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

DSLN Playbook

2ND SPRINT OF THE DATA SCIENCE LEADERS NETWORK

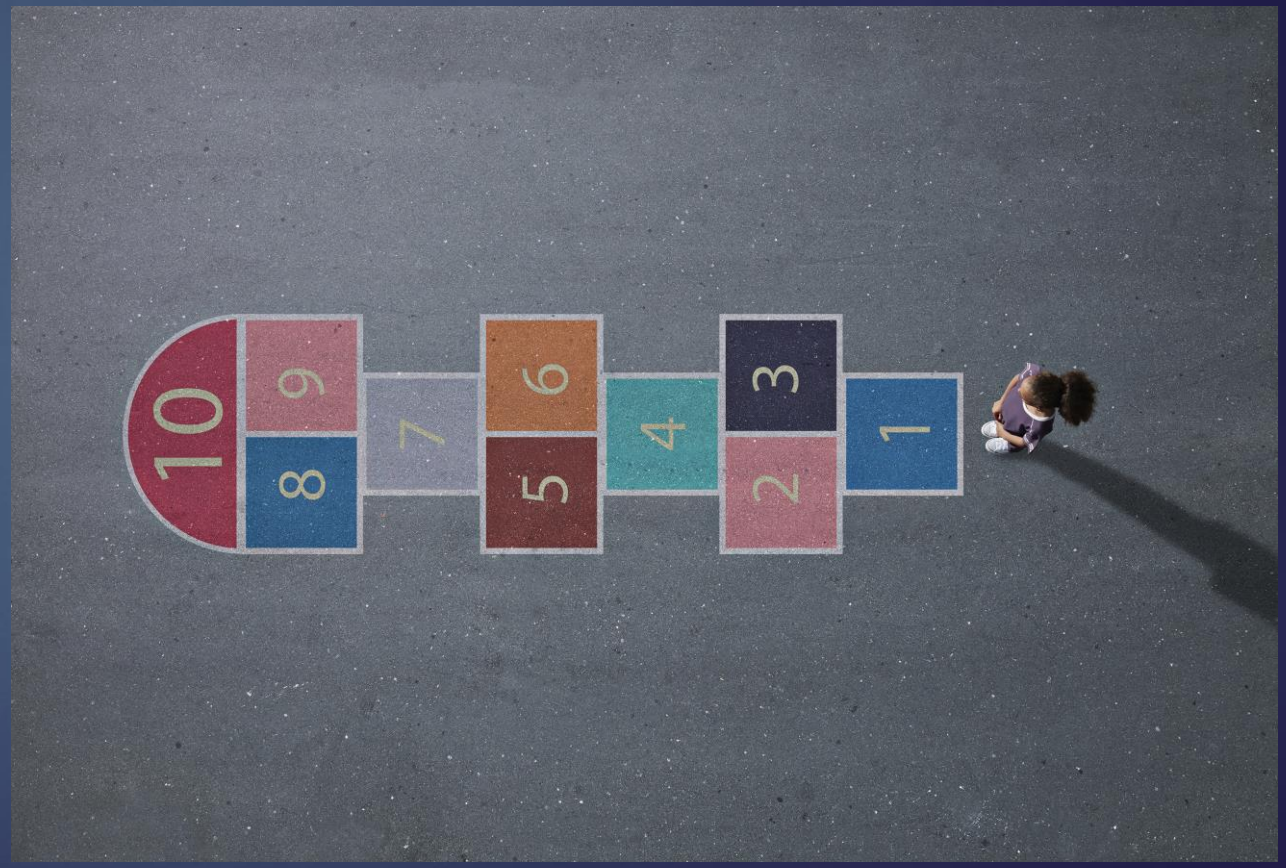
SESSIONS 1 AND 2
(23 JANUARY 2024)

How did we get here?

- **1st DSLN Sprint:** Strategic recommendations
- **Bureau meeting:** Endorsed idea of the playbook
- **Webinars**: Provided initial ideas for structure and content

Why a Data Science “playbook”?

- **Action-oriented**
- **Scalable and customizable**
- **Accessible**



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

Design principles for the playbook

- **Purpose and scope:** Clearly define what the playbook is intended for and its limitations
- **Multi-level approach:** Start with content suitable for a basic level of capability, with options to delve into more advanced topics.
- **Inclusive:** Make content approachable for both traditional statisticians and new data science practitioners from different national and regional contexts
- **Diverse NSO experiences:** Highlight experiences from different-sized NSOs, acknowledging those already working with data science.
- **Learning from failures:** Share lessons from significant failures to provide comprehensive insights.

Design principles for the playbook

- **Persona-focused:** Tailor content to different personas involved in the data science life cycle. Outline user-journeys to guide readers through the playbook based on specific roles.
  - Subject matter expert / analyst
  - Data scientist
  - Data engineer
  - Statistical programme / business line manager
  - Senior manager (e.g., chief statistician)

Types of content and format

- **Use cases:** Incorporate practical use cases to illustrate concepts.
- **General resources:** Include concise summaries for senior policymakers and materials catering to university students.
- **Practical resources:** Include links to practical resources like code examples in GitHub, Kaggle.
- **Summaries and quick wins:** Provide summaries with main findings and examples of quick wins for easy reference.
- **Digital format:** Develop the playbook as an online, digital product, allowing for iterative updates and improvements (e.g., using

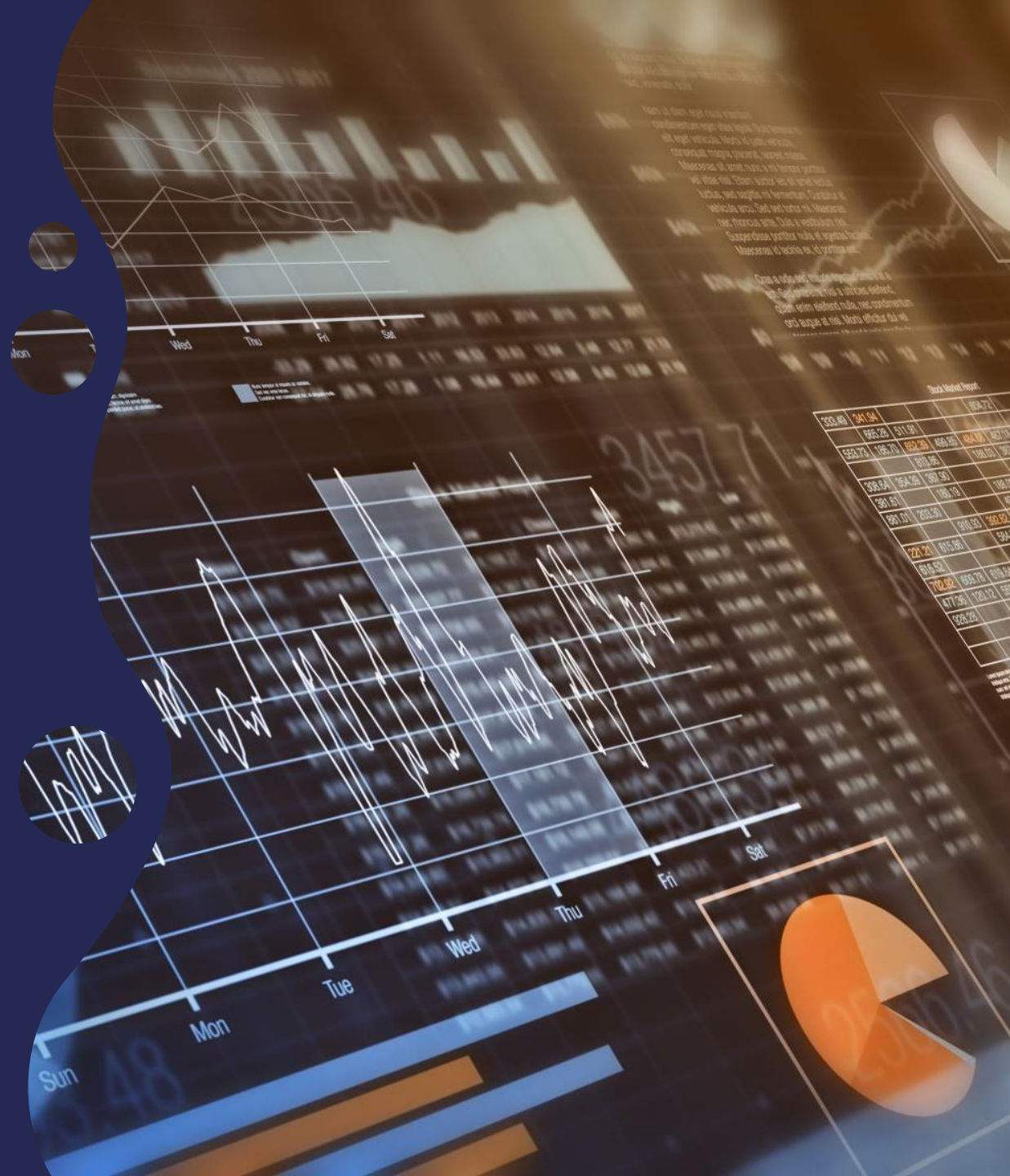
Organization of the sprint

- **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 1: Efficiency gains and automation



1.1 Introduction to data science for NSOs

- Definition of data science within the NSO context
- Role of data science in enhancing statistical processes and outputs
- Critical gaps in current NSO operations where data science can help

1.2. Automation and process optimization

- How to develop automated pipelines and data workflows to streamline NSO operations
- Automating data collection, processing and dissemination.

1.3 Adopting new tools and technologies

- Selection and implementation of accessible data science tools in NSOs, including open-source software and light-weight analytics platforms.
- Importance of training and development for non-data scientists to enhance their ability to utilize these new tools.
- Guidance on the systematic identification of small-scale data science projects that can demonstrate quick value to NSOs.

1.3 Adopting new tools and technologies

- Selection and implementation of accessible data science tools in NSOs, including open-source software and light-weight analytics platforms.
- Importance of training and development for non-data scientists to enhance their ability to utilize these new tools.
- Guidance on the systematic identification of small-scale data science projects that can demonstrate quick value to NSOs.

1.4. Cultural and process change management

- Tackling resistance by team members towards process streamlining.
- Emphasizing the advantages of automation, such as quality assurance, process documentation.
- Clarifying the positive impact of automation on job security.
- Distinguishing between coding and hacking.

Section 1: Case studies

- Presentations
  - **Reproducible Data Pipelines**, by Jeroen Minderman, Senior Data Scientist (TP), Data Science Campus, ONS UK
  - **Data linkage project of post enumeration survey and census,** by Ivan Murenzi, Deputy Director General, National Institute of Statistics of Rwanda
- Discussion:
  - How could the basic data science tools used in this case study be adapted or expanded to address similar challenges in other NSO operations?
  - What are some potential obstacles when implementing these data science tools in NSO operations, and how might they be overcome?
  - Can you identify any opportunities for scaling or enhancing the efficiency gains witnessed in this case study to other areas of NSO work?
  - What lessons can be learned from this case study about integrating data science tools into traditional statistical processes?

Section 1: Breakout groups

1.1 Introduction to data science in NSOs

1.2 Automation and process optimization

1.3 Adopting new tools and technologies

1.4 Cultural and process change management

- 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 2: Generating insights in response to emerging needs

2.1 Rapid response and data integration

- Efficient data management strategies for NSOs
- Integrating multiple data sources for dynamic and timely statistical outputs.
- Importance of integrating geospatial data and other innovative types of data sources, for nowcasting and experimental statistics.
- Enhancing knowledge building and sharing in using non-traditional data sources.

2.2. Skill development and capacity building

- Essential data science competencies for NSO staff
- Skills required to adapt to rapidly changing data needs
- Strategies for skill enhancement and knowledge sharing
- Developing training plans for methodologists and statisticians.

2.3 Data science partnerships and collaboration

- How to develop partnerships and encourage collaboration beyond NSOs
- Engagements with the private sector and academia to foster knowledge exchange and innovation.
- How NSOs can be more aware of real-world needs.
- Fostering better cooperation with private sector data holders.

2.4 Quality frameworks

- Establishing robust quality management systems in NSOs
- Maintaining data integrity when introducing new data, methods and technologies.
- Establishing quality frameworks for rapid response, especially when dealing with non-traditional data sources.

2.5 Resource mobilization

- Strategies for securing funding and resources for data science initiatives
- Innovative approaches to resource mobilization.
- Engaging leadership and stakeholders to gain support and buy-in for data science initiatives
- Mobilize leadership support and appropriate resource allocation.
- Secure better access to data.

Section 2: Case studies

- Presentations
  - **Improving the quality of transport statistics,** by Bertrand Loison, Vice-Director General Switzerland)
  - **Web scrapping and NLP for Business Statistics,** by Barteld Braaksma, Innovation Manager, CBS Netherlands
- Discussion:
  - How did the integration of diverse data sources in this case studies enhance the ability to respond rapidly to new analytic requirements?
  - What challenges might arise when trying to replicate this approaches in different contexts or with different types of data?
  - In what ways could the methodologies used in these case studies be improved or innovated upon for even more effective outcomes?
  - How can the insights generated from these approaches inform policy decisions, and what are the potential limitations?

Section 2: Breakout groups

2.1 Rapid response and data integration

2.2 Skill development and capacity building

2.3 Data science partnerships and collaboration

2.4 Quality frameworks

2.5 Resource mobilization and leadership support

- 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?