CANADA FOUNDATION FOR INNOVATION

CFI Research Infrastructure & Shared Data Challenges

Mark Lagacé
Director of Programs

&

Claudia Fall
Associate Director for Research Facilities

INNOVATION
Canada Foundation for Innovation
Fondation canadienne pour l'innovation

Table of Contents

INNOVATION
Canada Foundation for Innovation
Fondation canadienne pour l'innovation

CANADA FOUNDATION FOR INNOVATION

Mandate

TO INCREASE THE CAPABILITY OF CANADA'S UNIVERSITIES, COLLEGES, RESEARCH HOSPITALS AND NON-PROFIT ORGANIZATIONS TO CARRY OUT HIGH QUALITY RESEARCH BY INVESTING IN RESEARCH INFRASTRUCTURE

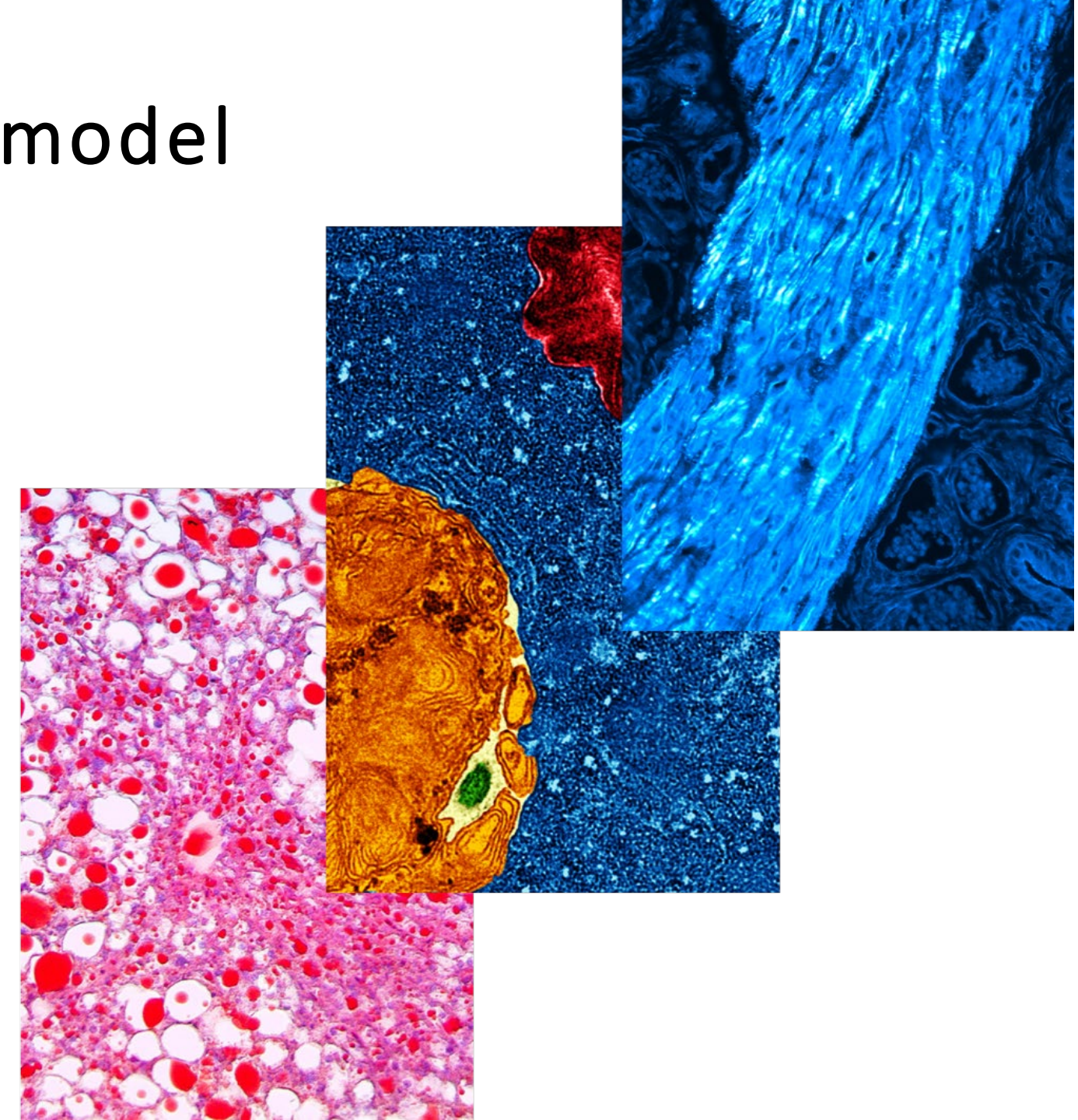
A unique funding model

CFI provides 40% of infrastructure costs (remaining 60% provided by provinces, institutions and private sector)

Since its creation in 1997, the CFI has committed more than $10 billion in support of 12,930 projects at 173 research institutions in 81 municipalities across Canada

Highly competitive funding requiring excellence and institutional commitment to maintaining the infrastructure over its useful life

CFI also provides funding for the operation and maintenance of infrastructure through the MSIF and IOF



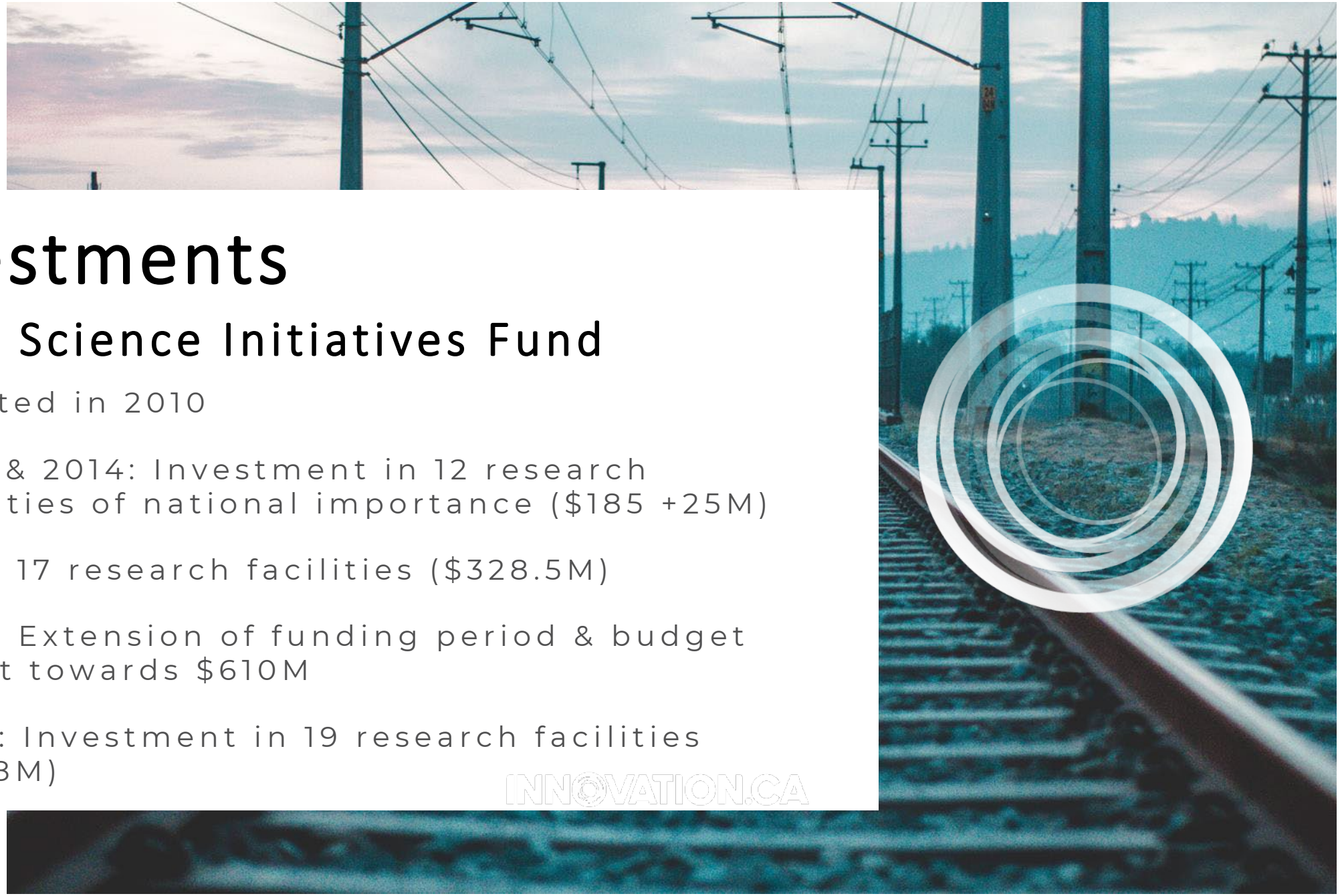
Major Science Initiatives Fund

Objectives

- Enable pan-Canadian research communities to undertake world-class research and technology development that lead to social, health, economic or environmental benefits for Canadians

- Enable facilities to operate at an optimal level to ensure the best use of their specialized equipment, services, resources, and technical and scientific personnel

- Promote responsible stewardship through the adoption of best practices in governance and management

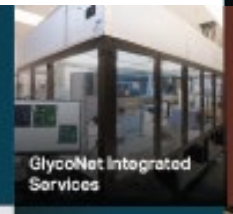


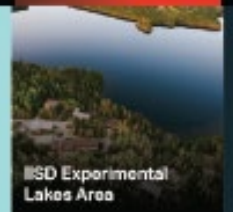
Investments

Major Science Initiatives Fund

- Created in 2010
- 2012 & 2014: Investment in 12 research facilities of national importance ($185 +25M)
- 2017: 17 research facilities ($328.5M)
- 2019: Extension of funding period & budget uplift towards $610M
- 2023: Investment in 19 research facilities ($628M)

Canadian research facilities of national importance



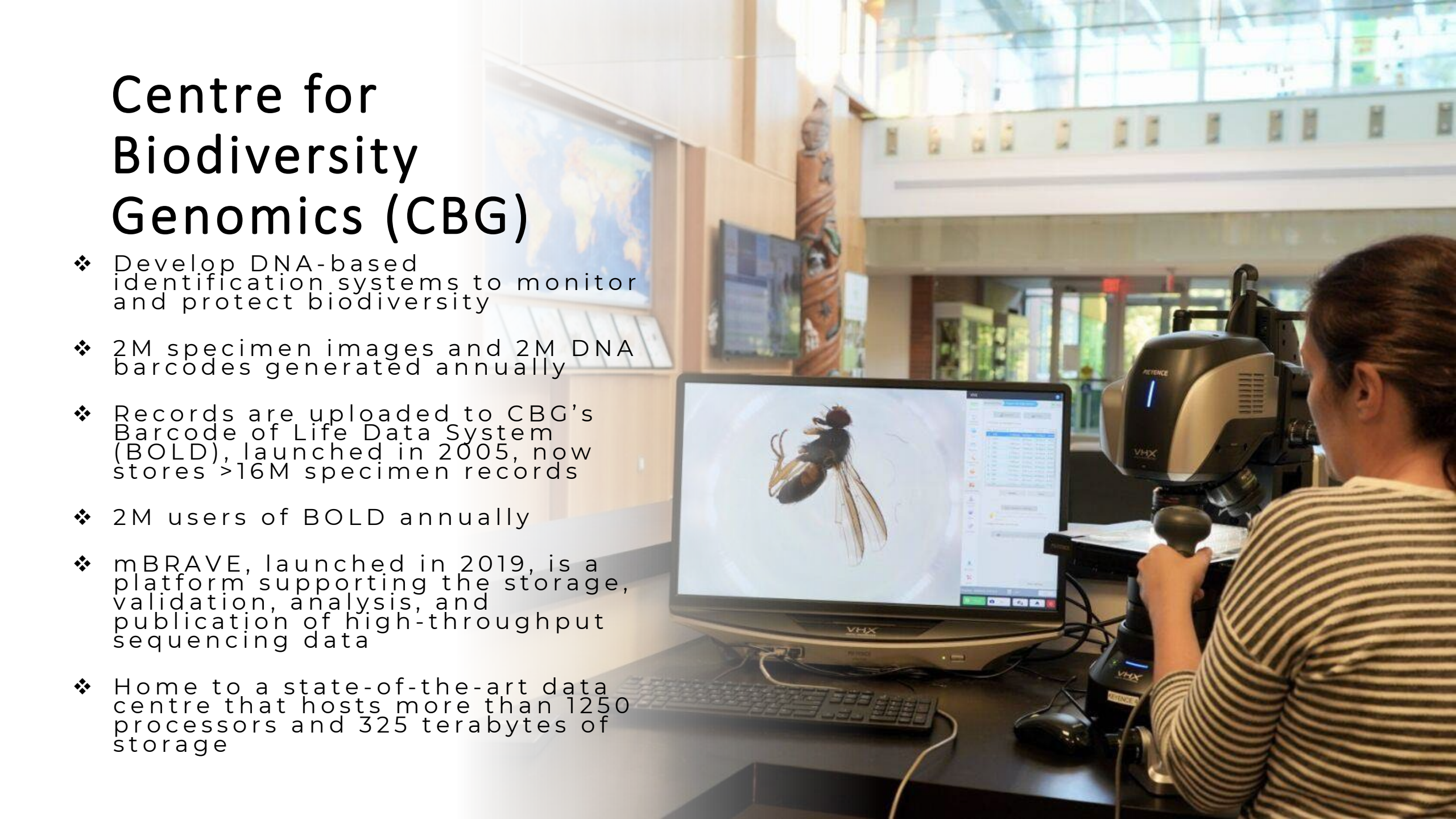


Ocean Networks Canada (ONC)

- Deep-sea, coastal, and land-based cabled, mobile, and autonomous observatories on all three coasts of Canada
- Real-time, long-duration data of high spatial and time resolution that serve public good (e.g. marine safety, hazard alerts)
- Empower Indigenous communities to collect their own data
- >32,000 data users accessing the >1,200 TB data archived by ONC, coming from >10,000 active sensors producing data
- Data is open access through Oceans 3.0

Centre for Biodiversity Genomics (CBG)

- Develop DNA-based identification systems to monitor and protect biodiversity
- 2M specimen images and 2M DNA barcodes generated annually
- Records are uploaded to CBG's Barcode of Life Data System (BOLD), launched in 2005, now stores >16M specimen records
- 2M users of BOLD annually
- mBRAVE, launched in 2019, is a platform supporting the storage, validation, analysis, and publication of high-throughput sequencing data
- Home to a state-of-the-art data centre that hosts more than 1250 processors and 325 terabytes of storage



CGEn – Canada's national platform for genome sequencing and analysis

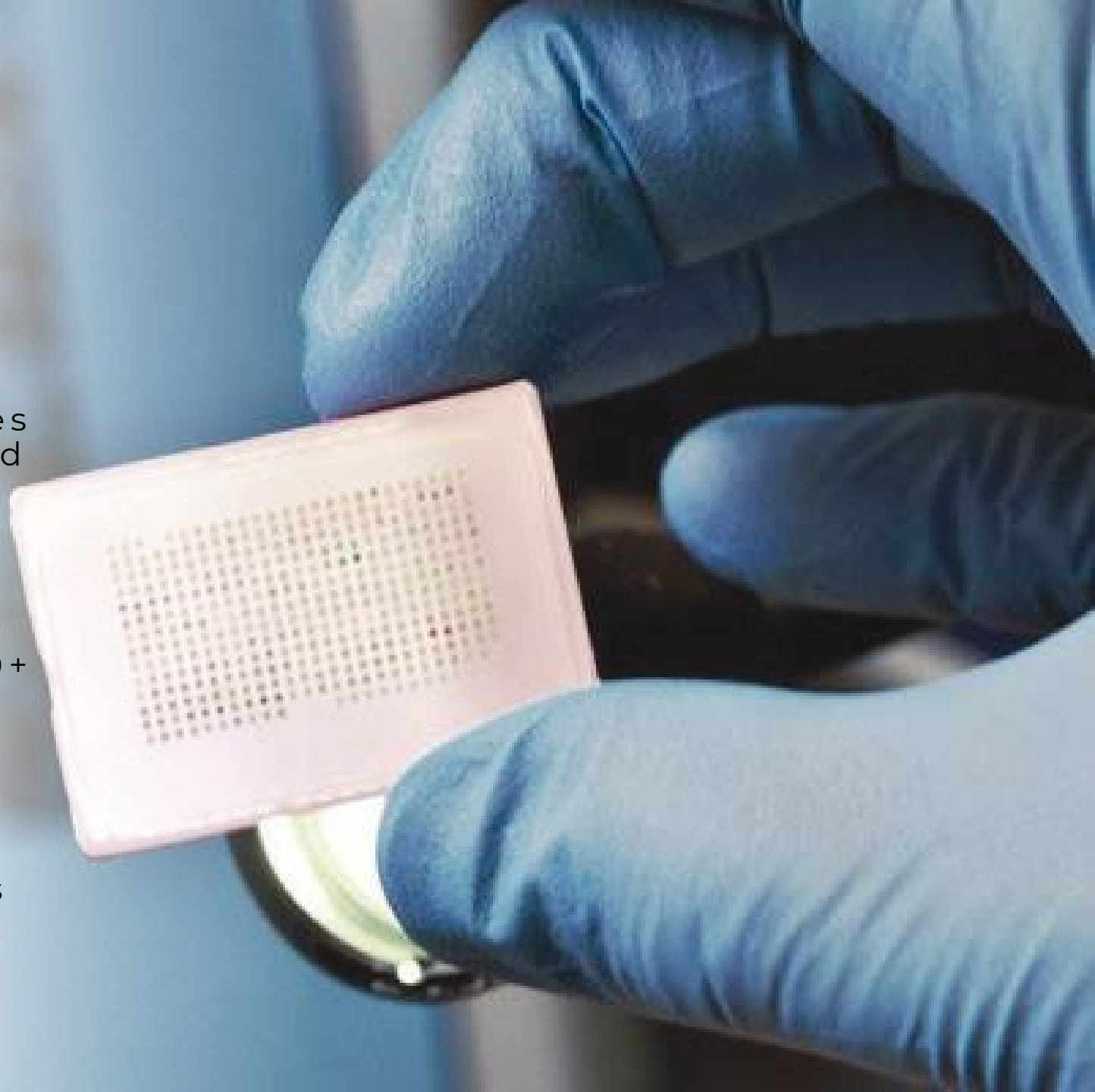
- Integrated high-throughput genomics platform, with nodes in Toronto, Montreal and Vancouver
- Produce 2.4 petabases of data annually
- Led delivery of national HostSeq initiative- databank of whole genome sequencing and matched clinical data > 11k COVID-19 patients
- Aim to establish a national genomics databank, representative of the diversity of Canada's population

Canadian Research Data Centre Network (CRDCN)

- Partnership between >40 Canadian universities & Statistics Canada
- Provides access to confidential Statistics Canada microdata (censuses, surveys, etc.)
- Supports >2,400 researchers in social and health sciences
- ~60 new datasets added annually
- National virtual platform for remote data access to approximately 80% of the microdata files is under development – significant cybersecurity considerations

Canadian Cancer Trials Group (CCTG)

- Development and conduct of cancer trials – early phase studies to large international randomized controlled phase III trials
- CCTG's Tumour Tissue Data Repository is the largest repository of tissues from cancer trial patients in Canada (300,000+ samples from 120+ trials)
- Users can access tissue and clinical data for their own research and secondary analyses
- Data Sharing Platform facilitates user access



CCGS Amundsen

- Icebreaker dedicated to carrying out cutting-edge scientific expeditions and data collection in the Arctic Ocean
- Used by oceanographers, paleoceanographers, marine geologists, geophysicists, ecologists and atmospheric scientists
- 1,800 researchers, technicians, students, professionals from 26 countries have been aboard
- Dedicated data management and communications unit responsible for overseeing the entire data lifecycle
- Part of Canadian Consortium for Arctic Data Interoperability and the Canadian Integrated Ocean Observing System
- Data are published in the open access Polar Data Catalogue
- 1000 data users annually

Canadian service providers

Digital Research Alliance of Canada

- Established in 2019, funded by Government of Canada
- Mandate: coordinate & fund Advanced Research Computing, Research Data Management and Research Software activities
- Provides researchers with access to digital tools, services & infrastructure
- Tools : Federated Research Data Repository , Lunaris, etc.

CANARIE

- Established in 1993, funded by Government of Canada
- Ultra-high-speed network
- Connects researchers to each other & to global data, technology
- Funds, implements & supports cybersecurity initiatives
- Identity management services to academia & cloud resources to Canada’s startups

Data management practices & challenges

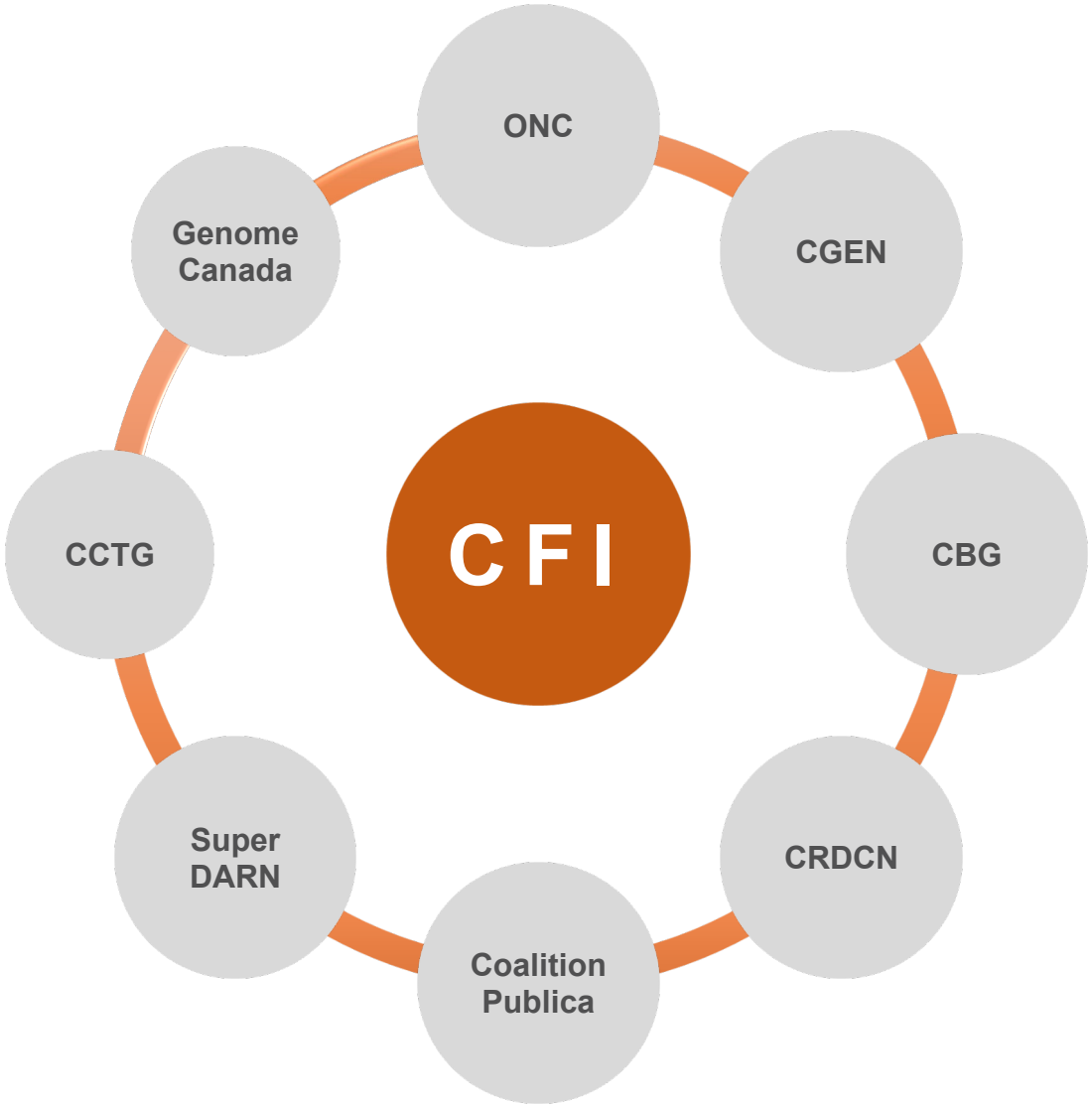
How

Organize annual workshop to

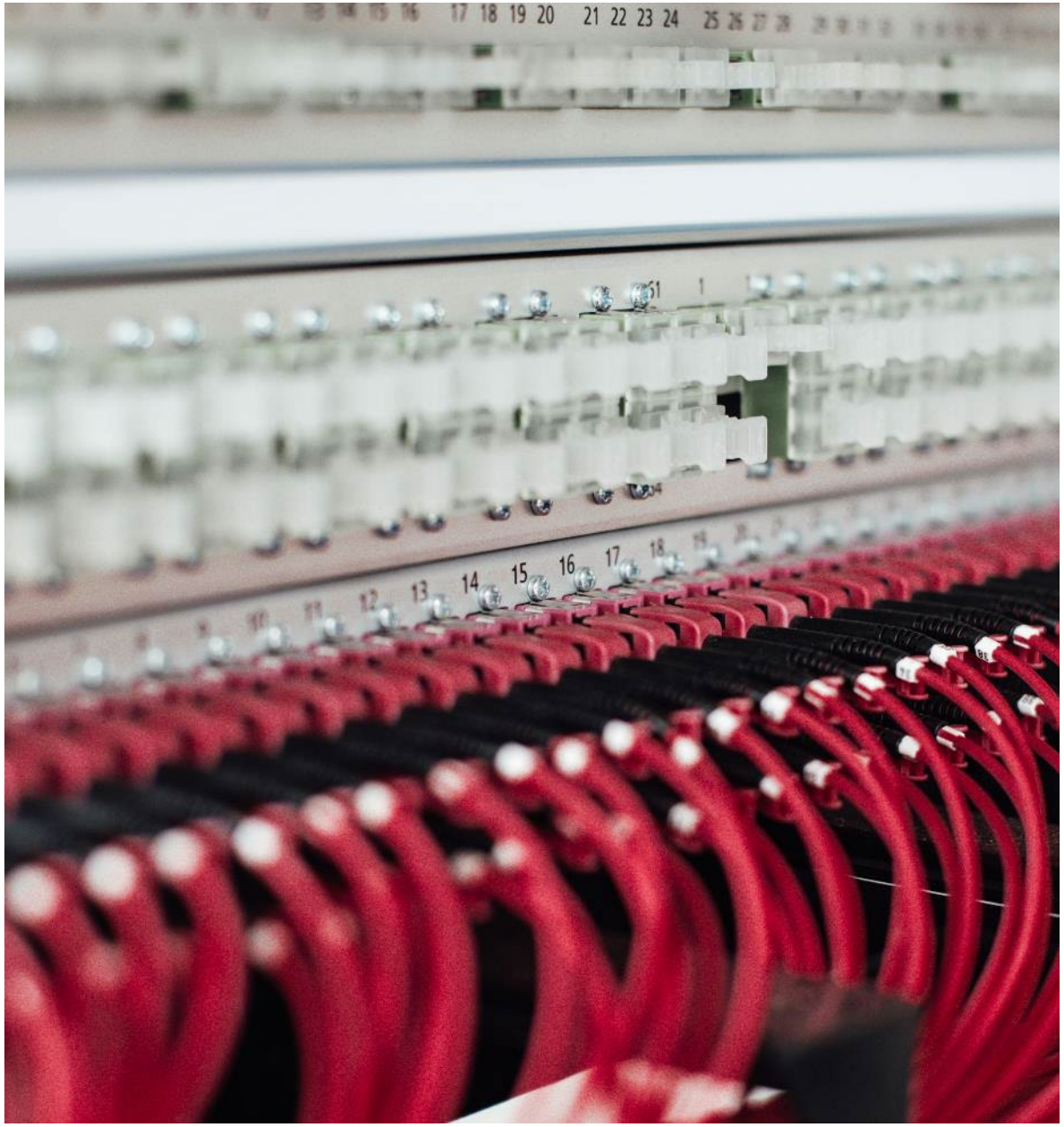
- Build community, share knowledge & discuss challenges
- Bring together staff from facilities, host universities, funding partners & special guests from international counterparts
- Identify need for action for more in depth discussions (e.g. rapidly changing technologies & research facilities in a data-intensive and open-science world)

Data Working Group

- Inform & advise the CFI
- Data management practices
- Creation & implementation of data lifecycle model for facilities
- Role of service providers
- Barriers to international data sharing
- Report back to community



Data Management Challenges

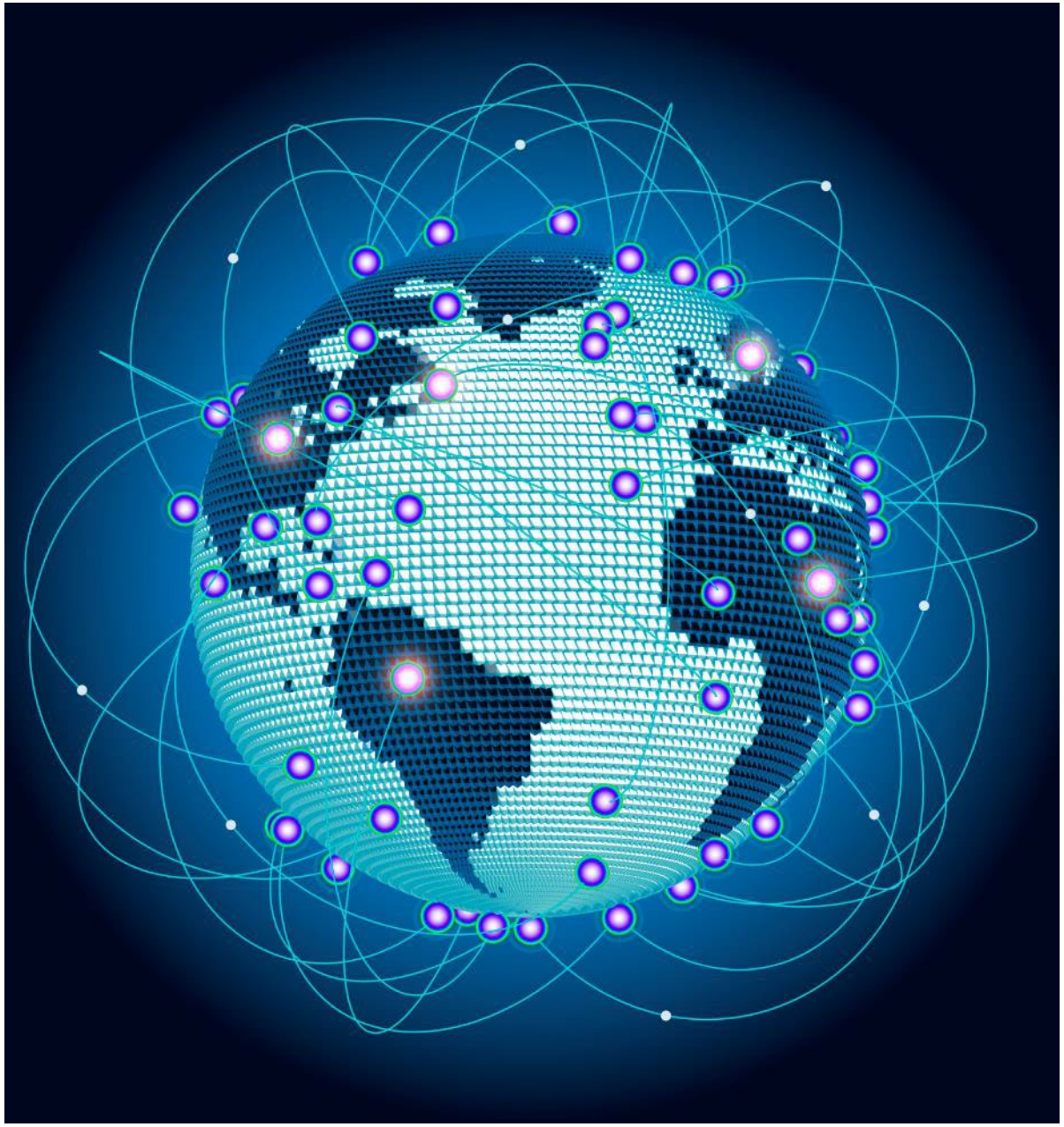


Long term data preservation & storage

- Improvements in technology have led to explosion in data outputs. How to store increasingly large volumes of data generated and for how long?
- Data pipelines within university become insufficient as facilities grow.
- Need for better public data storage platform.
- Research facilities do not have priority access to resources from research data service providers (the Alliance) – application process is separate, and they risk losing access to the services.

Sharing/enabling access to secure data

- For facilities that work with sensitive data (e.g. personally identifiable, health information), sharing this data with partners and keeping the data secure to the satisfaction of the primary data holder.
- A centralized repository with high availability and high bandwidth for content delivery is needed

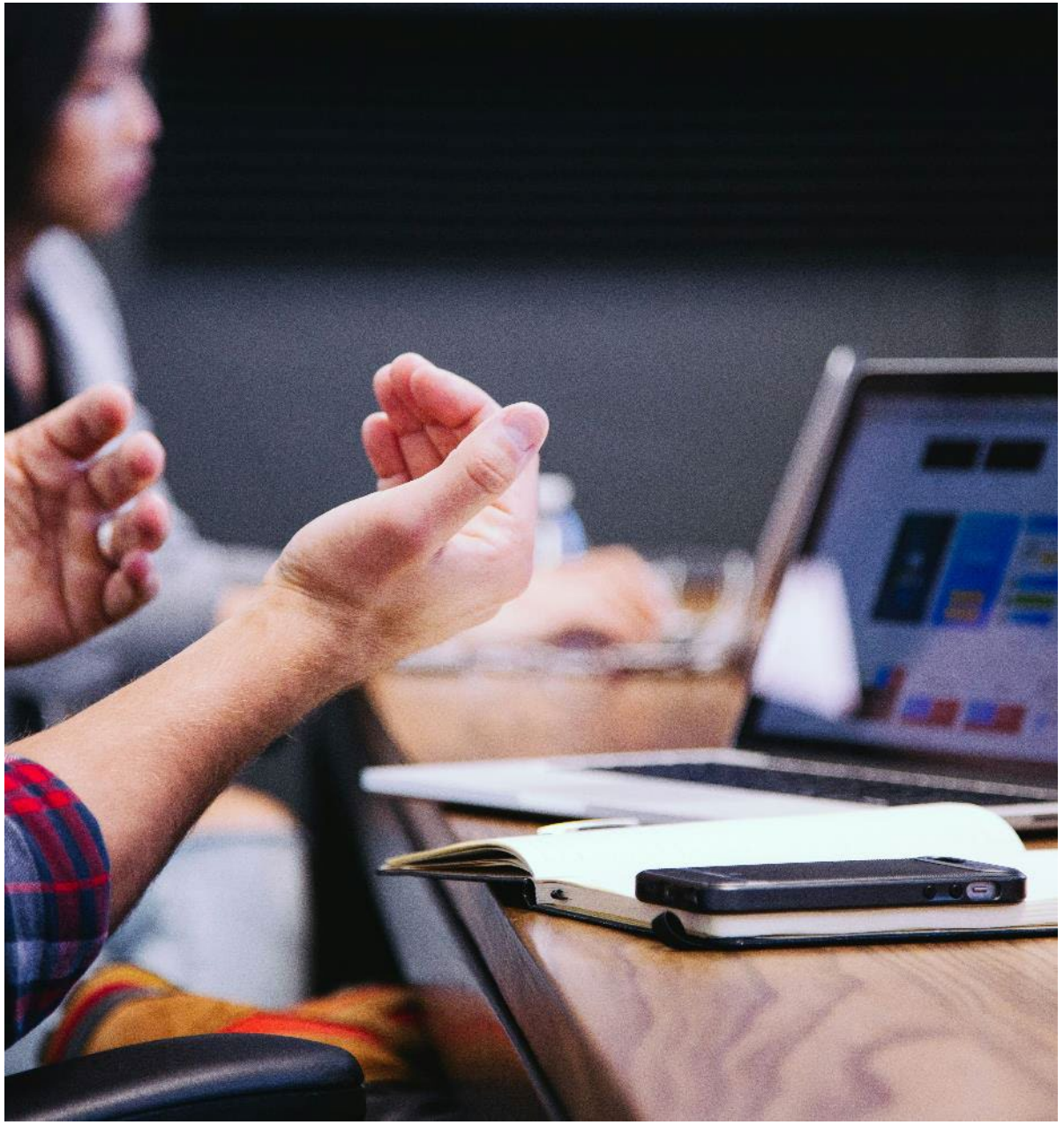


Data Curation

- Generating good metadata requires a lot of highly qualified personnel, which are difficult to retain.
- Ensuring that data is interoperable and reuseable (FAIR principles). Differing standards, nomenclature, regulatory environments.
- Curating data in a meaningful way for the general public

Training

- Users need training around open access and open science standards.
- User communities may not fully understand principles and processes of data lifecycle



Changing Environment

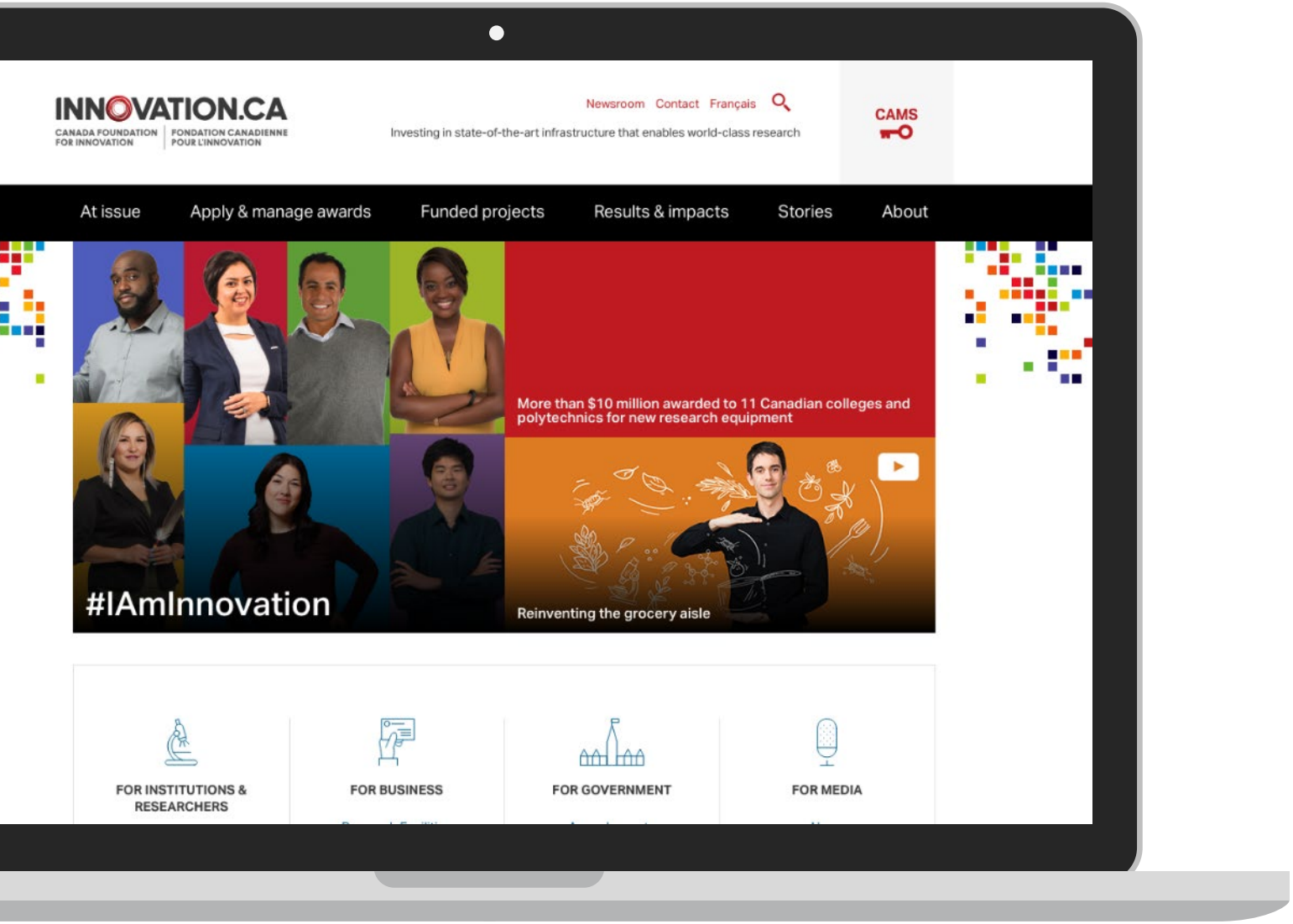
Research Security

- National Security Guidelines for Research Partnerships
- Policy on Sensitive Technology Research and Affiliations of Concern

Open Science

- Federal Roadmap for Open Science (2020)
- Chief science Advisor's Science Data Framework Advisory Committee
- Quebec Open Access policy (2021)

  Federal research granting agencies
  - Open Access Policy on Publications expected in 2025


INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION
Newsroom Contact Français
Investing in state-of-the-art infrastructure that enables world-class research
CAMS
At issue
Apply & manage awards
Funded projects
Results & impacts
Stories
About
More than $10 million awarded to 11 Canadian colleges and polytechnics for new research equipment
#IAmInnovation
Reinventing the grocery aisle
FOR INSTITUTIONS & RESEARCHERS
FOR BUSINESS
FOR GOVERNMENT
FOR MEDIA

Thank you!