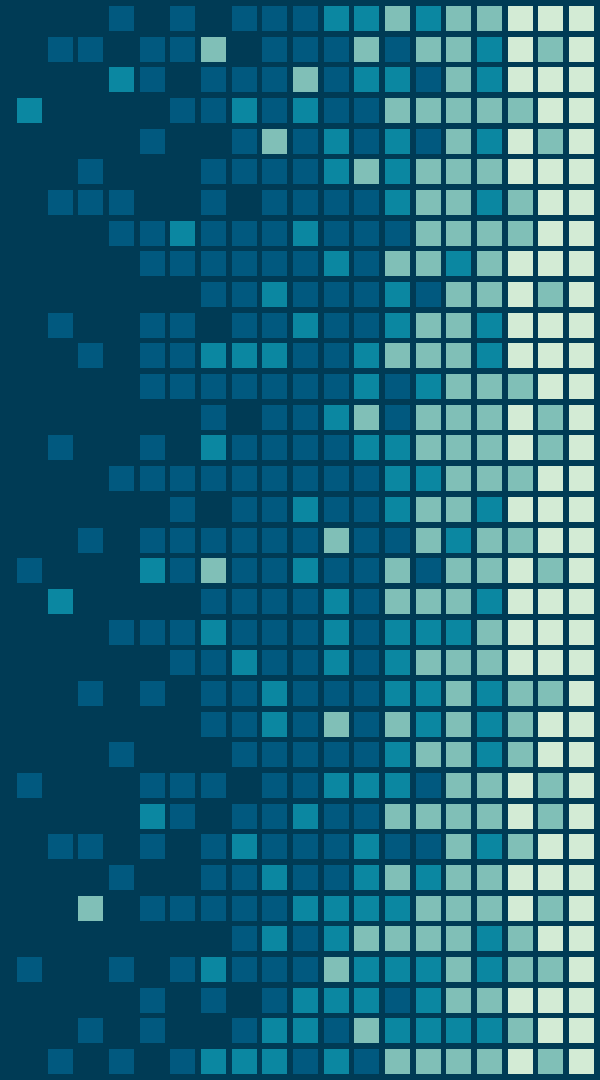


# Canadian Health Omics Repository, Distributed (CHORD)

David Bujold

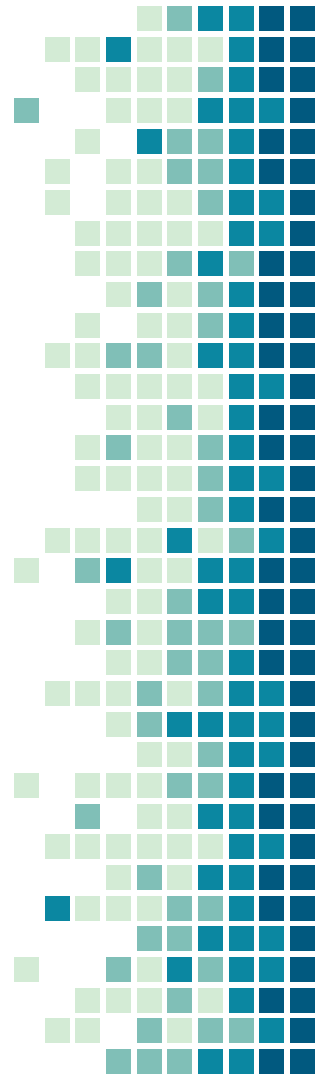
Canadian Centre for Computational Genomics

2019-01-24



# Sharing genomic data beyond a project's scope

- How can we share genomic datasets while ensure studies participants protection?
  - Enabling analysis of data from multiple studies
  - Taking into account the private nature of the genome
- Many of the existing solutions involve storing data in silos
  - E.g. Controlled Access Repositories
  - Obtaining and analyzing data difficult on large datasets



# CanDIG



- Funded 4 year CFI Cyberinfrastructure project
- Fully decentralized, distributed and federated
- Send analysis to the data
- Access to data through requests, either for data as it stands or for processing through some pipelines
- Local sites control access to their data
- Sites authenticate their users



# GA4GH 2017 Driver Projects



**All of Us Research Program**  
United States



**Australian Genomics**  
Australia



**BRCA Challenge**  
International



**CanDIG**  
Canada



**ClinGen**  
United States



**ELIXIR Beacon**  
Europe



**ENA / EVA / EGA**  
Europe



**Genomics England**  
United Kingdom



**Human Cell Atlas**  
International



**ICGC-ARGO**  
International



**Matchmaker Exchange**  
International



**Monarch Initiative**  
International



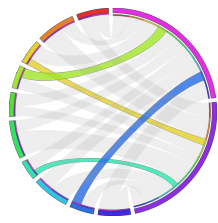
**National Cancer Institute  
Genomic Data Commons**  
United States

**TOPMed**  
United States

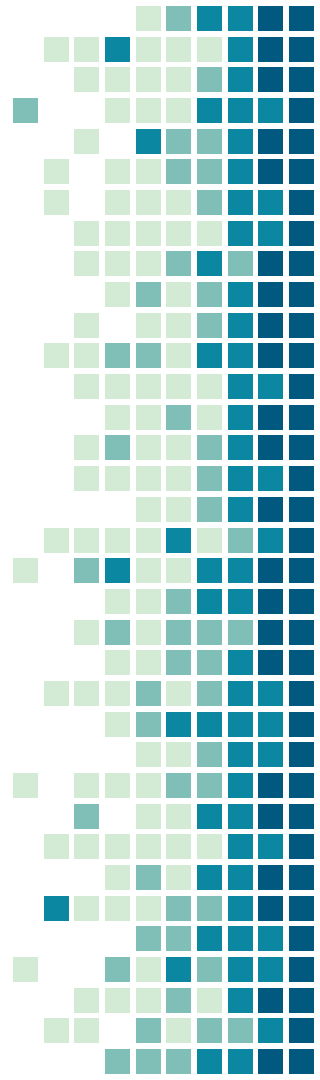


**Variant Interpretation  
for Cancer Consortium**  
International

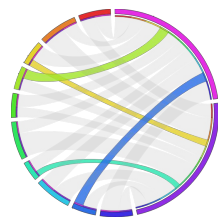
# CanDIG CHORD



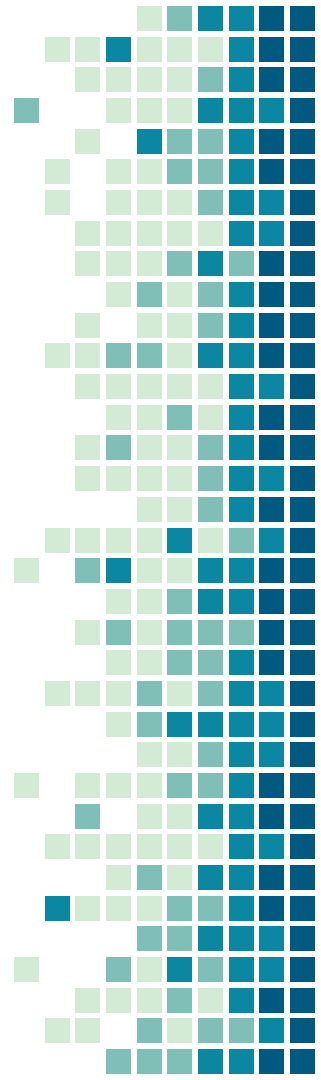
- CANARIE Research Data Management project
- Goal: Building software for a national data service for health genomic data
- Target community: Canadian researchers willing to make their genomic data available to the wider community



# CanDIG CHORD

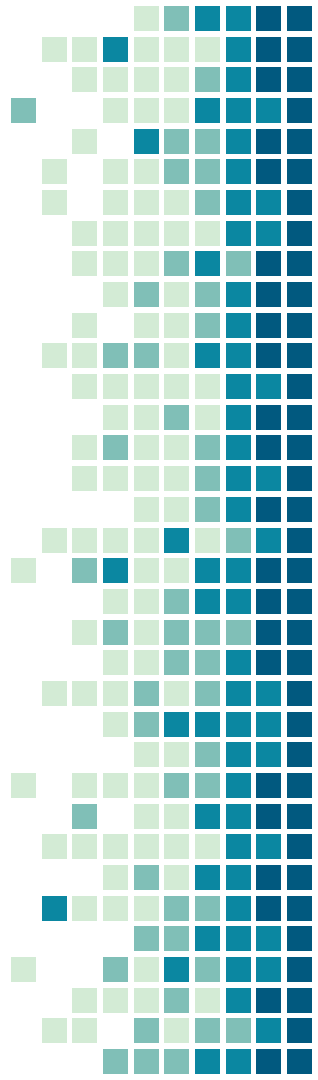


- Allow publishing genomic (meta)data, and make them available for analysis
- Enable searching for data
  - provide information about who to contact for authorization, when needed
- Support moving computation to private data, allowing analyses of sensitive data
  - combine intermediate results of computations between sites
- Facilitate the linking of data and literature by providing citable persistent identifiers

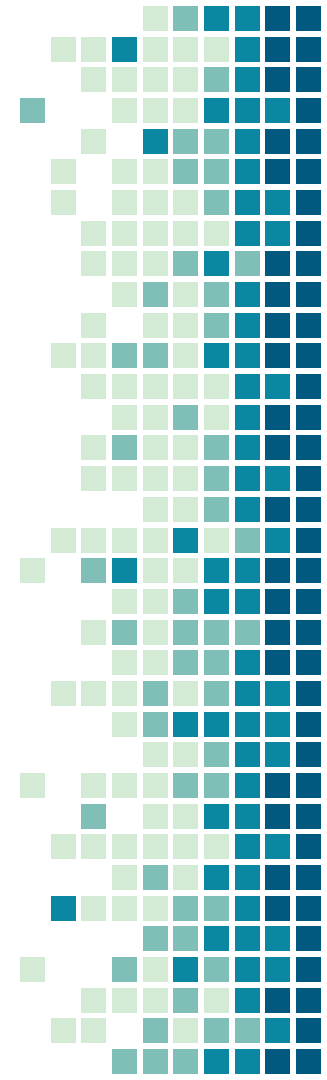
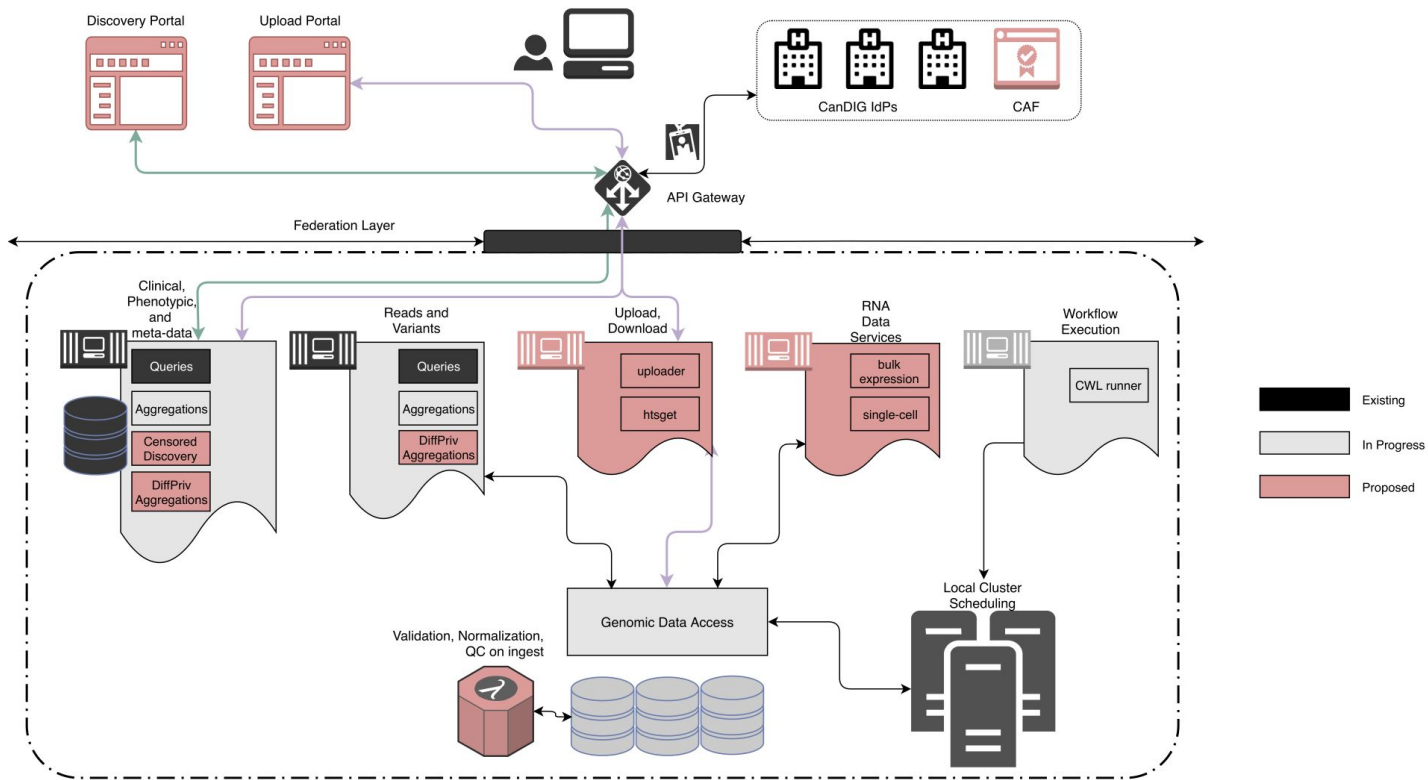


# CHORD deliverables

1. Data Publishing (upload, ingestion, UID)
2. Findability and Access (Data Portal)
3. Privacy-Preserving Reuse of Data
4. Expand CanDIG for better Health Data Support



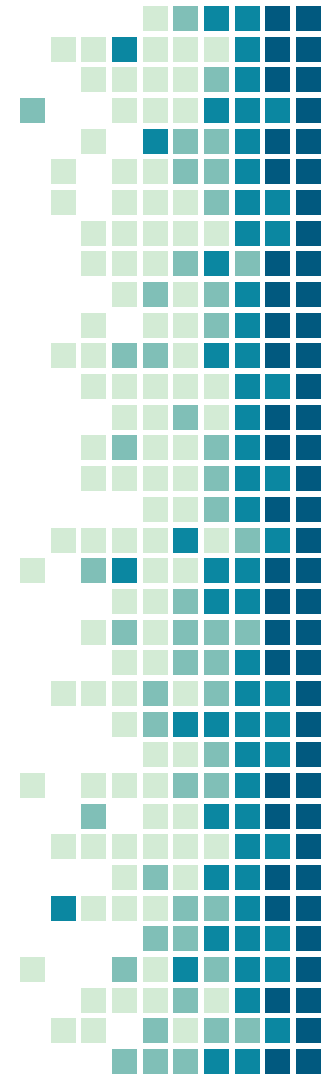
# CHORD System Architecture





# How will CHORD-deposited genomic data adhere to FAIR principles?

- **Findability:** Classify deposited datasets with unique IDs
- **Accessibility:**
  - Implement a "censored" data download service (as in non personally-identifiable)
  - Offer sensitive data download when authorized
- **Interoperability:** Make use of globally-accepted ontologies to classify data (e.g. DUO)
- **Reusability:** Use GA4GH emerging standard ADA-M to query data by allowable use





# The CanDIG Team

## TECHNICAL TEAM



**Zoltan Bozoky**  
Research Associate, BCGSC



**David Bujold**  
C3G McGill site technical  
lead



**Jonathan Dursi**  
Technical Lead



**Carol Gauthier**  
Systems, Sherbrooke



**Jimmy Li**  
Research Programmer,  
BCGSC



**Adam Lipski**  
Research Programmer,  
BCGSC



**Zhibin Lu**  
Bioinformatics, FMCC



**Julian Mazzitelli**  
Workflow Data Engineering,  
HSC



**Quan Nguyen**  
Systems, MUQJIC



**Brendan O'Huiginn**  
Systems Administrator,  
BCGSC



**Kat Pavlova**  
Clin/Pheno Data Modeler, HSC



**Shaikh Farhan  
Rashid**  
Architecture & Devops, HSC



**Matthew Wong**  
Waterloo Co-op, HSC



**Richard de Borja**  
UHN site technical lead

## LEADERSHIP TEAM



**Guillaume Bourque**  
Assoc. Professor - McGill,  
MUQJIC



**Michael Brudno**  
Prof, CS, U Toronto; Director,  
HPC4Health



**Pierre-Étienne  
Jacques**  
Asst. Professor, Biology,  
Sherbrooke



**Yann Joly**  
Assoc. Professor, Genomics  
and Policy, McGill



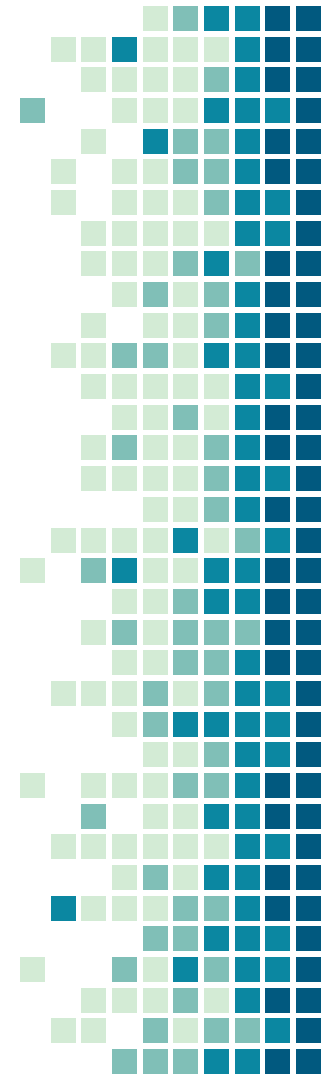
**Steven Jones**  
Assoc. Director, BCGSC;  
Professor, UBC & SFU



**Trevor Pugh**  
Asst. Professor, Medical  
Biophysics, U. Toronto



**Carl Virtanen**  
Director and Research Lead,  
UHN Digital



**Thank you!**

