Migrating legacy code to R: A team-based approach to open-source adoption

Joshua Gutoskie (he/him)
Senior Methodologist
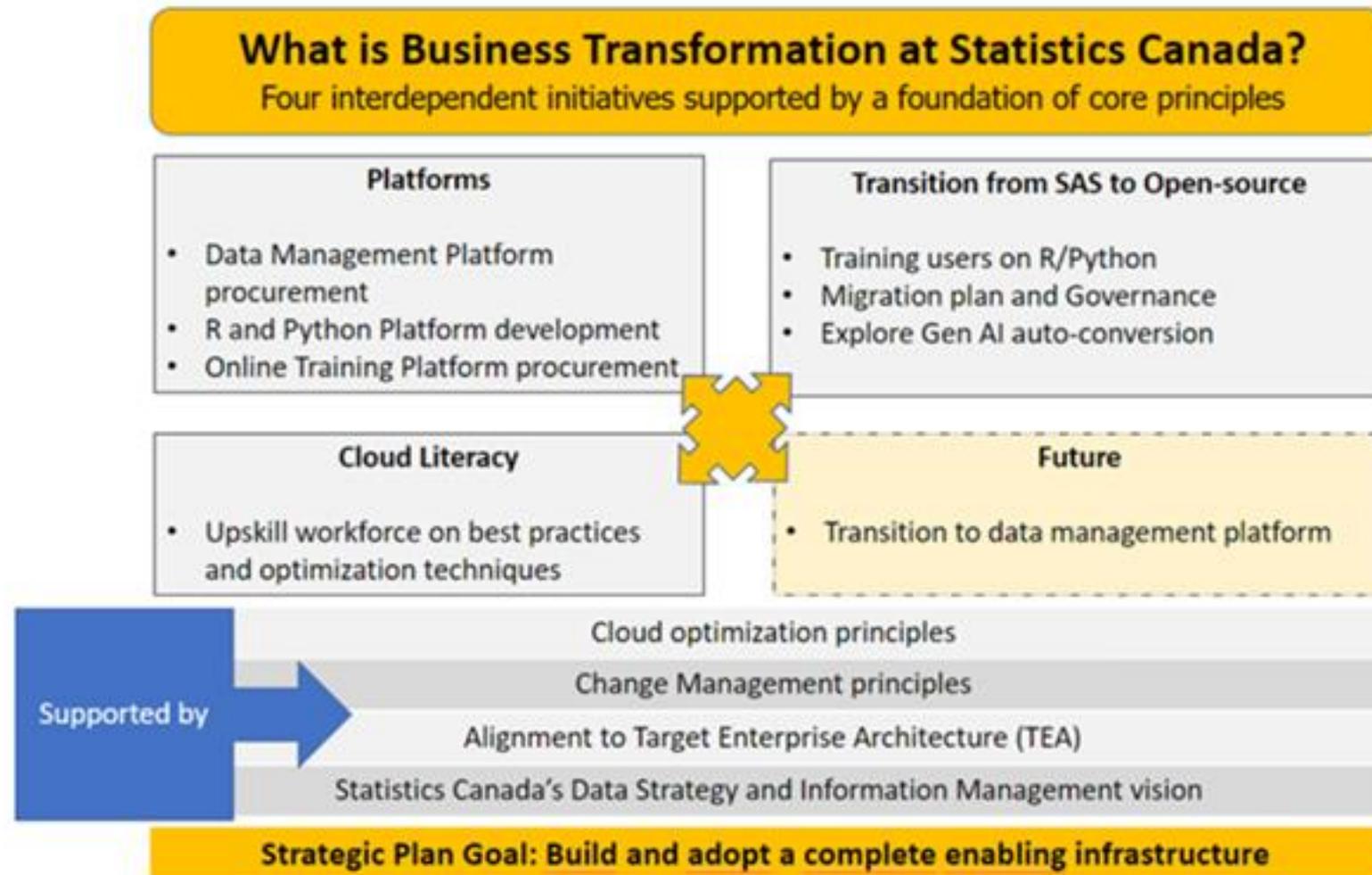Statistics Canada

Canada

Overview

- Open-source migration at Statistics Canada

- What is a survey methodologist?

- Migration philosophy and challenges

- Team-based code migration

- Lessons learned

Open-source migration at Statistics Canada

Statistics Canada

- Canada's central statistical agency

- More than 450 active surveys

- Deliver relevant information to Canadians

- Committed to protecting confidentiality of all information

Open-source migration at Statistics Canada

What is Business Transformation at Statistics Canada?
Four interdependent initiatives supported by a foundation of core principles

Platforms

- Data Management Platform procurement
- R and Python Platform development
- Online Training Platform procurement

Transition from SAS to Open-source

- Training users on R/Python
- Migration plan and Governance
- Explore Gen AI auto-conversion

Cloud Literacy

- Upskill workforce on best practices and optimization techniques

Future

- Transition to data management platform

Supported by

- Cloud optimization principles
- Change Management principles
- Alignment to Target Enterprise Architecture (TEA)
- Statistics Canada's Data Strategy and Information Management vision

Strategic Plan Goal: Build and adopt a complete enabling infrastructure

What is a survey methodologist?

Survey methodologist

Design & Implement Statistical Methods

▸ Surveys, Censuses, Alternative Data

Develop Innovative Solutions

▸ Big Data, Machine Learning, Modeling

Multidisciplinary Collaboration

▸ Cross-Functional Teamwork

Solve Complex Problems

▸ Analytical & Creative Solutions

Migration philosophy and challenges

Why we migrated to R?

- Cost savings

- Extensive package ecosystem

- Community-driven development

- Modern tool integration

Migration philosophy

- Match the output, not the code

- Prefer clarity over cleverness

- Find common approaches when applicable

- Treat migration as modernization

Challenges

- Data integrity
- Complex legacy logic
- Capacity building
- Organisational challenges

Team-based code migration

Team-based code migration

- Employee-led Planning
  - Employee-led workshops/planning sessions for common processes
  - What are we currently doing?
  - What works? What doesn't work?
  - What improvements would we like to see?

Team-based code migration

- Open communication

  - Use a common platform to communicate changes, updates, etc.

    - GitLab/GitHub

  - Hold open forums to discuss progress

    - Demos

  - Be mindful of open-source migration in different teams

    - Participate in other working groups

    - Share experiences

Team-based code migration

- Technical strategy

  - Modular code

  - Consistent standards/styles

  - Shared R packages

  - Robust testing framework

Team-based code migration

- Training and capacity building
  - Make training platforms available
  - Peer code review
  - Mentoring and user groups

Lessons Learned

Lessons learned

- Adaptation is essential

- Start migration efforts early

- Find a community

STAY CONNECTED

Eh Sayers podcast

StatsCAN Plus

The Daily

My StatCan

Website

Surveys and statistical programs

Data service centres

Questions? Contact us: infostats@statcan.gc.ca

STATISTICS CANADA

Delivering insight through data for a better Canada