Data Science and Engineering in the Real World

Kassy Raymond, MSc, PhD(c)
Data Engineer | Data & Analytics, Wellington-Dufferin-Guelph Public Health, Ontario, Canada
University of Guelph | Global Burden of Animal Diseases (GBADs)

Agenda

- Path from research to data engineering
- Real-world projects in One Health
- Lessons on collaboration

Academic Path

BSc: Biological Sciences

University of Guelph, ON, Canada

2014-2019

MSc: Bioinformatics

School of Computer Science,
University of Guelph

2019-2021

PhD: Computational Sciences

School of Computer Science,
University of Guelph

2021-present

Career Path

Research Assistant
Ontario Ministry of Agriculture, Food, and Rural Affairs
2017

Machine Learning Co-Lead
iris | University of Guelph
2020-2021

Technical Manager
Global Burden of Animal Diseases (GBADs)
2021-2024

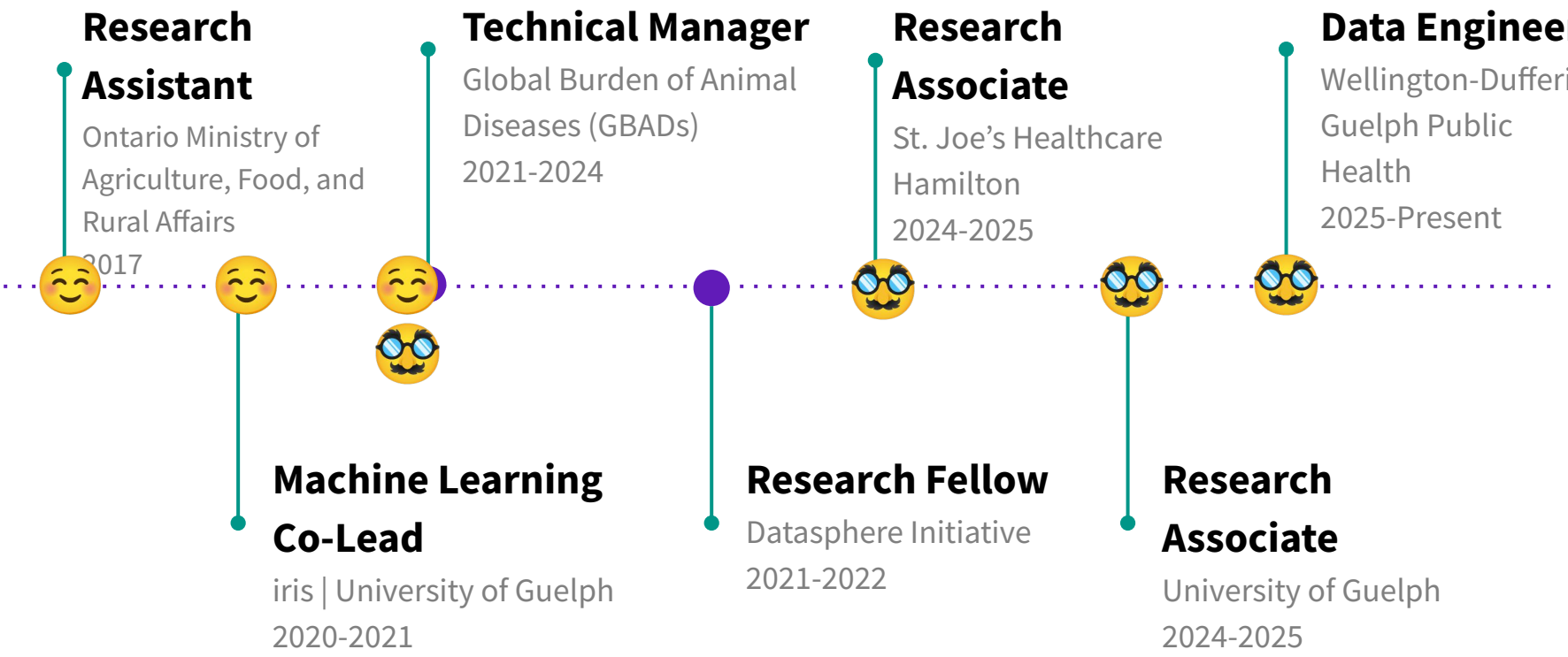
Research Fellow
Datasphere Initiative
2021-2022

Research Associate
St. Joe's Healthcare Hamilton
2024-2025

Research Associate
University of Guelph
2024-2025

Data Engineer
Wellington-Dufferin-Guelph Public Health
2025-Present

Career Path


Research Assistant
Ontario Ministry of Agriculture, Food, and Rural Affairs
2017
Machine Learning Co-Lead
iris | University of Guelph
2020-2021
Technical Manager
Global Burden of Animal Diseases (GBADs)
2021-2024
Research Fellow
Datasphere Initiative
2021-2022
Research Associate
St. Joe's Healthcare Hamilton
2024-2025
Research Associate
University of Guelph
2024-2025
Data Engineer
Wellington-Dufferin-Guelph Public Health
2025-Present

Data Science vs Data Engineering

Data Science

Analyzes and interprets data to extract insights and build predictive models

Works with **statistics, machine learning, and visualization** tools

Answers ***“What does this data tell us?”***

Delivers insights, reports, and models.

Data Engineering

Data Science vs Data Engineering

Data Science

Analyzes and interprets data to extract insights and build predictive models

Works with **statistics, machine learning, and visualization** tools

Answers ***"What does this data tell us?"***

Delivers insights, reports, and models.

Data Engineering

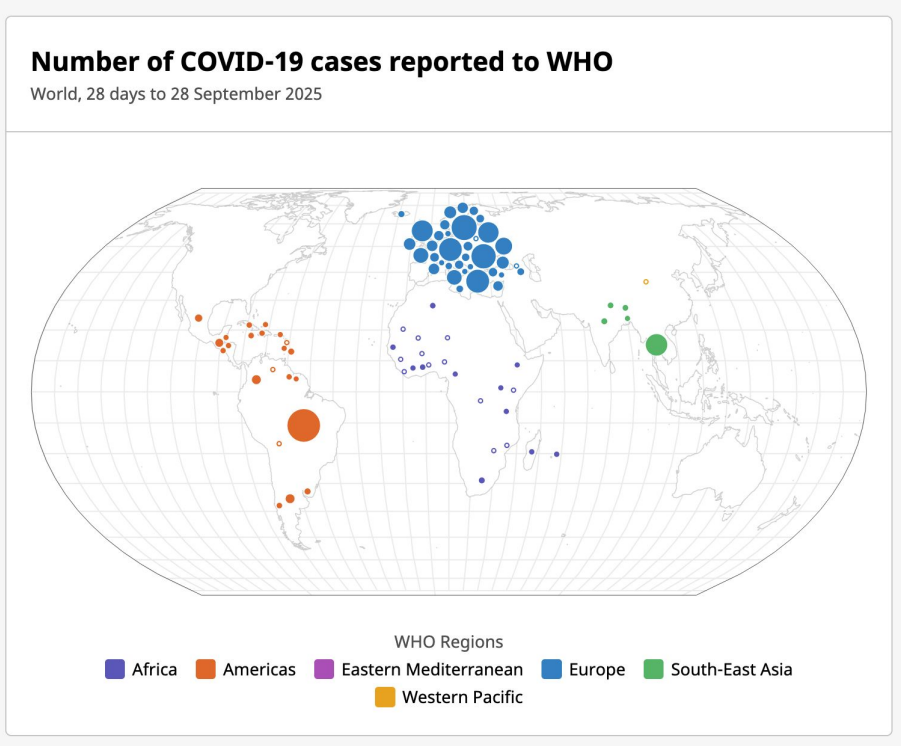
Builds **and maintains systems** that make data accessible, clean, and reliable.

Works with **data pipelines, databases, and cloud infrastructure**

Answers ***"How do we get this data where it needs to go?"***

Delivers data architecture and pipelines.

COVID-19 Dashboards



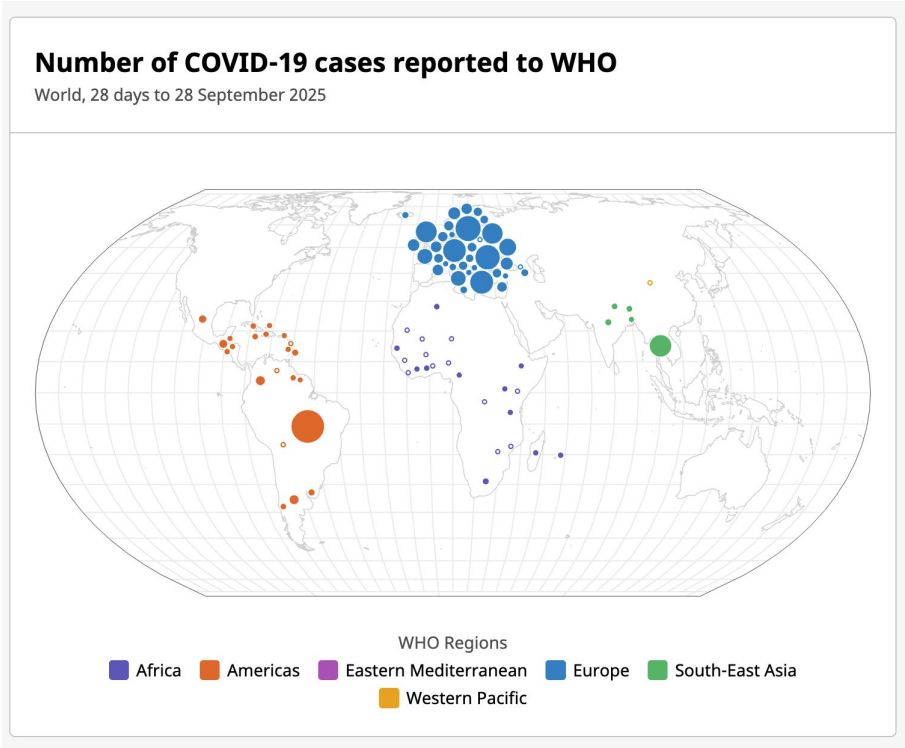
Data Engineering:

- Create data pipelines to pull case counts from hospitals or labs
- Cleaned data
- Data storage in databases

Prevent garbage in - garbage out!

https://data.who.int/dashboards/covid19/cases

COVID-19 Dashboards



Data Engineering:

- Create data pipelines to pull case counts from hospitals or labs
- Cleaned data
- Data storage in databases

Data Science:

- Forecast models
- Dashboards
- Data visualization

https://data.who.int/dashboards/covid19/cases

What is a Data Engineer?

Responsibility	What it means	Example
Collecting data	Bringing information together from many sources	Combining files from hospitals, schools, or surveys
Cleaning and organizing	Making sure data is accurate, consistent, and follows the same rules	Fixing typos, standardizing dates, removing duplicates
Building pipelines	Automating how data moves from place to place	Developing code and scripts that fetch, clean, and store data
Storing data	Choosing the best way to save data for easy and secure access	Designing cloud storage solutions and databases
Monitoring data quality	Checking for missing or unusual values	Logging errors and alerting when things go wrong
Collaboration	Working with analysts, epidemiologists, and IT staff	Helping them access clean, trustworthy data

Agenda

- Path from research to data engineering
- Real-world projects in One Health
- Lessons on collaboration

Agenda

Real-world projects in One Health

1. Active Surveillance and Immunization Notice Automation at WDGPH
2. FAIR Data for the Global Burden of Animal Diseases (GBADs)
3. Thermal Imaging for Metabolic Signal Re-creation

Vaccine Mandates: Brief History of Policy

Vaccine mandates have been part of public health for **over two centuries**

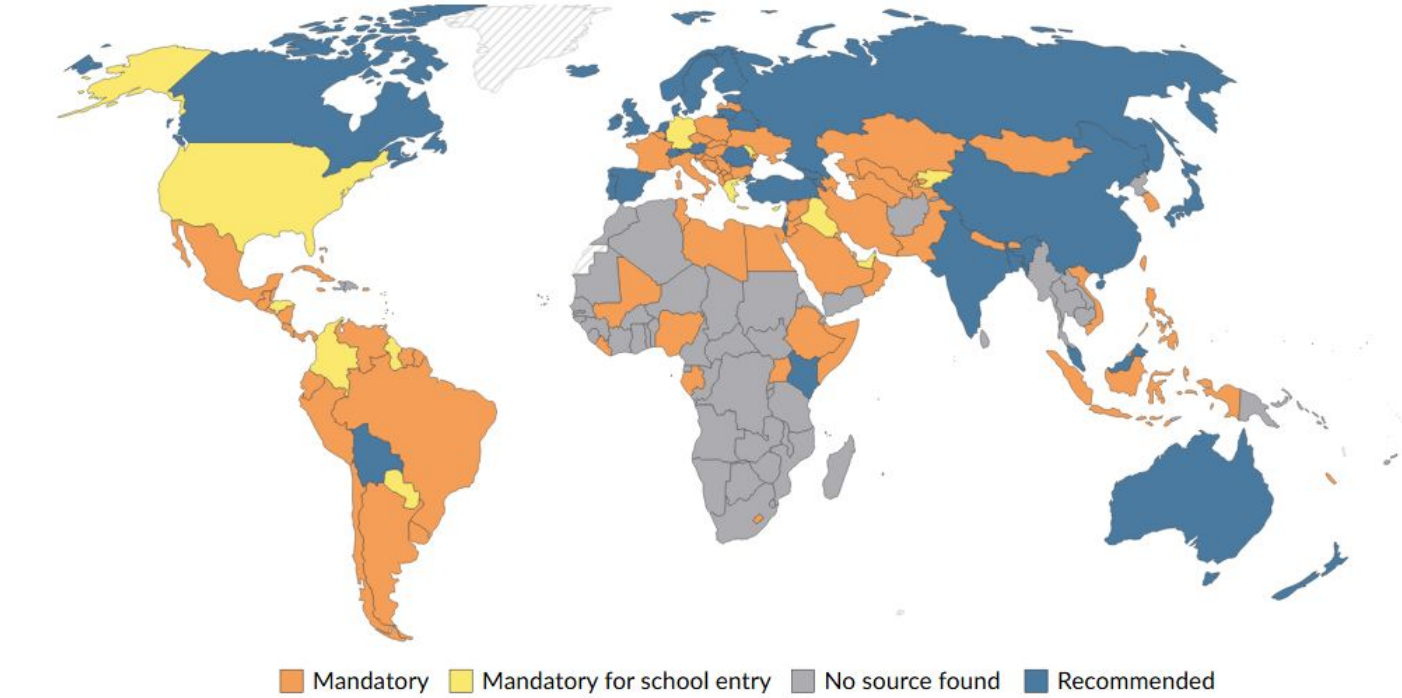
- Smallpox
  - 1777 - US Continental Army troops,
  - 1853 - England Vaccination Act - smallpox vaccine mandatory for infants and penalties for non-compliance,
  - 1855 - Schoolchildren in Boston
  - 1922 - US Supreme Court - schools may exclude unvaxed children, even outside of outbreaks

Which countries have mandatory childhood vaccination policies?

Countries are mapped based on having requirements or recommendations for at least one vaccine in 2019.

Our World in Data

Table | Map | Zoom to... | 2D | 3D | Select countries



Data source: Vanderslott & Marks (2021). Charting mandatory childhood vaccination policies worldwide. Vaccine. – Learn more about this data

Note: Policies can vary at the state level in some countries.

OurWorldinData.org/vaccination | CC BY

Vaccination Enforcement in Ontario

Report immunizations to your local public health unit

Under the *Immunization of School Pupils Act* (ISPA), you are responsible for reporting your children's vaccinations to your local public health unit so your child can attend school in Ontario. Report all immunizations as soon as you can to avoid possible disruptions to your child's school attendance.

Vaccination Enforcement in Ontario

Report immunizations to your local public health unit

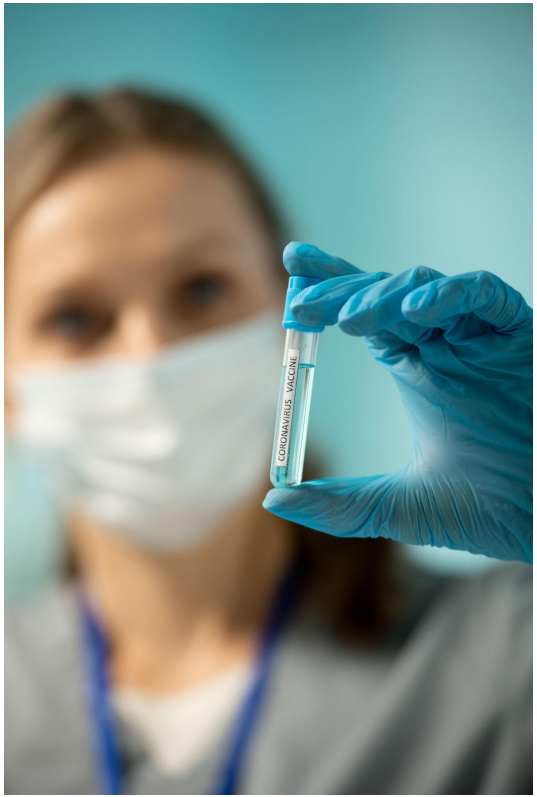
Under the *Immunization of School Pupils Act* (ISPA), you are responsible for reporting your children's vaccinations to your local public health unit so your child can attend school in Ontario. Report all immunizations as soon as you can to avoid possible disruptions to your child's school attendance.

ISPA diseases include:
diphtheria, measles, mumps, poliomyelitis, rubella, tetanus, and any other diseases prescribed by the Minister of Health and Long-Term Care

Managing Immunization Data

- Public Health Units **manage student vaccination data**
- Children who **do not have a complete immunization record** may be **suspended** from school until they provide **proof of vaccination** or **an exemption**

Children overdue for diseases are sent letters; parents mandated to submit proof of vaccination.

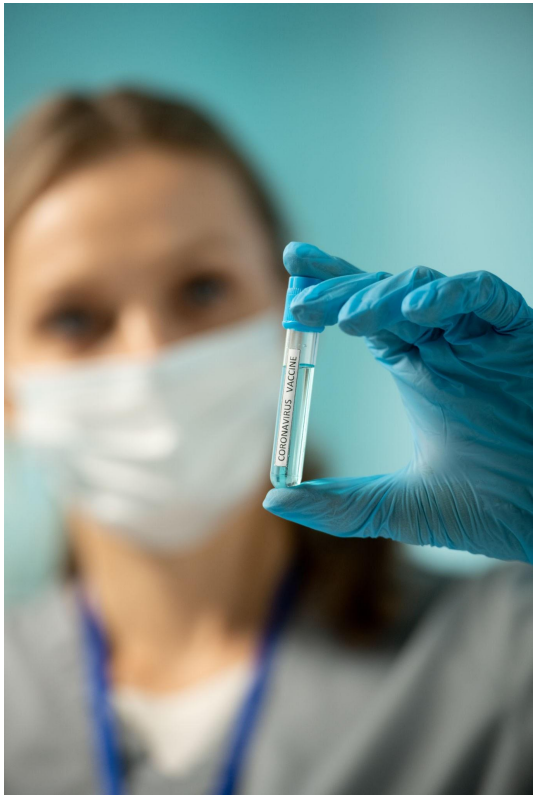


Managing Immunization Data

- Public Health Units **manage student vaccination data**
- Children who **do not have a complete immunization record** may be **suspended** from school until they provide **proof of vaccination** or **an exemption**

Children overdue for diseases are sent letters; parents mandated to submit proof of vaccination.

Active Surveillance: actively reviewing, collecting, and updating records for every student enrolled in our jurisdiction



Vaccine Notice Automation at WDGPH

Workflow Challenge:

- Notice generation is time consuming and tedious
- Legacy tools do not support customization of letter format

Our Solution

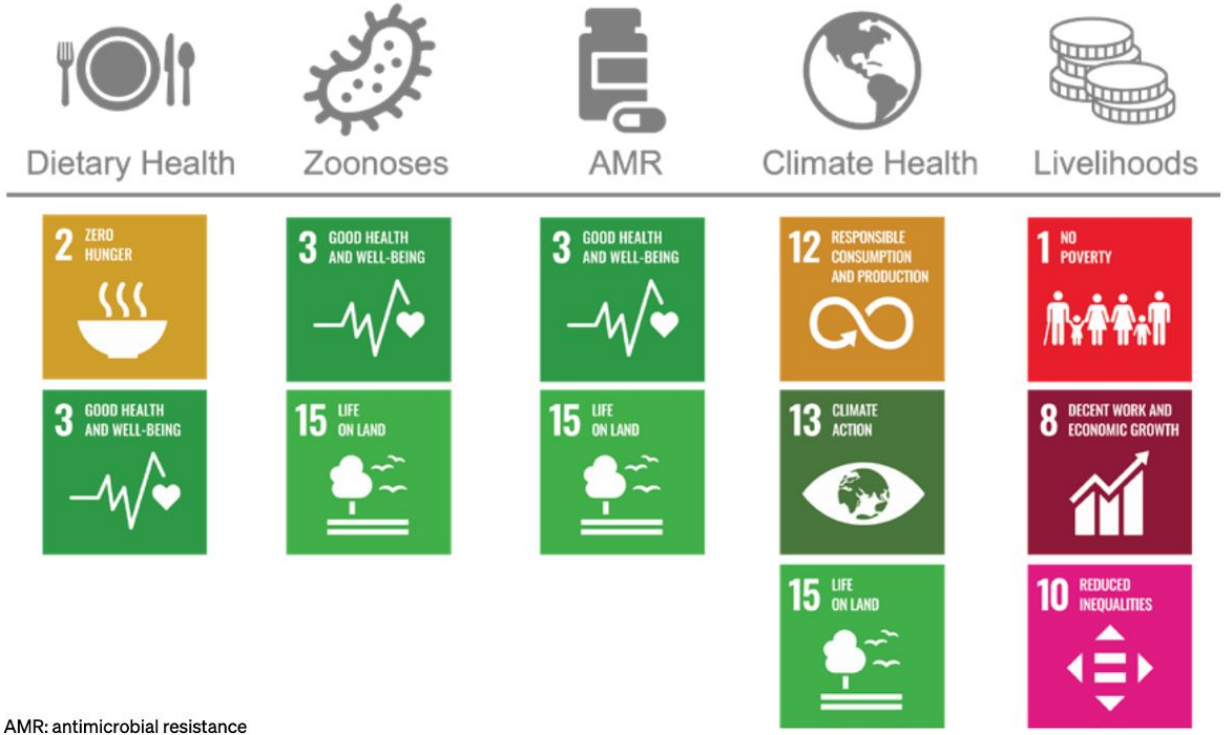
- Automated pipeline for generation of ISPA and measles notices
- Open-Source
- Ability to generate ~50,000 notices/hr

Agenda

Real-world projects in One Health

1. Active Surveillance and Immunization Notice Automation at WDGPH
2. **FAIR Data for the Global Burden of Animal Diseases (GBADs)**
3. Thermal Imaging for Metabolic Signal Re-creation

GBADs generates estimates of the social and economic burden of animal diseases



AMR: antimicrobial resistance

Figure 2

The Global Burden of Animal Diseases (GBADs) programme and the United Nations Sustainable Development Goals

The Sustainable Development Goals [37] to which the GBADs outputs contribute

Raymond K, BenSassi N, Patterson GT, et al. Informatics progress of the Global Burden of Animal Diseases programme towards data for One Health. Revue Scientifique et Technique (International Office of Epizootics). 2023 May;42:218-229. DOI: 10.20506/rst.42.3365. PMID: 37232302.

Importance of Livestock Population Data

Intergovernmental Organizations

FAO FIAT PANIS

eurostat

World Organisation for Animal Health
Founded as OIE

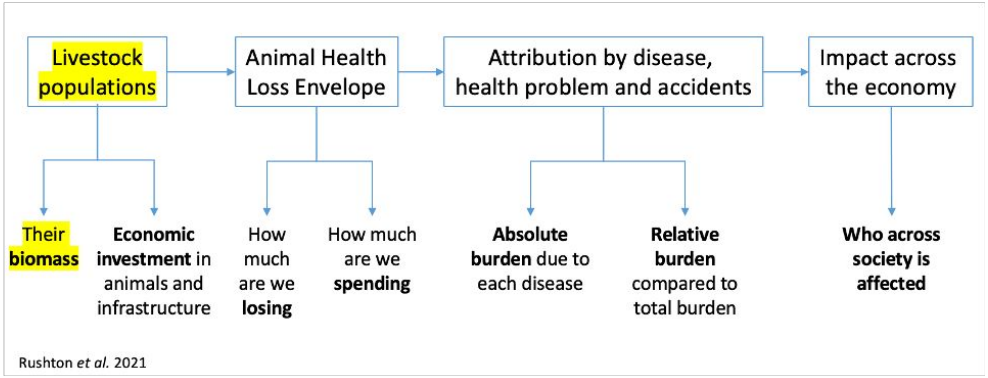
National Statistics Organizations

CENTRAL STATISTICS AGENCY
ኢትዮጵያ

USDA

Statistics Canada

Statistique Canada



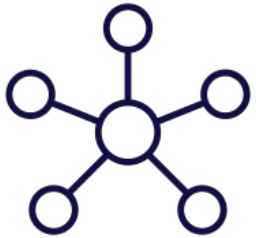
Challenges with Livestock Population Data

- No central access point for data
- Lost time spent looking for data
- Metadata not provided or without level of detail needed
  - Species
  - Age/sex/production system

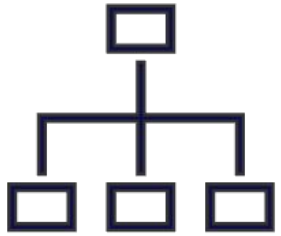
Objectives

1. Requirements for interoperability

2. Improve access and discovery of livestock data



3. Develop a library of livestock classifications



Graph Databases

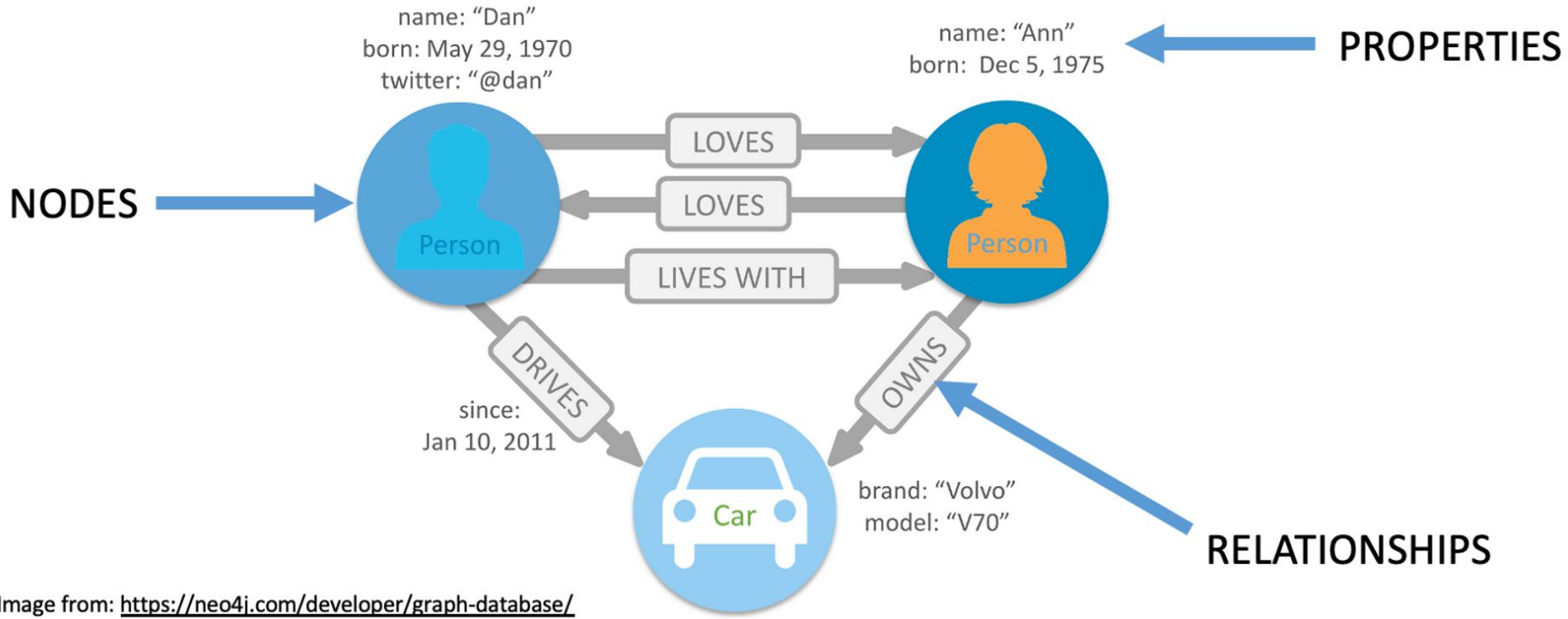


Image from: https://neo4j.com/developer/graph-database/

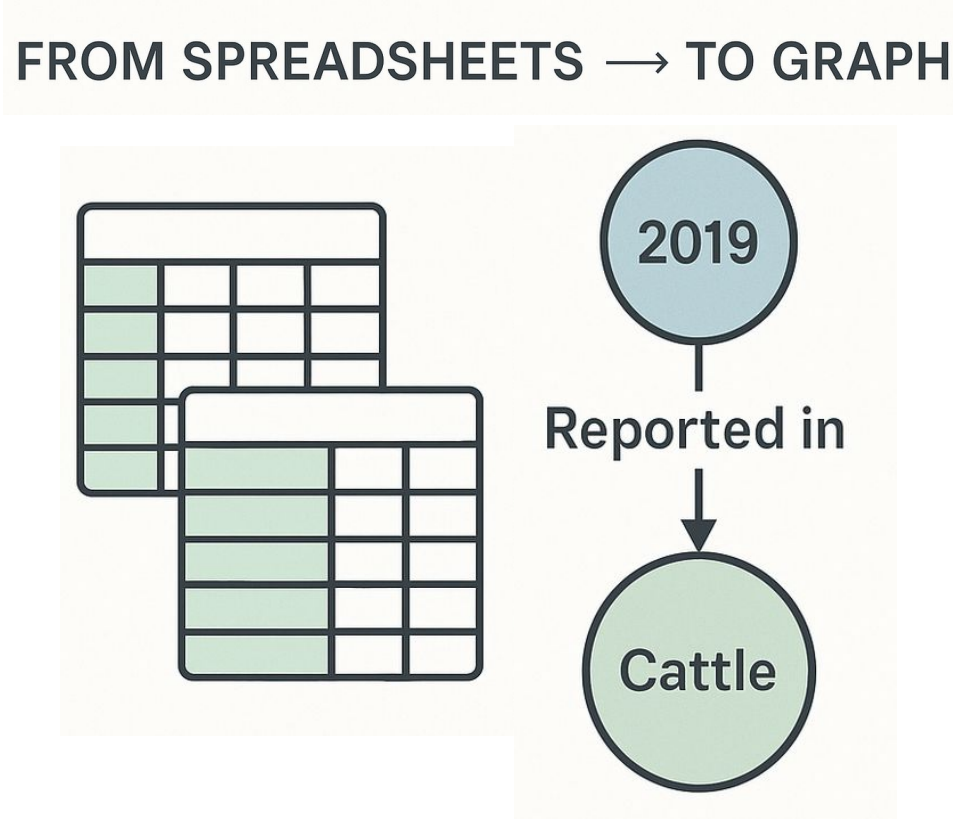
Data Engineering Solution

Collect: Livestock data gathered from multiple countries and agencies (e.g., FAO, national surveys).

Clean and Standardize: Python scripts extract metadata, fix inconsistencies, and align fields like *species*, *year*, and *country.*

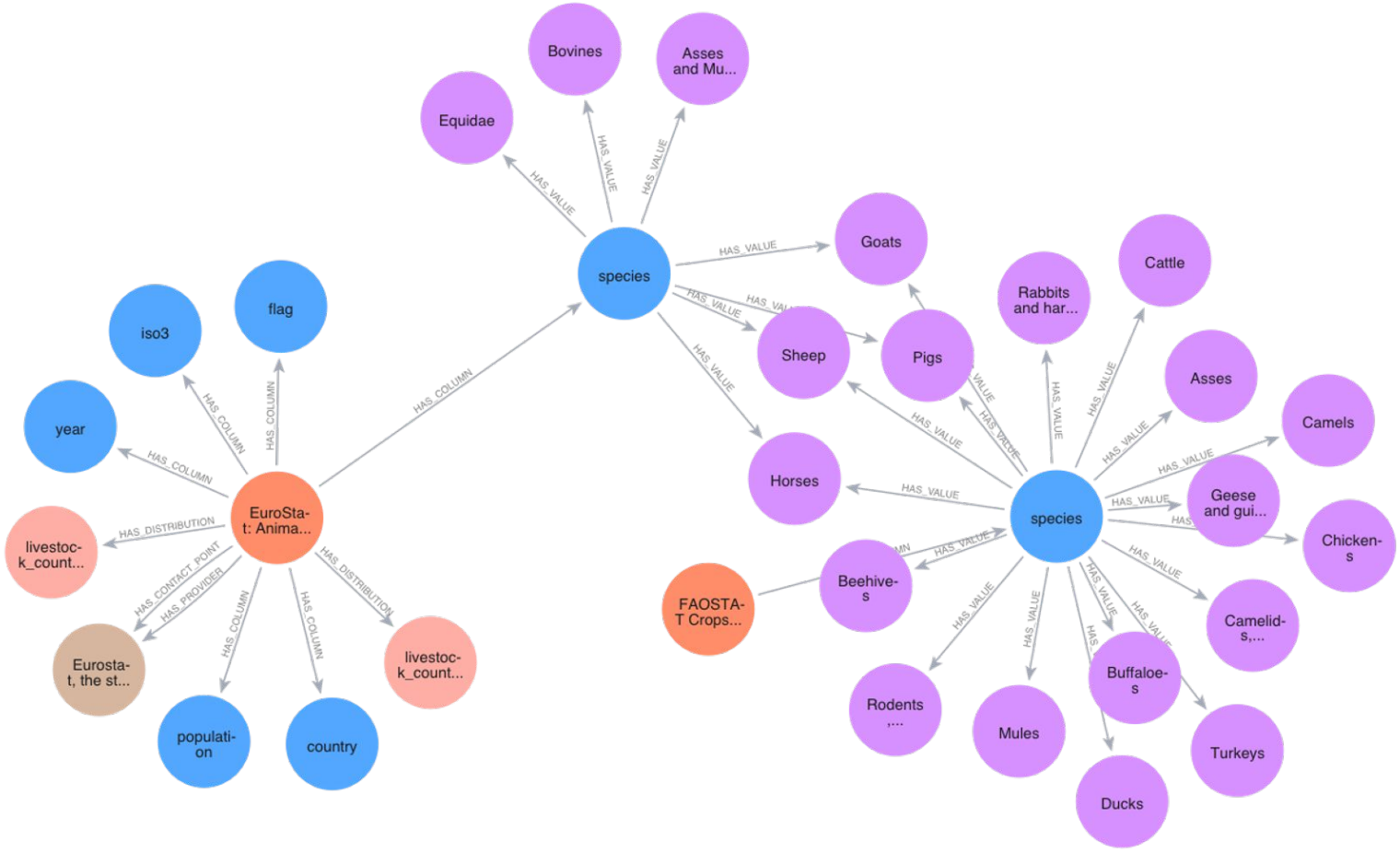
Connect & Query: Data loaded into a Neo4j graph, showing relationships like:

Cattle → Reported in → Ethiopia → 2019.



Supported Queries

1. Get a list of all species and species categories used for reporting data
2. Get a list of all countries reporting data
3. Return datasets that report a given a specie(s), countrie(s)


Bovines
Asses and Mu...
Equidae
HAS_VALUE
HAS_VALUE
HAS_VALUE
species
HAS_VALUE
Goats
HAS_VALUE
HAS_VALUE
Sheep
Pigs
HAS_VALUE
HAS_VALUE
Horses
HAS_VALUE
iso3
flag
HAS_COLUMN
HAS_COLUMN
HAS_COLUMN
year
HAS_COLUMN
EuroSta-t: Anima...
HAS_DISTRIBUTION
livestoc-k_count...
HAS_CONTACT_POINT
HAS_PROVIDER
Eurosta-t, the st...
HAS_COLUMN
HAS_COLUMN
HAS_DISTRIBUTION
populati-on
country
livestoc-k_count...
Rabbits and har...
Cattle
HAS_VALUE
HAS_VALUE
Asses
HAS_VALUE
Camels
HAS_VALUE
Geese and gui...
HAS_VALUE
Chicken-s
HAS_VALUE
species
HAS_VALUE
Camelid-s,...
HAS_VALUE
HAS_VALUE
Buffaloe-s
HAS_VALUE
Turkeys
HAS_VALUE
Ducks
HAS_VALUE
Mules
HAS_VALUE
Rodents ,...
HAS_VALUE
Beehive-s
FAOSTA-T Crops...

Summary

Researchers and policymakers couldn’t easily find or compare datasets, even when they described the same animals or countries.

- Users can **discover datasets faster**, even across institutions.
- The system reveals **relationships and gaps** in global livestock reporting.
- Enables **FAIR data principles**: Findable, Accessible, Interoperable, and Reusable.

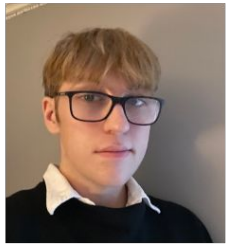
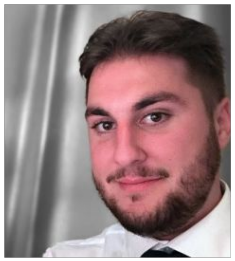
GBADs Informatics Team

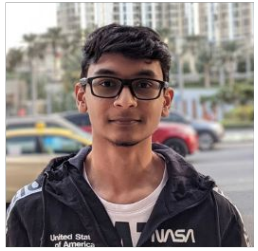










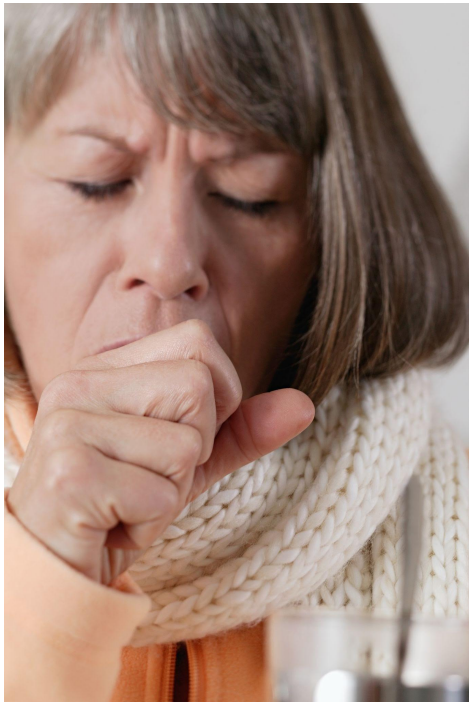

Agenda

Real-world projects in One Health

1. Active Surveillance and Immunization Notice Automation at WDGPH
2. FAIR Data for the Global Burden of Animal Diseases (GBADs)
3. **Thermal Imaging for Metabolic Signal Re-creation**

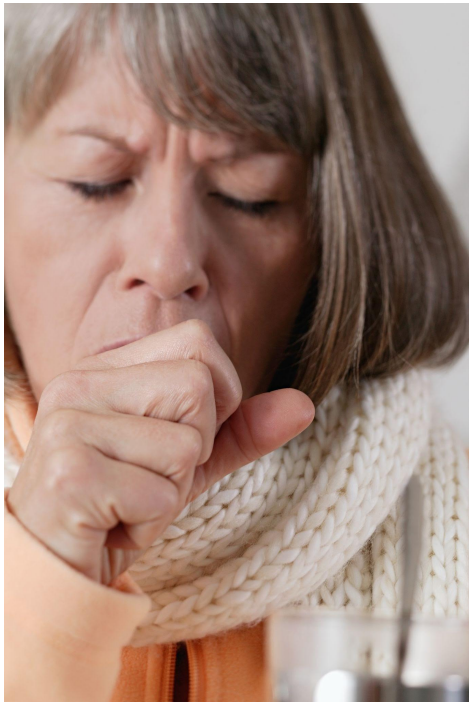
Respiratory Infections in Long-Term Care Homes

- Older adults in long-term care homes (nursing homes) vulnerable to **Respiratory Syncytial Virus (RSV)** and **COVID-19**



Respiratory Infections in Long-Term Care Homes

- Older adults in long-term care homes (nursing homes) vulnerable to **Respiratory Syncytial Virus (RSV)** and **COVID-19**
- **Symptoms of RSV:** Runny nose, fever, cough, short periods without breathing (apnea), wheezing, breathing faster than normal, trouble breathing



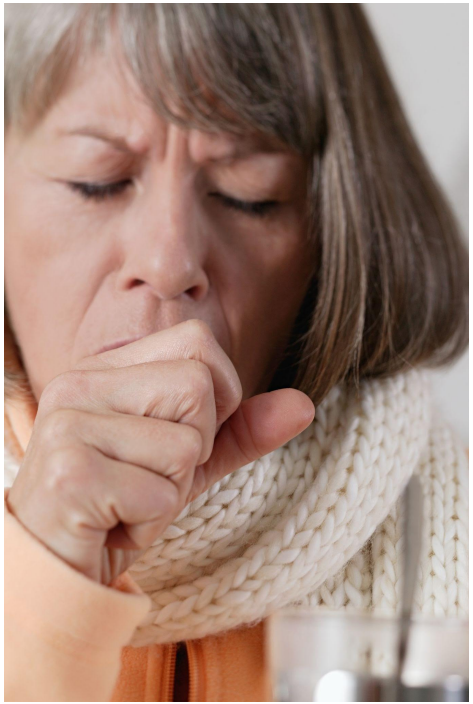
Respiratory Infections in Long-Term Care Homes

- Older adults in long-term care homes (nursing homes) vulnerable to **Respiratory Syncytial Virus (RSV)** and **COVID-19**

Table 2a: Confirmed outbreaks in LTCHs among residents, by pathogen, Ontario: August 27, 2023 – March 2, 2024

Measure	COVID-19	Influenza	RSV
Total number of cases among residents	**17,747**	**1,885**	**697**
Median outbreak attack rate (IQR)	25.0% (12.5-41.6%)	14.7% (9.0-22.6%)	15.4% (9.5-24.0%)
Total number of cases hospitalized	**558**	**156**	**30**
Median hospitalization rate (IQR)	4.0% (0.0-11.1%)	0.0% (0.0-11.3%)	0.0% (0.0-0.0%)
Total number of deaths	**396**	**58**	**13**
Median case fatality rate (IQR)	2.5% (0.0-6.7%)	0.0% (0.0-0.0%)	0.0% (0.0-0.0%)

IQR: Interquartile Range; --: Refers to where data are not reported or are not presented due to low counts.



Infection Prevention and Control of RSV

Infection Prevention and Control (IPAC) refers to the set of **practices**, **policies**, and **systems** designed to prevent the spread of infections among residents, staff, and visitors

Infection Prevention and Control of RSV

Infection Prevention and Control (IPAC) refers to the set of **practices**, **policies**, and **systems** designed to prevent the spread of infections among residents, staff, and visitors

Goals:

1. **Prevent infections before they occur**
    - Vaccines, hand washing, environmental cleaning, and **surveillance**
2. **Identify and control outbreaks quickly**
    - Detect unusual illness patterns early to reduce spread
3. **Protect high-risk populations**
    - Minimize exposure for residents with weaker immune systems or particular care needs
4. **Ensure compliance with provincial and federal public-health standards**

Surveillance in Long-Term Care Homes

Surveillance → watching for infections before they are spread

Traditional Methods - Staff Monitoring for RSV symptoms:

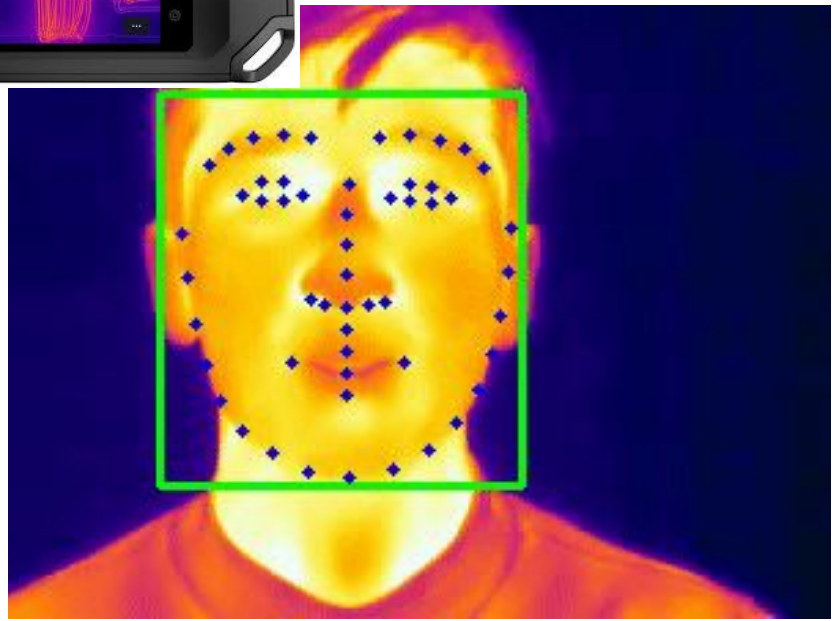
- Fever, cough, shortness of breath
- Changes in appetite, or alertness
- Gastrointestinal illness
- Skin changes

Surveillance in Long-Term Care Homes

Traditional Methods - Staff Monitoring for RSV symptoms limitations:

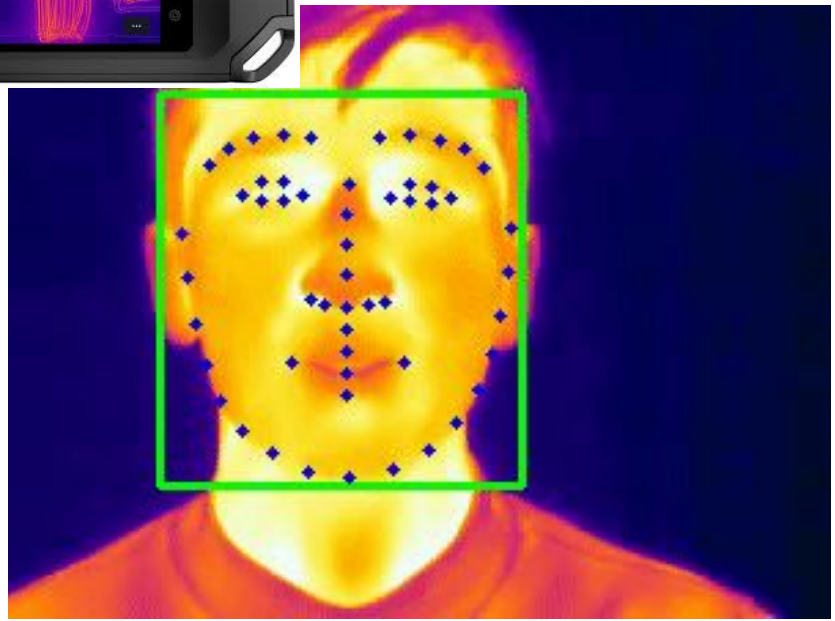
- Reliance on visible symptoms
- Infrequent checks
- Resource intensive
- Limited data and documentation

Thermal Imaging for Long-Term Care Homes



Thermal imaging is a non-contact technology that detects and visualizes temperature changes on objects and people

Thermal Imaging for Long-Term Care Homes

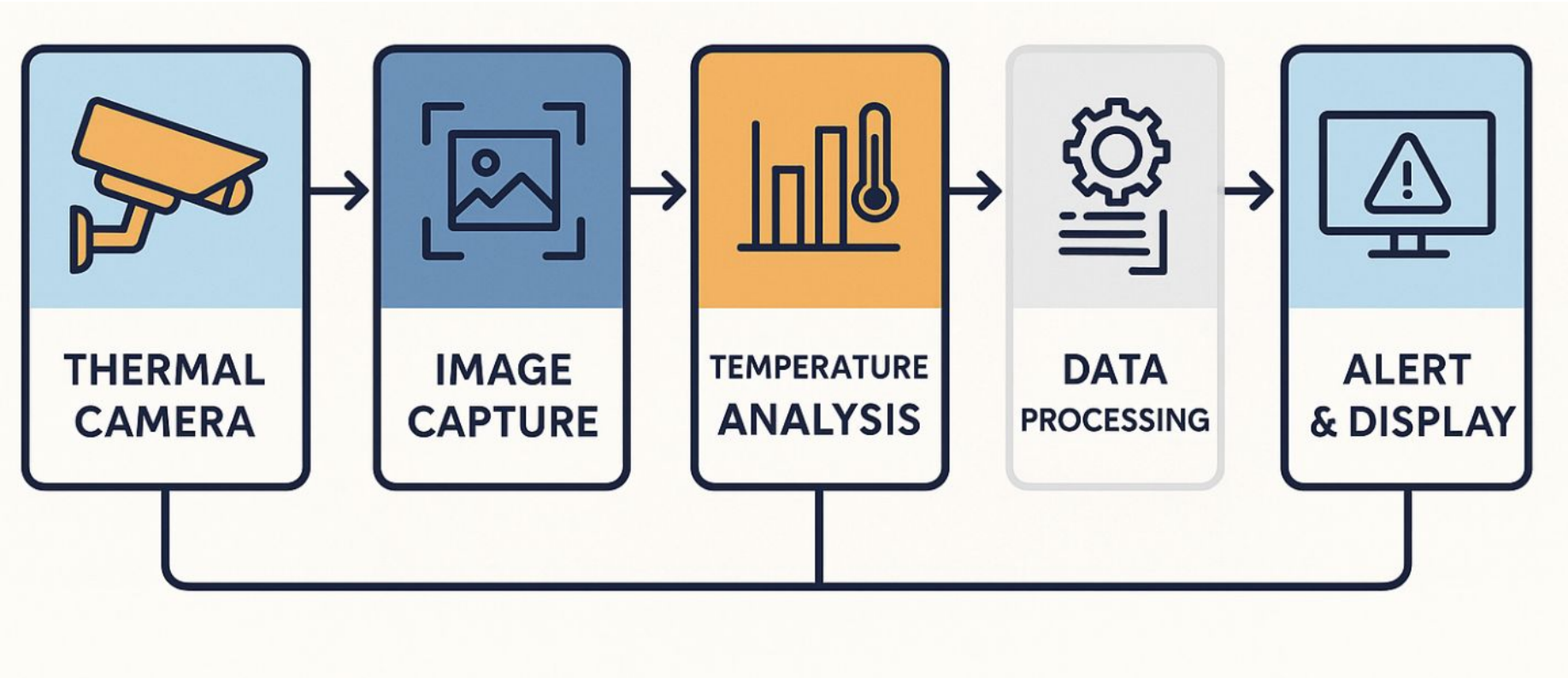


Thermal imaging is a non-contact technology that detects and visualizes temperature changes on objects and people

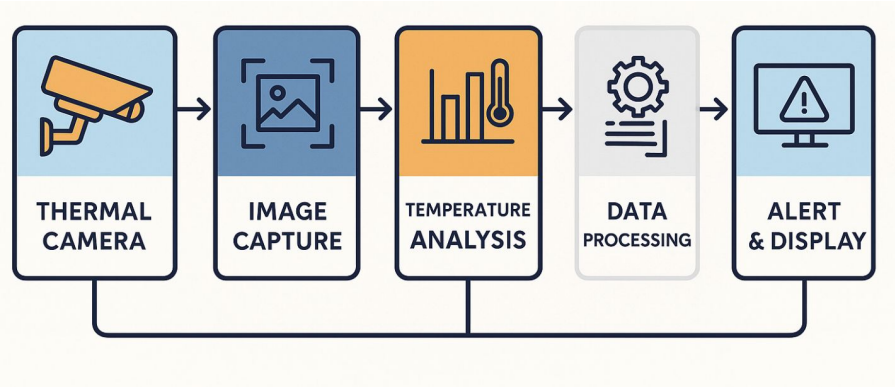
Benefits:

- Early detection of illness
- Continuous non-invasive monitoring
- Detect infectious individuals early

Thermal Imaging for Surveillance



Data Engineering and Data Science



Data Engineering:

- Capture and store data
- Clean and organize
- Deliver data to modelling tools

Data Science:

- Algorithms detecting breath, temperature, heart rate
- System flags anomalies (elevated temp or resp. rate)

AIR-ALERT Team

- **Dr. Andrew Hamilton-Wright** (University of Guelph, School of Computer Science, Canada)
- **Dr. Dan Perri** (St. Joesph's Hospital, Hamilton, Canada)
- **Sarah Culgin** (Project Manager; St. Joesph's Hospital, Hamilton, Canada)
- **Kelly Waters** (St. Joesph's Hospital, Hamilton, Canada)
- **Kassy Raymond** (University of Guelph, School of Computer Science, Canada)
- **Sayana Varughese** (University of Guelph, School of Computer Science, Canada)
- **Cathrine Nayrouz** (University of Guelph, School of Computer Science, Canada)

UNIVERSITY of GUELPH

St. Joseph's Healthcare Hamilton

Agenda

- Path from research to data engineering
- Real-world projects in One Health
- Lessons on collaboration

Lessons on Collaboration

1. Understanding different perspectives
    - Understand data workflows and work with domain-experts
2. Speak different languages
3. Respecting expertise

Lessons on Collaboration

1. Understanding different perspectives
2. Speak different languages
    - Understanding different terms across disciplines (e.g. spreadsheet vs. database)
3. Respecting expertise

Lessons on Collaboration

1. Understanding different perspectives
2. Speak different languages
3. Respecting expertise
    - Everyone brings important expertise when it comes to data!
    - Listen and ask questions

Contact

Kassy Raymond
kassyraymond2@gmail.com
kassyray.github.io