October 16-17, 2024
Hyatt Regency Toronto
Toronto, ON
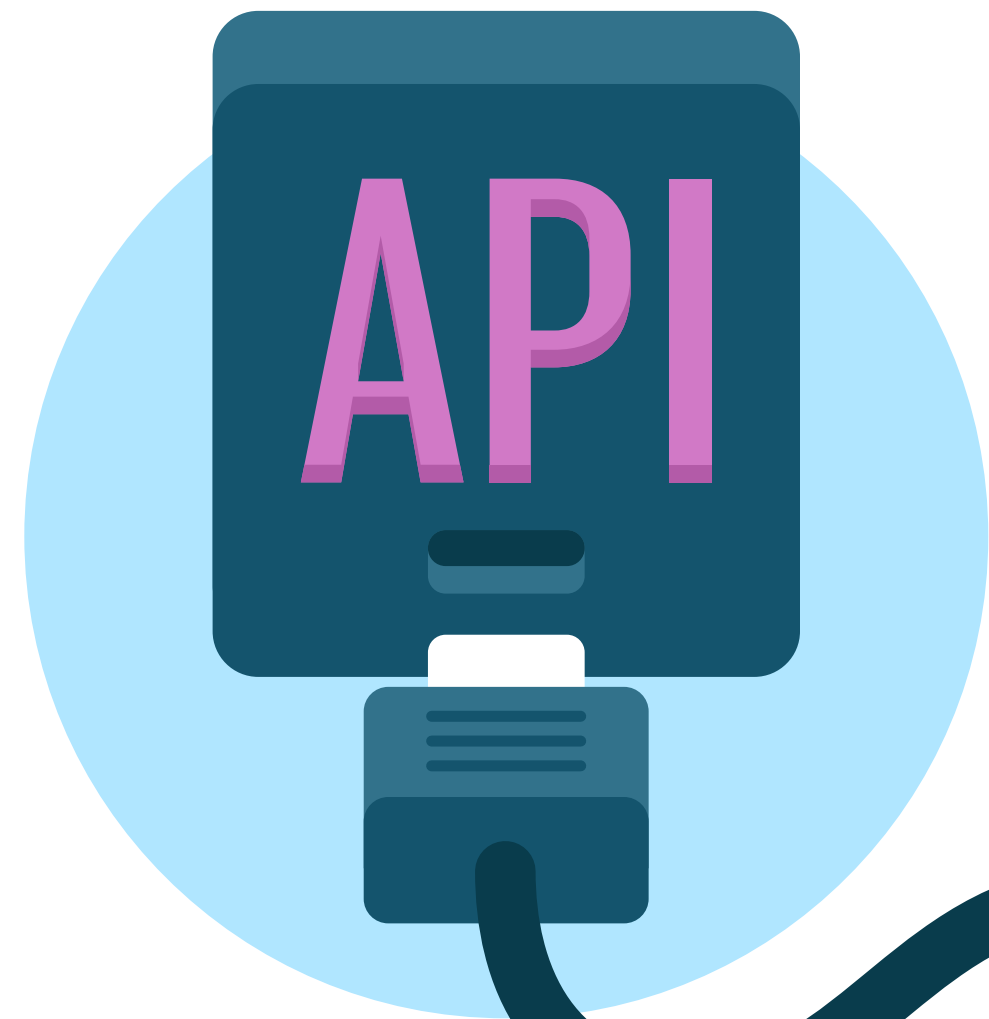
Don't Share Your Data: Applications of Federated Data Architecture in Canadian Public Health

Christopher Allison
Directory General and Chief Data Officer
Public Health Agency of Canada

1

Stop **SHARING DATA** ...

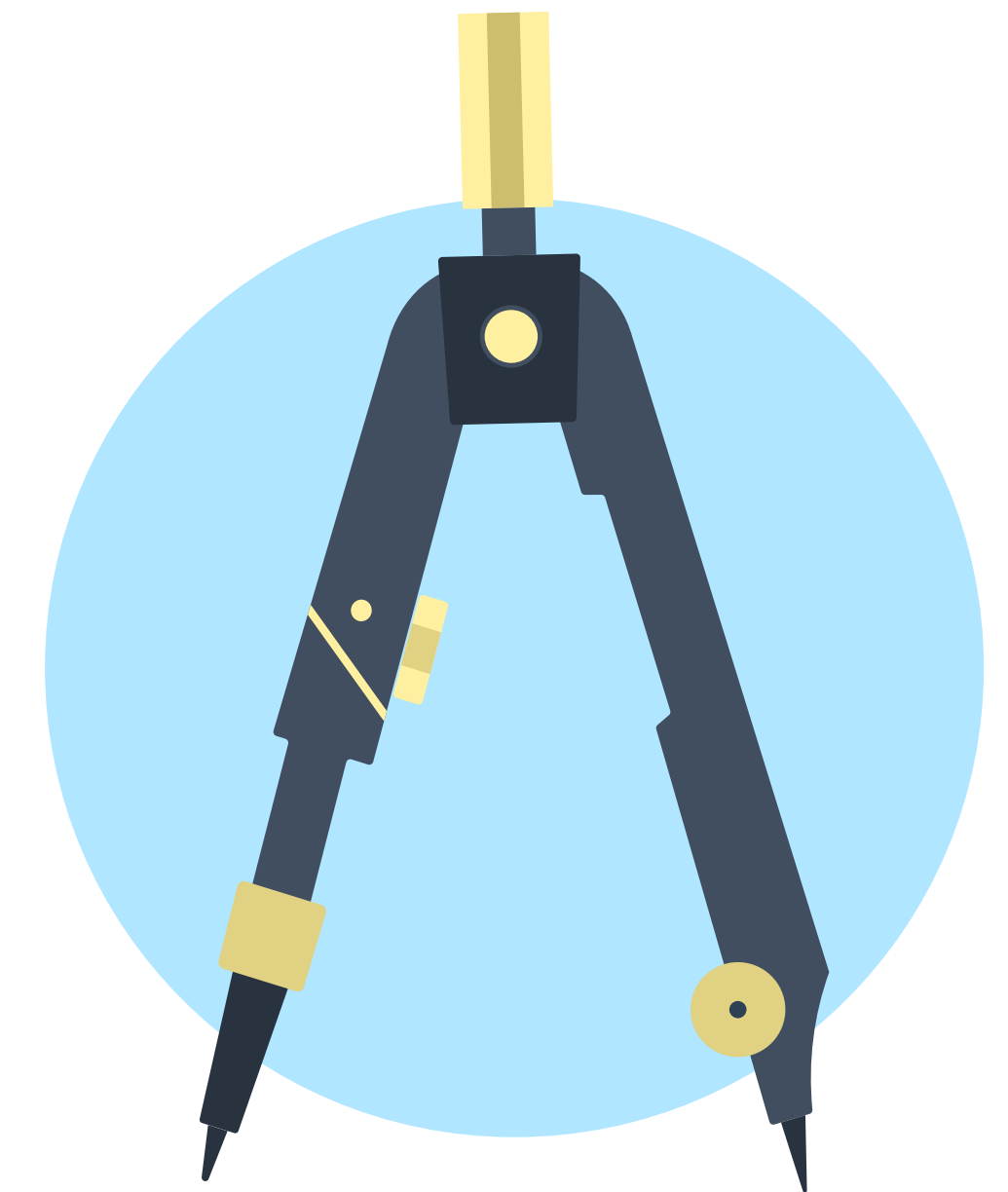... enable **DATA ACCESS** instead

2

DATA STEWARDS do not just wrangle data ...

... they **CREATE VALUE** for downstream data users

3

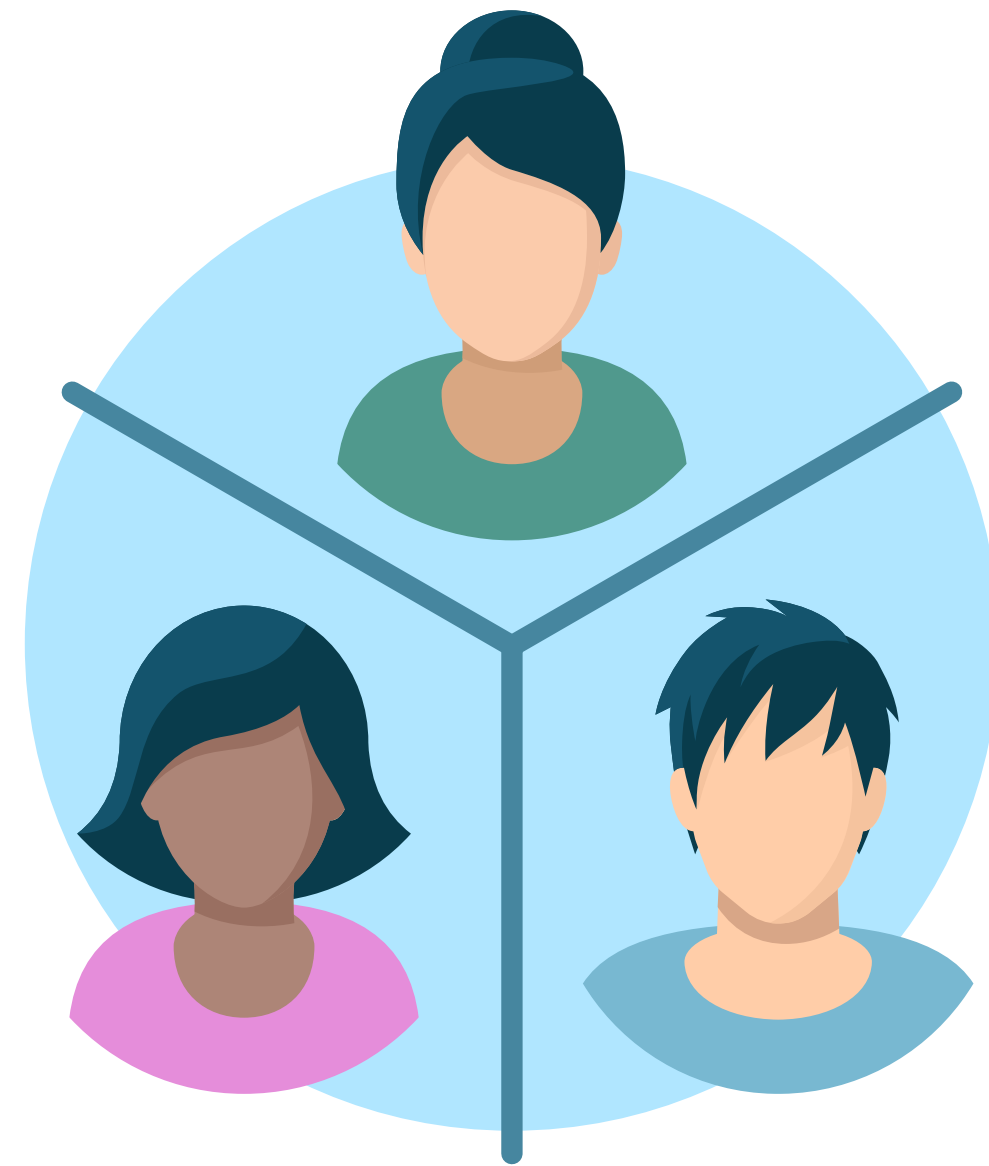Many **THREATS TO DATA** that we fixate on ...

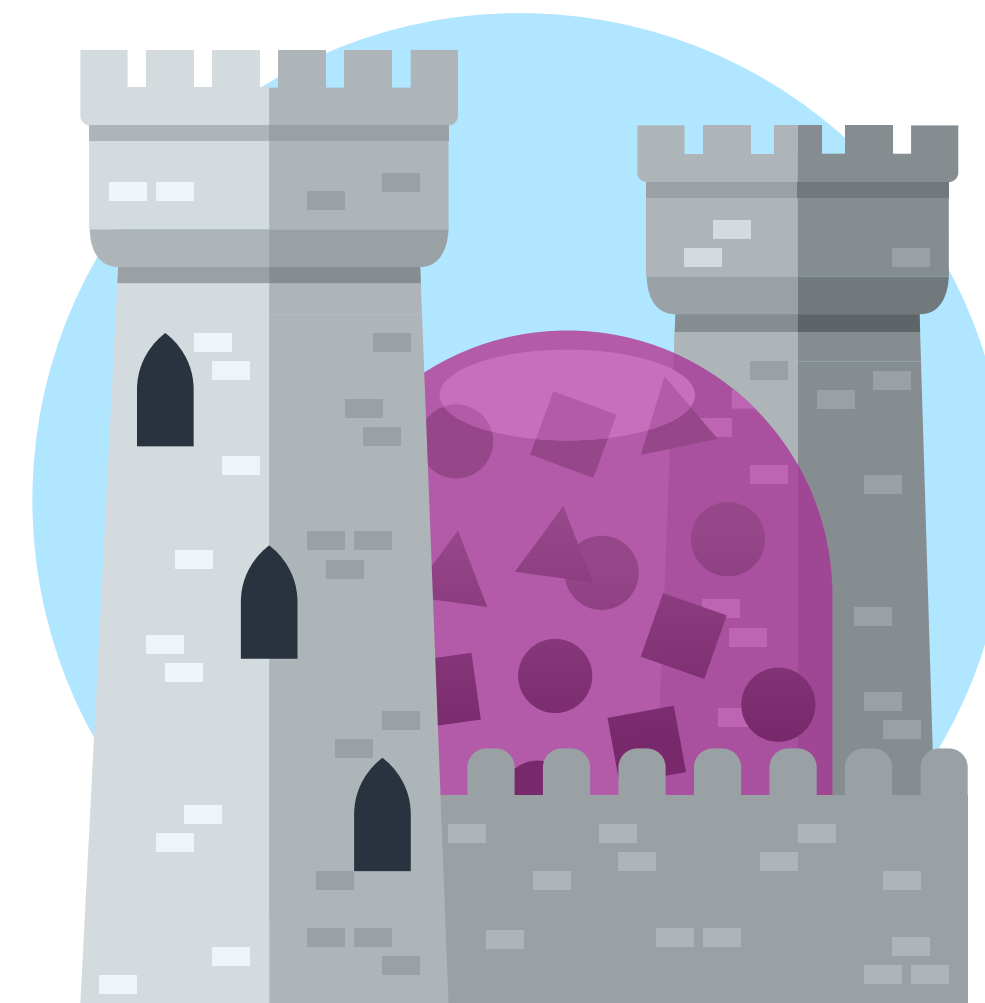... can be side-stepped with the right **ARCHITECTURE**

HEALTH DATA CHALLENGES

Making the most of health and medical data has been undermined by a number of challenges
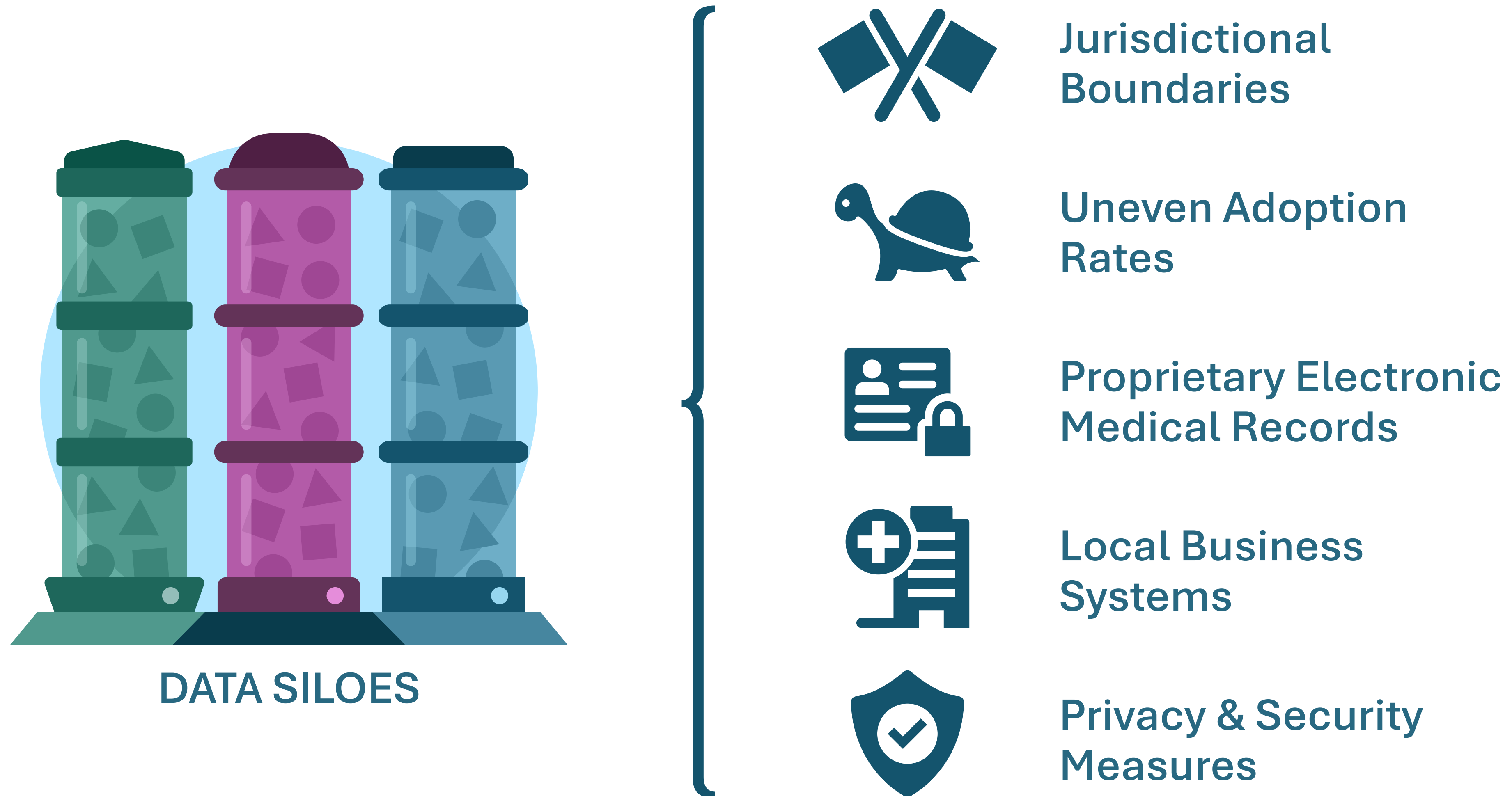

ORGANIZATIONAL
SILOES

PROFESSIONAL
FRAGMENTATION

WALLED-OFF
DATA ACCESS

LACK OF
STANDARDS

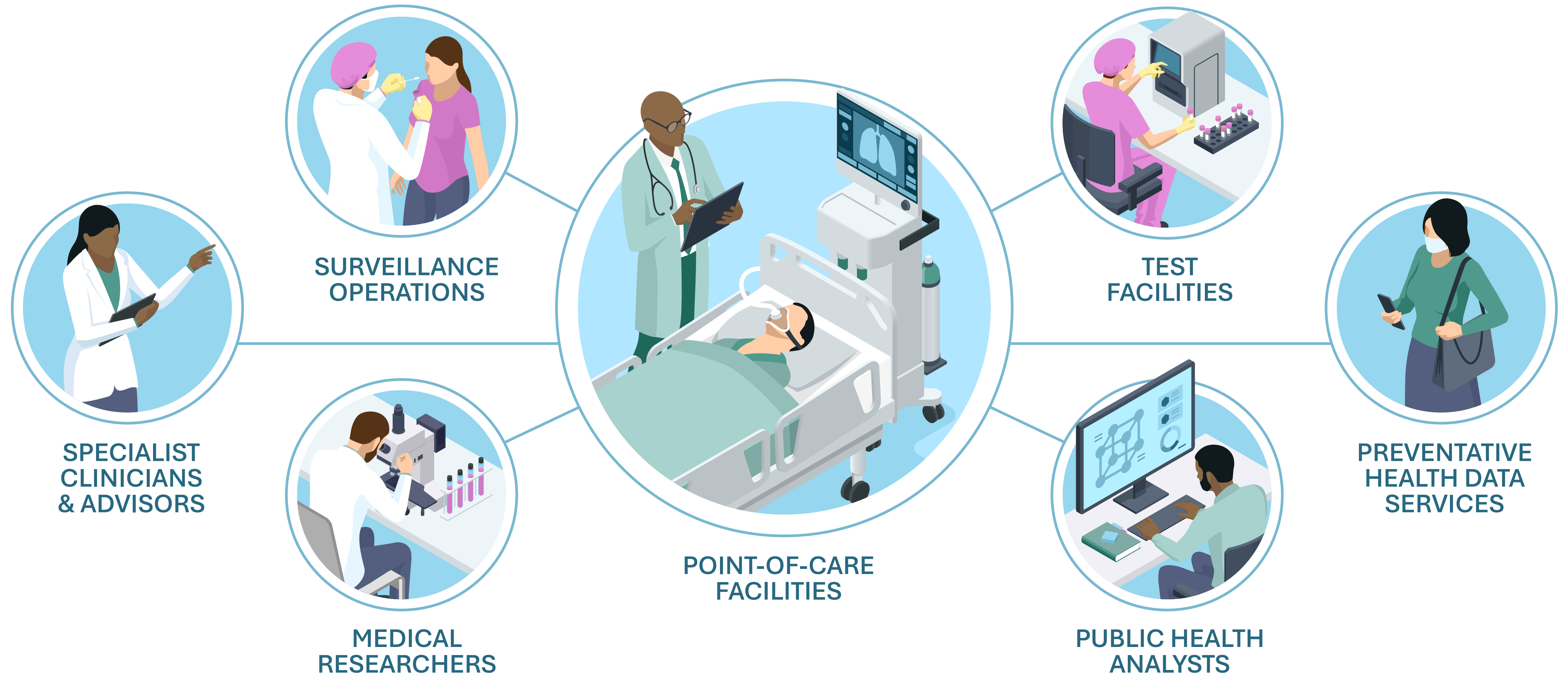# WHY SILOES?

Data becomes siloed as data owners fixate on their own operations without regard to the larger data ecology

DATA SILOES

Jurisdictional Boundaries

Uneven Adoption Rates

Proprietary Electronic Medical Records

Local Business Systems

Privacy & Security Measures

INTEGRATED PATIENT CARE

Data is the lynchpin of modern medicine but is difficult to gather together and synthesize in a timely way

SURVEILLANCE OPERATIONS

SPECIALIST CLINICIANS & ADVISORS

MEDICAL RESEARCHERS

POINT-OF-CARE FACILITIES

TEST FACILITIES

PUBLIC HEALTH ANALYSTS
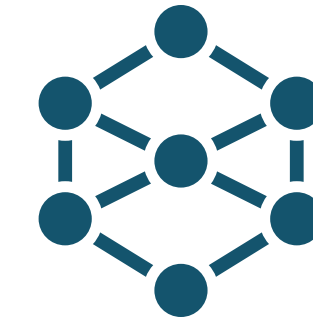
PREVENTATIVE HEALTH DATA SERVICES

DILEMMAS FOR HEALTH CARE

Data siloes are ill suited to looking at patients and public health challenges holistically
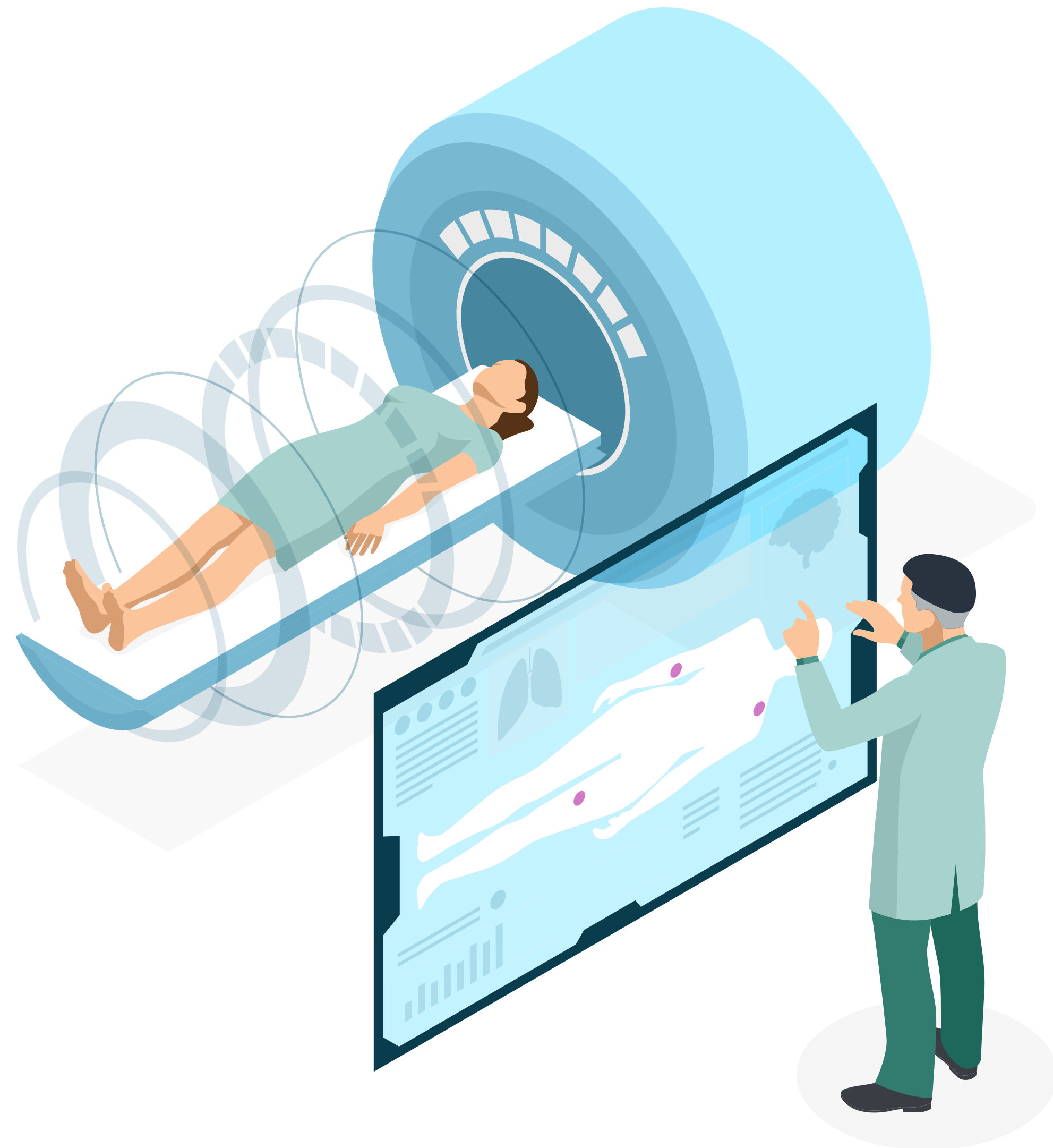
- Humans are Integrated Systems
- Need for Knowledge Synthesis Across Specializations
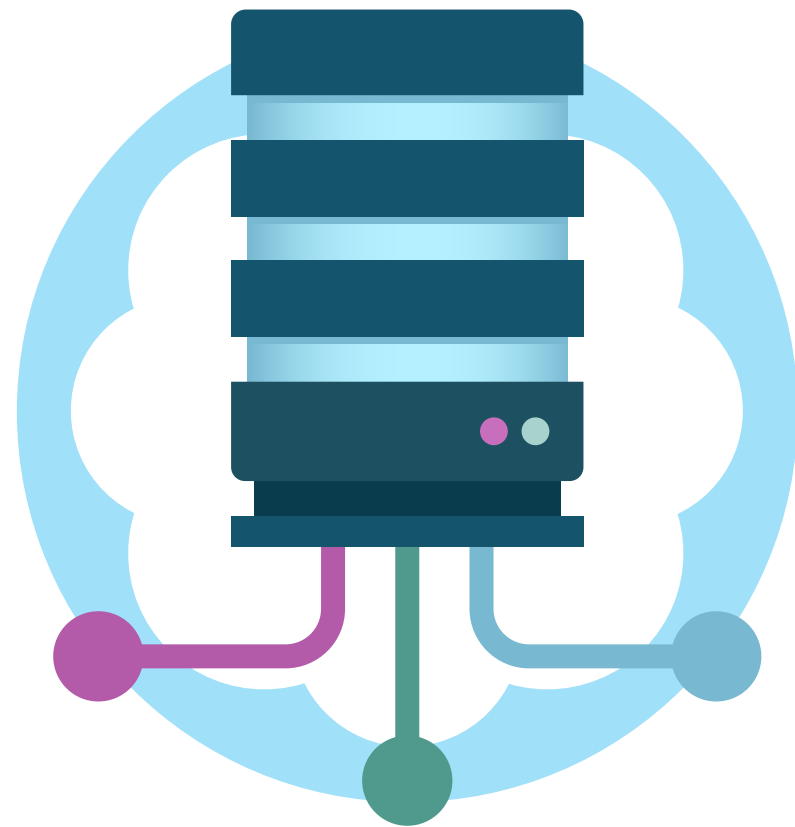- Treatment Delays caused by Untimely Data
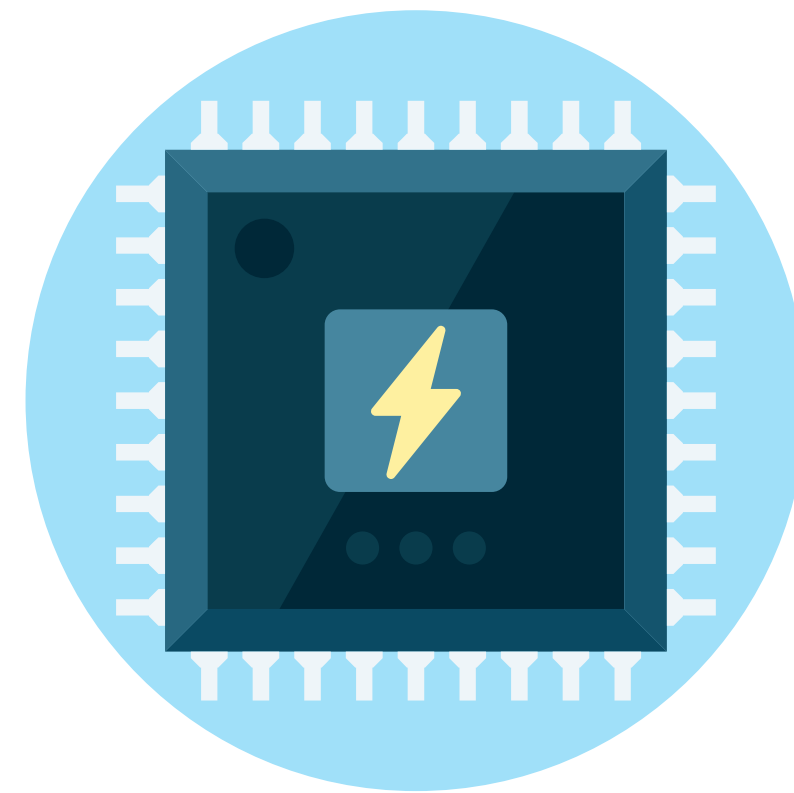- Misdiagnosis and Inappropriate Treatments
- More Hoops to Jump Through (Patient Burden)

THE CHANGING ENVIRONMENT

Technology and governance are evolving in ways that fundamentally change the way data is managed
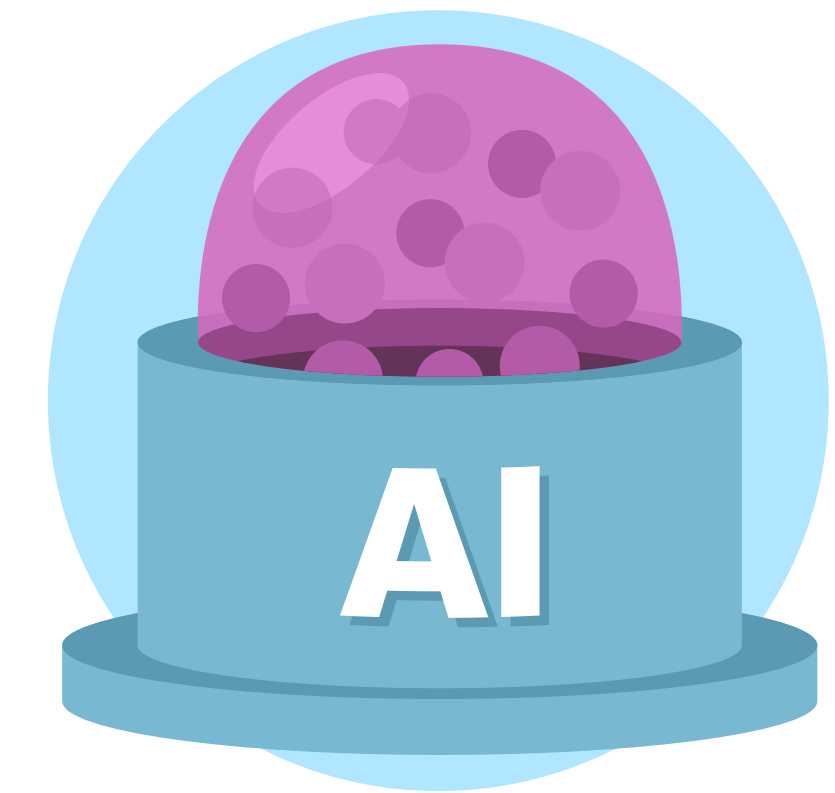
CLOUD
COMPUTING

COMPUTING
POWER
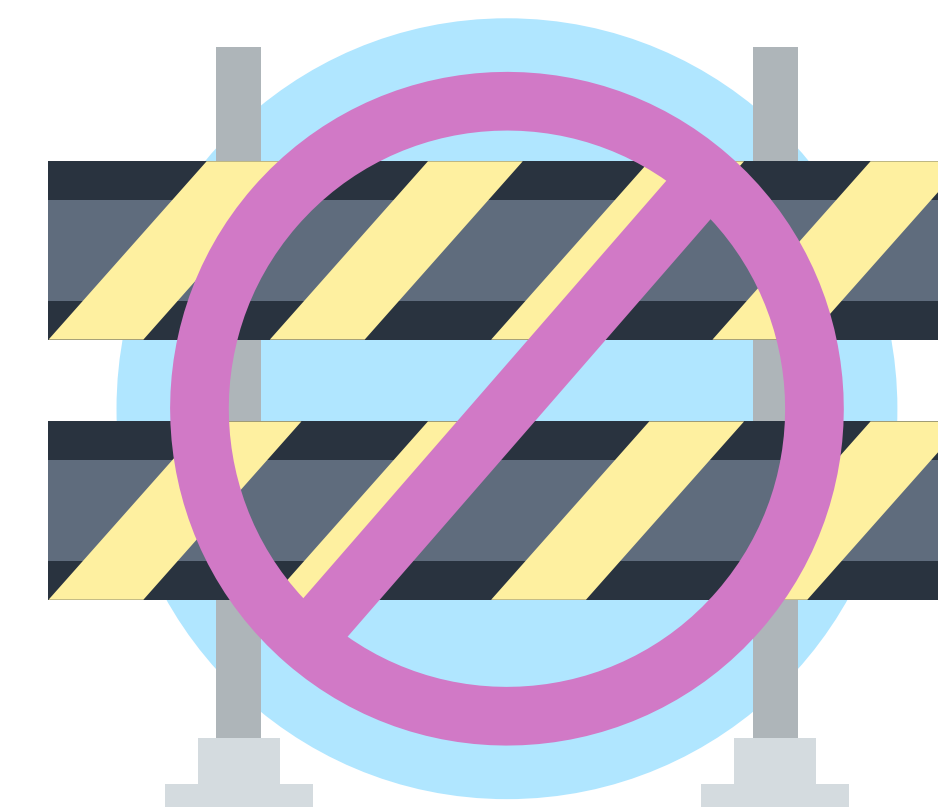
ADVANCED/OPEN
ANALYTICS

MACHINE
LEARNING & AI

DATA STANDARDS
MOVEMENT

DATA-SHARING
AGREEMENTS

DATA BLOCKING
PROHIBITIONS

WHAT ARE THE OPTIONS?

There are a number of approaches to overcoming these challenges

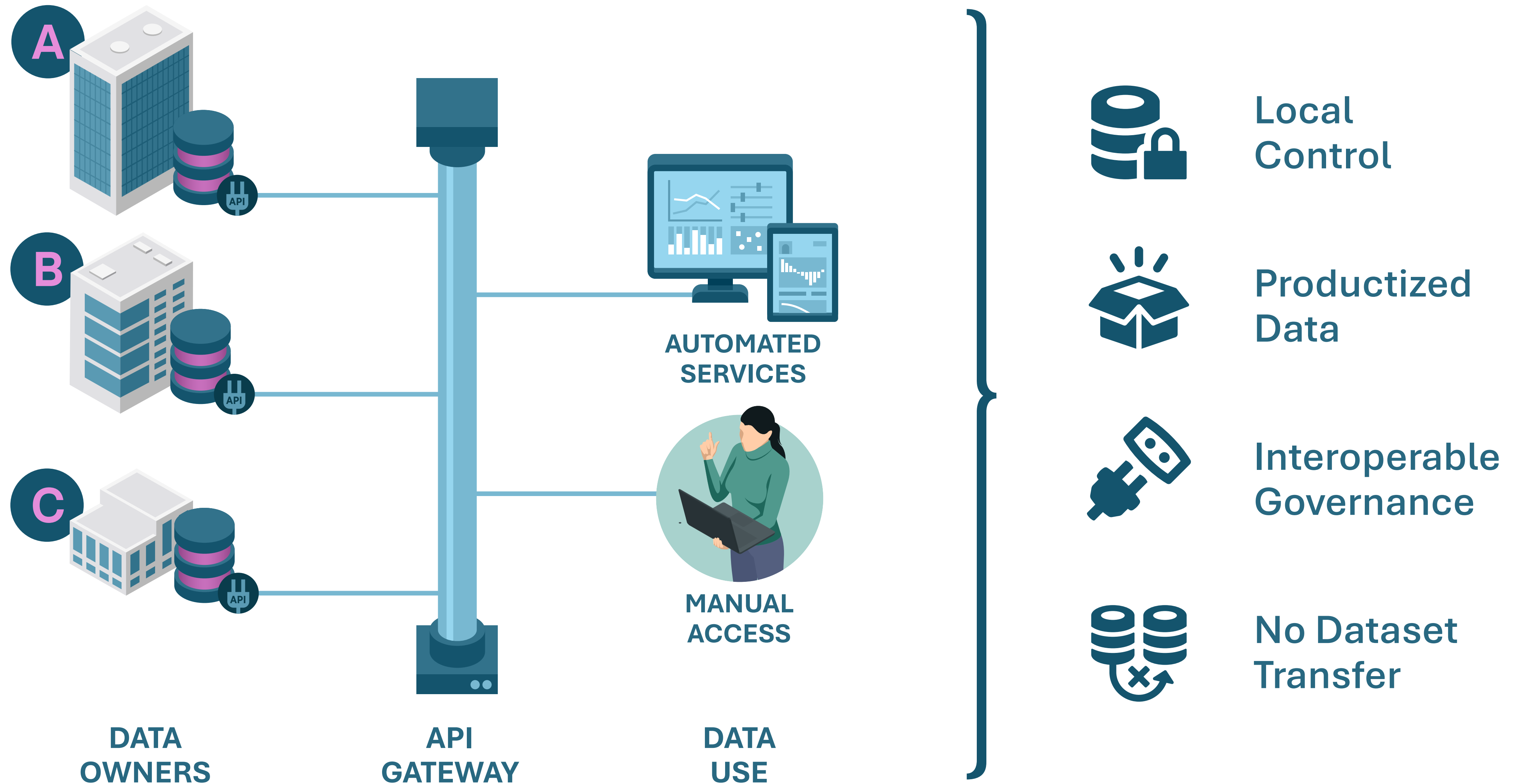| DATA TRUSTS | DATA INTEGRATORS | API-BASED DATA PLATFORMS | FEDERATED DATA SYSTEMS | AI-POWERED DATA MAPPING |

FEDERATED DATA

Data owners retain control of their data but make streams available to services through a common gateway

A

B

C

DATA OWNERS

API GATEWAY

AUTOMATED SERVICES

MANUAL ACCESS

DATA USE

Local Control

Productized Data

Interoperable Governance

No Dataset Transfer

DATA AS A PRODUCT

Data owners retain decentralized control in exchange for the obligation to make data available to others

- Always-on Cloud Database
- Self-serve Data Streams
- Self-describing Data (Meta Data)
- Managed using FAIR Principles
  - Findability
  - Accessibility
  - Interoperability
  - Reusability
- Flow-control Code (Rule Enforcement)

WHY FEDERATION?

Federated data systems sidestep many of the big-data problems that inflict centralized architectures

EASE OF SCALING

SINGLE SOURCE OF TRUTH

LESS MANUAL COORDINATION
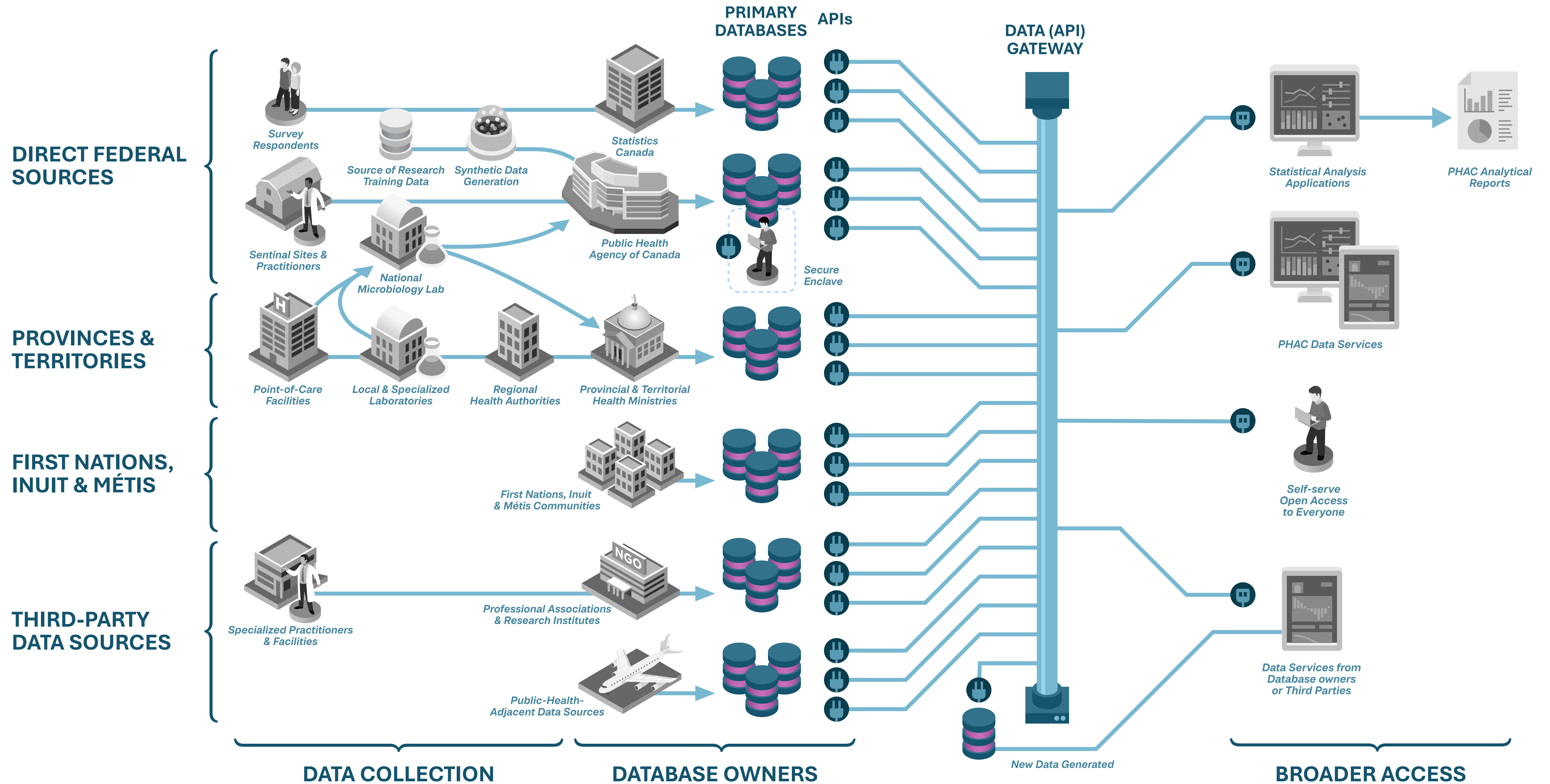
REAL-TIME DATA MORE VIABLE

**MORE SECURE**

BETTER PRIVACY

DEFUSE POLITICAL TENSIONS

CANADIAN HEALTH SURVEILLANCE

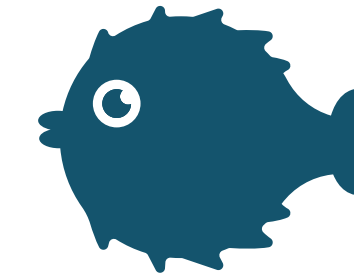*Federated data architectures best cater to the highly distributed nature of data ownership in Canada*

PRIMARY DATABASES

APIs

DATA (API) GATEWAY

DIRECT FEDERAL SOURCES

Survey Respondents

Source of Research Training Data

Synthetic Data Generation

Statistics Canada

Sentinal Sites & Practitioners

National Microbiology Lab

Public Health Agency of Canada

Secure Enclave

PROVINCES & TERRITORIES

Point-of-Care Facilities

Local & Specialized Laboratories

Regional Health Authorities

Provincial & Territorial Health Ministries

FIRST NATIONS, INUIT & MÉTIS

First Nations, Inuit & Métis Communities

THIRD-PARTY DATA SOURCES

Specialized Practitioners & Facilities

Professional Associations & Research Institutes

Public-Health-Adjacent Data Sources

Statistical Analysis Applications

PHAC Analytical Reports

PHAC Data Services

Self-serve Open Access to Everyone

Data Services from Database owners or Third Parties

New Data Generated

DATA COLLECTION

DATABASE OWNERS

BROADER ACCESS

WHERE TO START?

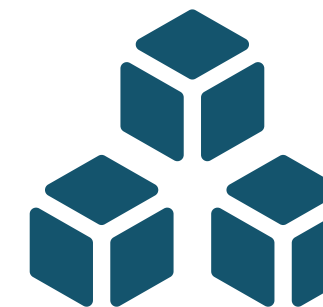*Our approach involved involving stakeholders, laying the foundations, and building a proof-of-concept*
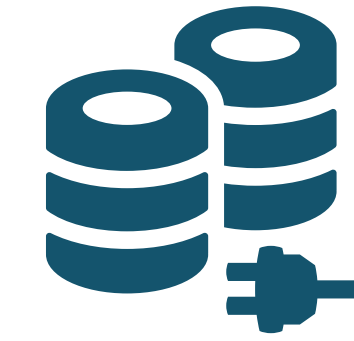
- Explaining Federated Data
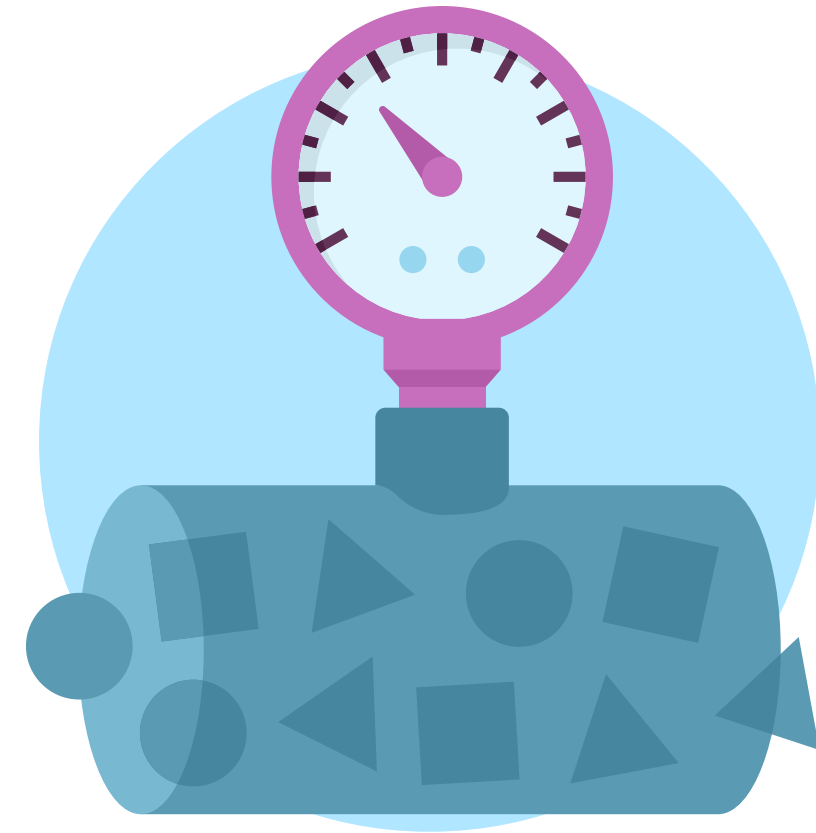- Storytelling Analogies
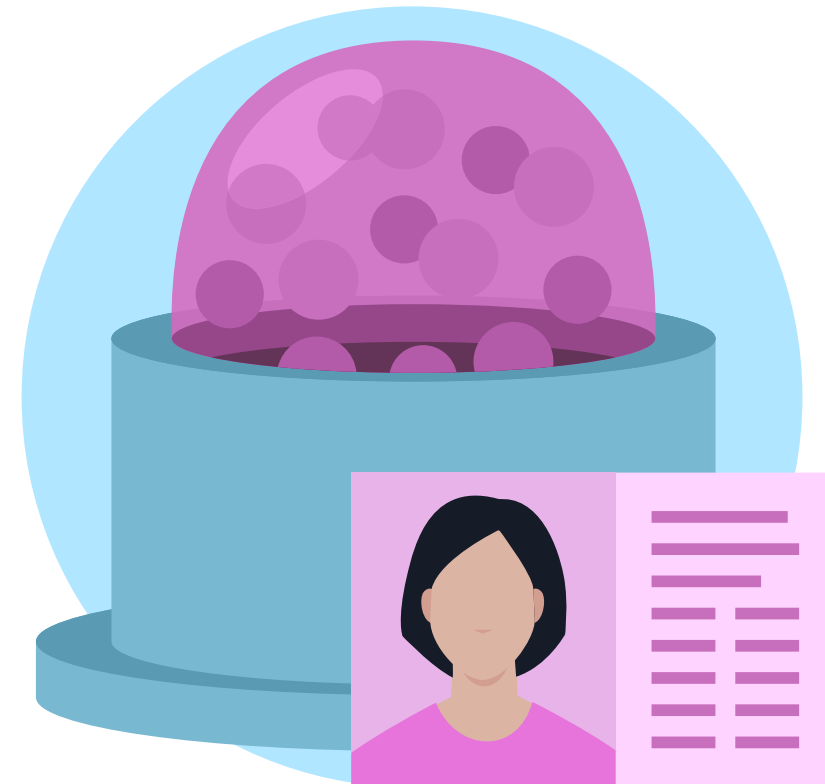- Thinking Through the Ecosystem
- Identifying Components
- Data Stewardship Principles

PROOF OF CONCEPT

We have built a proof of concept using vaccination data with several powerful features
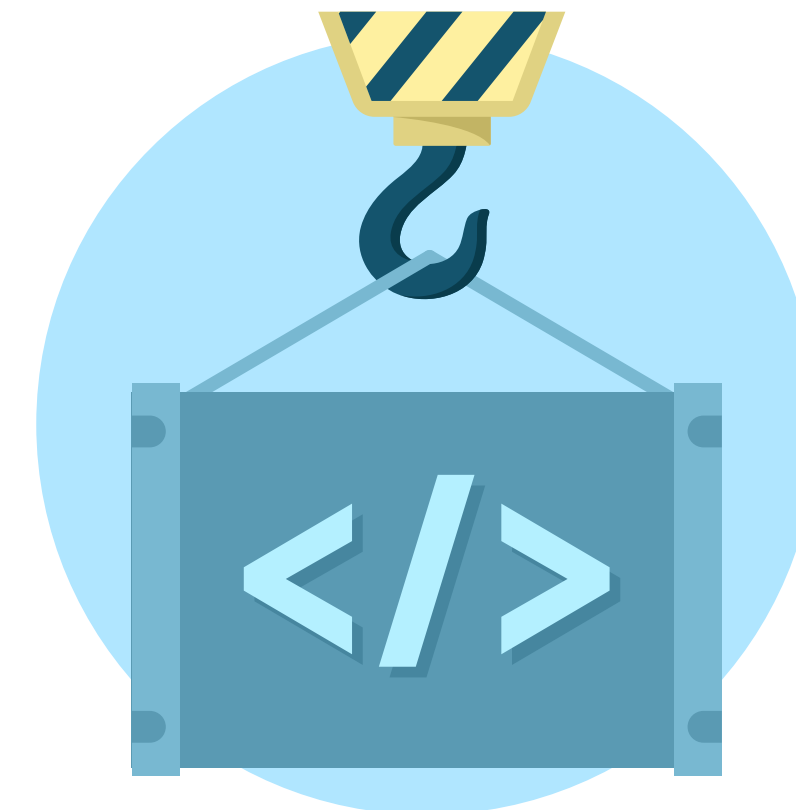
**REAL-TIME
DATA STREAMING**

**SYNTHETIC
PATIENT DATA**

**GOVERNANCE
GUI**

**ANALYTICS
SUITE**

**INFRASTRUCTURE/
GOVERNANCE AS CODE**

CHALLENGES

Federated data is not a plug'n'play solution and there is a need for some humility

**TECHNICAL
BAR TO ENTRY**

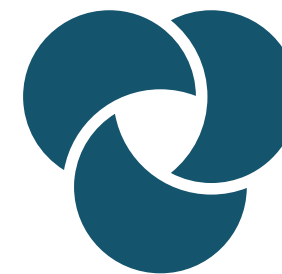**DATA QUALITY
CONTROL**

COMPLICATIONS

A variety of factors add further wrinkles to data stewardship
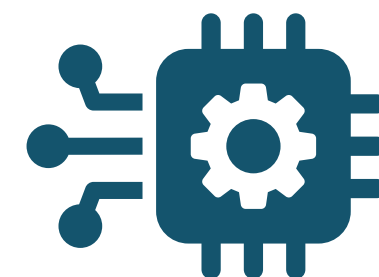
Data Collection at Source

Connect Before You Collect

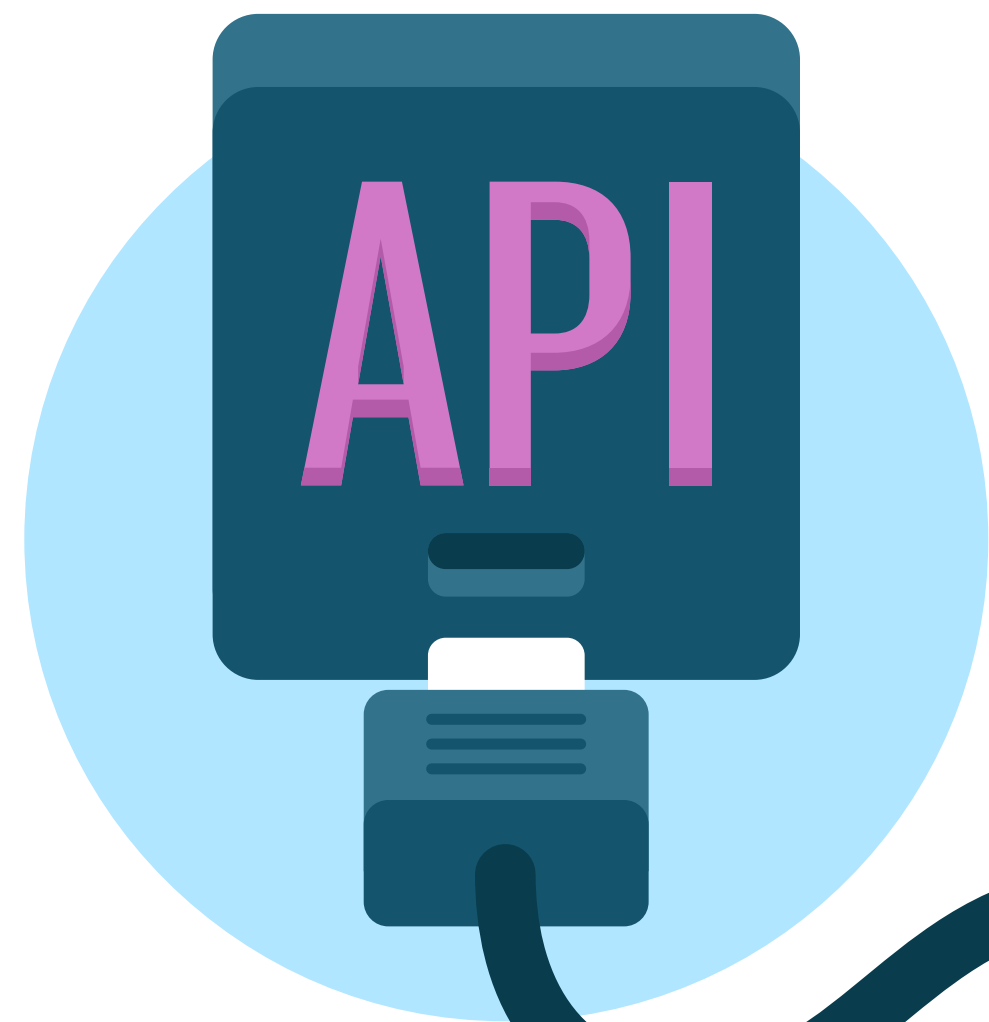Intersectionality

Power in Asking Questions

Community Data Rights/ Indigenous Data Sovereignty

Intelligent AI Usage

1

Stop **SHARING DATA** …
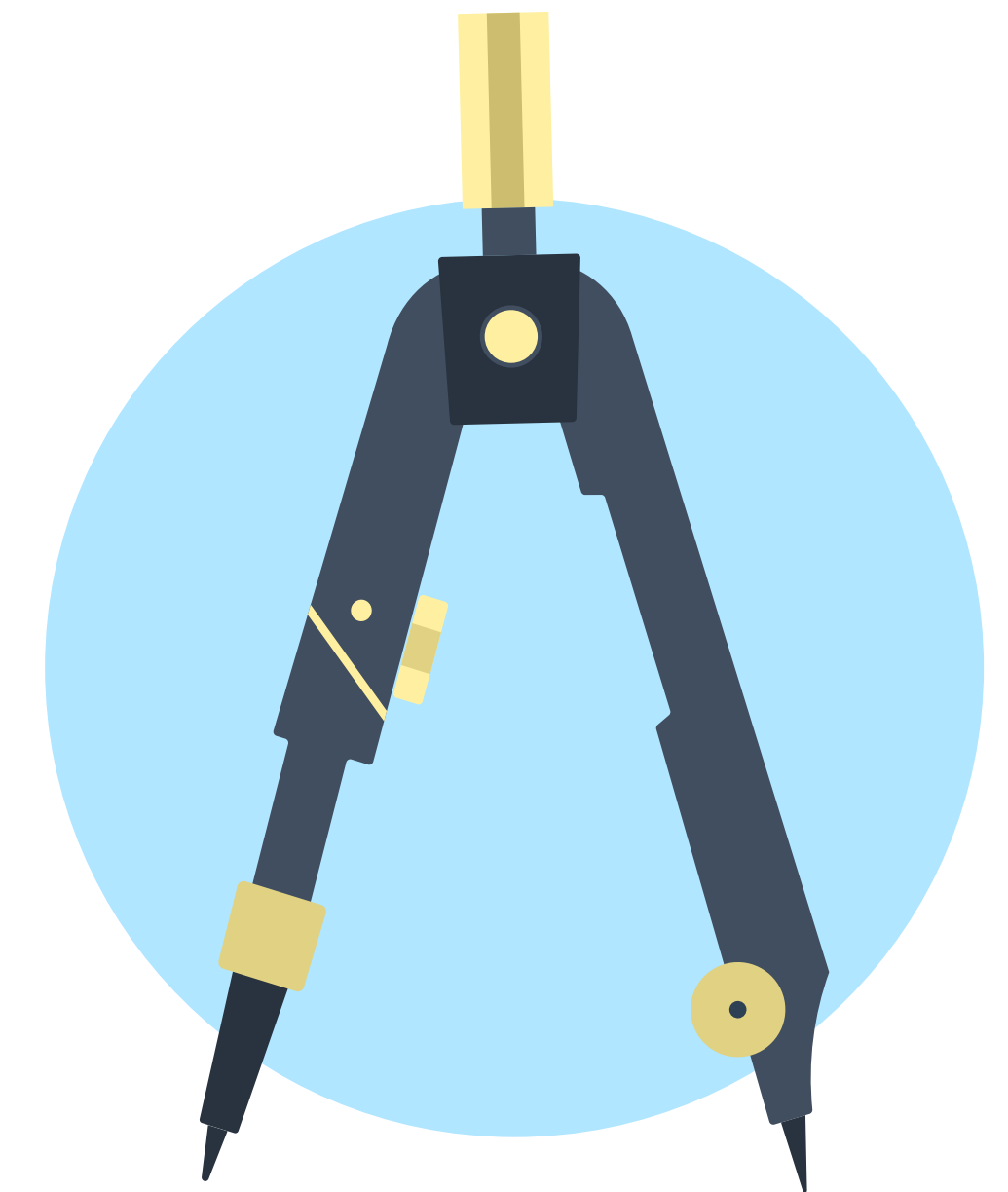

… enable **DATA ACCESS** instead

2

DATA STEWARDS do not just wrangle data …

… they **CREATE VALUE** for downstream data users

3

Many **THREATS TO DATA** that we fixate on …

… can be side-stepped with the right **ARCHITECTURE**