

Parallel session of at the
International Seminar on Data Science for the
Statistical and Transport Communities

DSL Playbook

2ND SPRINT OF THE DATA SCIENCE LEADERS NETWORK

SESSIONS 3 AND 4
(24 JANUARY 2024)

Overall structure of the playbook

- ▶ **Section 1:** Leveraging basic tools of data science for immediate efficiency gains in NSO operations
- ▶ **Section 2:** Generating additional insights in response to emerging needs
- ▶ **Section 3:** Full transformation of official statistics through digitalization
- ▶ **Section 4:** Cross-sectional themes

Organization of the sprint

▶ Sprint Sessions 1-3:

- ▶ Brief introduction by Facilitator
- ▶ 2-3 selected case studies + 15 min discussion
- ▶ Breakout groups
 - Key concepts and methodologies to be covered
 - Ideas for resources and references
 - Suggestions for practical examples
 - Contributors

Organization of the sprint

- ▶ **Sprint Session 4:**
 - ▶ Summary of cross-cutting issues identified in previous sessions
 - ▶ Case study presentations
 - ▶ Plenary discussion
- ▶ **Takeaways and Way Forward**





Section 3: Full transformation of official statistics through digitalization



3.1 Strategic leadership and support

- ▶ Securing top-level endorsement and political support for digital transformation initiatives within NSOs.
- ▶ Importance of forming strategic partnerships, such as with tech companies, for a successful digital transformation

3.2 Advanced data science tools and methods

- ▶ Identifying the right data science and big data tools for enhanced statistical production.
- ▶ Exploring new methods and sources for data collection.
- ▶ Understanding data science capabilities from a subject matter perspective.

3.3 Technology implementation and integration

- ▶ Principles of modern data architectures
- ▶ Building scalable and resilient systems
- ▶ Mitigating disruptions to data existing statistical programmes when introducing new infrastructure, technical capacities and resources
- ▶ Managing compatibility with legacy systems during the digital transformation process

3.4 Data quality management and security

- ▶ Management of potential bias in advanced AI and ML algorithms
- ▶ Strengthening cybersecurity measures in NSOs.
- ▶ Development of sociotechnical models for secure collaborative data science work, ensuring safe application of data science methods

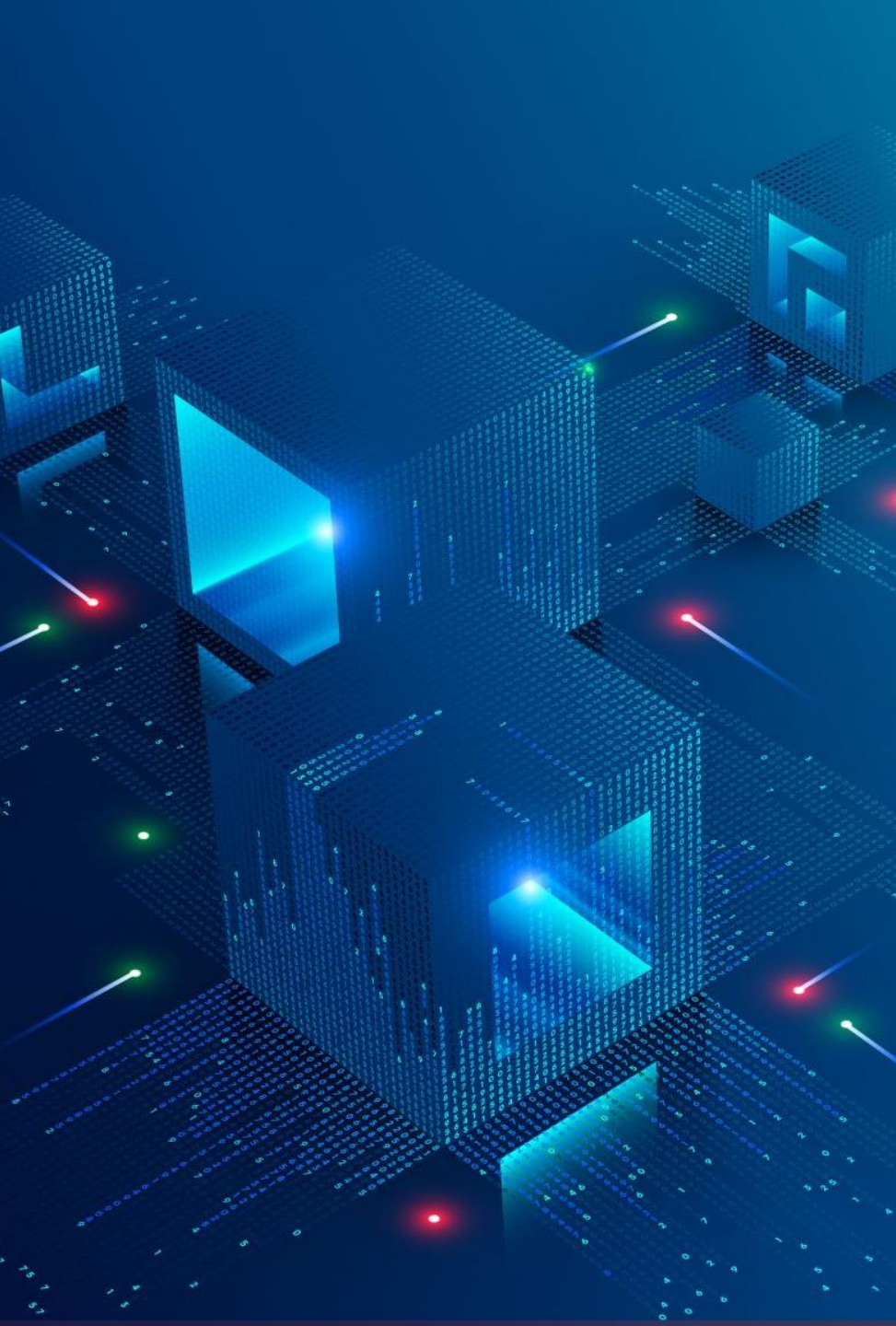
Section 3: Case studies

- ▶ Presentations
 - ▶ **Transformation of the Canadian Consumer Price Index (CPI)**, by Wesley Yung, Acting Director-General, Modern Statistical Methods and Data Science Branch, Statistics Canada
 - ▶ **Full transformation of official statistics through digitalization in Poland**, by Dominik Rozkrut, President, Statistics Poland
 - ▶ **Transformation of the enterprise data architecture and modernization of the IT infrastructure**, by Elio Villasenor, Data Science Lab Director, INEGI, Mexico
- ▶ Discussion:
 - ▶ What were the key factors that contributed to the successful transformation of the NSO in this case studies?
 - ▶ How did the use of advanced data science methods overcome specific challenges in the transformation process?
 - ▶ What could be the long-term implications of such a transformation on the quality and utility of official statistics?
 - ▶ What are the critical considerations for maintaining data security during a full-scale transformation?

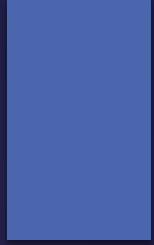
Section 3: Breakout groups

1. Strategic leadership and support
2. Advanced data science tools and methods
3. Technology implementation and integration
4. Data quality management and security

- ▶ What are the most **critical concepts or methodologies** that must be included in this chapter? Why are they essential?
- ▶ Can you suggest any **authoritative materials** that provide comprehensive insights into these topics?
- ▶ Are there any **real-world applications or case studies** that exemplify these concepts effectively? How did they contribute to the success of the project?
- ▶ What would be the most logical and coherent **structure for this chapter**? How can we best present these ideas to ensure clarity and engagement?
- ▶ **Who would like to contribute** to the drafting and editing of this chapter, and what specific aspects are you interested in or experienced with?



Section 4: Cross-cutting issues



Recap of cross-cutting issues identified in previous sessions

- ▶ Human resources and skills building
 - ▶ A successful data science team requires a diverse set of both technical and “soft” skills
 - ▶ Given the fast pace of progress in methods and technology, continuous updating of technical skills is crucial
 - ▶ All members of a data science team need to have basic understanding of a wide set of technical areas, such as data engineering, IT infrastructure, statistics, GIS analysis, data modelling, coding...
 - ▶ It is equally important that all team members have skills in modern project management approaches (e.g., Agile), fundraising, and communication, to ensure effective collaboration and coordination with each other and with external stakeholders.

Recap of cross-cutting issues identified in previous sessions

- ▶ Change management
 - ▶ Introducing data science in official statistics requires significant organizational changes, including new business processes, roles and responsibilities
 - ▶ Data science requires a new emphasis on interdisciplinary collaboration and an agile mindset.
 - ▶ Leaders / senior management need to model desired behaviors
 - ▶ Staff members' goals and incentives need to be re-aligned with new responsibilities and expectations

Recap of cross-cutting issues identified in previous sessions

- ▶ Data privacy and security
 - ▶ Data science projects need strict data privacy and security protocols in place, especially when linking records from different data sources through common identifiers of individuals, households, businesses or geographies.
 - ▶ This is fundamental to avoid any reputation damage and ensure NSO retains the trust of all stakeholders

Recap of cross-cutting issues identified in previous sessions

- ▶ Data quality as a driver of trust in data science for official statistics
 - ▶ It is crucial to establish trust in the results of data science projects, particularly from key users such as policy and decision makers.
 - ▶ The reliability of the sources and methods involved in the use of data science within NSOs is central to their widespread acceptance
 - ▶ This highlights the importance of transparent and participatory validation
 - ▶ Data science outputs should be explainable and reproducible by independent reviewers

Recap of cross-cutting issues identified in previous sessions

- ▶ Communicating results of data science programmes in official statistics
 - ▶ Importance of communicating early and frequently the applicability, value and limitations of data science in official statistics
 - ▶ Need to avoid the risk that the outputs of data science projects in NSOs remain underutilized due to inadequate communication
 - ▶ Developing communication strategy can maximize the reach and impact of statistical outputs based on data science
 - ▶ Ensure that the results are accessible to a diverse audiences with varying levels of technical expertise.

Section 4: Case studies

- ▶ Presentations
 - ▶ **Cultural change**, by Yves Jaques, Chief of Frontier Data and Technology Unit, UNICEF
 - ▶ **Institutional arrangements**, by Gogita Todradze, Executive Director, National Statistics Office of Georgia
- ▶ Discussion:
 - ▶ What were the key factors that contributed to the successful transformation of the NSO in this case studies?
 - ▶ How did the use of advanced data science methods overcome specific challenges in the transformation process?
 - ▶ What could be the long-term implications of such a transformation on the quality and utility of official statistics?
 - ▶ What are the critical considerations for maintaining data security during a full-scale transformation?

Section 4: Plenary discussion

- ▶ What are some emerging or **underrepresented cross-cutting themes** that deserve a dedicated chapter in the Playbook?
- ▶ **Why are these themes crucial** for the effective integration of data science into official statistics?
- ▶ Can you provide **specific examples or case studies** that effectively illustrate these themes?
- ▶ What **external resources, experts, or institutions** could we leverage to enrich the content and perspective of each themes?

Take aways
and way
forward



Tour-de-table

- ▶ **Personal Takeaways:**

What is your key takeaway from this sprint, and how has it shaped your perspective on data science in official statistics?"

- ▶ **Reflections on Improvement:**

Are there aspects of the sprint that could be improved in future sessions, particularly concerning sustainability, inclusivity, or other areas?"

- ▶ **Volunteering and Contribution:**

Which specific section or chapter of the Playbook are you interested in contributing to, and in what capacity (e.g., writer, editor, researcher...)?



Next steps
and closing
remarks.