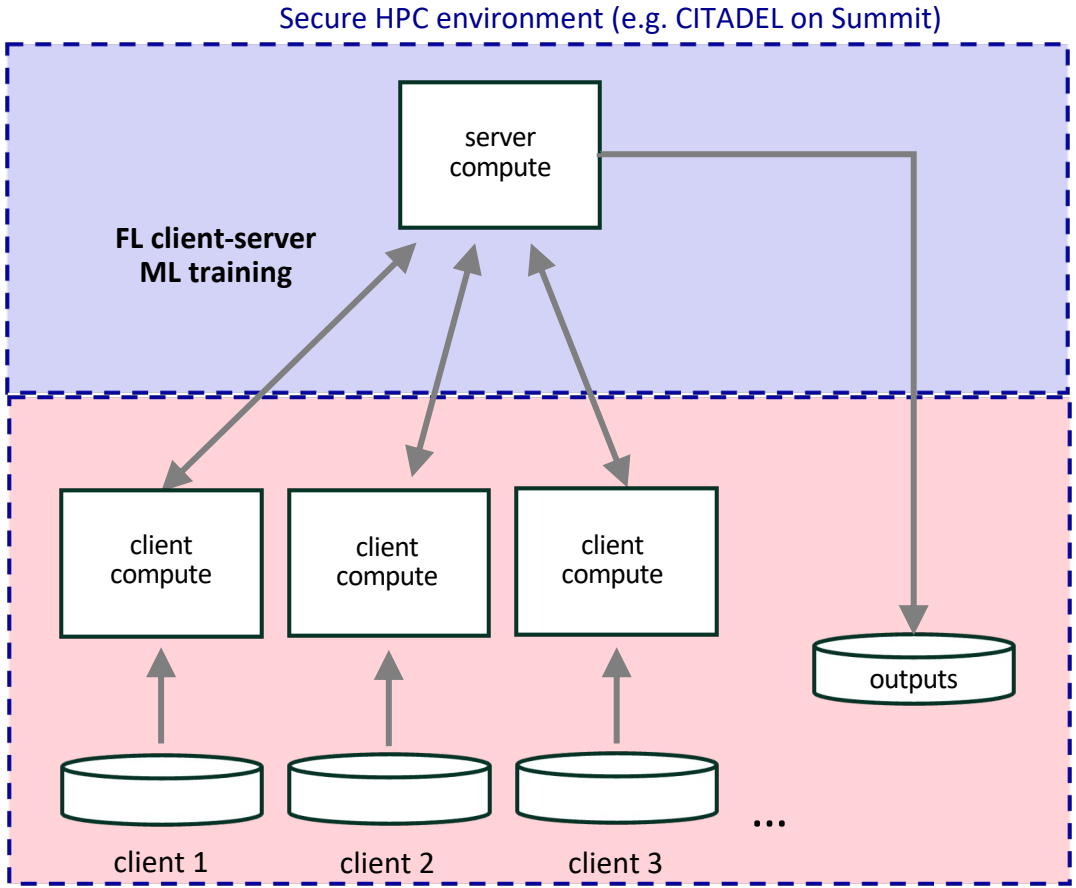
Benefits

- Local control of data
- Addresses privacy and security
- Enables the ability to analyze larger datasets

Constraints

- Potential privacy issues
- Data deidentification
- Difficult to change data structures

Federated Learning (FL) on HPC



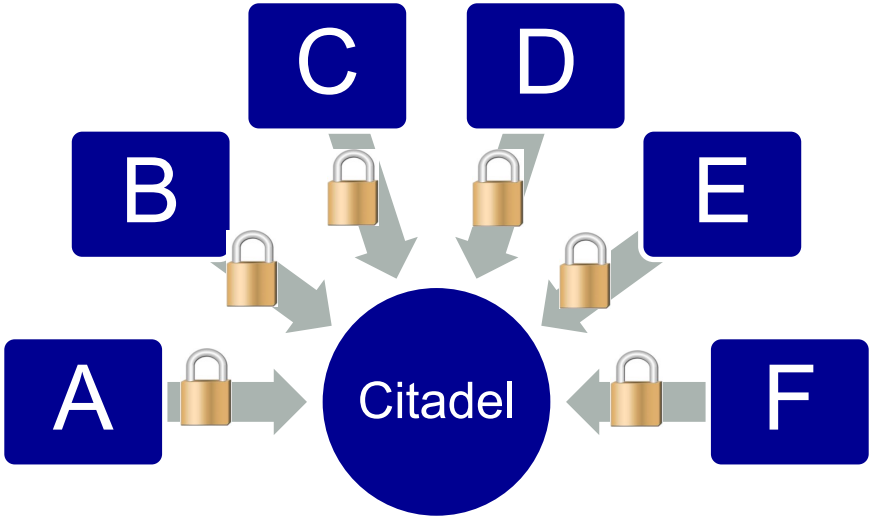
Benefits

- Agile data engineering
- Near real-time learning across organizations

Constraints

- Data heterogeneity
- Statistical heterogeneity
- Privacy
- Expensive communication

Near Real-Time Health Data Analytics with Privacy Aware Federated Learning

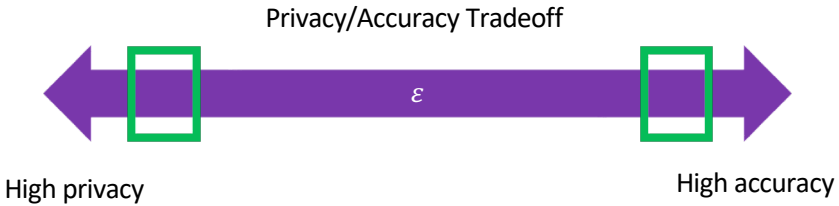


Our current implementation

- Cross Silo
- Horizontal
- Model Centric – Need for Cooperative Agreement between institutions/participants

Trusted Host

Differential Privacy
"Epsilon Indistinguishability"

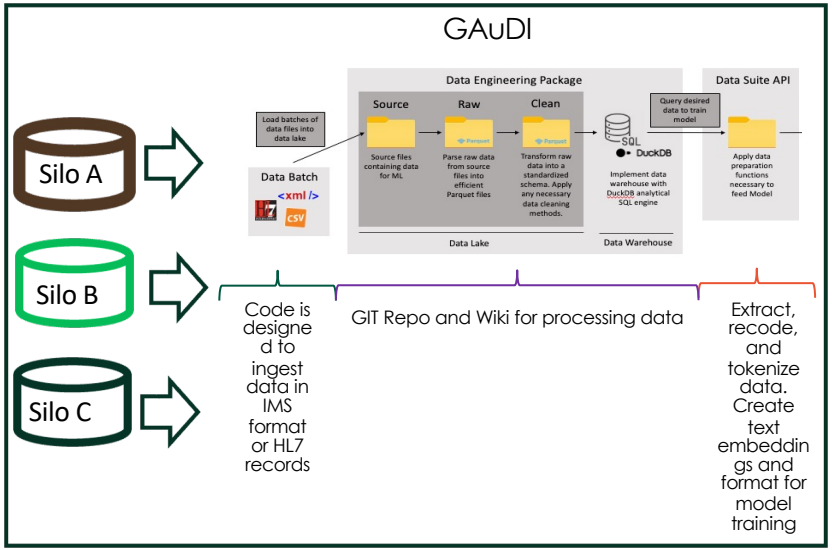


Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14

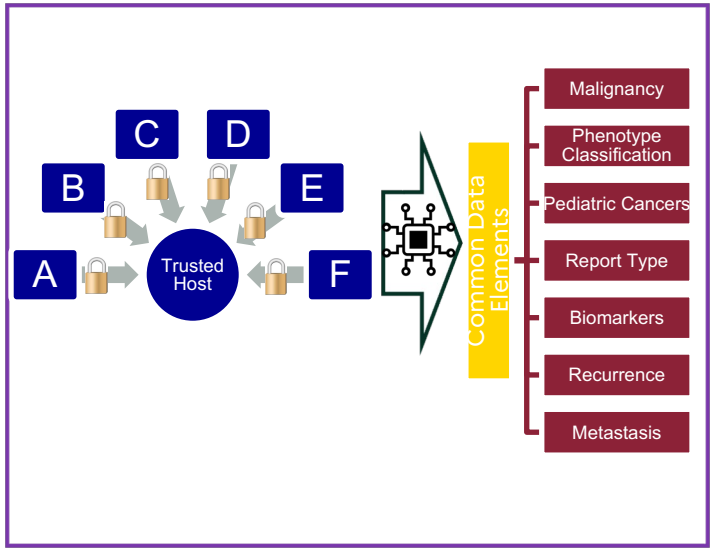
Charting a Path for Real World Federated Learning

Integrated Tools for AI at Scale

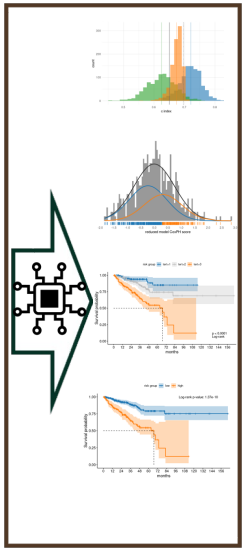
Standard workflows across siloed systems for ingesting and harmonizing unstructured text



FL for extracting common data elements



FL for downstream analysis



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INSTITUT NATIONAL DU CANCER

NIH NATIONAL CANCER INSTITUTE

NIH NATIONAL CANCER INSTITUTE

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INSTITUT NATIONAL DU CANCER

Thank you! Merci!

Heidi Hanson

Oak Ridge National Laboratory

hansonha@ornl.gov

Paris Conference for an International Childhood Cancer Data Partnership
Newcap Event Center (Paris, France), November 7-8, 2023

A global common oncology data standard:

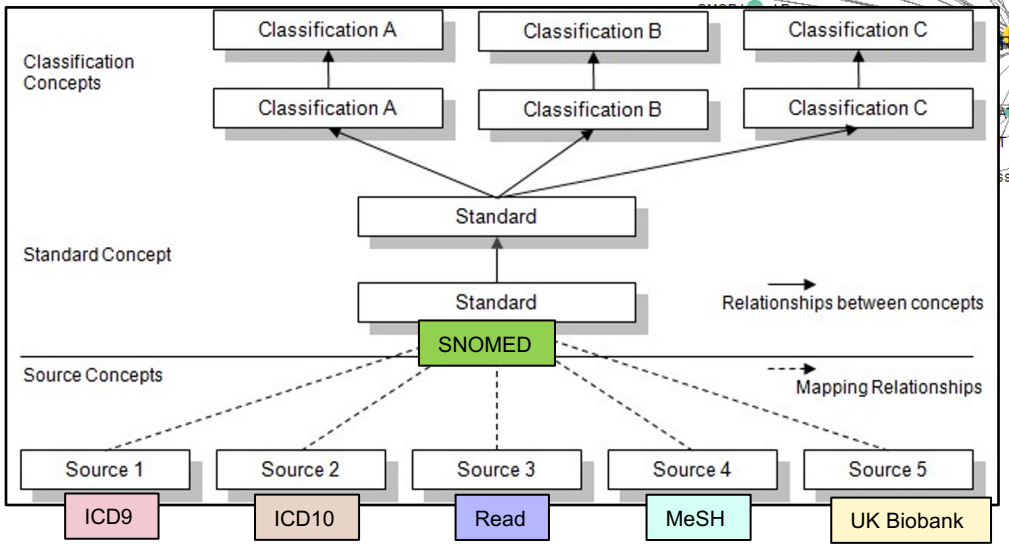
Challenges, Requirements & Opportunities

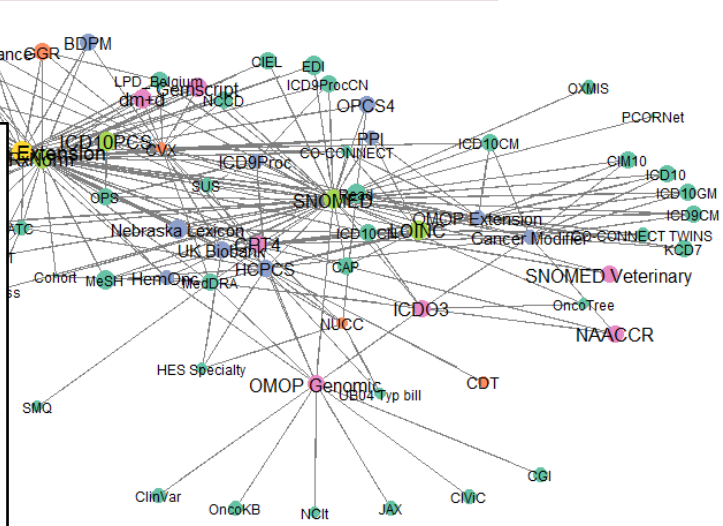
Robert Miller
Minderoo Foundation

Common Data Model (CDM): Core Components

1) Schema…

2) Ontologies



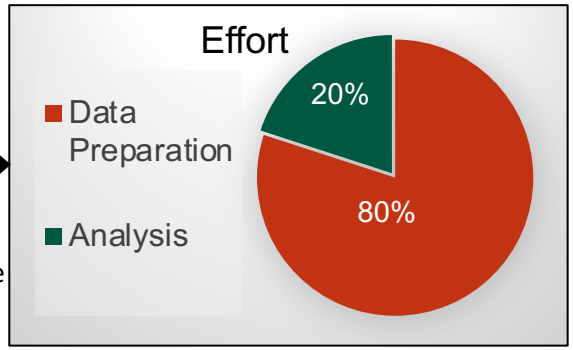


Semantic Interoperability

Power of CDMs

Efficiency

80:20 "Rule" of Data Science



In a CDM, the data only needs to be prepared once
*(and updated periodically)

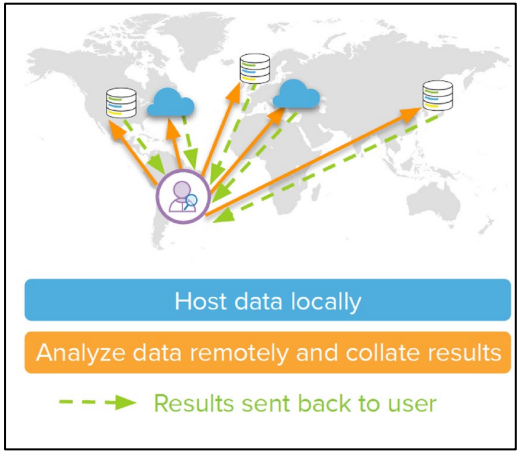
Tooling

A common platform enables collaborative tooling. Write once, use many.

- Cross-site queries, study packages, data quality routines, shared ETLs, regimen detection, NLP, genomic parsers, etc.

Enables Collaboration

- Establishes semantic interoperability
- Enables federation
  - A "decentralised" approach
  - Overcomes ethical, legal, privacy & ownership barriers
  - Limits risk
- "Opt-in" policy for research
  - Never required to transfer patient data
  - Maintain privacy and ownership
  - Decide when and how data is used

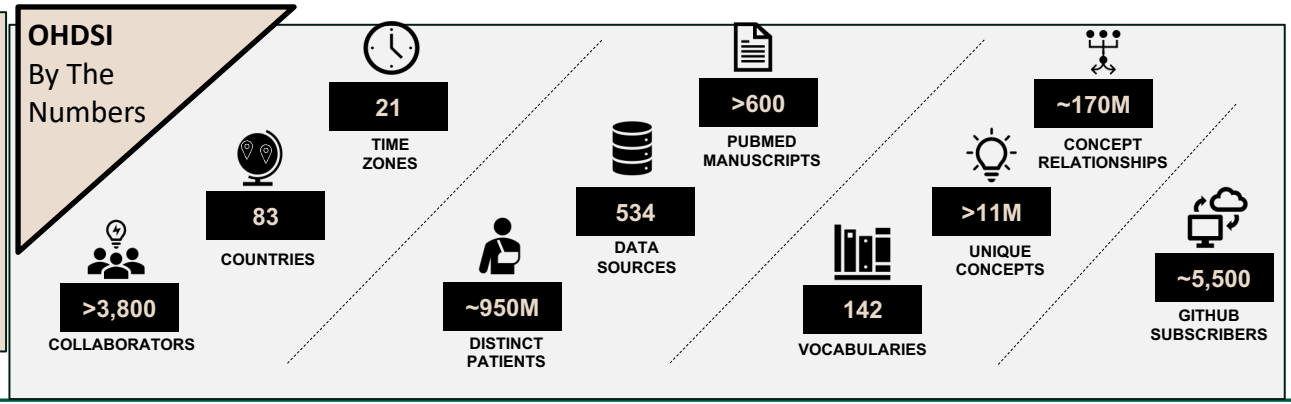


OMOP Common Data Model

Most widely adopted open health common data model globally

Persists data in a **de-identified** state and is federated by design

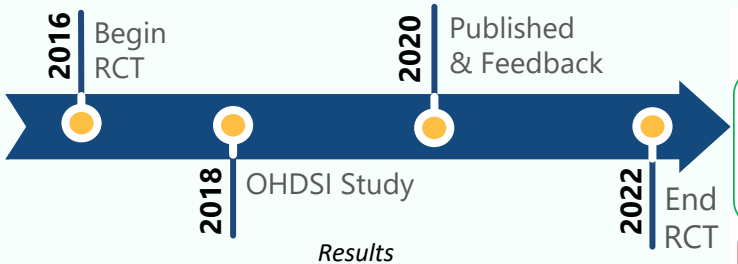
Maintained by the Open Health Data Sciences and Informatics **(OHDSI) community** - a volunteer, global open-health science collaboration



Hypertension Study

US Clinical Practice Guideline for Hypertension:

Drug A is preferred (over drug B) for proven **reduction of CVD**



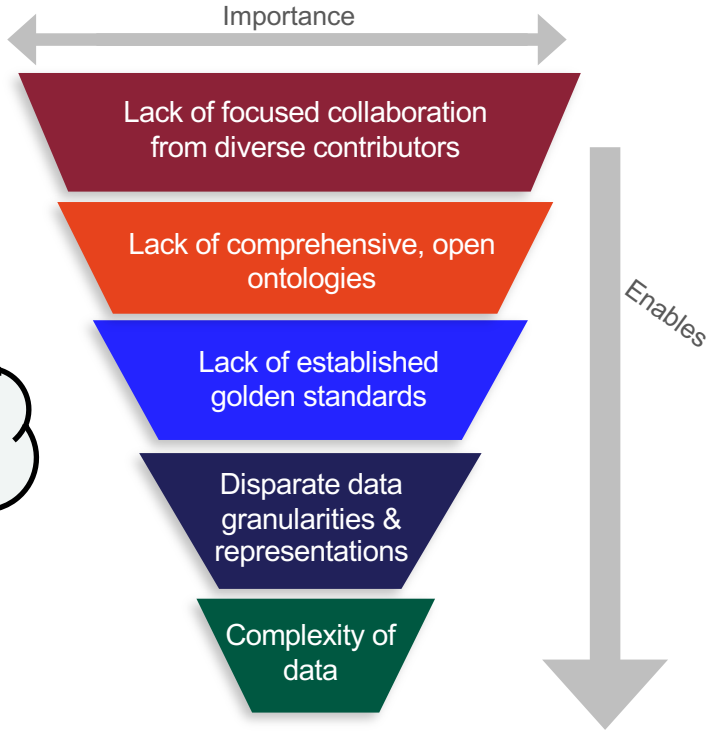
Results

Drugs have **same risk of CVD** and **Drug A** has **higher risk of hypokalemia**

	OHDSI's LEGEND in 2018/2020	Diuretic Comparison Project RCT in 2022
Cardiovascular events	1.00 (0.85-1.17)	1.04 (0.94-1.16)
Hospitalization for Acute myocardial infarction	0.92 (0.64-1.31)	1.02 (0.80-1.28)
Hospitalization for Stroke	1.10 (0.86-1.41)	1.00 (0.74-1.36)
Hospitalization for Heart failure	1.05 (0.82-1.34)	1.04 (0.87-1.25)
Hypokalemia	2.72 (2.38-3.12)	1.38 (1.19-1.60) $p < 0.001$

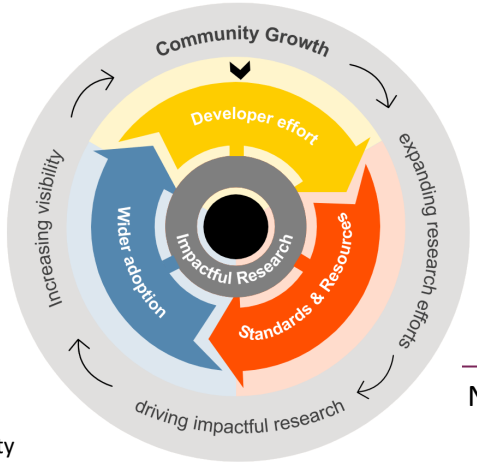
Challenges

- *Currently*, there is no open standard for comprehensive cancer patient data
- There has been a historical tendency towards “centralized”, “domain specific” and “ad hoc” approaches for large oncology efforts
- Lacking clarity and visibility of incentives and assurances of the approach
- Licensing and other usage restrictions by organizations developing vocabularies
- Absence of sufficiently visible successes in the space



Path Forward

"Enable and accelerate global federated oncology research"



Strategy to date:

- Build the community & mature the oncology standards in OMOP, notably around ontologies, to help establish a de-facto global standard, through:
  - Extensive outreach and curation of issues and gaps
  - Facilitating an open maturity effort, prioritizing transparency and asynchronous community contributions

    Managed in Github: t.ly/jbbdn
  - Early investments into standards development, pilot projects and SMEs

Next steps:

1. Continue to work with open-source communities to mature oncology standards
2. **Explore additional opportunities to leverage philanthropic funding towards:**
   - Improvement of standards and ease of adoption
   - Facilitating international collaborations and community growth
   - Enabling and accelerating impactful oncology research

Recommendations:

- Our solution must be **dynamic** requiring:
  - An active, aligned community
  - Prioritization of extensible design
  - Evolution through iterations of stable releases
- Prioritize an open and impact-driven international community
- Keep these conversations going and establish new ways of working
- Add visibility into the incentives and assurances of the approach
- Do pilot studies to demonstrate value and impact
- Choose solutions that are extensible and reusable rather than bespoke repetitions

rmiller@minderoo.org