



# Data Science at Statistics Canada

Sevgui Erman

November 2021



Delivering insight through data for a better Canada



Statistics  
Canada

Statistique  
Canada

Canada

# The Start: The Data Science Accelerator for Economic Statistics (formally created in March 2018)

# Data Science Division (since September 2019) with a corporate mandate



August 2019



February 2021

# DATA SCIENCE CAPACITY AT STATISTICS CANADA



## HUB AND SPOKE MODEL

- Central expertise and guidance
- Collaboration across the agency to leverage expertise



## CENTRE OF EXPERTISE

- Champion DS adoption and integration
- Expertise in AI/ML research and development



## SCIENTIFIC BACKBONE FOR DS

- Build the scientific backbone for DS through focus on quality, rigour and ethics



## LEADING-EDGE METHODS

- Explore new ways of making statistical programs more effective



## DATA SCIENCE IN ACTION

- Large spectrum of AI/ML applications
- Expansion of DS to support COVID-19 response

# VALUE PROPOSITION

Deliver timelier products to Canadians

Implement a new scale of knowledge work automation and cost-efficient operations

DS will bring value to policymakers, federal, provincial and territorial client departments and Canadians in two ways:

1. DS enables the creation of high-value, high-quality, trusted data products that are used by decision makers to make evidence-based decisions
2. Leading-edge DS tools and platforms enable more effective service delivery with a new scale of knowledge work automation and cost-efficient operations

Reduce response burden on households and businesses

New data collection tools

Produce more granular, complete statistics

Agile, user-centric way of working

Enhance tools for privacy and confidentiality

Produce new insights

Meet evolving, data-driven needs in an accountable, highly entrepreneurial manner

# Types of data science applications

100

Natural Language Processing  
(NLP for short)

Image Processing

Predictive Analytics

Outlier and Error Detection

Data Engineering and Pipeline

PDF Extraction and Optical Character Recognition (OCR)

Epidemiological/ Agent-based Modelling

Privacy Preserving Technologies

Web Information Retrieval

Spatiotemporal Modelling

Survival Analysis

Network Analysis

Responsible/ Explainable ML

Chat bot systems

Meta-data management

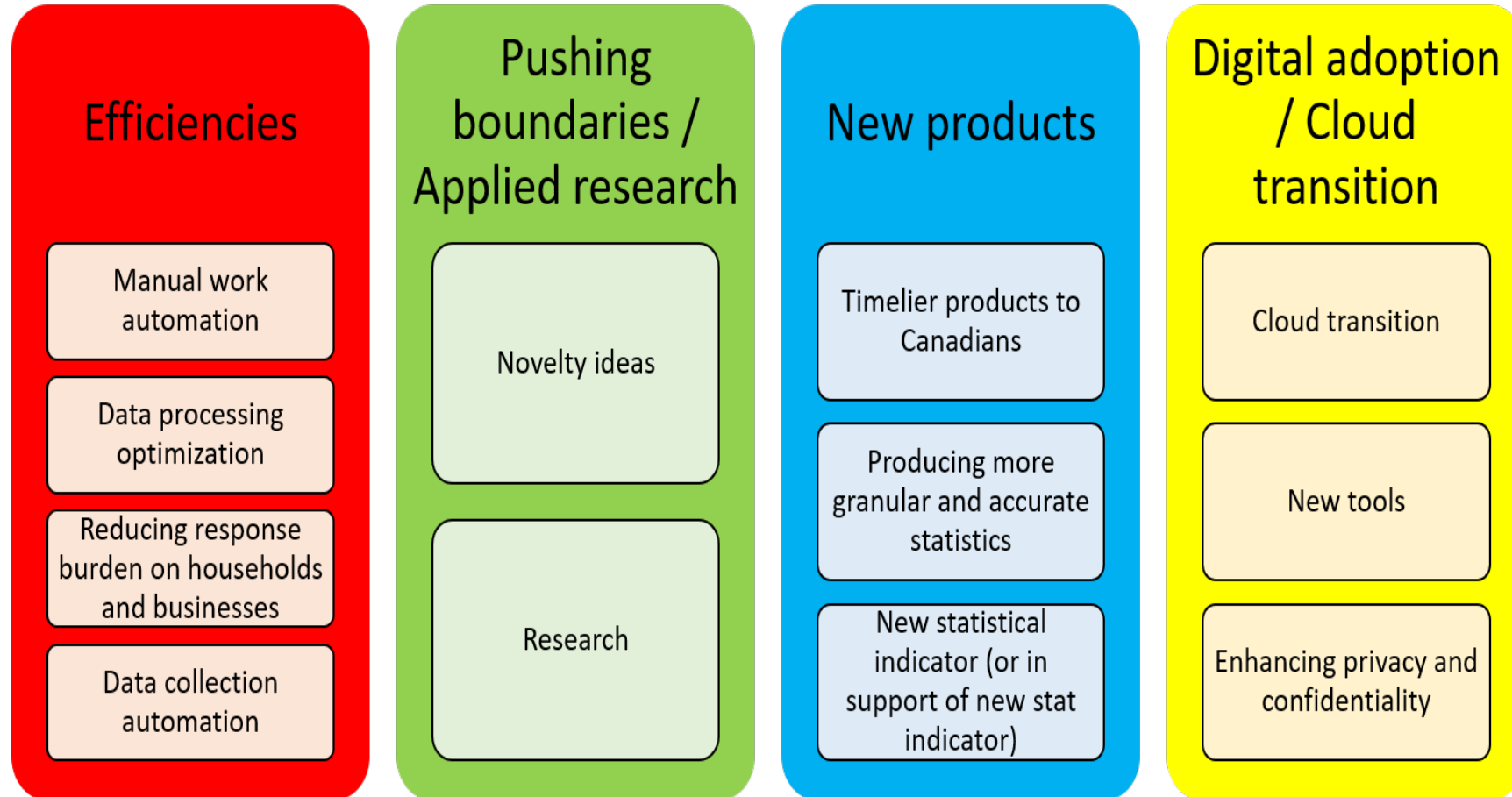
Research projects



# Types of Data



# The need for prioritization and Key Performance Indicators



# Statistics Canada's Framework for Responsible Machine Learning (Principles)

100

## RESPECT FOR PEOPLE

- Value to Canadians
- Prevention of harm
- Fairness
- Accountability

## RESPECT FOR DATA

- Privacy
- Security
- Confidentiality



## SOUND APPLICATION

- Transparency
- Reproducibility of process and results

## SOUND METHODS

- Quality learning data
- Valid inference
- Rigorous modeling
- Explainability

Trustworthy insight from responsible machine learning processes

## The Government of Canada Directive on Automated Decision-Making and Algorithmic Impact Assessment (AIA)



# Keys to building data science capacity

## TRUST

Deliver concrete results while adhering to high ethical standards at all times to build trust in data science methods



## QUALITY

StatCan's data science methods follow rigorous practices, including internal reviews of projects, to ensure high-quality results and valid statistical inference



## INNOVATION

StatCan's data scientists are committed to identifying and adopting the latest data science practices to deliver fast results



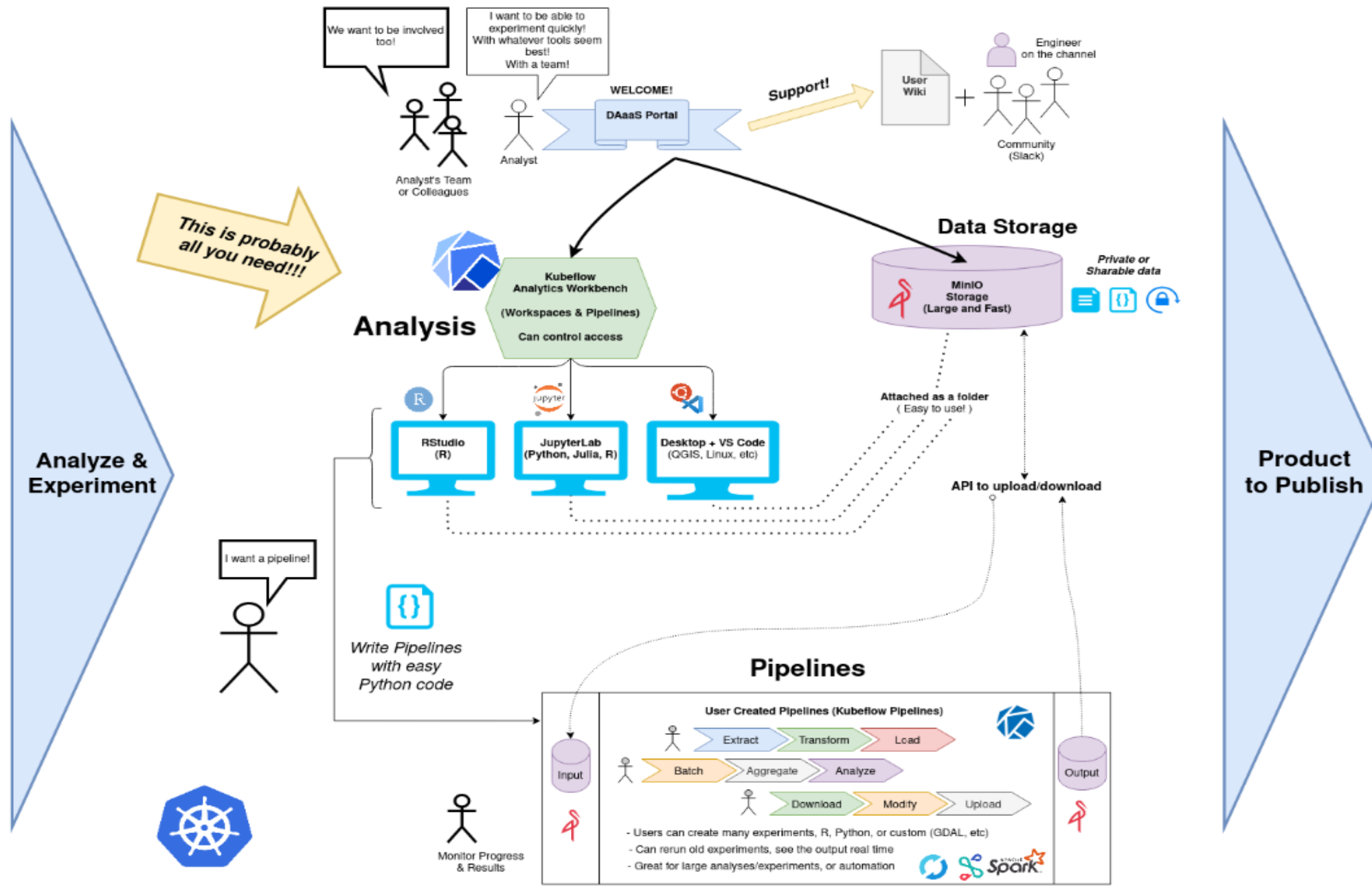
## COLLABORATION

StatCan is working with partners across the Government of Canada, academia, international partners and other members of the data science community to learn from one another and share leading-edge data science methods



# Enablers: Co-Creation & Cloud Environment

*DAaaS (Data Analytics as a Service)  
Advanced Analytics  
Workspace  
Infrastructure*



# Enablers: Data Science Network for the Federal Public Service

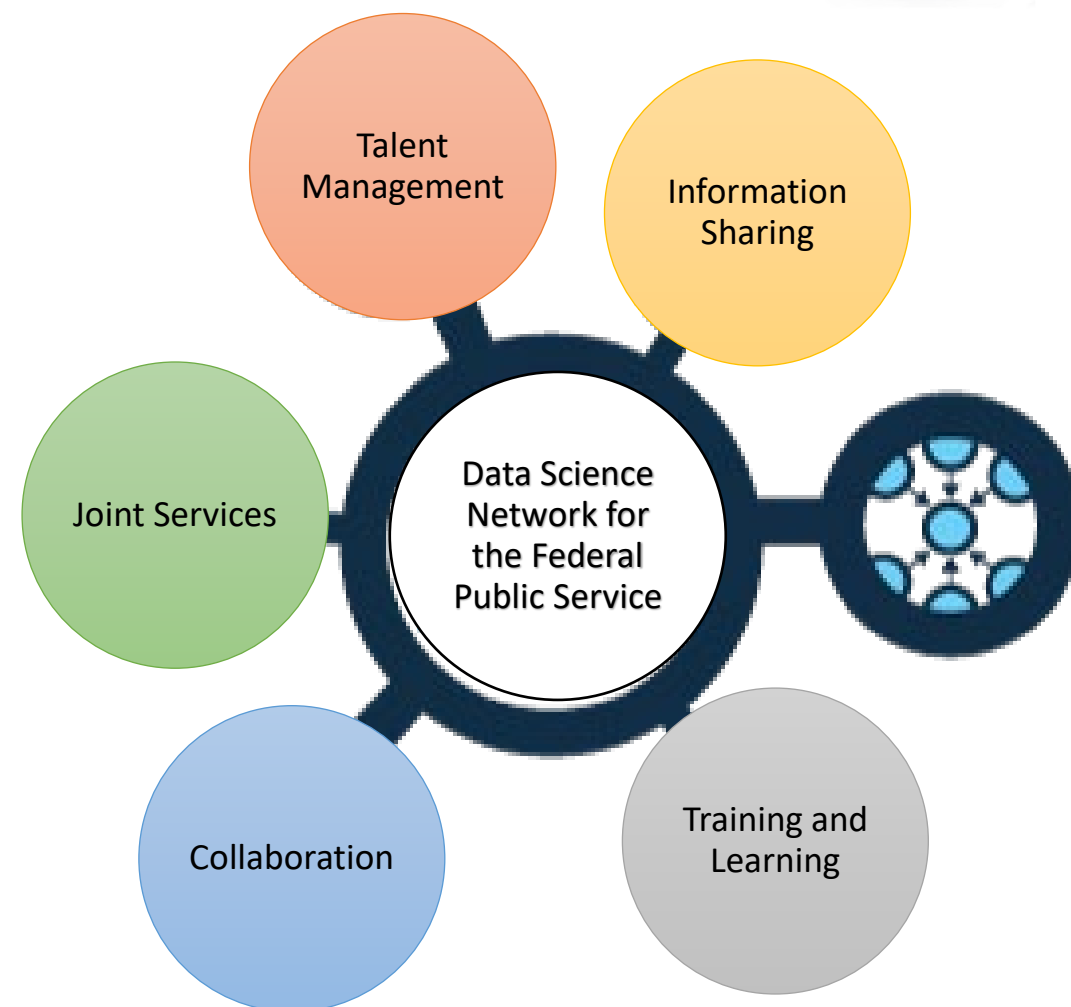
Goal: Establish the foundations of a public-service-wide data science ecosystem to take advantage of synergies, solve concrete problems, share knowledge and expertise.

The network has grown to have more than 2100 members representing more than 480 departments, educational institutions, and private organizations.

You can find Data Science Resources and sign up for our newsletter at

<https://www.statcan.gc.ca/eng/data-science/network>

to get the latest training news, tools you may find helpful, communities of practice and more!



POWERED BY  
STATCAN STEWARDSHIP

**Merci!**  
**Thank You!**

**Questions?**

[sevgui.erman@statcan.gc.ca](mailto:sevgui.erman@statcan.gc.ca)

