Digital Research Alliance of Canada

Services and Operations

HEPiX, University of Victoria, October 2023

Patrick Mann, Director of Operations

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Abstract

The Digital Research Alliance of Canada is a new organization, replacing the earlier organization Compute Canada, that provides compute and storage to Canadian researchers. The Alliance provides resources for the particle physics community and operates the ATLAS Tier-2 facilities as well as providing compute and storage capacity for other national and international experiments. This talk will provide an overview of the services and resources that currently make up the National Platform together with a brief introduction to our operational management and support practices.

Who We Are

- A member-based, not-for-profit organization
- Designed to simplify and streamline the process to fund and deliver digital research infrastructure (DRI) services to the research community
- Mandate:
  - **Strategy**: long-term direction of DRI in Canada
  - **Funding**: investment in capital infrastructure and operational expenses to maintain DRI
  - **Operations**: delivering and coordinating DRI services to researchers

History I - Compute Canada

Prehistory	Loose federation of 4 regional Computing organizations ● **WestGrid:** BC, AB, Saskatchewan, Manitoba ● **Compute Ontario:** 3 consortia ● **Calcul Québec:** 3 consortia ● **ACENET:** Atlantic Canada: Newfoundland, New Brunswick, Nova Scotia ~12 regional systems + USask Data Silo Funded through the Canada Foundation for Innovation (CFI) and Provincial/Institutional match
2010-2022	**Compute Canada** ● Created by the regions and member institutions ● CFI requests for a single point-of-contact with a single funding proposal ● Need for central coordination, Resource Allocation, etc.
2015-2019	**CFI Cyberinfrastructure Program** ● Aggregation into 5 sites with new funding
2019	**Industry, Science and Economic Development (ISED) Expansion Program** ● Additional expansion funding for the 5 sites

History II - Digital Research Alliance of Canada

2017	**Leadership Council for Digital Research Infrastructure (DRI)** ● Responding to a clear and significant need for a national DRI strategy. ● Community-based process ● 3 position papers: Research Data Management (RDM), Advanced Research Computing (ARC), Coordination ○ Extended into Research Software (RS) ● Successful proposal to Federal Government!
2019	**Digital Research Alliance of Canada** created (“The Alliance”) ● RDM, RS and ARC ● ~$350M over 5 years (to Mar 31, 2025) ● +40% match from provinces/institutions.
Apr 1, 2022	**The Alliance took over** ● Compute Canada dissolved

Advanced Research Computing (ARC) in the Digital Research Alliance of Canada

Canada's Advanced Research Computing Platform

Digital Research Alliance of Canada

- *Compute Resources*
- *Storage Resources*
- *Cloud Resources*
- *Support and Training*

Cluster Capabilities

Cluster	Cores	GPUs	Project Storage	Nearline Tape Storage	CPU Jobs Submitted Jan-Mar 2023	GPU jobs submitted Jan-Mar 2023
General Purpose (diverse nodes with general purpose interconnect topology)						
Béluga	32,080	688	17 PB	10 PB	2.1M	213K
Cedar	92,528	1,352	19.5 PB	18 PB	8M	836K
Graham	36,432	540	10 PB	20 PB	2.8M	162K
Narval	75,584	648	14 PB	5 PB	2.4M	379K
Large Parallel (consistent nodes with special purpose interconnect topology)						
Niagara	80,960	-	7 PB	45 PB	388k	23K
Total	**217,584**	**3,228**	**68**	**98 PB**		

Arbutus Community Cloud

Arbutus Cloud is Canada's largest research cloud system located at University of Victoria, providing virtual CPU's, virtual GPU's and massive storage to enable collaboration platforms, customized websites, machine learning/artificial intelligence, and big data.

- Virtual Machines, including CPU, GPU and Block Storage
- Shared Object and Filesystem services

https://docs.alliancecan.ca/wiki/Cloud_resources

Volume and Snapshots	5.3 PB on Ceph
Object storage and Shared File system	12 PB on Ceph
dCache	13.7 PB

vCPUs	33,662
vGPUs	416

Infrastructure Refresh 2022-2025

Equipment is getting old!

- Capital funding allocated in the contribution agreement.
- Needed a formal proposal to the ministry (ISED)

September 2022	**Multi-Year Funding Proposal (MYFP) to the Federal government.** • Various major projects proposed • Including ARC (HPC and Cloud) refresh
May 2023	**Approved** • $117.5M from Federal Government for the ARC refresh • Plus match from institutional/provincial sources: 50:50 • Total $225M
May 2023 - April 31, 2025	**Procurement and Implementation** (next slide)

ARC Refresh Current

Each existing National Host Site allocated funds to refresh their system.

- **Node-for-node approach** used to estimate the required budget.

Call for Proposals	Summer 2023 (almost complete)	Formal Call for Proposal to each site: Sites provided a detailed proposal based on the node-for-node concept ● Review panel with internal and external members
Contracts	November, 2023	Formal contracts with each host site. ● **Key dependency: match funding**
Request for Proposals	January, 2024	Each site undertakes their internal procurement/RFP process
Procurement and Installation	2024/25 ● to Mar 31, 2025	Systems installed and pass RFP requirements. ● Handoff to systems teams.
Configuration and Migration	Summer, 2025	Configured to Alliance standards and users migrated.

A very exciting couple of years ahead!

Current ARC Services, Organization and Governance

Central ARC Administrative Services

Accounts & Resource allocation	CCDB central database managed by the CCDB-dev team (Alliance): https://ccdb.alliancecan.ca
Middleware - Authentication and Authorization	Managed by the *Infrastructure Operations National Team* with staff support from the Alliance ● LDAP for national host sites ● Central Identity Provider for other services
National Helpdesk (ticketing system)	Managed by the *Research Support National Team*
Central Monitoring	Managed by the *Data Analytics National Team*
Grant administration	Managed by the Alliance
Central Software Provision	Managed by the *Research Support National Team*

Central Software Distribution: CVMFS for software

- Decouple software from system
  - update and distribute separately
- Maintain just 1 central repository
- Updates are automatic, atomic, and fast
  - ~ 5 min
- Original use-case: cloud computing (CernVM)

~200 packages installed
- https://docs.alliancecan.ca/wiki/Available_software

Virtual Clusters or Servers can also load the stack.

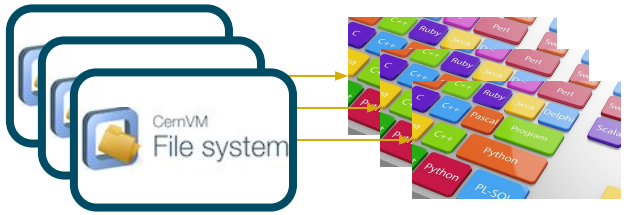
software packages

VM images

maintain each package in each image

use CVMFS

maintain each package and each image separately



14

Thanks to Ryan Taylor for providing slides

dCache Storage Service

dCache is supported on the Arbutus Cloud

dCache available	13.7 PB
dCache requested RAC 2023	14.1 PB (expected to increase substantially for RAC 2024)

This audience knows a lot more than I do about dCache!

Magic Castle: A Virtual Cluster Provisioning Tool

Developed and managed by Félix-Antoine Fortin, (CQ)

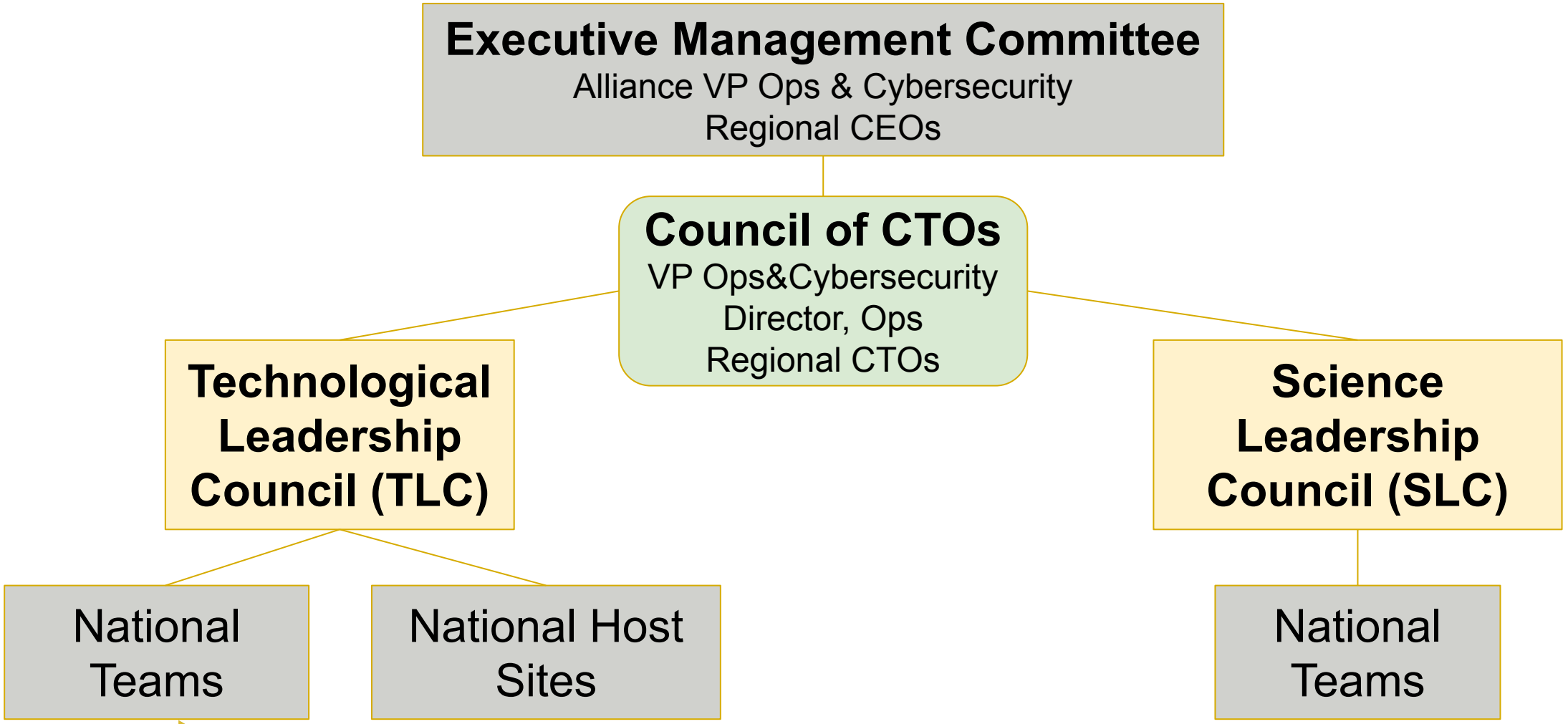
- Loads the CVMFS software stack
- Identical to an Alliance cluster!
- Can be provisioned in any cloud (including commercial)

Currently use extensively for training:

- create a temporary cluster in the Arbutus cloud

Supercomputing 23 - Magic Castle: Terraforming the Cloud to Teach HPC

ARC Operational Governance



National Teams - TLC

The Technology Leadership Council (TLC) is a coordinating body. All national and site teams that have direct responsibility for coordinating and delivering operational components of the national advanced research computing (ARC) platform are represented on the team.

- Accelerators Working Group
- Authentication and Authorization National Team
- Automation and Configuration National Team
- Bioinformatics National Team
- Cloud National Team
- **CVMFS National Team**
- Data Analytics National Team
- Globus National Team
- Networking National Team
- Storage National Team

- Resource Allocations Competition Administration Team
- Research Data Management
- Infrastructure Operations National Team
- Middleware Infrastructure Team
- Molecular Modelling and Simulation National Team
- Monitoring National Team
- Scheduling National Team
- **Subatomic Physics National Team**
- Training Coordination National Team

National Host Site teams (UVic, SFU, UToronto, UWaterloo, McGill)
National Security Council
Science Leadership Council

SLC National Teams

The Science Leadership Council (SLC) is a coordinating body to discuss strategic directions for research support, assessment of researcher needs, exchange of best practice, and innovative service development.

- Accelerators Working Group
- Bioinformatics National Team
- Grant Administrative Committee
- Humanities and Social Sciences
- Molecular Modelling and Simulation National Team
- Research Support National Team
- Resource Access Program Administrative Committee (RAPAC)
- **Subatomic Physics National Team**
- Training Coordination National Team
- Visualization National Team

ARC Training

Extensive training events and resources are available generally through the regions

- Various links are at https://docs.alliancecan.ca/wiki/Getting_started

WestDRI (Western Canada Research Computing covering both BC and the Prairies regions)	• Training Materials website • UAlberta ARC Bootcamp
SHARCNET	• Training Events Calendar • Youtube Channel • Online Workshops
SciNet	• SciNet Education Site • SciNet YouTube Channel
Calcul Québec	• Workshops • Training Events and Resources
ACENET	• Training information

- Basic introductions to the systems
- Basic linux, command-line, best practices, etc.
- Advanced Cluster usage
- Advanced programming, including parallelization and GPU use
- Intro to cloud services
- Discipline-specific: AI, Molecular Dynamics, CFD, Bioinformatics, Humanities and Social Sciences, etc.
- Schools (intro, job submission, basic programming, etc.)
- Special topics…

Getting Started

Any faculty member at a recognized institution can create a PI or "Project" account

- The PI can then sponsor team/user accounts

Includes default ("Rapid Access Service") access to the GP clusters, and a small cloud allocation

- The PI and sponsored users can immediately start submitting jobs or create virtual machines.
- Approximately 20% of resources are reserved for default use

Getting started details	https://docs.alliancecan.ca/wiki/Getting_started
Technical documentation	https://docs.alliancecan.ca
Support	support@tech.alliancecan.ca

Resource Allocation Competition (RAC)

Additional resources (beyond default) are allocated through the Resource Allocation Competition

- Annual competition
- https://www.alliancecan.ca/en/services/advanced-research-computing/accessing-resources/resource-allocation-competition

Sept 2023	RAC 2024 announced and application portal is open
Nov 7, 2023	**Deadline for submitting an application to RAC 2024**
Dec 2023 - Feb 2024	Science and Technical reviews and final allocations
Mar 2024	Awards announced
April 2024	Allocations implemented

Research Data Management (RDM)

Thanks to Lee Wilson, Director, RDM for providing slides

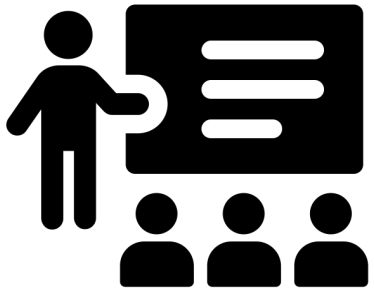
infrastructure by Nithinan Tatah from the Noun Project
computing by Adrien Coquet from the Noun Project
tools by tanu doank from the Noun Project
training by Adrien Coquet from the Noun Project

Research Data Management
Gestion des données de recherche

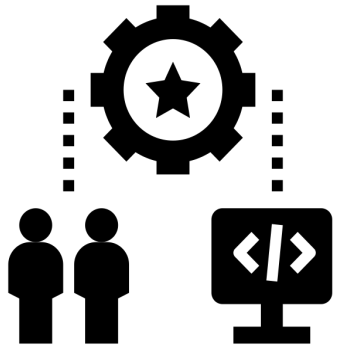
150+ Experts
70+ Organizations

Alliance RDM

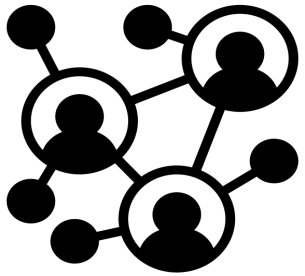
Researcher-centric & Service-oriented



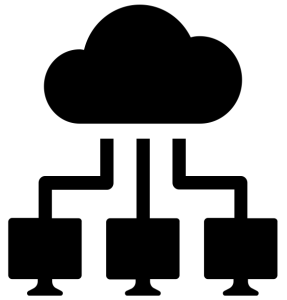
Training & Outreach
Formation et sensibilisation



Services, Policies, & Best Practices
Services, politiques et meilleures pratiques



Network of Experts
Réseau d'experts



Infrastructure Platforms
Plateformes d'infrastructures

Tools
Outils

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

National RDM Services & Programs

DMP ASSISTANT

PGD

FRDR
Federated Research Data Repository

DFDR
Dépôt fédéré de données de recherche

lunaris
Discover Canadian Research Data

borealis
The Canadian Dataverse Repository
Le dépôt Dataverse canadien

Canadian Research Knowledge Network

Réseau canadien de documentation pour la recherche

doi

ORCID

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Alliance and Partners Cybersecurity Program

- As per the Digital Research Infrastructure Contribution Agreement Cybersecurity Framework: *The Recipient will, in consultation with other DRI partners, develop and implement a cybersecurity framework, covering governance, policies, technology and operations.*

- The Cybersecurity Program was co-built with our DRI partners and was approved by ISED in April 2022.

Thanks to Masood Akhtar, Director, Cybersecurity for providing slides

NIST Cybersecurity Framework (CSF) Core

- **Cybersecurity Strategy and Roadmap**
  - All 5 NIST Functions: Identify, Protect, Detect, Response and Recover.
  - 23 Categories of NIST
    - 64 Subcategories of NIST
  - 57 activities/actions
- **Major Working Groups:**
  - *Identity & Access Management*
  - *Vulnerability Management*
  - *Training and Awareness*
  - *Security Information and Event management system*
  - *Security Operations*
  - *Incident Management*
  - *Asset Management*
  - *Policy & Governance*

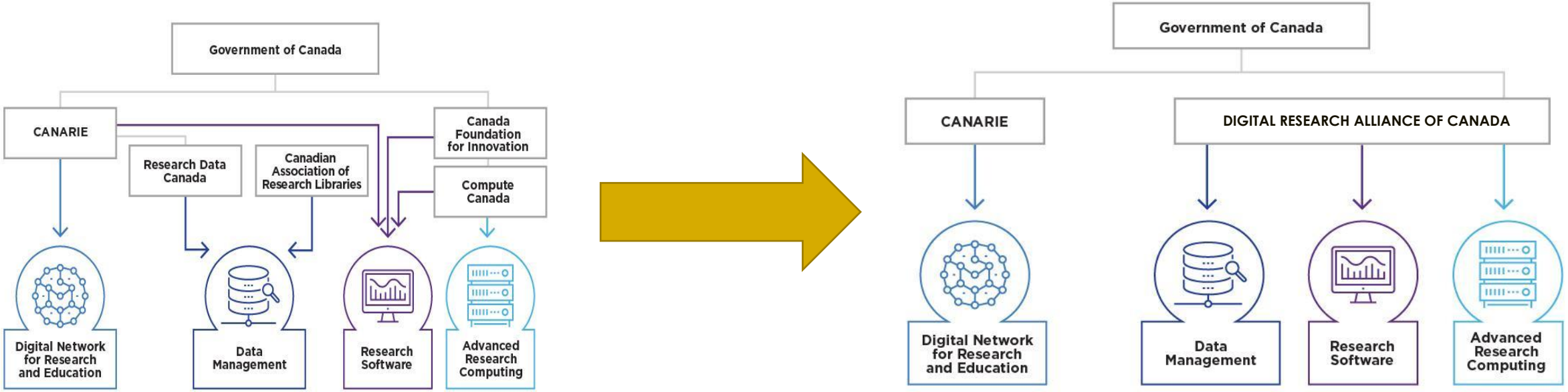
Cyber Security Long Term Action Plan (LTAP) Project Roadmap			
FUNCTION	CATEGORY	SUBCATEGORY	Deliverables for LTAP Mar 2022-Mar2024
IDENTIFY (ID)	Asset Management (ID.AM)	AM-1	Implement a national asset inventory solution for use across the Federation covering a majority of existing assets.
		AM-2	Implement software inventory providing a record of critical applications and packages in the shared environment.
		AM-3	Develop network architecture diagrams across the federation
		AM-5	Develop a standard for classifying assets based on risk and incorporate classification data into the national asset inventory.
		AM-6	Collect, develop and publish Security RACI and Security Org Chart for the federation.
		GV-1	Continued development of the Cyber Security Framework documents based on existing priorities. Which includes an information security policy
	Governance (ID.GV)	GV-2	Coordinate with the security awareness and training program around governance policy, standards, and procedures. Socialize security resource map developed in AM-6
		GV-3	Discussion within Alliance leadership team and legal; To identify legal, privacy and regulatory requirements within the cybersecurity domain. (this item needs followup) (cyber insurance?)
		GV-4, RM-1	Socialization and implementation of risk management process and risk register
	Risk Assessment (ID.RA)	RA-1	See PR.IP-12 Develop an implementation plan for application penetration testing of AF assets and services.
		RA-2	Implement Cyber Intelligence platform/store. Develop Threat Fusion Process. Develop a threat awareness program/content
		RA-3-6	Follow-on activity from risk management process implementation effort
	Risk Management Strategy (ID.RM)	RM-2	Socialize risk policy

Questions?

patrick.mann@alliancecan.ca

Extra Slides

Compute Canada Transition to the Alliance



Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Community Cloud

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

The NIST Definition of Cloud Computing: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Clusters

Narval (McGill) November 2022

An HPC cluster consists of **multiple high-speed computer servers networked together**, with a **centralized scheduler** that manages the parallel computing workload. The computers, called nodes, use either high-performance multi-core CPUs or, more likely today, GPUs (graphical processing units), which are well suited for rigorous mathematical calculations, machine learning models and graphics-intensive tasks. A single HPC cluster can include 100,000 or more nodes. https://www.ibm.com/topics/hpc

RDM Network of Experts

Curation
Curation

Research Intelligence
L'Intelligence en recherche

Sensitive Data
Données sensibles

Training
Formation

Data Management Planning
Planification de la gestion des données

Dataverse North
Dataverse Nord

Data Repositories
Dépôts de données

Preservation
Préservation

Discovery & Metadata
Découverte et les métadonnées







Plan
Planification

Create
Créer

Process
Nettoyez

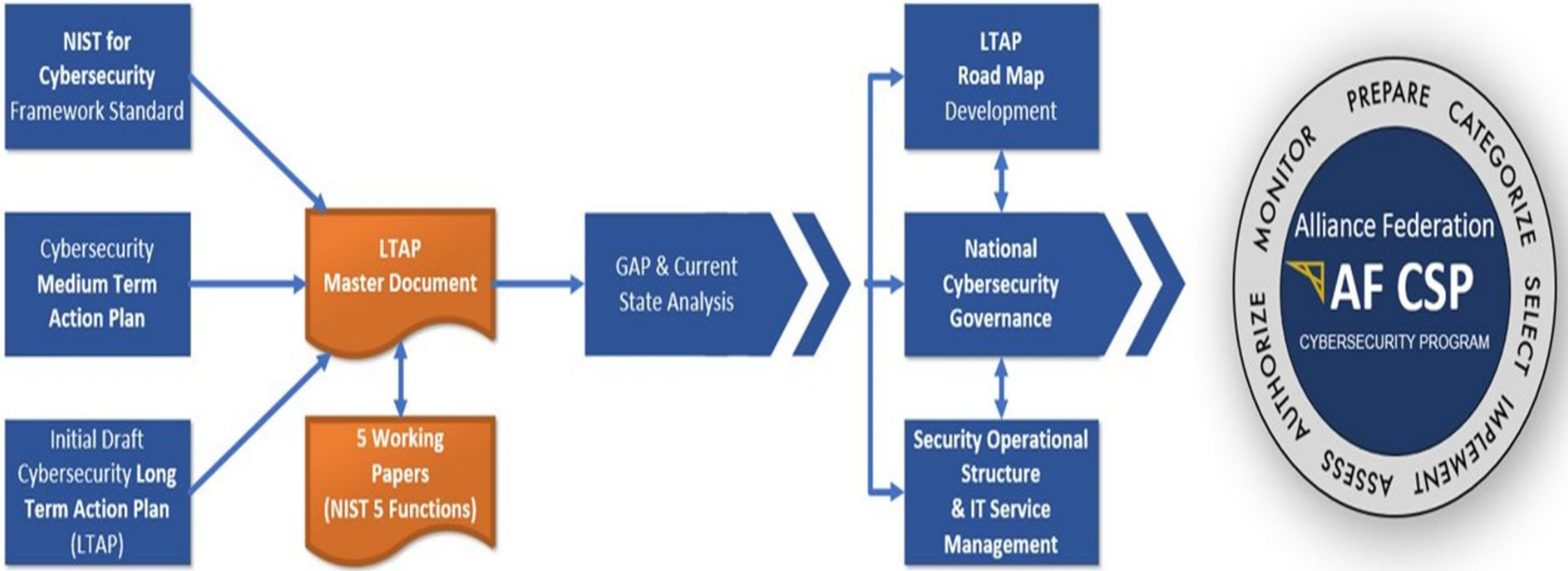
Analyze
Analysez

Disseminate
Disséminer

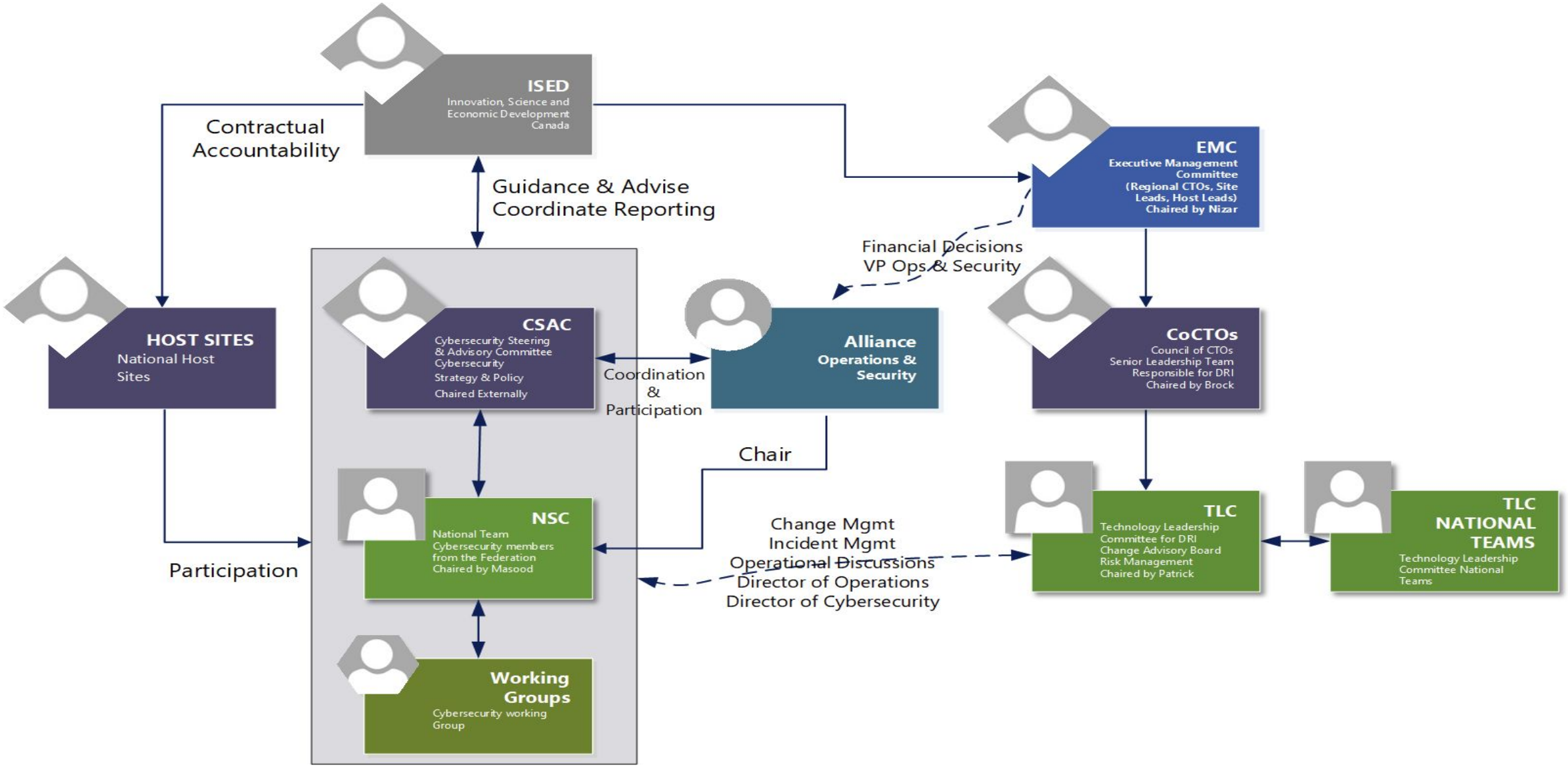
Preserve
Préserver

Reuse
Réutiliser

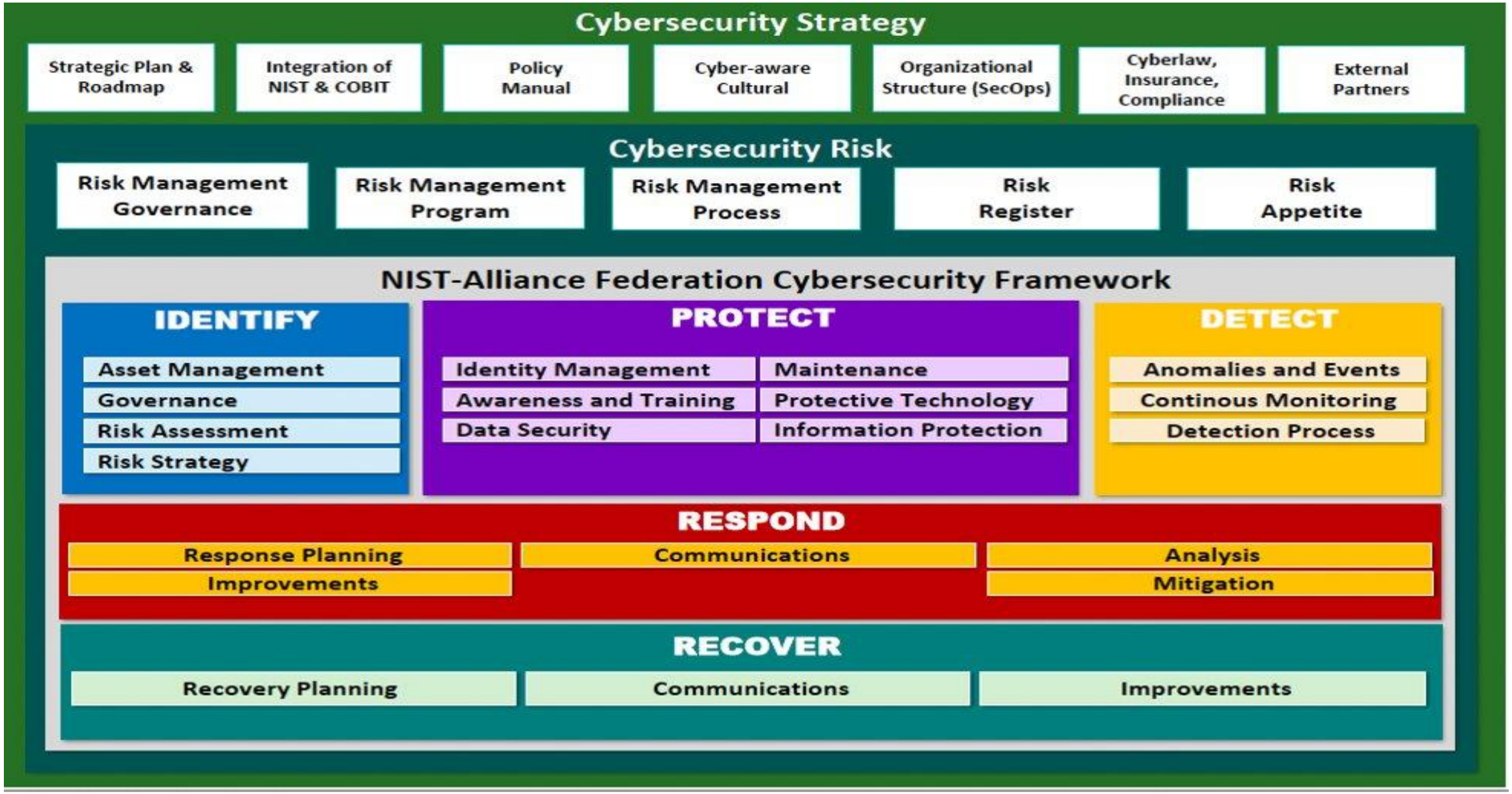
Cybersecurity Program Development



Alliance Federation Current Operating Model



Cybersecurity Strategy Reference Architecture



The Research Data Life Cycle

Le cycle de vie des données de recherche

Plan

Planification

Reuse

Réutiliser

Create

Créer

Preserve

Préserver

Process

Nettoyez

Disseminate

Disséminer

Analyze

Analysez