OM1
Interview Talk

March 30, 2022

Tim Farrell

Product Manager

Day Zero Diagnostics (DZD)

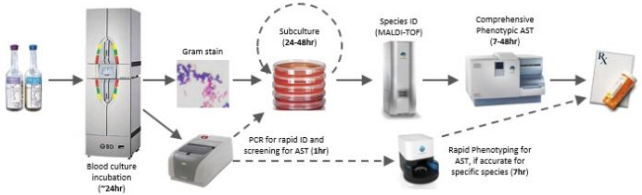
Agenda

1. **DZD**
2. **MicrohmDB**: Data platform for genomics + ML
3. **epiXact**: Outbreak detection for health systems

Day Zero Diagnostics (DZD)

The Future of Microbiology is Digital

Analog, culture-based diagnostics



Slow	**Days to weeks** to diagnose
Low Resolution	Fails to grow 40% of the time, requires multiple assays for full diagnosis
Limited	Optimized for core diagnostic use-cases

Digital, sequencing-based diagnostics

Fast	Answers within **hours**
High Resolution	High sensitivity, full genomic interrogation in a single diagnostic
Comprehensive	Capable of comprehensive organisms, HAI & epidemiology, forecasting

Day Zero Diagnostics (DZD)

Advanced ML Algorithms

Complete diagnosis from genomic data

Keynome® ID
Comprehensive, accurate pathogen detection

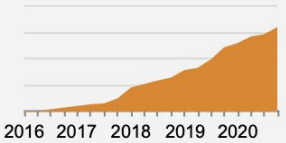
+

Keynome® *g*-AST
Comprehensive AST profiling directly from genomic data

MicrohmDB® proprietary dataset for ML Training

One of the world's largest dataset combining pathogen sequences & AMR profiles

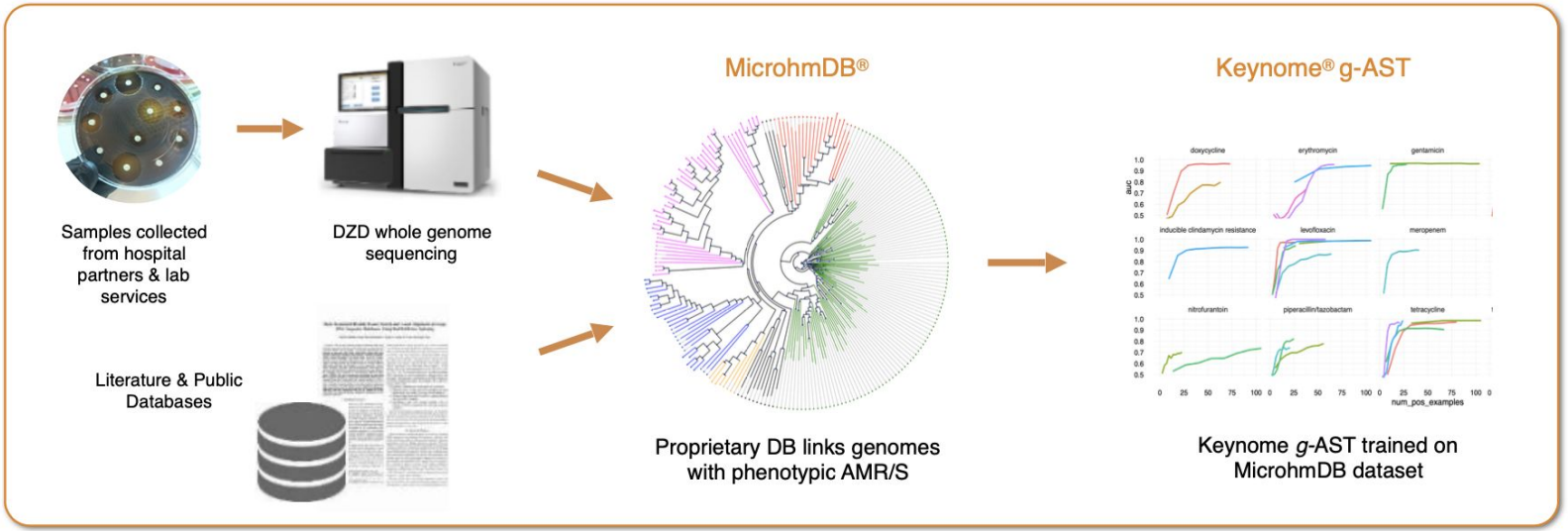
50,000+ Samples and Growing



Day Zero Diagnostics (DZD)

Keynome® g-AST

DZD Algorithms Are Trained on MicrohmDB for AST Calls



MicrohmDB: Cloud migration and versioned releases

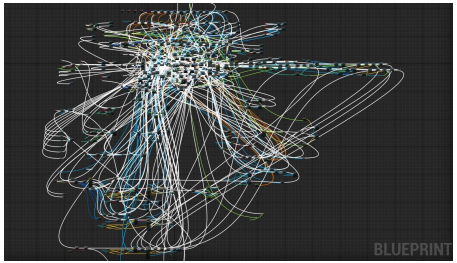
When I took this over in late 2019, it was poorly designed and technical debt was slowing data integration and sourcing..

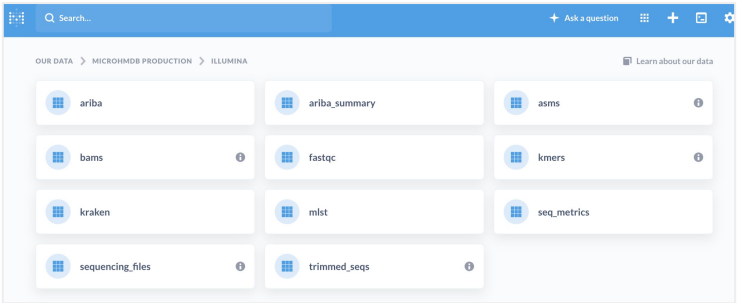
Legacy system

- Data not stored in cloud
- Compute not scalable
- Analysis not reproducible
- Data poorly modeled
- Data not versioned

MicrohmDB v1.0

- Data stored in cloud
- Scalable compute
- Analysis reproducible
- Simper, cleaner design
- Data releases quarterly

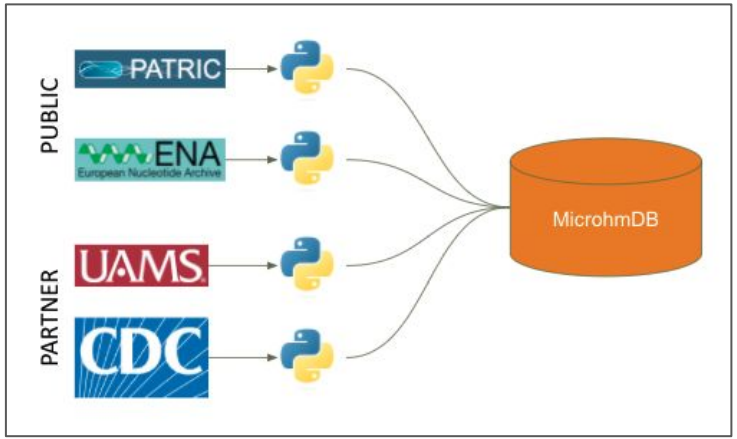




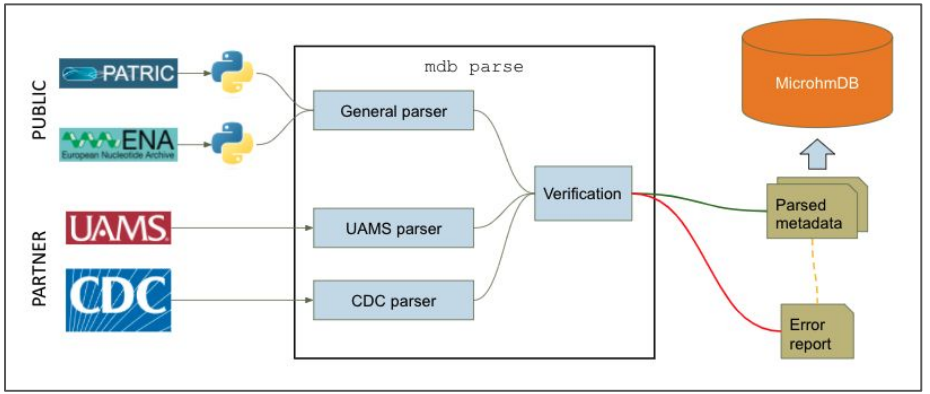
MicrohmDB: Data integration and QC

.. as we rapidly integrated new data sources, we also wanted to put data quality checks/ verification in place to ensure we were only including high quality data

Legacy system



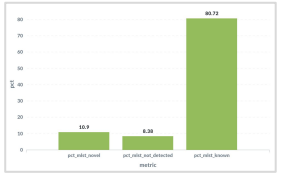
MicrohmDB v1.0

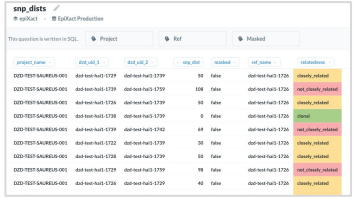


MicrohmDB: Data platform architecture

Partners and vendors can transfer sequencing data directly over FTP

Users can interact with data through BI dashboards, making plots and saving common queries





Sequencing data

Couchdrop

Users

Metabase

Sequencing data

mdb.upload

Google Cloud Storage

mdb.import

Postgres DB

mdb.process

Software tooling (`mdb`) that allows for data upload, data import into the database and data processing

nextflow

docker

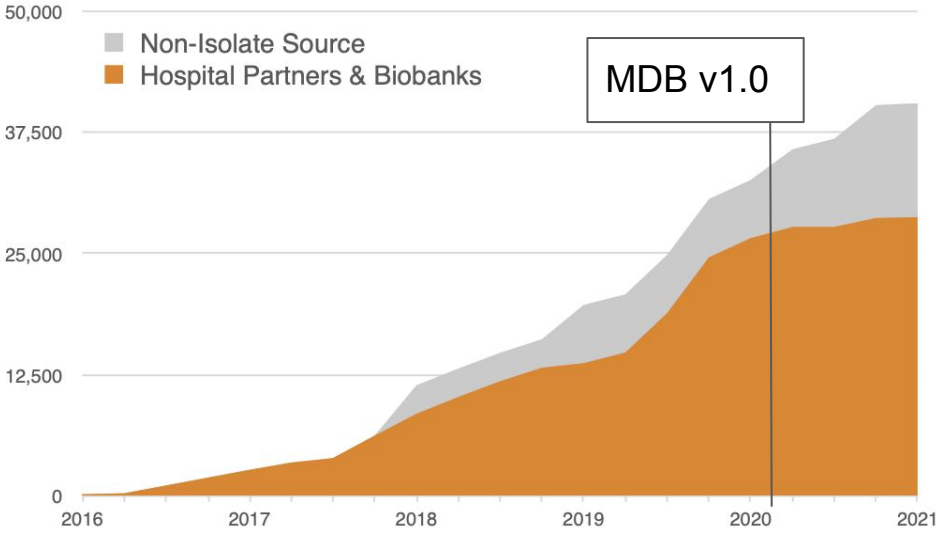
Nextflow pipelines–powered by Google Life Sciences Compute API–facilitate scalable, reproducible compute.

Bioinformaticians write modules within the pipelines and the platform executes them. They don't have to worry about scalability, etc.

MicrohmDB®

Large Scale Dataset of Pathogen Genomes and AMR Profiles

MicrohmDB: 50,000+ and growing



Customers, Contributors & Collaborators

MASSACHUSETTS GENERAL HOSPITAL

UAMS University of Arkansas for Medical Sciences

UCIrvine

BRIGHAM AND WOMEN'S HOSPITAL

ihma

NEW ENGLAND BAPTIST HOSPITAL

CANADIAN ANTIMICROBIAL RESISTANCE ALLIANCE CARA

ARLG Antibacterial Resistance Leadership Group

bei RESOURCES SUPPORTING INFECTIOUS DISEASE RESEARCH

New partners since 2019

MicrohmDB: Challenges + future directions

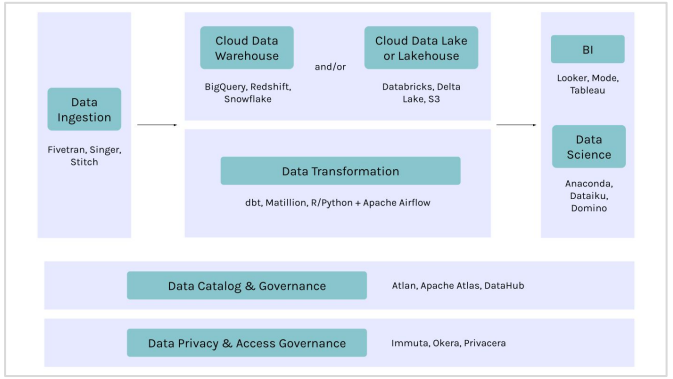
Data versioning is hard

- Use data lakehouse architecture, but data lakes are more conducive to AI R&D workflows and make data versioning easier
- Plan to fully transition to dake lake

Data cataloging and provenance is also hard

Plan to transition from ETL → ELT

- Increases development/ data velocity
- Data lake architecture helps this as well



Architecture of modern data platforms [Source]

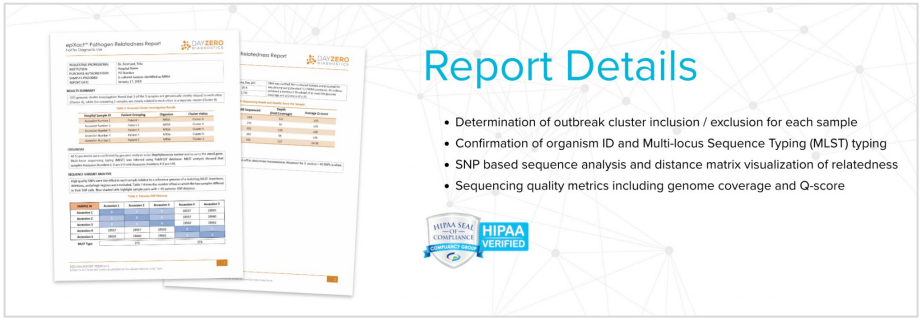
epiXact: Hospital-acquired infection (HAI) outbreak detection for health systems

- In Fall 2020, became technical and product lead for epiXact*, the company's first commercial offering
- Process:
  - Requested by email or on our website
  - Samples collected from patients suspected of HAI outbreak infection and sent to DZD
  - Sequencing => Analysis => Report

*in addition to MicrohmDB

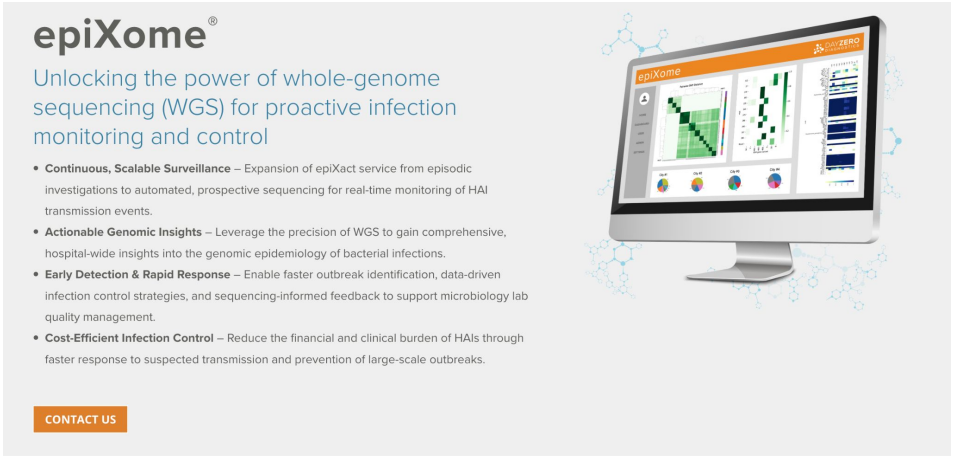
epiXact: Key results

- Co-led Clinical Laboratory Improvement Amendments (CLIA) certification with Lab Director, which enabled epiXact to be ordered as a lab test
- Automated analysis and reporting to shorten turnaround time from ~1 week to less than 48 hours, which became an SLA we promised customers
    - Speed was critical for our customers because each day increased likelihood that more patients would be infected by an outbreak
- Updated design of our report for CLIA certification and also to improve CX/ UX
- Projects/ deals grew from ~1 per quarter to ~3-4 per quarter



epiXome: New product development

- Worked with our BD team, Marketing and CEO to develop new commercial offerings built on top of epiXact technology
- Part of the business challenge with epiXact is that it was “spiky”, and we often struggled for customers to do regular/ repeat deals with us
- epiXome would be a continuous surveillance service for health systems to proactively detect and control for outbreaks
- I led the new product development process from concept/ design to launch



Thank you!