



HAND IN 01101000 01100001 01101110 01100100 WITH THE MACHINE: A ROADMAP TO QUALITY

Caroline-Soledad Mallette, Director of Innovation,
Government of Canada's Translation Bureau



Public Services and
Procurement Canada

Services publics et
Approvisionnement Canada

Canada 



BACKGROUND



THE CANADIAN LANGUAGE INDUSTRY

- Has an estimated value of USD 1.2B.
- Employs an estimated 27,500 Canadians on a part-time, contract or as-needed basis.
- About 75% of its businesses have fewer than 10 employees; 1% have 100 employees or more.

THE TRANSLATION BUREAU: THE GC'S CENTRE OF EXCELLENCE

- Provides optional translation, interpretation and terminology services in official, Indigenous, foreign and signed languages.
- Serves Parliament, the judiciary and federal departments and agencies, mostly on a cost-recovery basis.
- Is ranked 15th on CSA Research's Top 100 Language Service Providers list for 2022, with USD 154M in revenues.
- Outsources ~45% of its business volume.



**BOASTING AN ORGANIZATIONAL
INNOVATION TEAM:
LICENSED TO TRY!**

1,300 EMPLOYEES
NATIONWIDE

28,000 HOURS OF
INTERPRETATION IN 2021-2022

360M WORDS TRANSLATED
IN 2021-2022



WHERE WE WERE A DOZEN YEARS AGO



Aging request management system



Data silos



Fractured TM (~600 textbases)

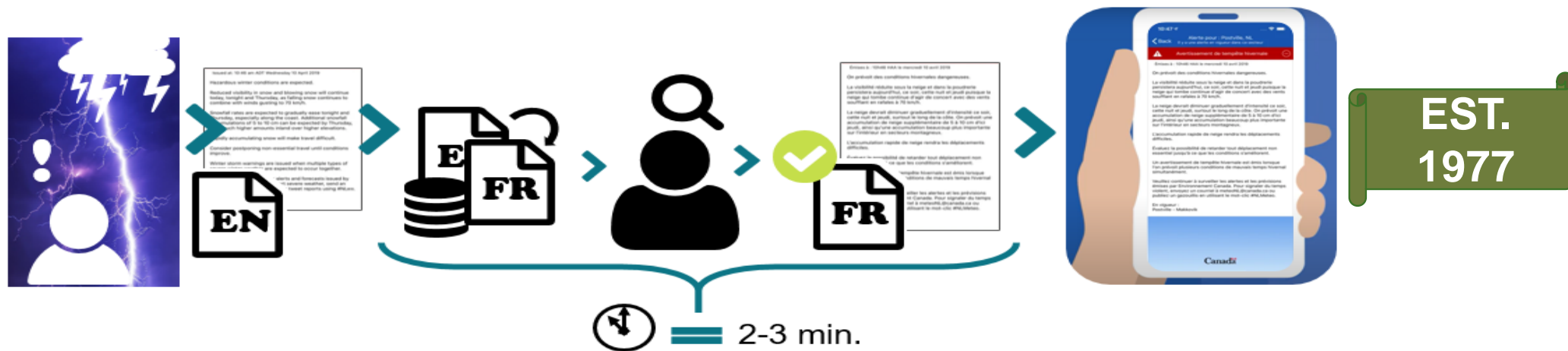


Inconsistent, human-intensive workflows



Lack of technological know-how and skillsets

A BRIGHT SPOT: MACHINE TRANSLATION OF WEATHER ALERTS





BY LEAPS AND BOUNDS

2010

Creation of a unified “megacorus” with a standard workflow and custom analysis tools.



2016

Launch of the *Language Comprehension Tool*, a machine translation tool for federal public servants.



2018

Launch of trials with clients and benchmarking pilots using commercial NMT tools.



2019

Procurement of GClingua, a COTS, cloud-based, holistic request management solution.



2020

Implementation of structured proof-of-concept projects for in-depth analysis of various approaches.



A charted path to the future



4



Public Services and Procurement Canada

Services publics et Approvisionnement Canada

Canada










FRUITFUL PROOFS OF CONCEPT

King Kong (7M words)		<ul style="list-style-type: none">Manual processing—baseline
Turkish Delight (3M words)		<ul style="list-style-type: none">Optimized processing, advanced analytics, automated packaging (daily workload)Traditional TM
Scientific abstracts (700 docs)		<ul style="list-style-type: none">Intento NMT hubCustom-trained domain NMT only
1mill22 (1M words)		<ul style="list-style-type: none">Optimized processing, advanced analytics, automated packagingDomain TM, Intento NMT hub
Legal (700,000 words)		<ul style="list-style-type: none">Domain identification, metadata enrichment, content sectioningCustom TMs, MS Collab custom NMT and Intento NMT hub, with advanced analytics and processing, packaging, terminology extraction and Termium sync
Immigration and Refugee Board of Canada (10M classified words)		<ul style="list-style-type: none">Custom anonymization of training data to train MS Collab custom NMT



LESSONS LEARNED: GOVERNMENTS HAVE SPECIFIC ADMINISTRATIVE HURDLES THAT MUST BE OVERCOME



-  **Budget:** Strong advocacy is needed to put innovation at the forefront of the spending agenda.
-  **Procurement:** An agile approach is needed to keep pace with progress (today's best-in-class is tomorrow's straggler).
-  **Governance:** Working-level SMEs must be empowered to make decisions or supported by a nimble decision-making structure.
-  **Security:** Specific safeguards must be put in place to protect the public and national interest without impeding innovation.
-  **Workflows:** Processes must be aligned as closely as possible with the private sector to make the most of COTS solutions.
-  **HR:** Public servants with the right skillset for AI innovation are scarce, and attracting the best and brightest is difficult.
-  **Culture:** Strong change management is needed to ensure buy-in.

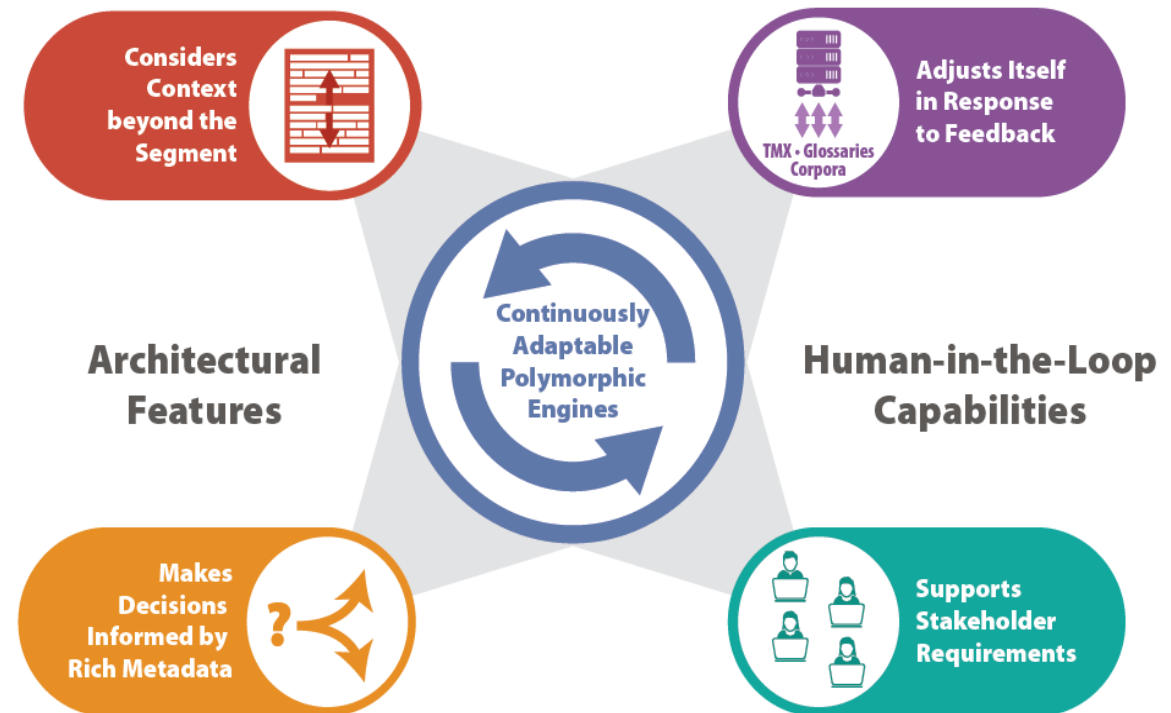




LESSONS LEARNED: ONE SIZE DOES NOT FIT ALL

- Specialization is as much of an asset for MT as it is for translators.
- Generic MT systems show limited efficiencies in specialized domains.
- We need a responsive approach to MT.

Responsive MT Inputs and Capabilities



© CSA Research



Public Services and
Procurement Canada

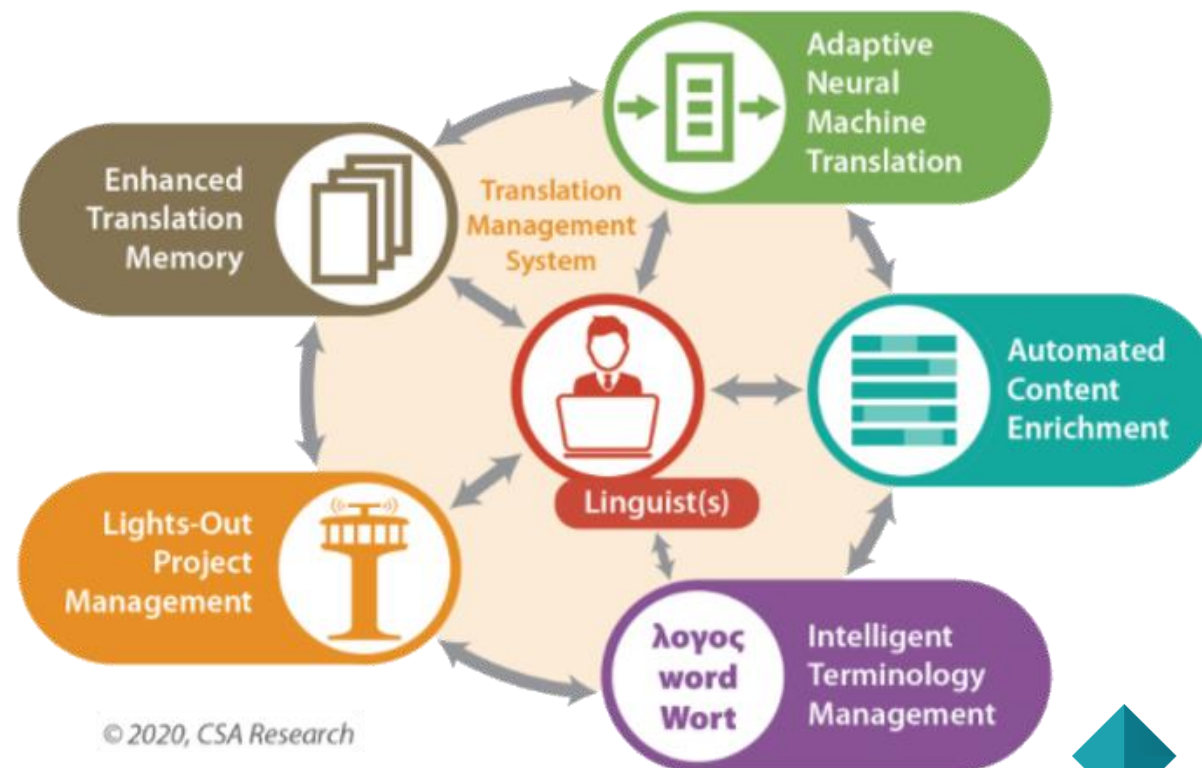
Services publics et
Approvisionnement Canada

Canada



LESSONS LEARNED: KEEP HUMANS IN THE LOOP

- Equal content and quality in English and French is an obligation for the Government of Canada.
- Volumes far exceed human capacity, yet machines cannot provide sufficient quality → a hybrid model is required.
- Humans work with technology but remain in control of the process. They focus on the difficult aspects that machines cannot handle, rather than on low-value tasks.



LESSONS LEARNED: PROVIDE OUR PEOPLE WITH THE RIGHT TOOLS




PROJECT MANAGERS HAVE THE FEWEST TOOLS AVAILABLE TO THEM



The industry has focused on tools for language professionals...



...leaving PMs ill-equipped to make pivotal decisions at the start of a project.



PMS MUST BE PROVIDED WITH TOOLS TO SUPPORT...

- Clean import and segmentation
- Content identification
- Content distribution
- Workflow selection
- Lossless slicing and packaging
- Fluid timeline planning and adjustment
- Assignment of tasks to the most suitable resources (humans or machines)
- Real-time status updates
- Real-time communication
- Detailed reporting
- Improved efficiency

LESSONS LEARNED: PROVIDE OUR PEOPLE WITH THE RIGHT INFORMATION



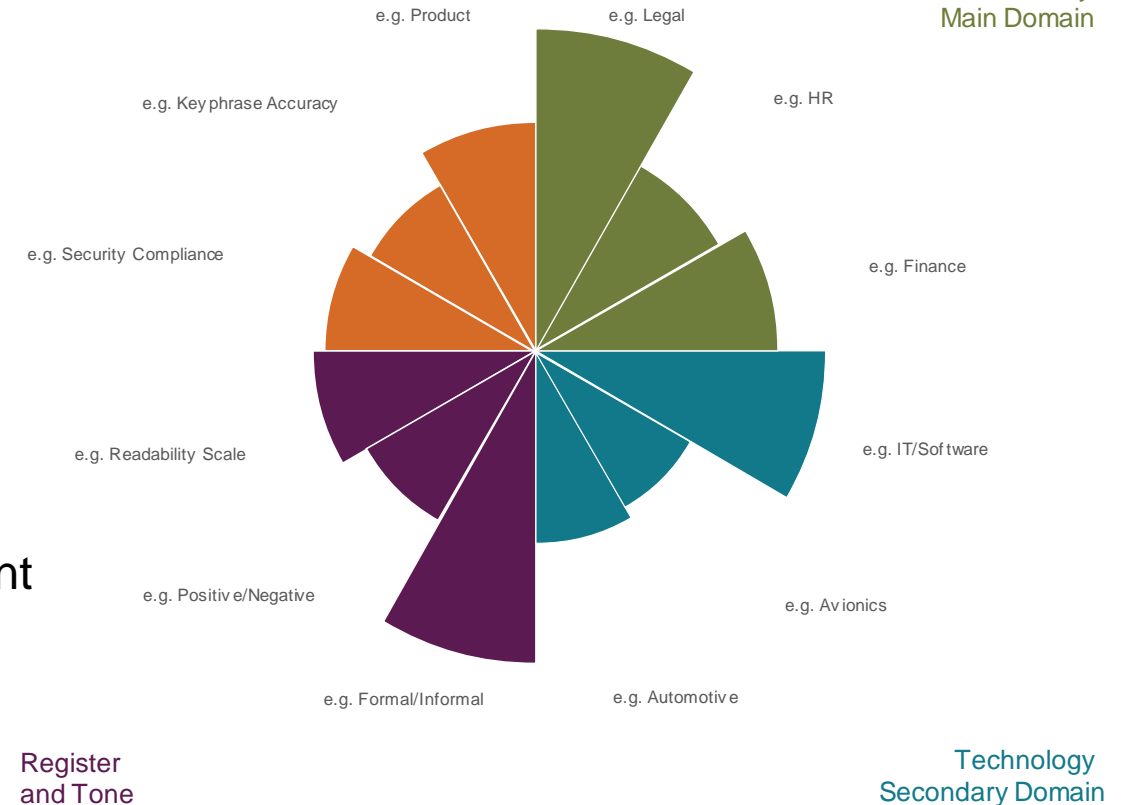
FIRST: PROPERLY AND THOROUGHLY IDENTIFY THE SOURCE CONTENT

- By segment NOT by project
 - Domain and subdomains
 - Register and tone
 - Client or product-specific preferences

THEN: STOP THE DATA LOSS

- Properly tag and add metadata to the source content, and enrich as we go.
- Make sure all data remains in the target document
 - For information distribution
 - reusable metadata in source AND in target
 - For content searches and identification
 - For NMT training

Criteria Set 3

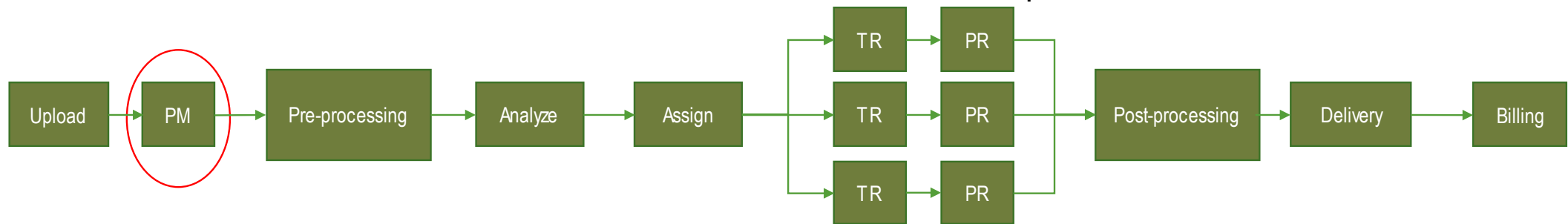


LESSONS LEARNED: PROVIDE OUR PEOPLE WITH THE RIGHT TRAINING AND MINDSET



1) KNOWING ONE'S PLACE IN THE WORKFLOW

- What are the tasks assigned?
- What is my own role?
- What effect does my work have downstream?
- What can I do to make the next step easier?



2) KNOWING HOW TO USE THE MACHINE



3) KNOWING WHEN TO USE THE MACHINE

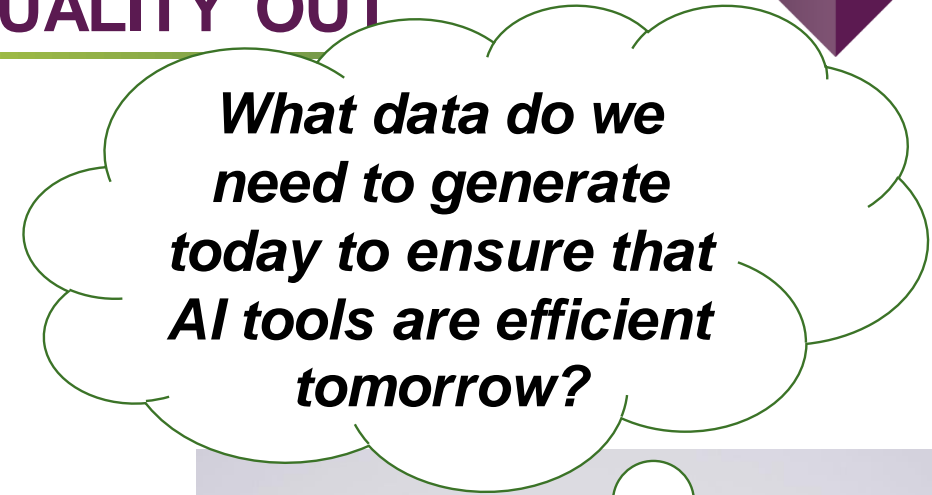
- What type of text is it?
- What research will need to be done?
- What is the visibility/lifespan of the document?
- Could undetected errors have serious consequences?
- Is the source text well written? Is adaptation required?
- What is the deadline?

4) THINKING FORWARD, STAYING AGILE AND BROADENING ONE'S HORIZONS



LESSONS LEARNED: QUALITY IN, QUALITY OUT

- If we instruct post-editors to focus on “good enough” quality, how good will our training data for NMT be 5 years from now?
→ degraded NMT quality by design.
 - Overedit by design to have better quality for tomorrow.
 - Enhance data collection and stop the data loss (enrich metadata).
 - Optimize corpora to retain only data that will not mislead the AI (e.g. remove single-word segments).

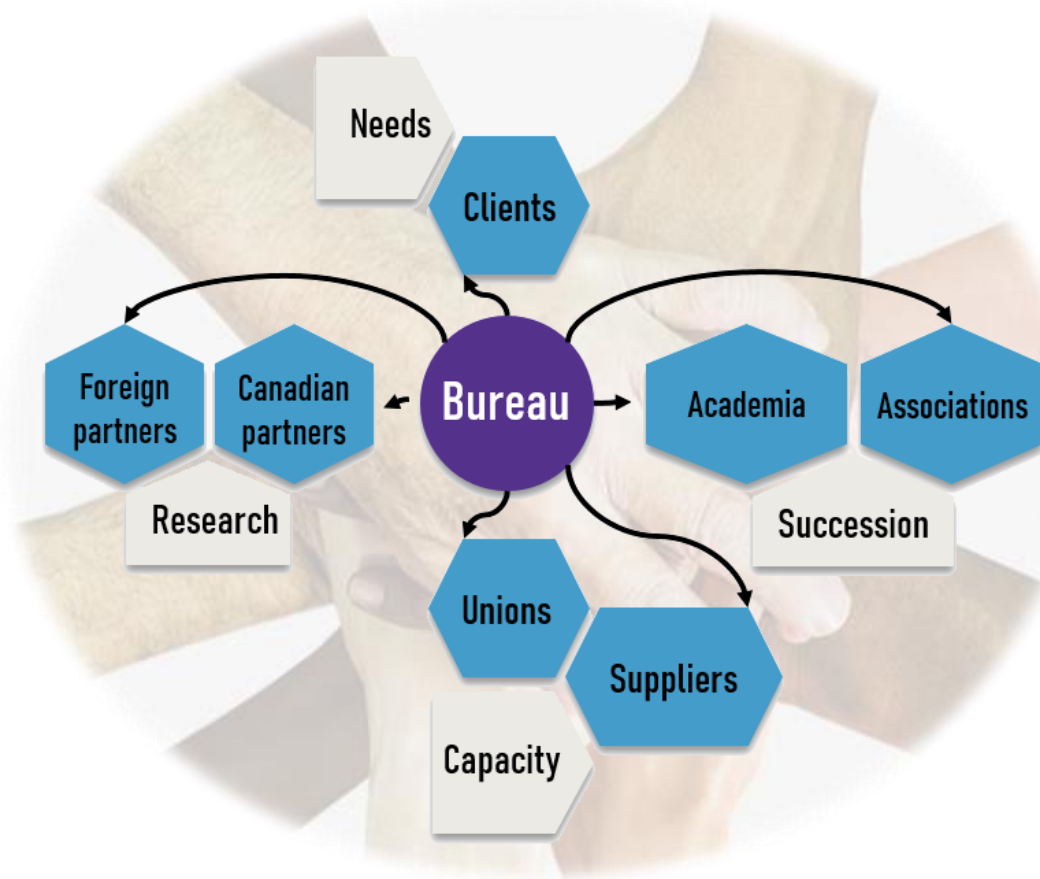


One overarching goal: QUALITY





LESSONS LEARNED: NETWORKING IS KEY



The Translation Bureau is applying a teamwork approach and actively reaching out to partners domestically and abroad to:

- Identify needs
- Share expertise
- Find innovative solutions to challenges
- Plan for the future



QUESTIONS OR COMMENTS?



Caroline-Soledad.Malette@tpsgc-pwgsc.gc.ca



14



Public Services and
Procurement Canada

Services publics et
Approvisionnement Canada

Canada 

APPENDIX / SOLID FOUNDATIONS: A COMPREHENSIVE AND FORWARD-LOOKING STRATEGY FOR AI



AMBITIOUS GOALS

Centre of excellence in leveraging AI for quality
Better, faster and cheaper services
Support for OGDs with AI projects
Creating trust throughout the continuum
Alignment with GC statutes, policies and priorities

GUIDING PRINCIPLES

Support (rather than replace) humans
Invest in people to create trust
Prepare our employees for the future of work
Start small: experiment and build
Partner with leaders in the field
Plan for bumps but don't wait for perfection

FOUR PILLARS

CONTENT AND
DATA

WORKFLOW

USER TRUST

PARTNERSHIPS

Decisions cannot be made without having specific human intervention points during the decision-making process, and the final decision must be made by a human.

Requirement of the *Directive on Automated Decision-Making* for decisions with a high impact on the rights of individuals

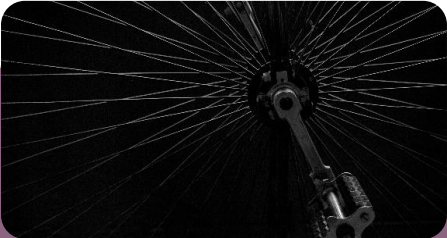


APPENDIX / SOLID FOUNDATIONS: A STRONG DATA MANAGEMENT STRATEGY



PEOPLE AND CULTURE

- Identify the Bureau's data requirements
- Establish a data governance committee
- Form a network (hub) of data people
- Assess data literacy and recommend training



ENVIRONMENT AND DIGITAL INFRASTRUCTURE

- Assess the required data environment and digital infrastructure
- Build a data warehouse
- Develop a cloud-based infrastructure
- Develop a modern analytics system for near-real-time reporting



DATA AS AN ASSET

- Build a data analytics centre to support fast and confident decision-making
- Develop and train AI tools for cost reduction
- Generate new data for holistic reporting and quality improvement

APPENDIX / HOW TO DO BUSINESS WITH THE BUREAU



Translation services in Canada's official languages

- The Translation Bureau has a permanent Request for Supply Arrangements process posted on Buyandsell.gc.ca under which suppliers can apply at any time in order to qualify to fulfill requirements in official languages translation. Arrangements received over a calendar year are evaluated quarterly. Details are available at buyandsell.gc.ca/procurement-data/tender-notice/PW-ZF-526-40507.

Terminology services, translation services in Indigenous and foreign languages, and interpretation services in Canada's official languages, Indigenous languages, foreign languages and signed languages

- Visit the Translation Bureau Supplier Info website at <https://www.tpsgc-pwgsc.gc.ca/bt-tb/services/accueil-home-eng.html> for specific guidelines.

The Canadian Content Policy applies—see
<https://buyandsell.gc.ca/policy-and-guidelines/supply-manual/annex/3/6>.



17



Public Services and
Procurement Canada

Services publics et
Approvisionnement Canada

Canada