Statistics Poland

The evolution of data stewardship in official statistics

governance, capacity, best practices and innovation

Dominik Antoni Rozkrut
President
Statistics Poland

Doha Data Forum, 29 October 2023

President of **Statistics Poland**
Adjunct Professor at **University of Szczecin**
President of **International Association for Official Statistics** (2023-2025)
Vice-President of the **Polish Statistical Association**
A member of the Econometrics and Statistics Committee and the Demographic Sciences Committee of the **Polish Academy of Sciences**
A member of **ISI Committee on Agricultural Statistics** (ISI-CAS)

UNECE Conference of European Statisticians (CES), Bureau Vice-Chair
Co-Chair of the UN Working Group on Data Stewardship
Involved also in various other UN groups
- UN Friends of the Chair Group on the Fundamental Principles of Official Statistics
- UN Committee of Experts on Big Data and Data Science for Official Statistics
- UN Network of Economic Statisticians
- UN Friends of the Chair on Social Statistics

Previously a member of
- EU Commission Business-to-Government Data Sharing Expert Group (2019)
- EU Commission Expert Group on Facilitating the Use of New Data Sources for Official Statistics (2021)

Dominik Antoni Rozkrut
President of
Statistics Poland

OPEN DATA WATCH

ODIN
OPEN DATA INVENTORY
by OPEN DATA WATCH

Global Rankings

87 ODIN SCORE

Poland

2nd GLOBAL RANK OUT OF 195

81 COVERAGE SCORE OUT OF 100

92 OPENNESS SCORE OUT OF 100

	Country	Region	Overall	Coverage	Openness
1	Singapore	South-Eastern Asia	90	77	100
2	Poland	Eastern Europe	87	81	92

Statistics Poland

Context

- The main drivers
  - Digital transformation, datafication, new data ecosystems, more data for everyone
- New reality
  - Avalanche of data regulations, data governance, data altruism, data collaboratives, data stewardship
- Challenge
  - Disinformation, decline in trust, budget challenges, and increased complexity of public problems require data innovation

Fundamental Principles of Official Statistics

Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation.

To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Fundamental Principles of Official Statistics and Data Stewardship

- FPOS enshrine well the concept of data stewardship
  - Principle 1
- FPOS didn’t exist before 1990s
  - still things were done in the same spirit then
- The concept of data stewardship emerged into focus recently
  - similarly, hard to say it’s new

Brief history of data stewardship in official statistics

1. Production, publication, dissemination,
2. International comparison, harmonization, information standards
3. Administrative sources, domestic coordination
4. Evidence-based policies, stakeholder engagement
5. Capacity building, development aid

Brief history of data stewardship in official statistics

6. Open data, data sharing
7. Data services, support for data usage, partnerships
8. Statistical literacy, data literacy, education
9. Fostering data ecosystems, national data governance

Data ecosystem

- The data (including statistical data), along with the data subjects, a broad range of stakeholders and data users, capacities, processes, policies, and infrastructure used to capture and analyze data

Data governance

- A system of decision rights and accountabilities for managing the availability, usability, integrity, and security of the data and information and the resulting regulations, policies, and frameworks

Data management

- Involves the development, execution, and supervision of plans, practices, concepts, programs, and the accompanying range of systems that contribute to the organization and maintenance of data processes to meet ongoing information lifecycle needs

Data stewardship in official statistics

- Definition
  - The roles, functions, and activities that enable the re-use of data for public benefit in a systematic, sustainable, and responsible way through data collaboration
- Mission
  - Official statistics should deliver responsible data leaders in the public sector seeking new ways to create public value through cross-sector data collaboration

Data stewardship activities

- Collaborate
  - Working with others to unlock the inherent value of data when it serves the public good
- Protect
  - Managing data ethically and preventing harm to all whose data may be shared
- Act
  - Proactively identifying partners who can unlock value and insights

Data stewardship in the context of official statistics

1. Data governance, management, and strategies
2. Maintaining information standards
3. Coordination of public information systems
4. Quality assistance and assurance
5. Data access, acquisition, and sharing
6. Partnerships and community engagement

Data stewardship in the context of official statistics

7. Nurture data collaboratives
8. Overseeing data life cycle management
9. Assuring ethics, value, and risk assessment
10. Supervising security and promoting transparency
11. Internal coordination and staff engagement
12. Dissemination, communication, and education

Time to act

Statistics Poland

Need to take advantage of the advent of data governance policies

- These are the times of setting up data governance everywhere around the world
- We see an avalanche of legislative acts in the EU, and other countries are following
- It is crucial to take advantage of the opportunity and not to miss the critical moment when the foundations for the functioning of data ecosystems are laid
- Setting up a national and international data governance system is a challenging task; we still learn the language to be used, especially in a legislative context, so I anticipate a lot of difficulties

Need to take advantage of the advent of data governance policies

- Therefore, politicians should work together hand in hand with official statisticians
- Data should be used for the public good, and this must be taken into account in the design of modern data governance systems at the state level
  - Ubiquitous data should serve not only as a fuel for the development of economies but also serve the public good through the availability of fast, reliable, and detailed official statistics, providing insight into the nature and progress of socio-economic processes
- This was the principle that guided the creators of the official statistics system from the very beginning, but new circumstances and the explosion in the amount of data pose new challenges

Need to use privately-held data to its full potential for creating the public good

- Need to access to create a common, public good
- One of the premises of official stats is access to data to come up with valuable in-depth insight into the nature and progress of society and economy in their respective environments
- Need for the support of policymakers to ensure that access to privately-held data for official statistics is guaranteed in the newly established national data governance systems

Need to fight the barriers to data innovation

- Barriers for innovation in statistics
  - Legal obstacles
  - Privacy awareness, public acceptability
  - Poor technological infrastructures
  - Lack of skills and competencies
  - Poor user stakeholder dialog
  - Isolation, no partnerships

Need for innovation in statistical production

- Innovation infrastructures
- Innovation culture (ability to innovate)
- Clear strategic direction (striving for innovation, fostering innovation)
- Supply-based innovation (exploratory based: applied research and experimental development)
- Demand-based innovation (user-driven)

Need for infrastructures for innovation in statistics

- Data access infrastructure (privacy-preserving techniques)
- Data processing infrastructure (micro-data modeling, data linking)
- Human infrastructure: skills and competencies for capacity development (context analysis, learning facilitation, change management, data collaboration and stewardship)
- Management infrastructure (reorientation towards a systematic approach to dealing with the transition or transformation of an organization's goals, processes or technologies, to implement strategies for effecting change, controlling change and helping people to adapt to change)
- Legal framework (hopefully not restraining innovation)

Statistics Poland
Data Science Academy

Statistics Poland

Internal capacity building program

DATA SCIENCE ACADEMY

Build a modern organization, able to anticipate and target fast-changing information needs by tapping into a multitude of data sources

Increase the integration of administrative and new data sources in statistical production – experimental & official

1. Employee development and retention

- talent recruiting (students, academia, research)
- general and specialized trainings
- maintaining competency model
- designing individual development paths
- personnel succession program
- mentoring

2. Sharing knowledge and experience

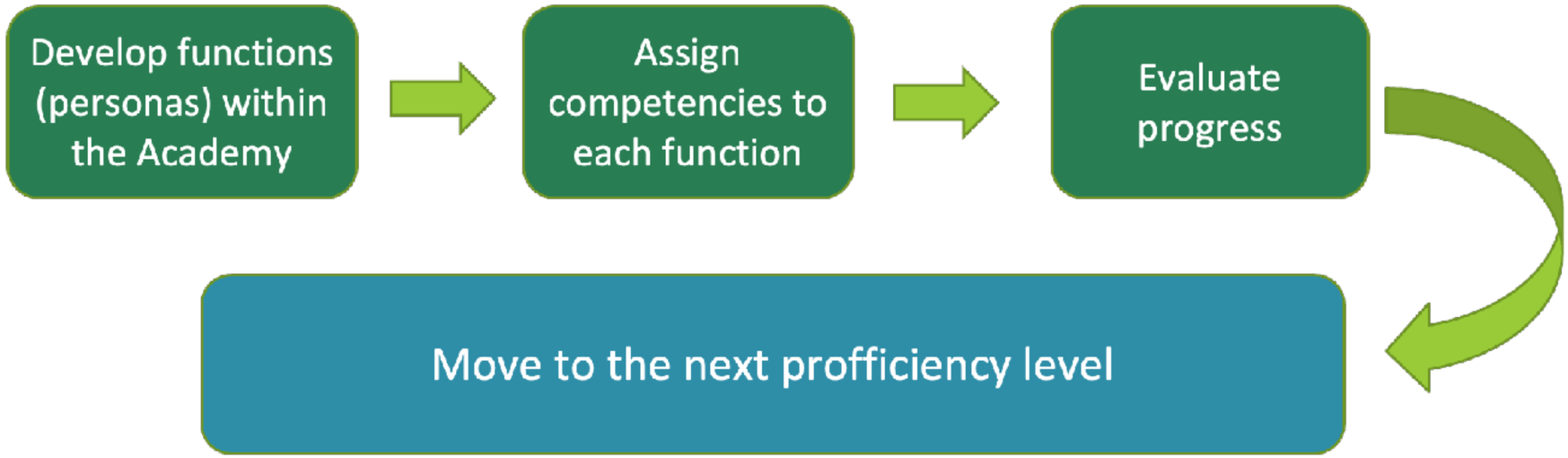
- seminars and meetings on general topics;
  proposals for meetings submitted by Academy members
- Data Science Talks
  presenting selected problems and issues
- regular meetings of
  the data science community

3. Building a data science community platform

- intranet site as a space to share knowledge and experience
- information about events, conferences and meetings, organized by domestic and foreign institutions, hackathons, etc.
- training area, information about planned training courses, catalogue of big data methods and techniques
- database of good practices, e.g., solutions to legal issues
- library of articles, reports and other materials that can be helpful in using big data sources
- discussion zone that will enable joint problem solving, discussions,
  and exchange of experiences

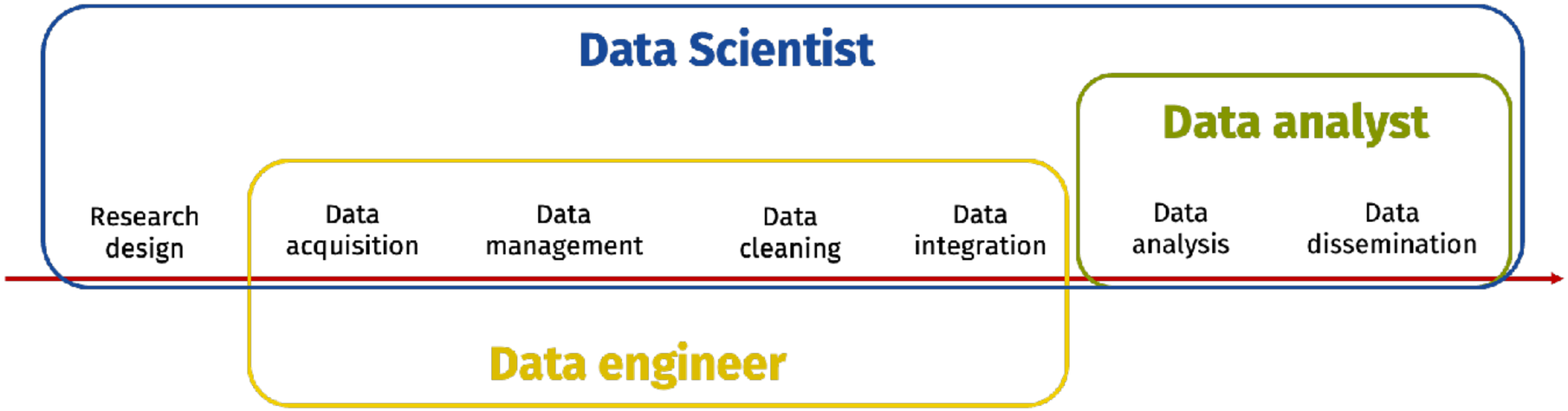
Poland

Data Science Academy at Statistics Poland



Poland

Data Science Academy at Statistics Poland



Poland

Data Science Academy at Statistics Poland

Data engineer	Data acquisition	Data management	Data cleaning	Data integration
Foundation	Skill 1	Skill 1	Skill 1	Skill 1
	Skill 2	Skill 2	Skill 2	Skill 2
	Skill 3		Skill 3	
Intermediate	Skill 1	Skill 1	Skill 1	Skill 1
	Skill 2	Skill 2	Skill 2	Skill 2
	Skill 3	Skill 3	Skill 3	Skill 3
	Skill 4			
Advanced	Skill 1	Skill 1	Skill 1	Skill 1
	Skill 2	Skill 2	Skill 2	Skill 2
	Skill 3		Skill 3	Skill 3

Entry assessment

Entry assessment

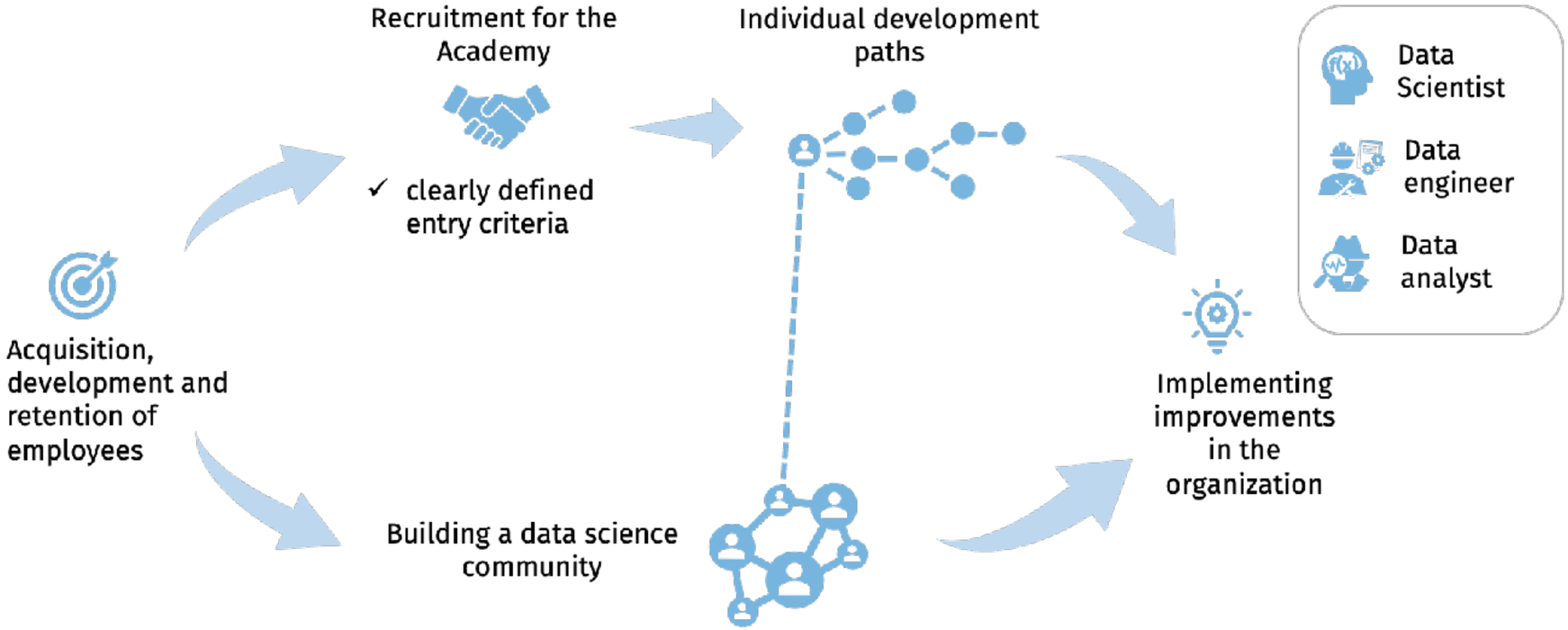
Entry assessment

Evaluation

Evaluation

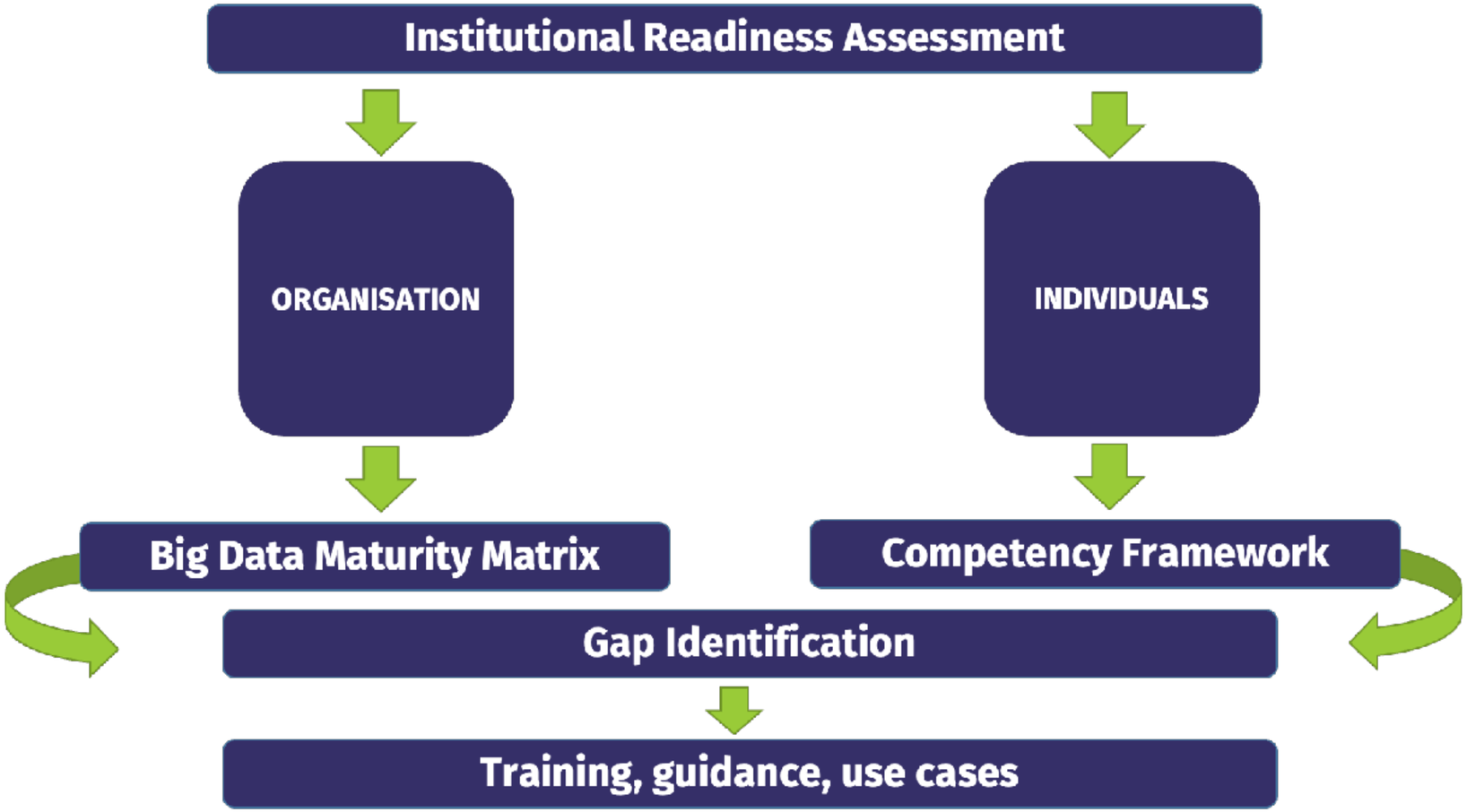
Poland

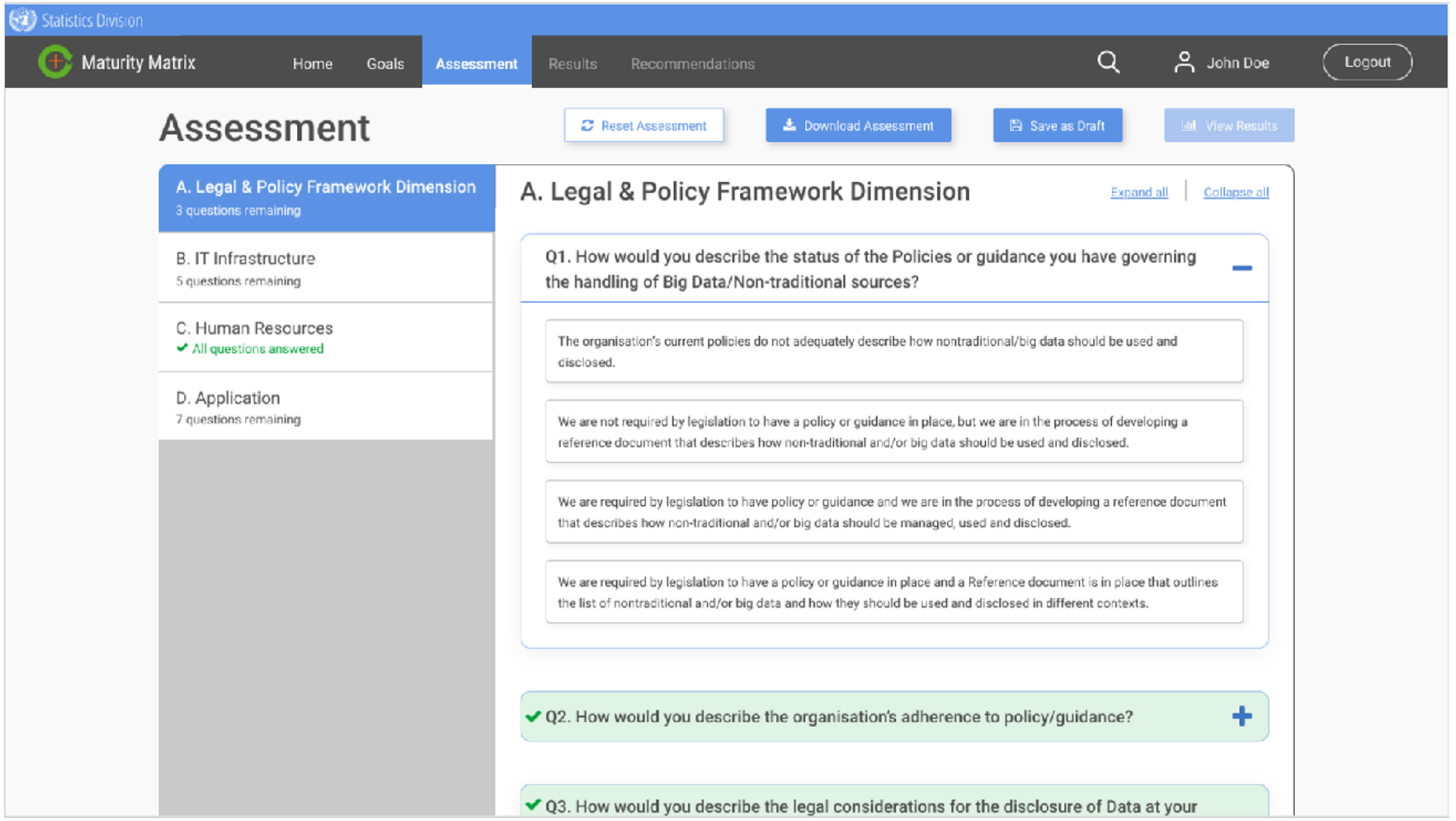
Data Science Academy Ecosystem

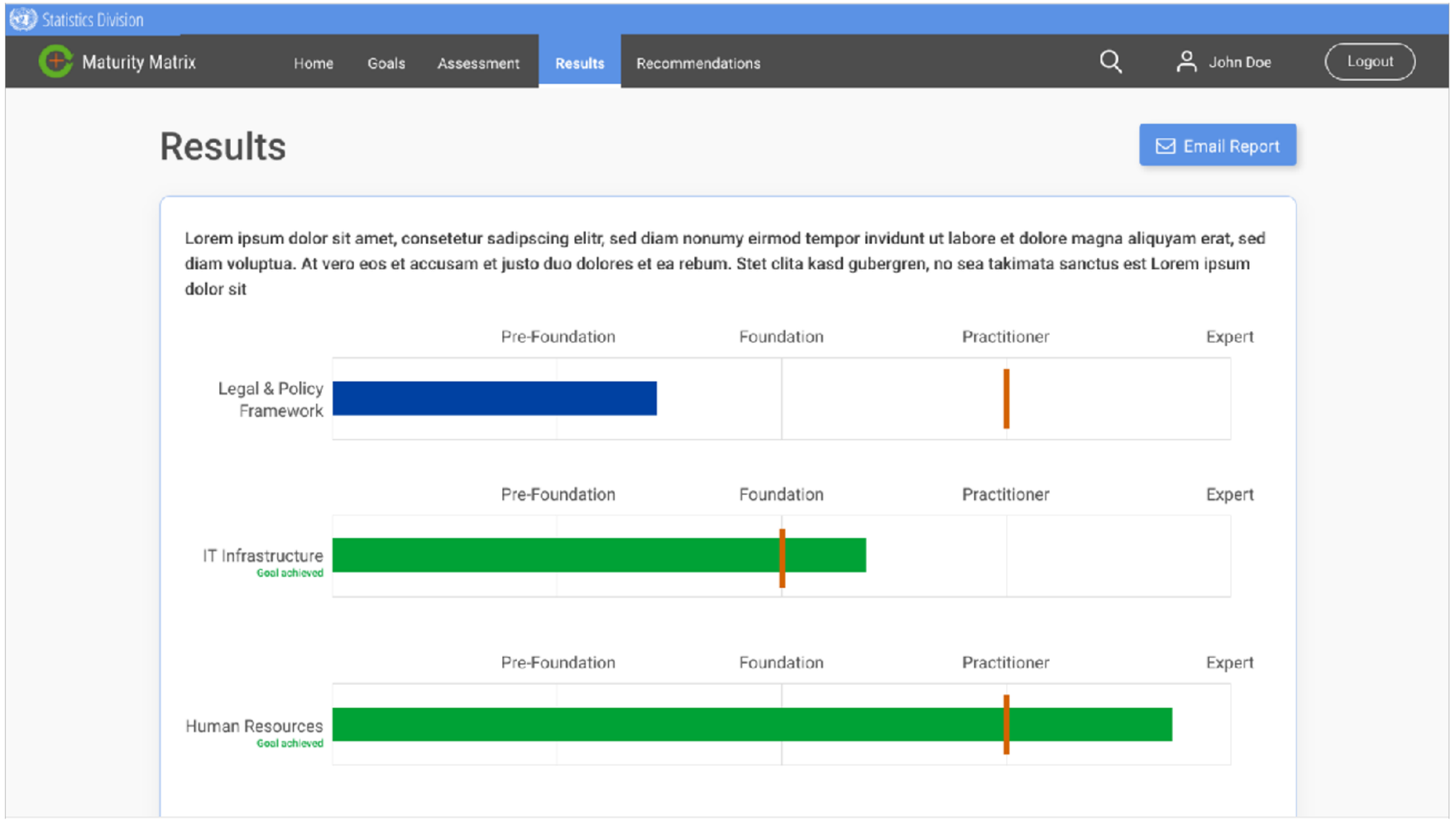


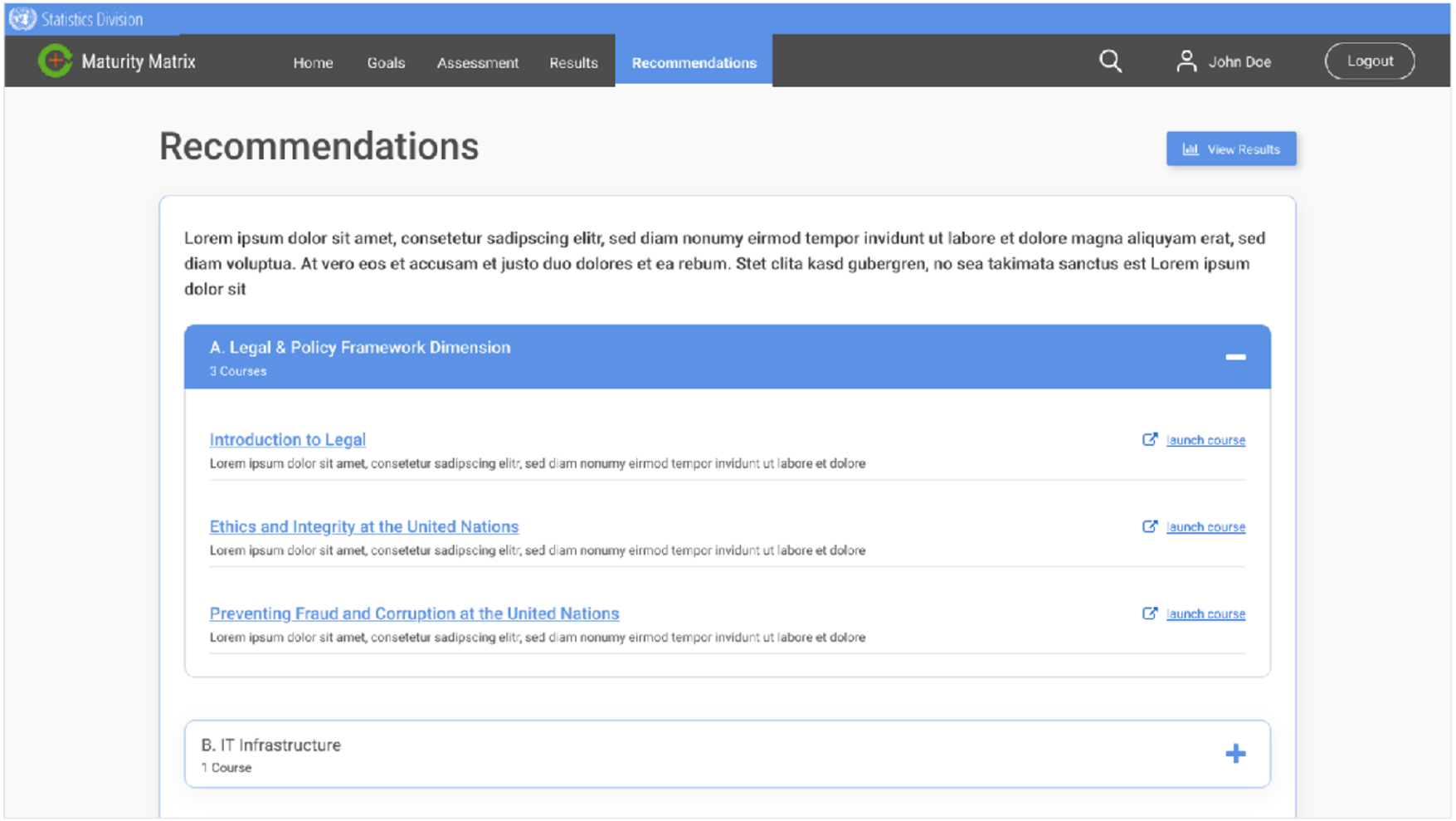
Advantages of our approach (compared to other)

- Organically growing human resources
- Injecting data science across the whole organization
- Distributed and embedded in nature
- Infinite degrees of freedom
- Fostering experimentation and innovation
- Seamlessly transforming organization


Institutional Readiness Assessment
ORGANISATION
INDIVIDUALS
Big Data Maturity Matrix
Competency Framework
Gap Identification
Training, guidance, use cases

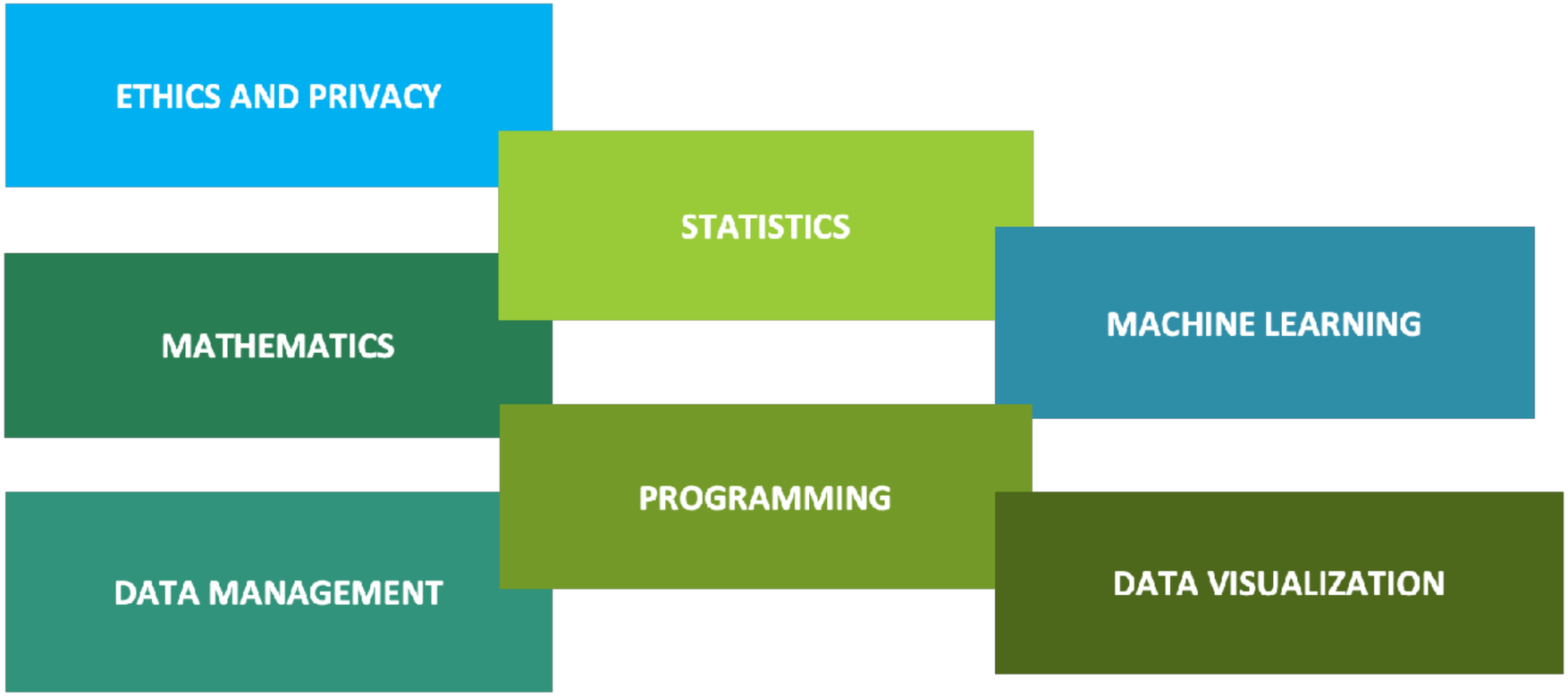

Statistics Division
Maturity Matrix
Home
Goals
Assessment
Results
Recommendations
John Doe
Logout
Assessment
Reset Assessment
Download Assessment
Save as Draft
View Results
A. Legal & Policy Framework Dimension
3 questions remaining
B. IT Infrastructure
5 questions remaining
C. Human Resources
All questions answered
D. Application
7 questions remaining
A. Legal & Policy Framework Dimension
Expand all
Collapse all
Q1. How would you describe the status of the Policies or guidance you have governing the handling of Big Data/Non-traditional sources?
The organisation's current policies do not adequately describe how nontraditional/big data should be used and disclosed.
We are not required by legislation to have a policy or guidance in place, but we are in the process of developing a reference document that describes how non-traditional and/or big data should be used and disclosed.
We are required by legislation to have policy or guidance and we are in the process of developing a reference document that describes how non-traditional and/or big data should be managed, used and disclosed.
We are required by legislation to have a policy or guidance in place and a Reference document is in place that outlines the list of nontraditional and/or big data and how they should be used and disclosed in different contexts.
Q2. How would you describe the organisation's adherence to policy/guidance?
Q3. How would you describe the legal considerations for the disclosure of Data at your


Statistics Division
Maturity Matrix
Home
Goals
Assessment
Results
Recommendations
John Doe
Logout
Results
Email Report
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit
Pre-Foundation
Foundation
Practitioner
Expert
Legal & Policy Framework
Pre-Foundation
Foundation
Practitioner
Expert
IT Infrastructure
Goal achieved
Pre-Foundation
Foundation
Practitioner
Expert
Human Resources
Goal achieved


Statistics Division
Maturity Matrix
Home
Goals
Assessment
Results
Recommendations
John Doe
Logout
Recommendations
View Results
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit
A. Legal & Policy Framework Dimension
3 Courses
Introduction to Legal
launch course
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore
Ethics and Integrity at the United Nations
launch course
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore
Preventing Fraud and Corruption at the United Nations
launch course
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore
B. IT Infrastructure
1 Course

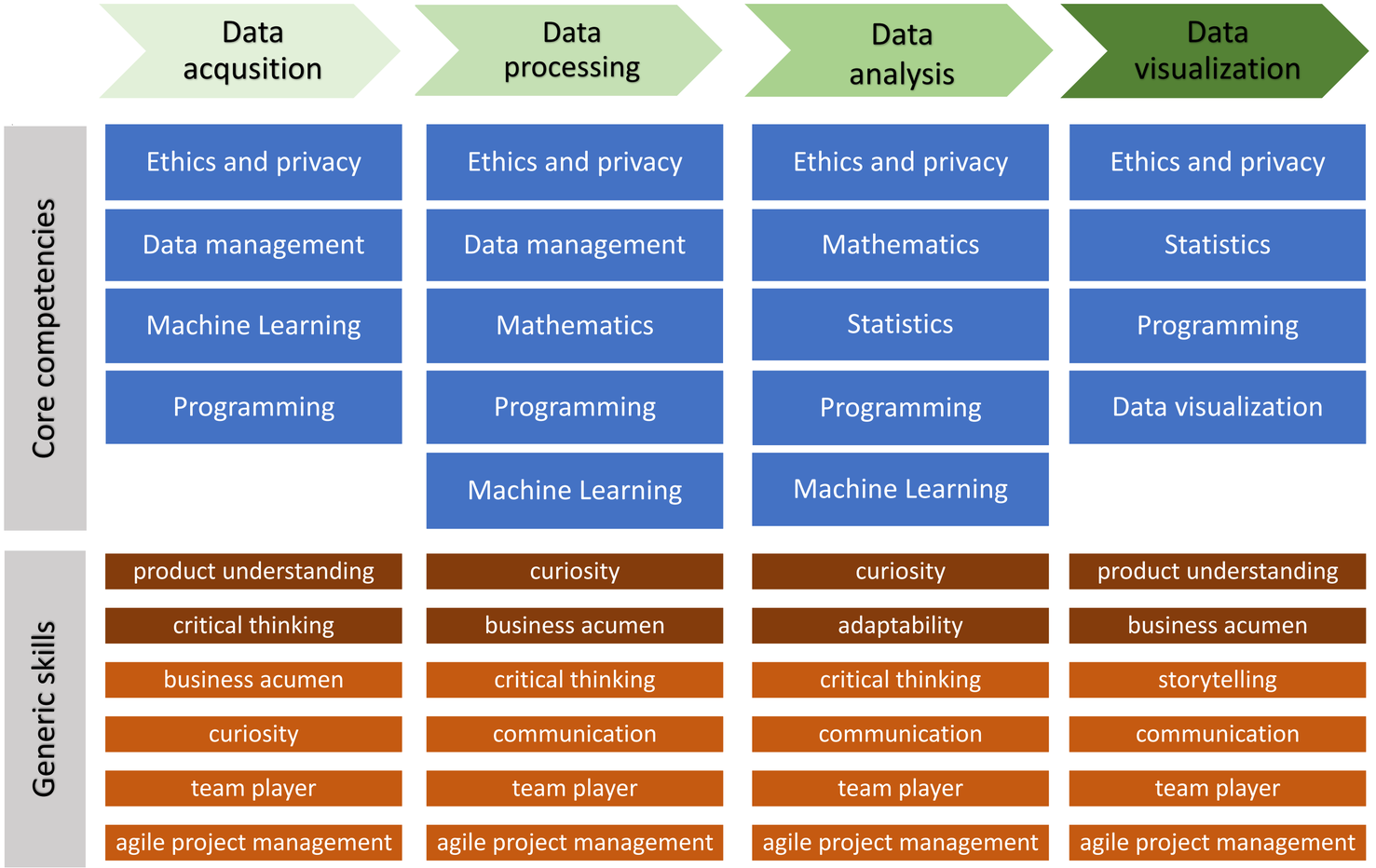
UNBigData

UN Big Data Competency Framework
Core Competencies



UN Big Data Competency Framework
Soft Skills

Big data competencies & statistical production process

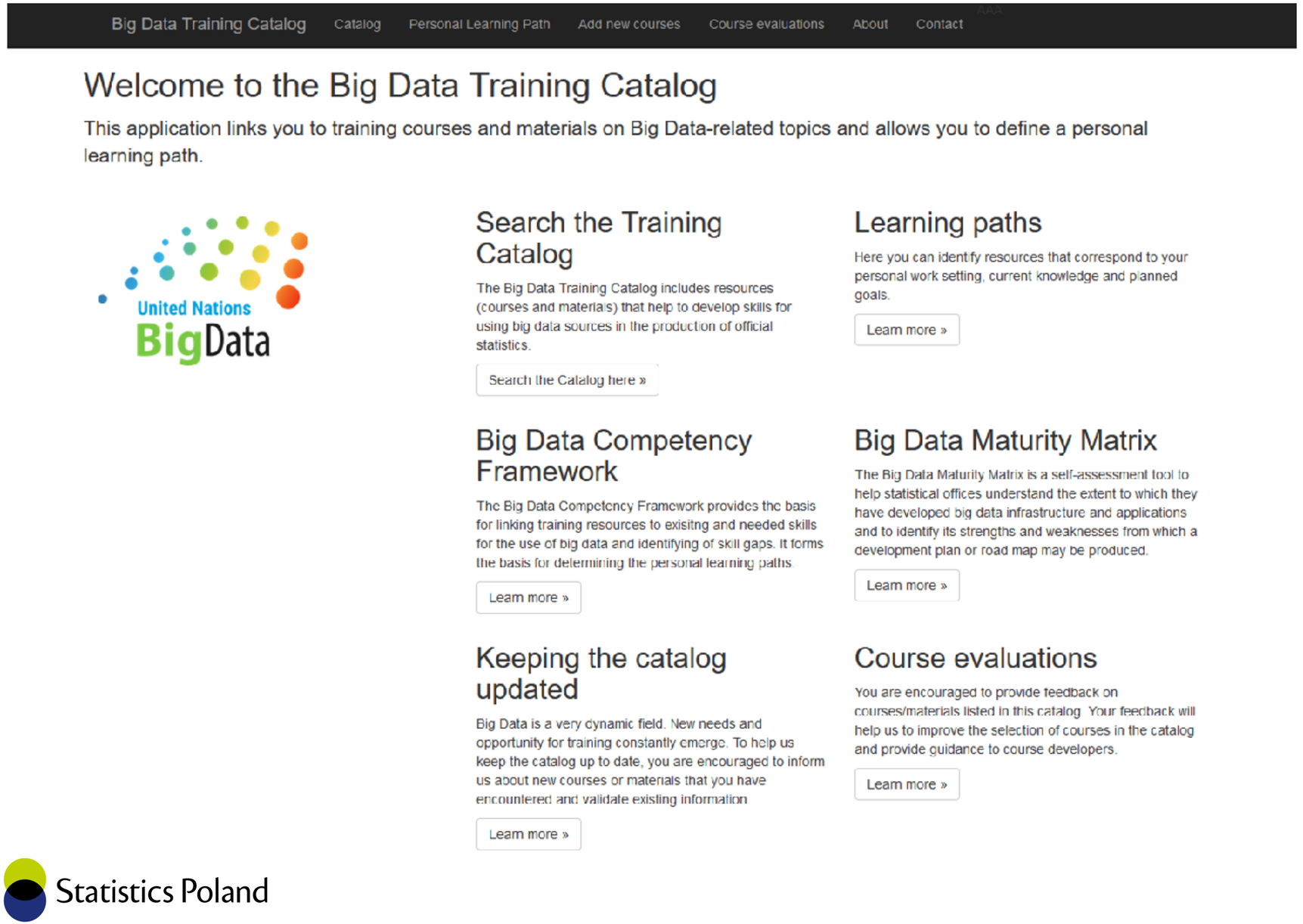


Competency Framework for big data acquisition and processing
UN Global Working Group on Big Data for Official Statistics
Task Team on Training, Competencies and Capacity Development

UN Big Data Competency Framework

- General guidance on big data knowledge and skills
- Knowledge, skills and attitudes for acquiring and processing big data
- Can be used to assess knowledge gaps, recruit and train staff at the NSO

Dimension 1			
Name of the area	Ethics and privacy		
Dimension 2			
Competence title and description	To possess a basic level of ethics and privacy knowledge in below-listed issues: 1) Basic definitions of issues related to the processing of big data (personal data and anonymous data, active and passive big data, dimensions of big data, consciously and not-consciously transferred data, etc.) 2) Philosophical aspects of collecting and processing big data (ethic control and a pragmatic view of the impact on the life of people and organizations: privacy, impact on personal capabilities and freedom, rights between data owner and data explorer) 3) Legal framework for management of big data (personal data processing steps and principles, privacy and transparency policy, data processing purposes) 4) Technical aspects of work with private customer and identity data (obtaining and sharing private information, transparent view of how our data is being used, openness of data)		
Dimension 3	A - Foundation	B - Intermediate	C - Advanced
Proficiency levels	Demonstrate knowledge and understanding of basic rules of philosophical, legal of collecting, processing and sharing of big data.	Demonstrate knowledge, understanding and putting into practice philosophical, legal and technical rules of collecting, processing and sharing of big data.	Thorough knowledge of the application of personal data protection law, proficiency in personal data management and skillfulness in performing operations on varied data sets respecting the law, ethical norms, while maintaining the highest technical standards. Advises others on the ethical and privacy considerations of data.


Big Data Training Catalog
Catalog
Personal Learning Path
Add new courses
Course evaluations
About
Contact
Welcome to the Big Data Training Catalog
This application links you to training courses and materials on Big Data-related topics and allows you to define a personal learning path.
United Nations
BigData
Search the Training Catalog
The Big Data Training Catalog includes resources (courses and materials) that help to develop skills for using big data sources in the production of official statistics.
Search the Catalog here »
Learning paths
Here you can identify resources that correspond to your personal work setting, current knowledge and planned goals.
Learn more »
Big Data Competency Framework
The Big Data Competency Framework provides the basis for linking training resources to exisitng and needed skills for the use of big data and identifying of skill gaps. It forms the basis for determining the personal learning paths.
Learn more »
Big Data Maturity Matrix
The Big Data Maturity Matrix is a self-assessment tool to help statistical offices understand the extent to which they have developed big data infrastructure and applications and to identify its strengths and weaknesses from which a development plan or road map may be produced.
Learn more »
Keeping the catalog updated
Big Data is a very dynamic field. New needs and opportunity for training constantly emerge. To help us keep the catalog up to date, you are encouraged to inform us about new courses or materials that you have encountered and validate existing information.
Learn more »
Course evaluations
You are encouraged to provide feedback on courses/materials listed in this catalog. Your feedback will help us to improve the selection of courses in the catalog and provide guidance to course developers.
Learn more »

Data stewardship at Statistics Poland

Statistics Poland

The way of the data steward

- Openness Declaration: Open by Default:
  Statistics Poland 5 O's Manifesto
  - Open data
  - Open algorithms
  - Open source (public funding - public code)
  - Open access
  - Open knowledge

Repository of Information Standards (RSI)

- A structured and organized resource of information about official registers and public administration information systems
  - A knowledge base containing metadata about official registers and public administration information systems
- It is updated by internal RIS administrators and external administrators of registers/infromation systems using dedicated web application
  - It is a tool for collecting, updating, sharing, and analyzing information on public administration information systems and for harmonizing/ agreeing on information standards

What’s inside?

- Information about resources in various public administration units place includes
  - the scope of data (information)
  - methodology
  - definitions of variables
  - classifications used
  - identifiers

Statistics Poland 2022 Statistical Program
usage of administrative data

- Statistics Poland directly used data from
  - **282** official registers and information systems
  - **84** administrators
- These data contributed
  - to **130** statistical studies
  - in **31** fields of statistics

Progress in using of administrative data
Statistics Poland 2022 Statistical Program

- Annual increase in the use of administrative sources
  - additional **65** administrative sources
  - **23** source owners

Progress in using of administrative data

Usage purpose of new sources in 2022

- data source for a study
  40 new sources
- source of data for creating and updating sampling frames
  34 new sources
- quality control of research data
  21 new sources
- imputation, data estimation
  17 new sources
- direct posting of information
  10 new sources

Progress in using of administrative data:
Statistics Poland 2023 Statistical Program

- Annual increase in the use of administrative sources
  - additional **36** administrative sources
  - **26** source owners

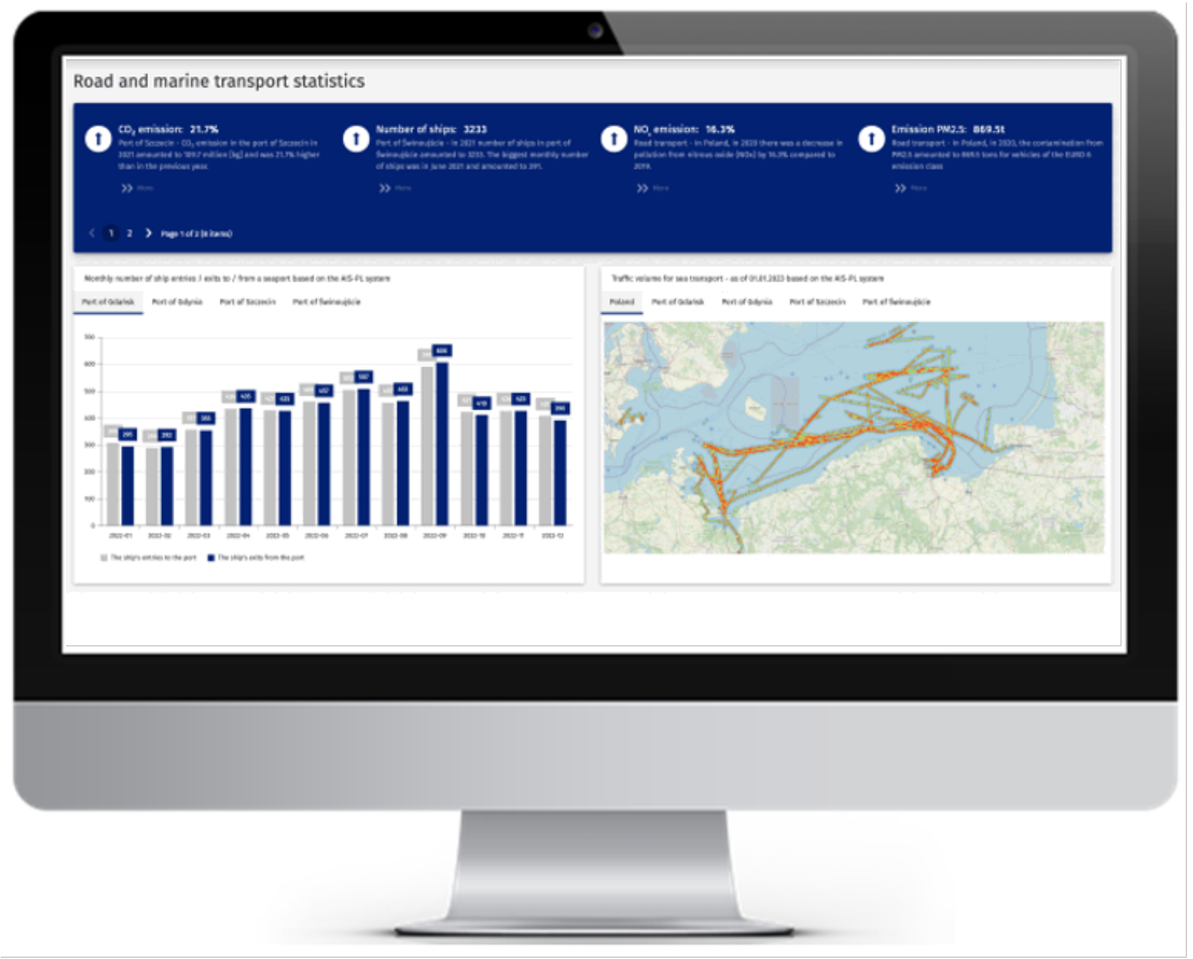
Innovation, examples

Statistics Poland

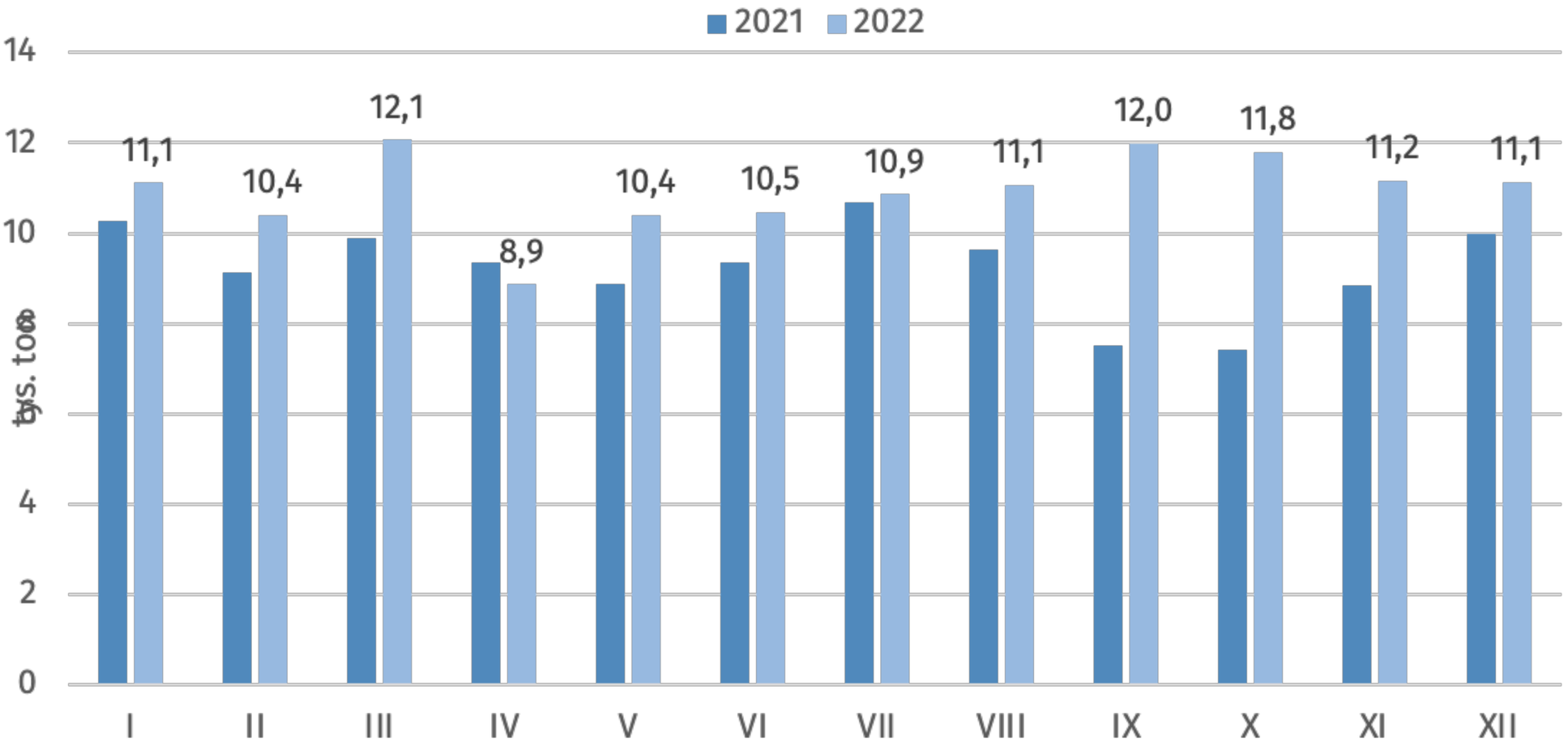
TranStat: Overview

- Getting access to sensor data:
  - Automatic Identification System (AIS)
  - e-TOLL electronic toll collection system
- Application of Big Data methods and tools
- Designing a methodology for traffic intensity, transportation volume and emissions estimations

https://transtat.stat.gov.pl



Monthly CO2 emissions in the port of Szczecin in 2021 and 2022



Statistics Poland

Statistics of transportation volume in maritime transport

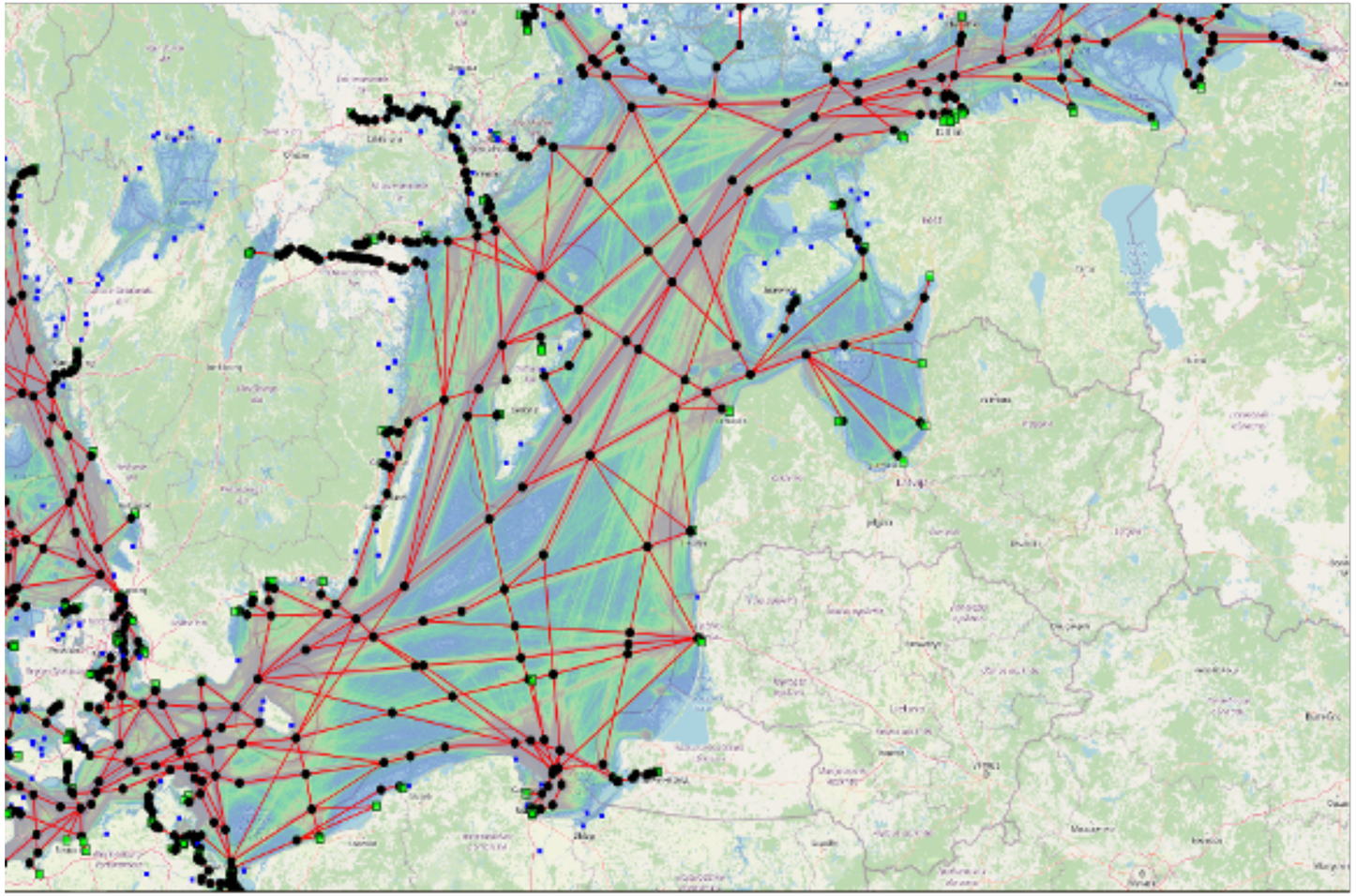


Figure 7. *Graph visualization for the Baltic Sea*

Source: Maritime University of Szczecin

VII Edition of the Statistical Olympiad

Statistics Poland

Since 2016, Statistics Poland and the Polish Statistical Association have been organizing the Statistical Olympiad.

The aim of the event is to disseminate knowledge and develop skills in the field of statistics in the field of socio-economic analyses.

Olimpiada statystyczna

schools

466

students

3 997

Reporting of social responsibility

Statistics Poland

Since 2018 we have been reporting CSR activities in public statistics

4

total numer of reports issued
for years: 2018, 2019, 2020, 2021

Reporting of social responsibility

Statistics Poland

We integrate employees into join activities

The effects of our activities from 2018

44
CSR’s actions

1500
people received support

20
hybrid meetings

2900
employees attended

28
pro-health campaigns

3600
participants

30
competitions/quis for employees and their family

6000
range of competitions

@StatPoland

@GlownyUrzadStatystyczny

Glowny Urzad Statystyczny

@gus_stat

Glowny Urzad Statystyczny GUS

Statistics Poland

ICC 2023

The Geostatistics Portal received 1st place in the digital products category

at the International Cartographic Exhibition during the 31st International Cartographic Conference in Cape Town.

Concluding remarks

OPEN DATA WATCH

ODIN
OPEN DATA INVENTORY
by OPEN DATA WATCH

Global Rankings

87 ODIN SCORE

Poland

2nd GLOBAL RANK OUT OF 195

81 COVERAGE SCORE OUT OF 100

92 OPENNESS SCORE OUT OF 100

	Country	Region	Overall	Coverage	Openness
1	Singapore	South-Eastern Asia	90	77	100
2	Poland	Eastern Europe	87	81	92

Statistics Poland

Thank you !

Dominik A. Rozkrut
President, Statistics Poland