TCAIREM Infrastructure and the Health Data Nexus

March 22nd, 2022

Temerty Centre for AI Research and Education in Medicine
University of Toronto

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Purpose

- **Provide an overview of T-CAIREM**
- **Describe the T-CAIREM Health Data Nexus**
- **Discuss future directions**

Background

- The Temerty Family provided a groundbreaking gift to the University of Toronto that provides the funding for: The Temerty Centre for Artificial Intelligence Research and Education in Medicine (T-CAIREM) is an interdepartmental Centre that serves as a focal point for collaboration
  - Members include healthcare providers, trainees, researchers, computer scientists, engineers, and industry

- T-CAIREM has four main goals:
  - Foster multidisciplinary, collaborative research in artificial intelligence in medical and health sciences and encourage clinical translation
  - Establish a leading education program in applied artificial intelligence for medical and health science professionals and learners
  - Establish a robust health data environment enabling timely access to high quality health data to fuel innovation and quality improvement and support education programs.
  - Create a multidisciplinary, inclusive community of engaged members, whose common goal is to advance healthcare through AI.

- There are three T-CAIREM themes: Research, Education, and Infrastructure

- To accomplish these goals, the Infrastructure theme is building the Health Data Nexus

T-CAIREM: Looking to the Future… Together

Muhammad Mamdani
T-CAIREM Director

Laura Rosella
T-CAIREM Education Lead

Felipe Morgado   Vinyas Harish
T-CAIREM Education Student co-Leads

Education

Research

Infrastructure

Anna Goldenberg   Mjaye Mazwi
T-CAIREM Research co-Leads

Alistair Johnson   Sean Hill
T-CAIREM Infrastructure co-Leads

https://tcairem.utoronto.ca/

https://twitter.com/UofT_TCAIREM

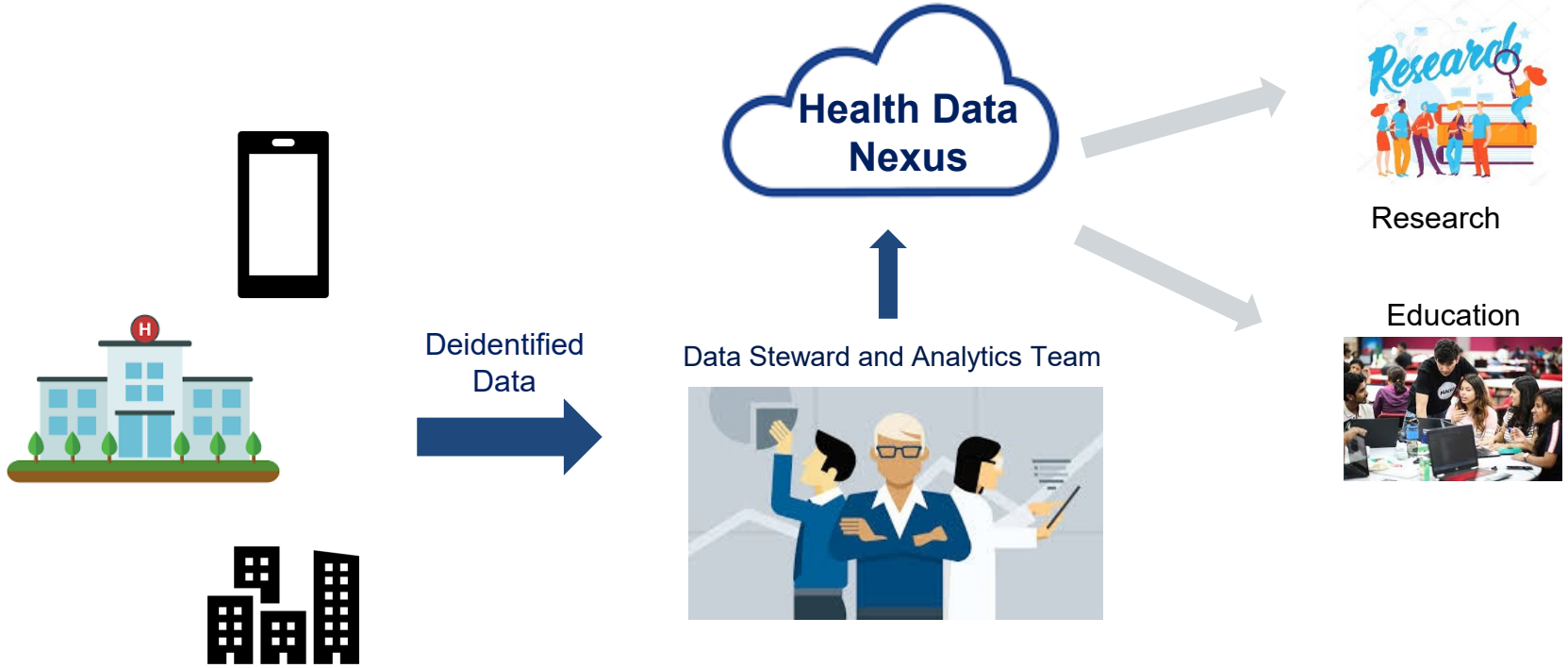https://www.linkedin.com/company/tcairem

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Infrastructure Theme Activities

Infrastructure Theme Activities

Health Data Nexus

➤ **Authorized users**
➤ **Streamlined data governance**
➤ **Scalable compute and storage**
➤ **Data stewardship**

Cloud-based data platform:
➤ Safe and secure controlled environment
➤ Hosted on Google Cloud Platform
➤ Tools for searching datasets
➤ Dashboards for visualizing datasets
➤ Integrated analytics environment (JupyterLab, RStudio) for conducting analysis without data exfiltration

➤ Initial "flagship" datasets available (pending final approval):
- St. Michael's Hospital datasets
  - General Internal Medicine
  - Medical Imaging in Cervical Spine Trauma
- Trillium Health Partners COVID-19 dataset
- MIMIC-IV clinical dataset
- Paediatric Intensive Care (PIC) clinical dataset

Infrastructure Theme Activities
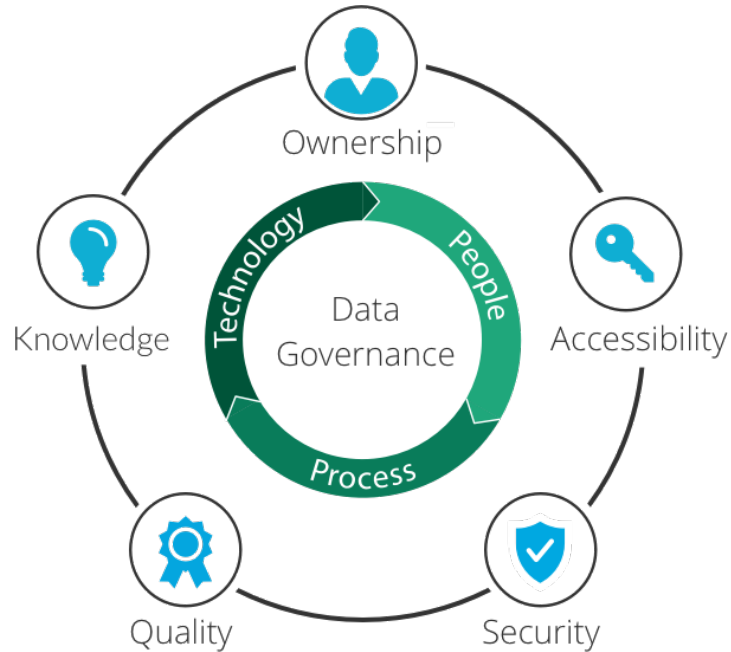
**Building Data Governance Framework and Establishing Best Practices**

➢ Creating a Data Governance Committee
➢ Data Sharing Agreement for contributors
➢ Authorization process for users
 ➢ Data Use Agreement
 ➢ Code of conduct
 ➢ Research training
➢ Zoned data access

Health Data Nexus Technical Infrastructure

➢ The platform consists of a front end for uploading and accessing the data and back-end for creating the analysis platforms
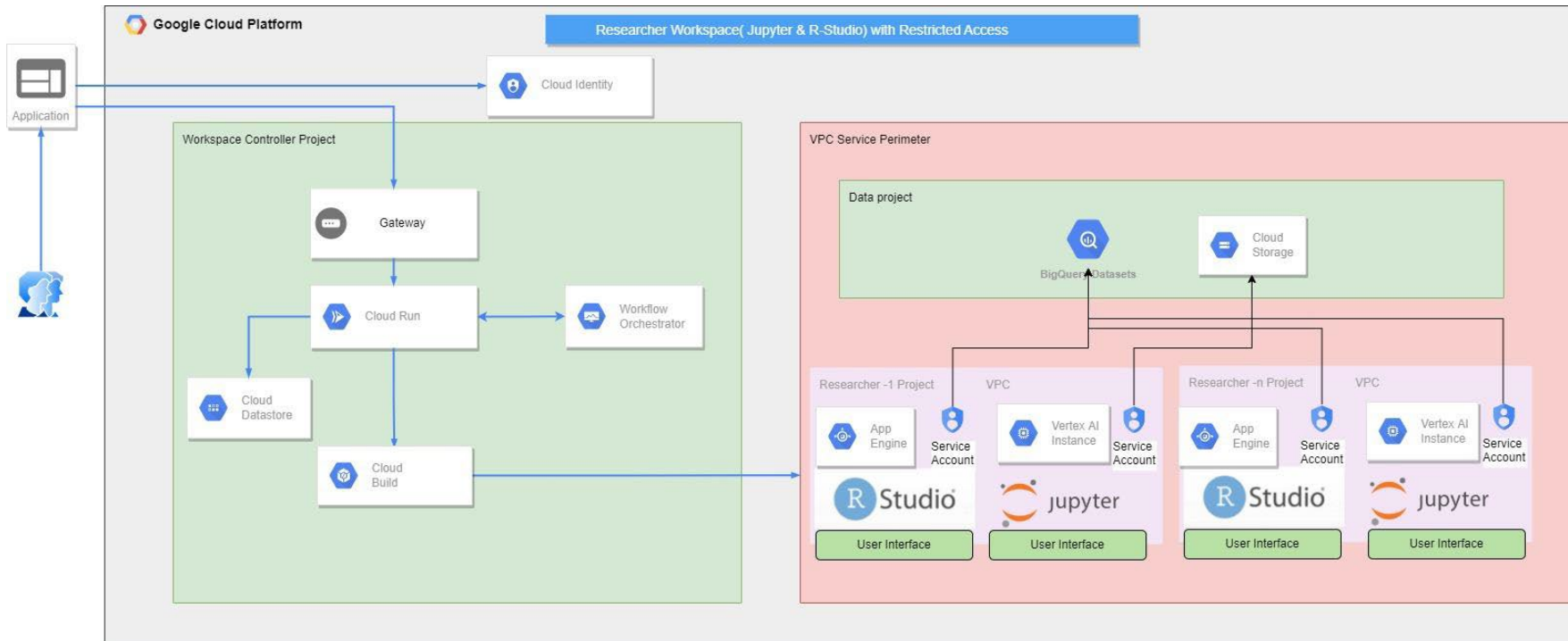
Front End:

➢ Based on the MIT-developed Physionet data platform

➢ Consists of a Django app (containerized in Docker) deployed on Kubernetes using Terraform templates

➢ Handles data uploading, access, and credentialing
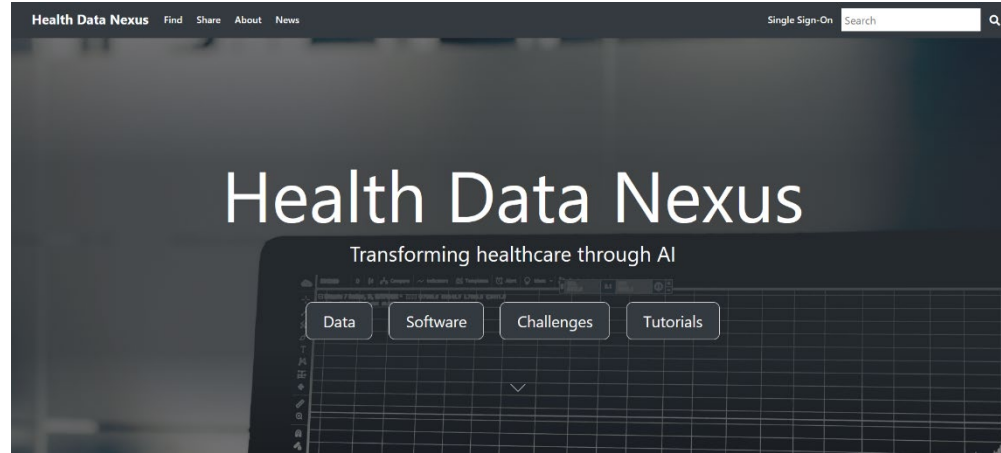
Health Data Nexus Technical Infrastructure

Back End:

Health Data Nexus Technical Infrastructure

Back End:

➢ Consists of a sequence of Google Cloud Platform API calls

➢ User identity is managed through Cloud Identity

➢ Data is stored in secure GCS buckets

➢ Secure analysis platforms (JupyterLabs and RStudio) are spun up as needed using Terraform templates built with Cloud Run and operating on Vertex AI or App Engine

➢ Data cannot be exfiltrated from the analysis platform

➢ Datasets cannot be linked (each workspace is associated with its own dataset)

Infrastructure Theme Future Directions
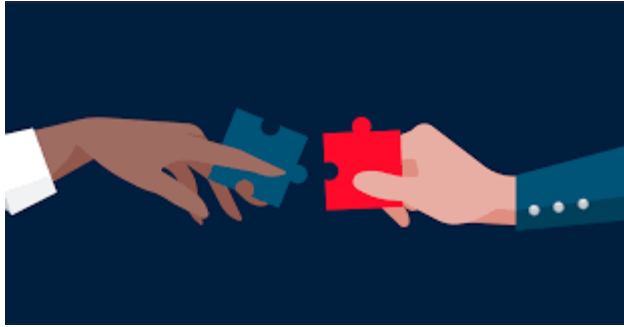
Establishing New Partnerships:

➤ Working to acquire new datasets (two new datasets each year):
  - GEMINI
  - UTOPIAN
  - National Trauma Databank

➤ Collaborate with Canadian Federated Research Data Repository
➤ Adopt standards for data use (e.g. Fast Healthcare Interoperability Resources (FHIR) for Electronic Health Record Data)
➤ Build additional tools to enhance user experience (improved analytics platform, journal integration)

Infrastructure Theme Future Directions

Partnership with Research and Education Themes:

➢ Health Data Nexus Utilization Grants: $10K to $50K.

➢ Enable training on real data through the platform: courses, workshops and datathon.