Statistics Poland

Digital transformation, data governance, data stewardship and official statistics

Pre-Conference Workshop of the
35th CIRET conference

Dominik A. Rozkrut

Data ecosystems
Data governance

Statistics Poland

Digital transformation

Statistics Poland

Digitalization

- Information and communication technologies (ICT), as general-purpose technologies, saturate all aspects of socio-economic life

- Progress in modern computer techniques and technologies has led to a revolution in collecting, processing, storing, and transmitting information

- The digitalization of the economy produces more and more data

- The increasing complexity of data ecosystems

Statistics Poland

Data ecosystems – a strategic perspective

- Technological changes in storing and processing the growing amount of data are a huge source of innovation

- The generation and acquisition of data cease to be the domain of only a narrow group of specialized entities.

  - These are not only IT enterprises anymore because digital technologies are increasingly used across almost all industries,

    - and beyond that, not only in businesses,

    - but also the public sector, NGOs, or simply citizens who can passively share their data, using modern devices and services, or actively participate in the citizen-generated data projects

Statistics Poland

Data ecosystems – a strategic perspective

- In February 2020, the European Commission published two documents, key to the data-based economy:

  - "Shaping Europe's Digital Future"

  - "A European Strategy for Data"

Statistics Poland

Data ecosystems – a strategic perspective

- The possibilities of data for innovative purposes are essential in the data-based economy.

  - Data availability is a crucial problem.

  - Even in cases where one may expect an abundance of data, a lack of sharing mechanisms may severely hamper the innovation potential.

  - Hence the need to unblock the flow channels is one of the critical elements of the European strategy.

Statistics Poland

Data ecosystems – a strategic perspective

- A common data ecosystem for the development of collaboration in the field of data sharing may be decomposed into four main spheres

  - B2B (Business-to-Business)

  - B2G (Business-to-Government)

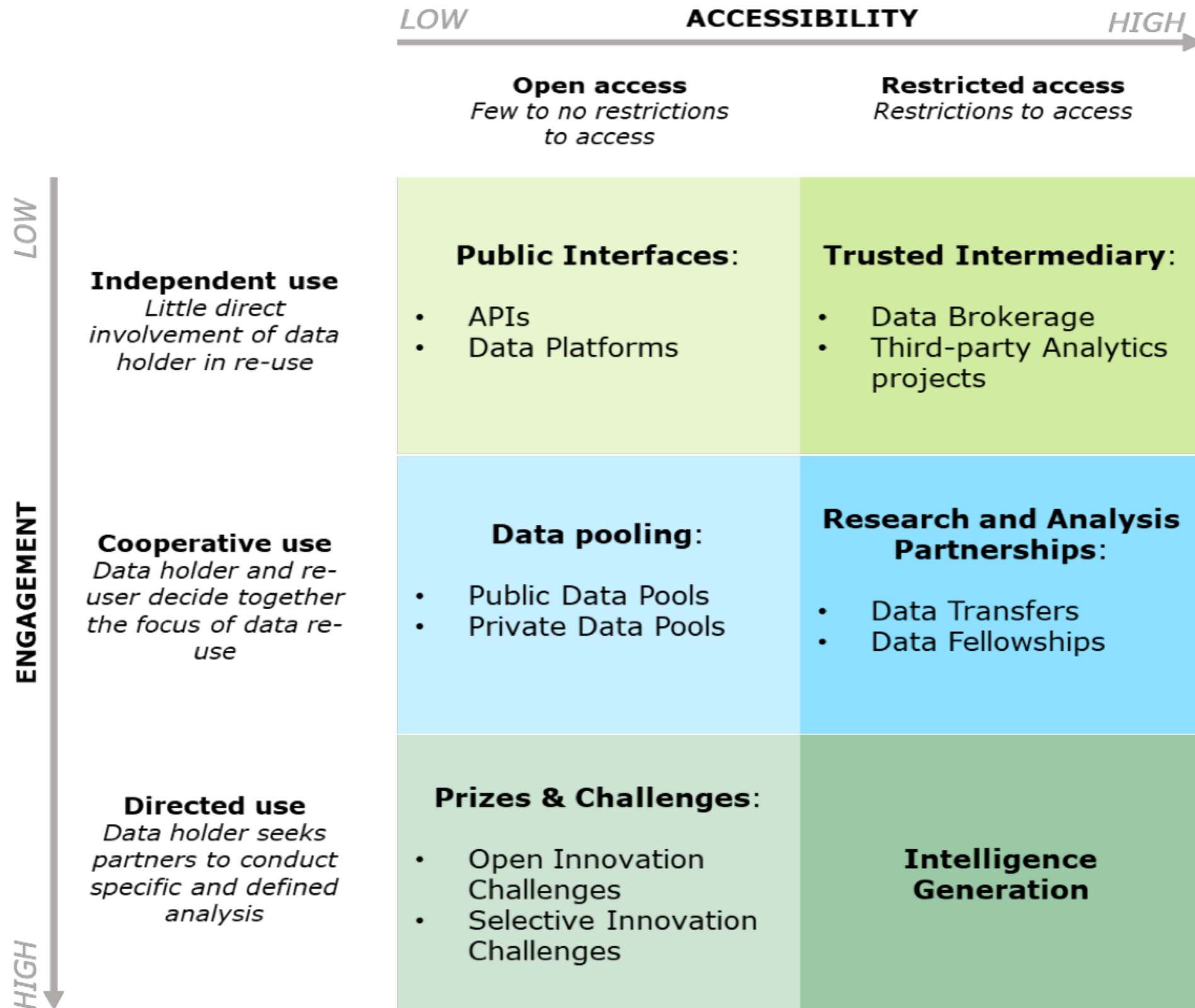 - G2B (Government-to-Business)

  - G2G (Government-to-Government)

Statistics Poland

Stakeholders in the new data ecosystem

- Producers

- Holders

- Intermediaries

- Users

Statistics Poland

Value creation between supply and demand

- Open data, public Interfaces (APIs)

- Data pooling

- Data collaboratives

- Trusted intermediaries

- Research & Analysis

- Intelligence generation

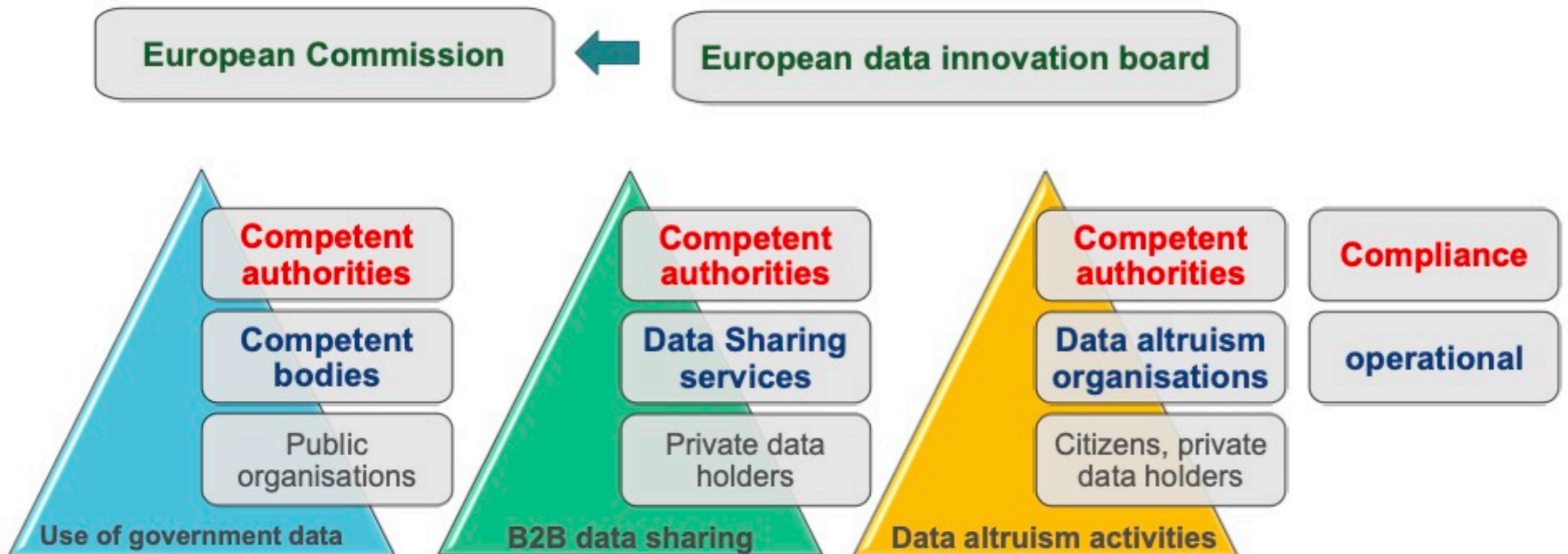Statistics Poland

Engagement versus accessibility in data collaboratives

ACCESSIBILITY

LOW → *HIGH*

	Open access *Few to no restrictions to access*	**Restricted access** *Restrictions to access*
Independent use *Little direct involvement of data holder in re-use* (ENGAGEMENT: LOW)	**Public Interfaces:** • APIs • Data Platforms	**Trusted Intermediary:** • Data Brokerage • Third-party Analytics projects
Cooperative use *Data holder and re-user decide together the focus of data re-use*	**Data pooling:** • Public Data Pools • Private Data Pools	**Research and Analysis Partnerships:** • Data Transfers • Data Fellowships
Directed use *Data holder seeks partners to conduct specific and defined analysis* (ENGAGEMENT: HIGH)	**Prizes & Challenges:** • Open Innovation Challenges • Selective Innovation Challenges	**Intelligence Generation**

ENGAGEMENT (LOW → HIGH)

Statistics Poland

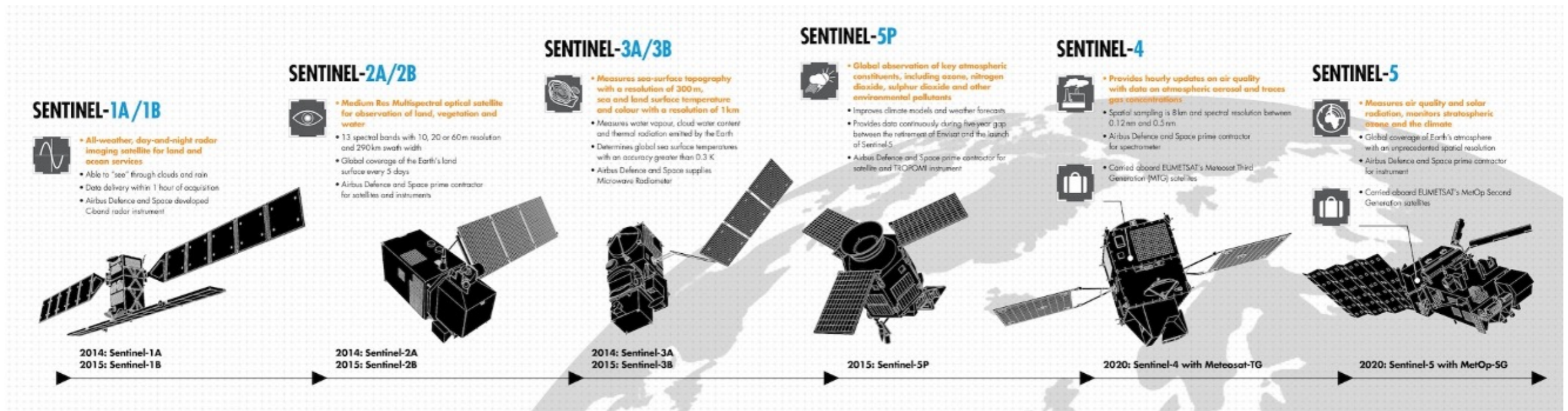# Data stewardship – new legal opportunities

- Policy makers reflection on data need in response to the developing global „datafication":

  - European data strategy

  - Data governance Act

  - Data act

  - Open data directive

  - High Value Datasets

  - (... GDPR, Digital Services, ePrivacy Regulation)

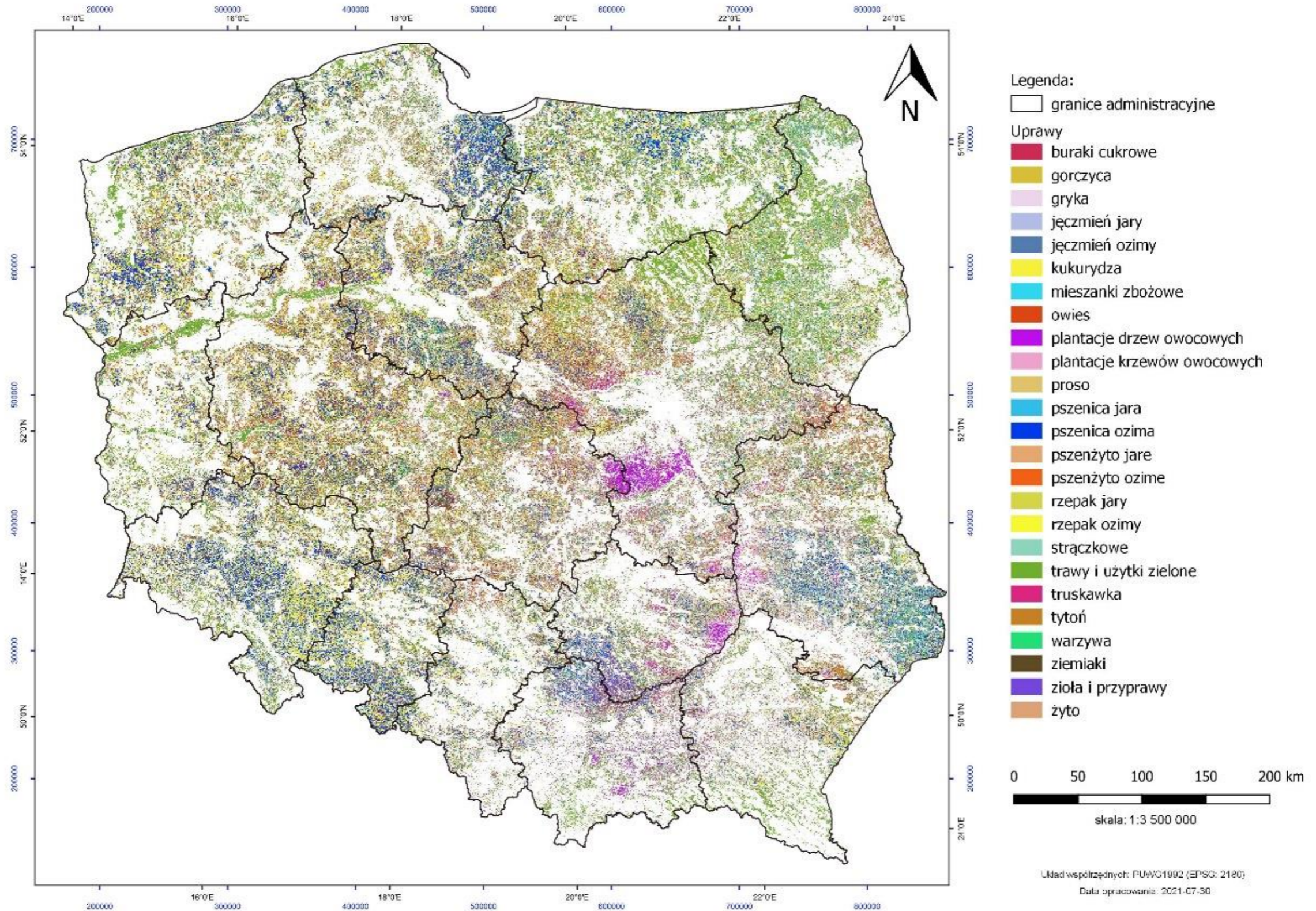- (How to ensure the European/official statistics role?)

Statistics Poland

Data Governance Act

European Commission ← European data innovation board

Use of government data
- Competent authorities
- Competent bodies
- Public organisations

B2B data sharing
- Competent authorities
- Data Sharing services
- Private data holders

Data altruism activities
- Competent authorities
- Data altruism organisations
- Citizens, private data holders

Compliance

operational

New data sources in official statistics

Statistics Poland

Copernicus

Crop estimates, 2021

Syst
tyfik
zwa
klas
dzia
dzie

Legenda:

☐ granice administracyjne

Uprawy
- buraki cukrowe
- gorczyca
- gryka
- jęczmień jary
- jęczmień ozimy
- kukurydza
- mieszanki zbożowe
- owies
- plantacje drzew owocowych
- plantacje krzewów owocowych
- proso
- pszenica jara
- pszenica ozima
- pszenżyto jare
- pszenżyto ozime
- rzepak jary
- rzepak ozimy
- strączkowe
- trawy i użytki zielone
- truskawka
- tytoń
- warzywa
- ziemiaki
- zioła i przyprawy
- żyto

0   50   100   150   200 km

skala: 1:3 500 000

Układ współrzędnych: PUWG1992 (EPSG: 2180)
Data opracowania: 2021-07-30

Statistics Poland

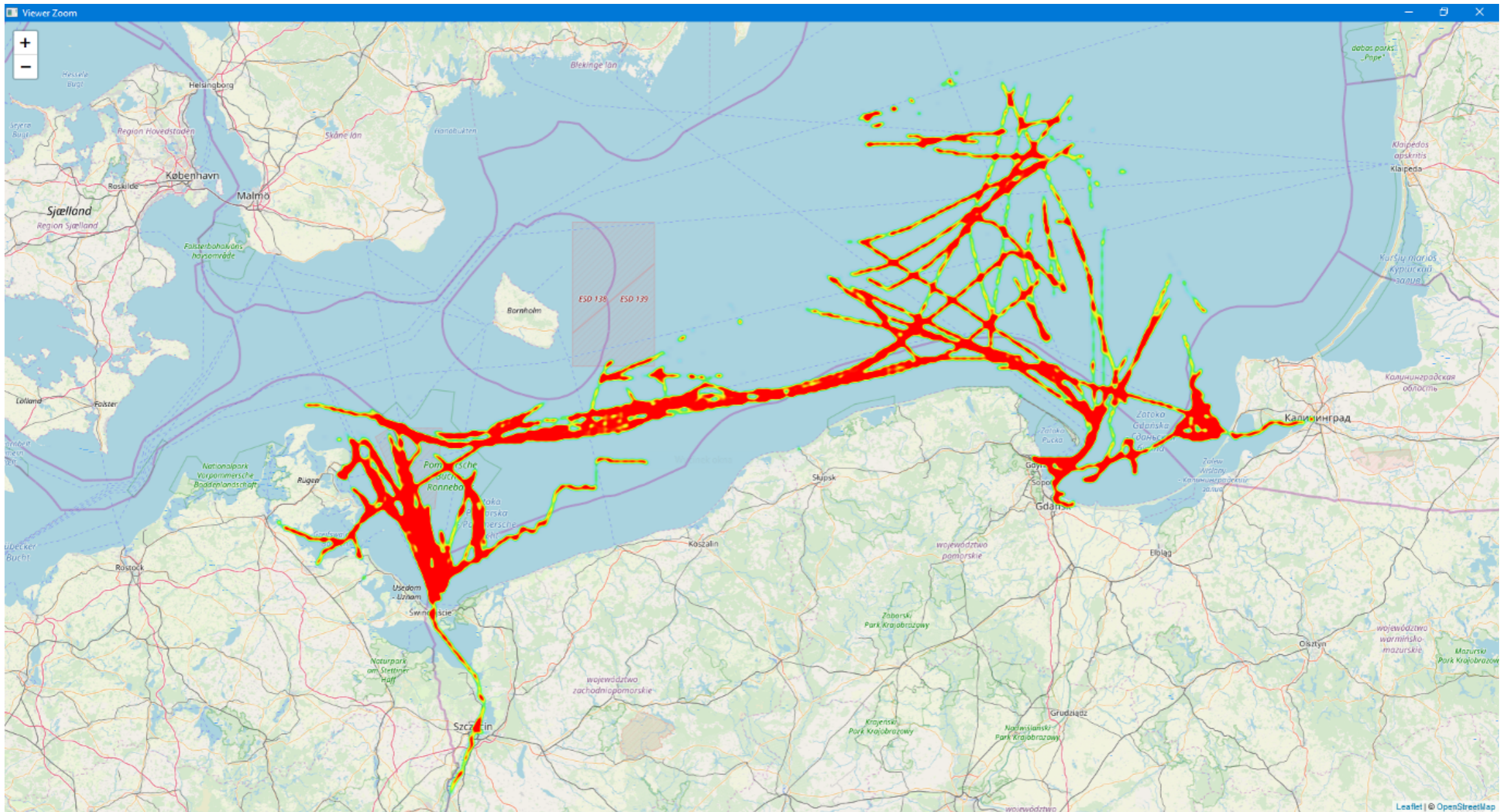# Web-scraping

noclegi.pl

booking.com

hotels.com

Web-scraping

Web-scraping

Web-scraping - tourism - impact of the pandemic

Daily Tourists
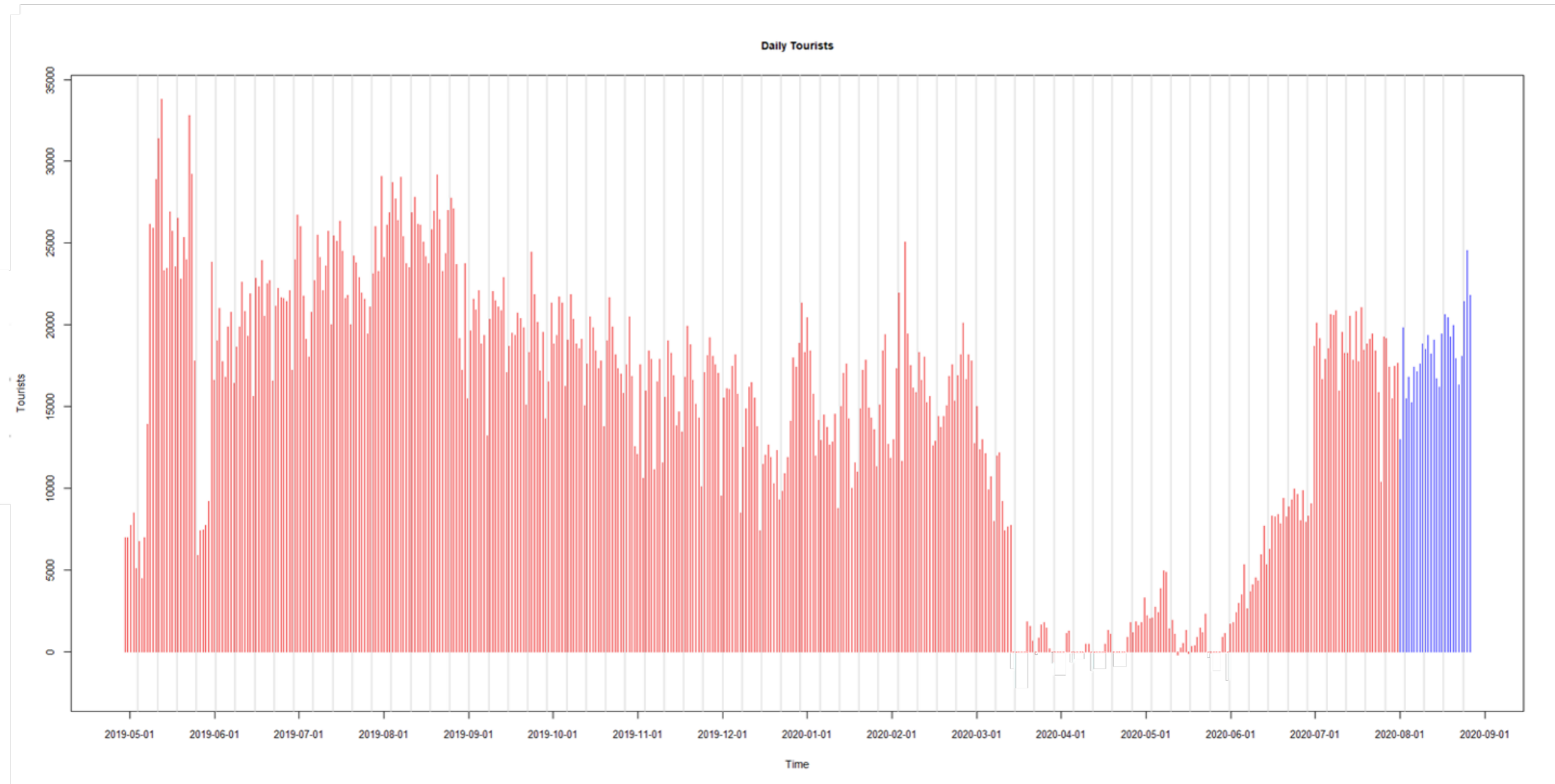
AIS

IT infrastructure for AIS-based statistics production
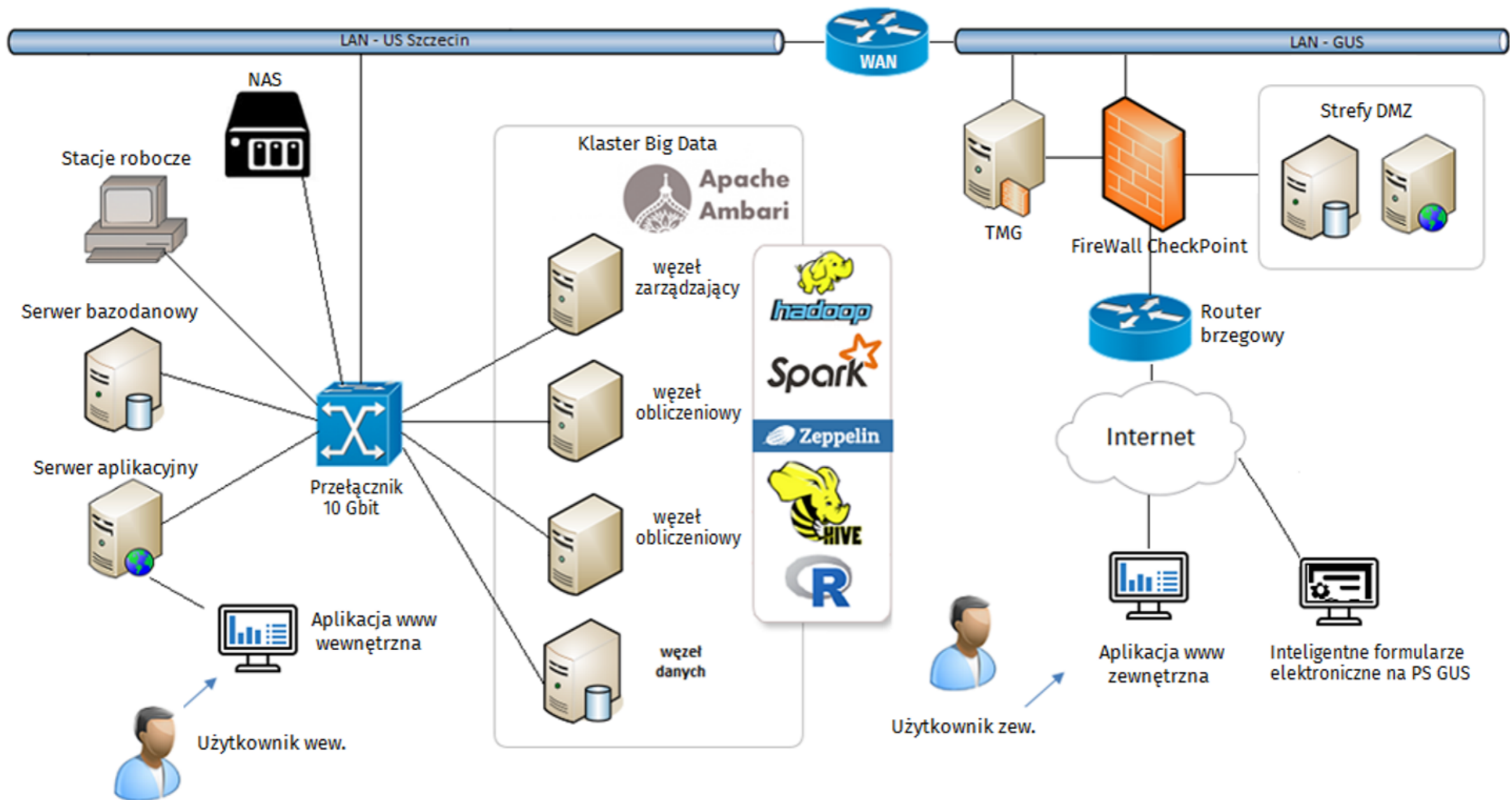
Fishing fleet statistics workflow

based on reference frame of fishing fleets

removing duplicates and incorrect mmsi/imo numbers

https://ec.europa.eu/fisheries/cfp/fishing_rules/fishing_fleet_en

web scraping methods to get missing information (e.g. mmsi numer)

Repozytorium kodów – Github:
https://github.com/AIS-data/WPE_fishing_fleet
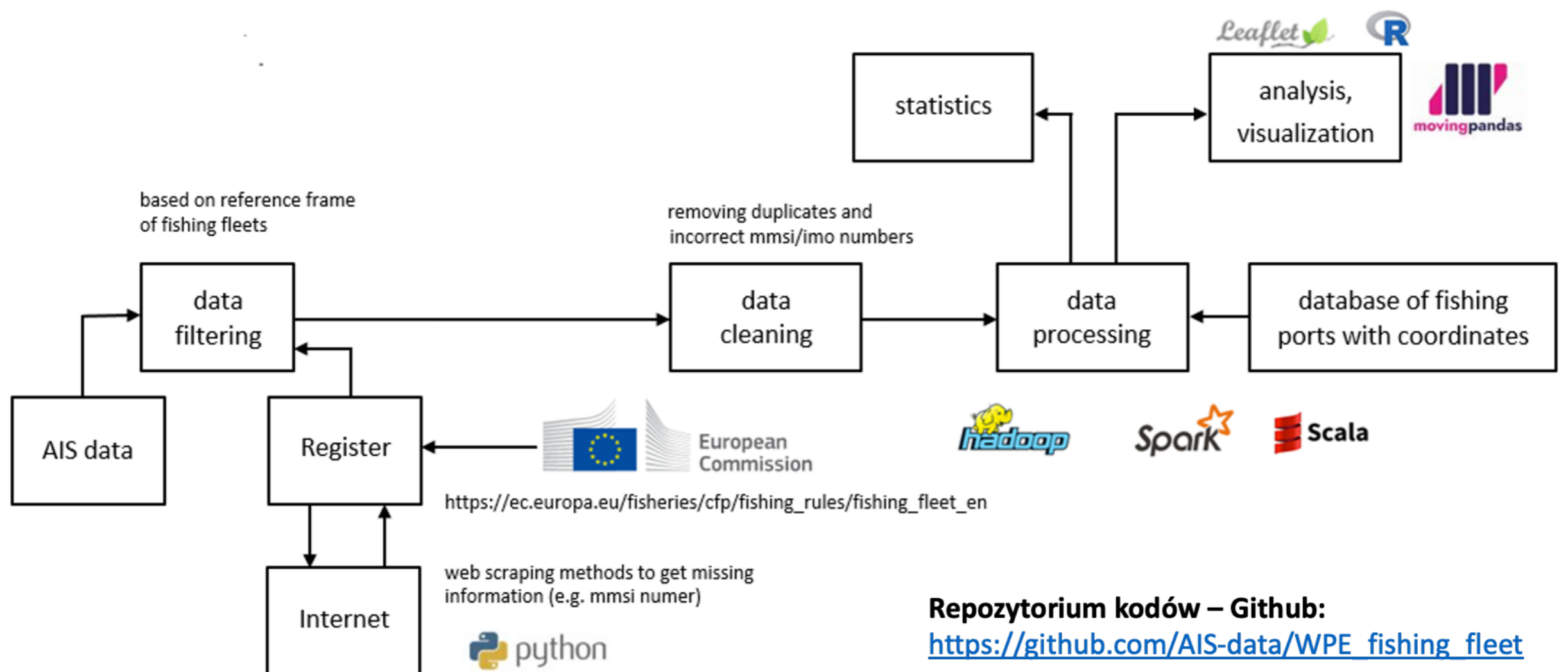
ViaToll/eToll
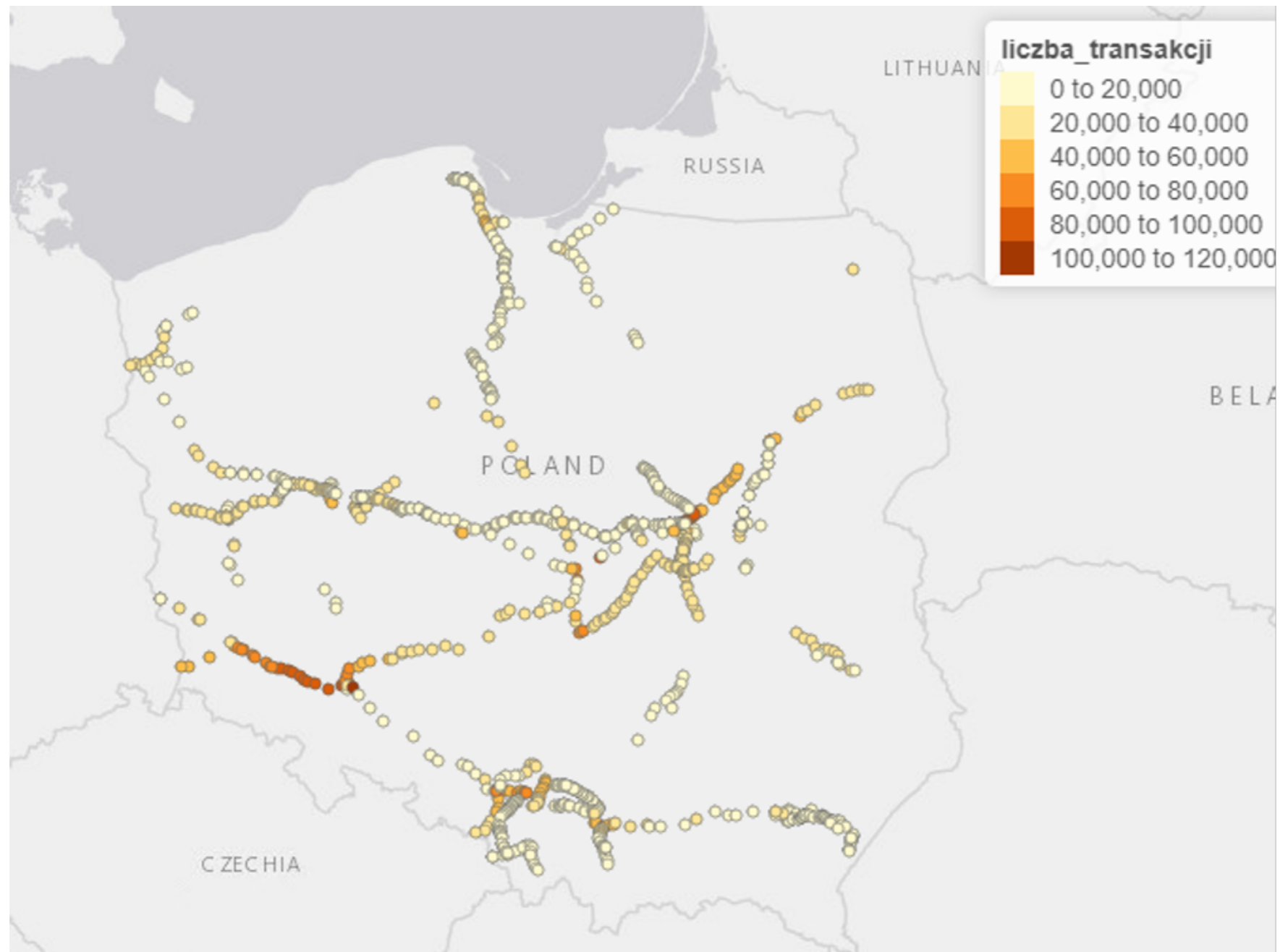
# Statistics Poland - international activities

- UN Global Working Group on Big Data for Official Statistics

  - Task Team on Training, Competencies and Capacity

  - Task Team on Big Data for Sustainable Development Goals

- Eurostat:

  - Steering Group Big Data włączone do DIME WG Methodology

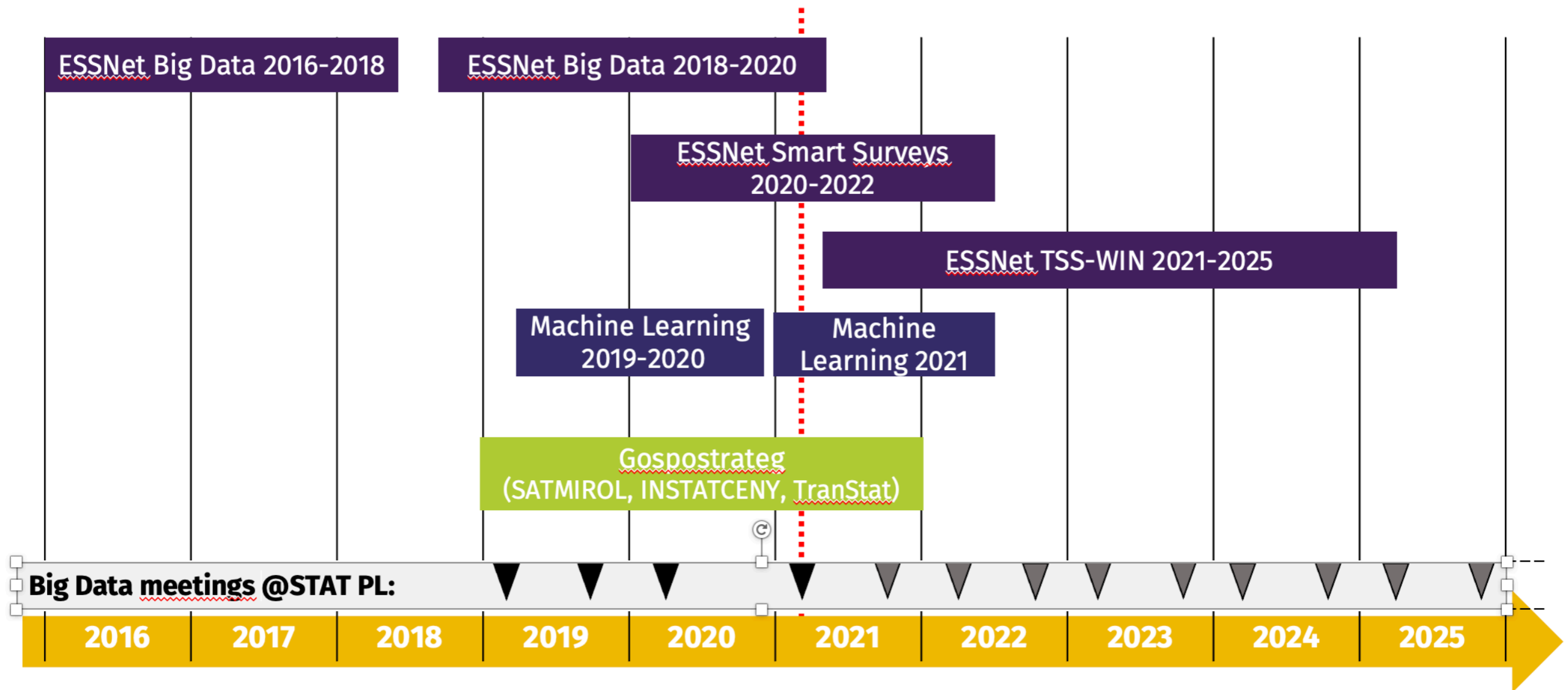 - Task Force on Trusted Smart Statistics

Statistics Poland

Statistics Poland - big data network
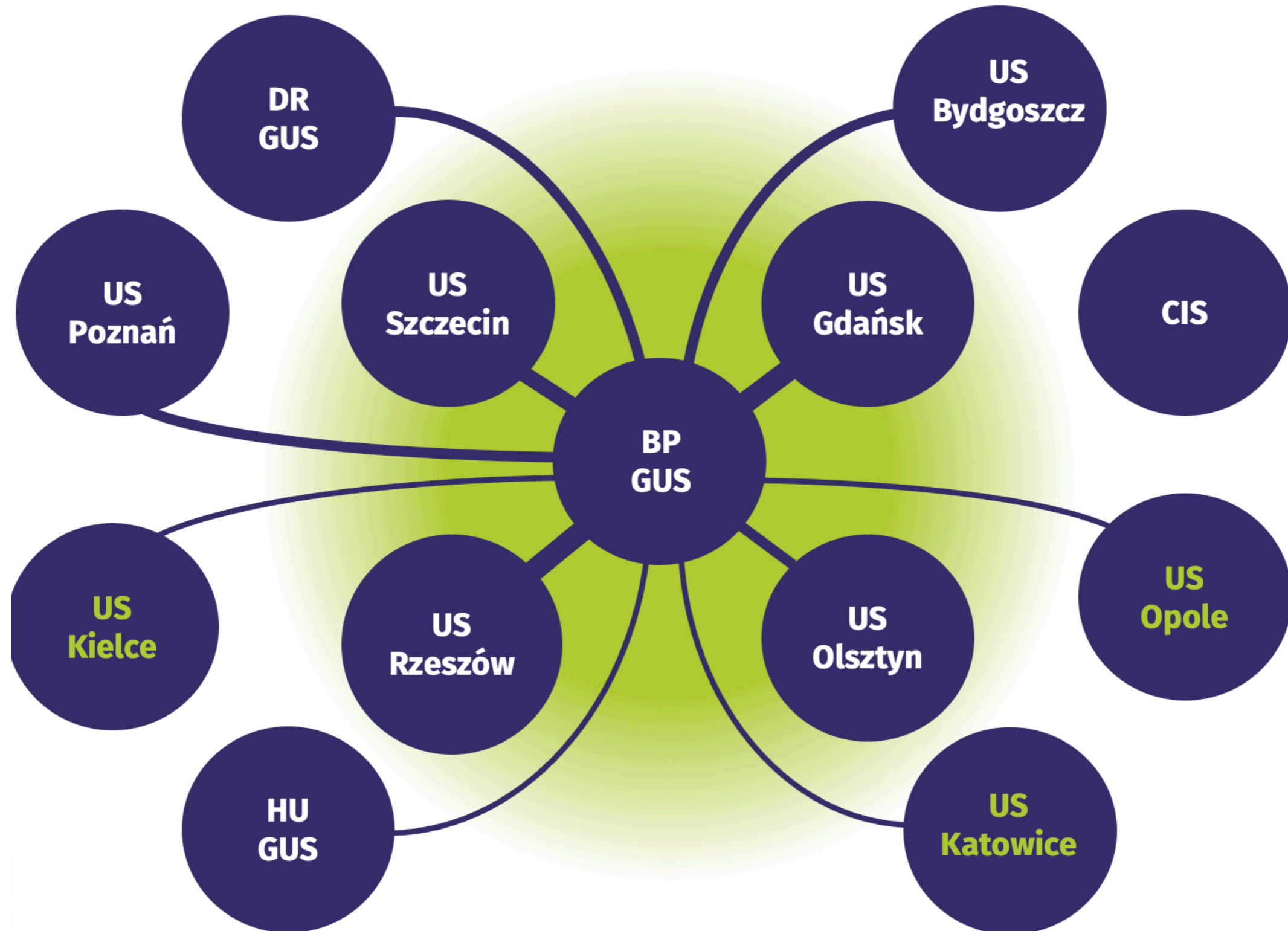
Statistics Poland - current projects

- ESSNet Big Data 2018-2020

- Gospostrateg (SATMIROL, INSTATCENY, TranStat)

- UNECE HLG-MOS – Machine Learning

- ESSNet Smart Surveys 2019-2021

- ESSNet Trusted Smart Statistics 2021-2025

Statistics Poland

Statistics Poland – current projects

ESSNet Big Data 2016–2018

ESSNet Big Data 2018–2020

ESSNet Smart Surveys 2020–2022

ESSNet TSS-WIN 2021–2025

Machine Learning 2019–2020

Machine Learning 2021

Gospostrateg (SATMIROL, INSTATCENY, TranStat)

Big Data meetings @STAT PL:

2016 2017 2018 2019 2020 2021 2022 2023 2024 2025

Statistics Poland

Accessing privately-held data (for official statistics)

- EC B2G Expert Group

- EC Expert Group on New Data Sources for Official Statistics

- UN Task Team on privately-held data

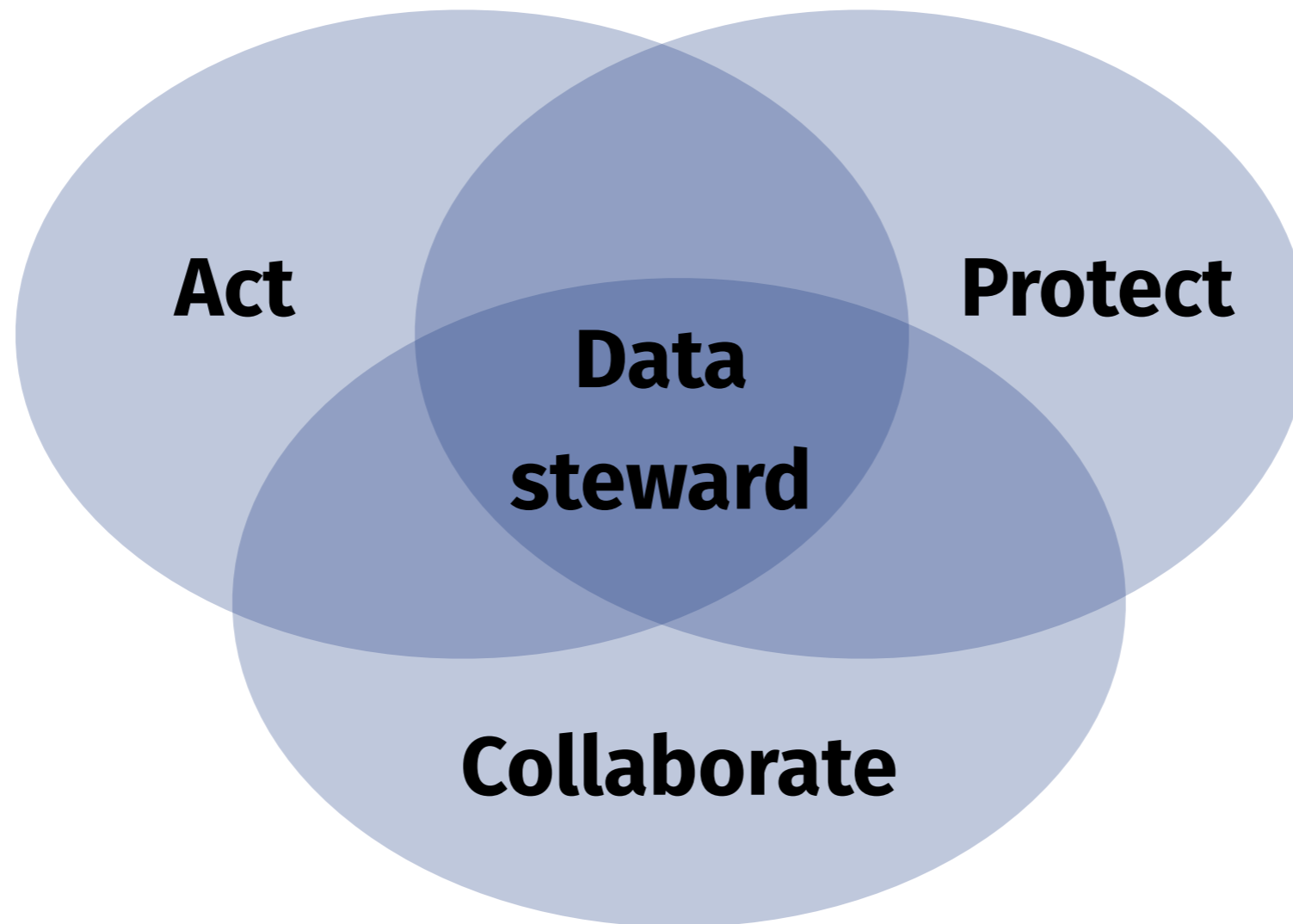- ESSC PG Task Force on accessing privately-held data

Statistics Poland

Making data open - yes but "both ways" access first, B2G work

- "Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing"

  - describes the data stewardship as one of the recommendation the creation of national structures and functions which is necessary for B2G data sharing

- The expert group recommends

  - to systematize process of partnering and help scale efforts

  - the promotion and recognition of data stewards

  - the development of digital skills and capacity building in order to ultimately create a data-sharing culture that cuts across the private and public sectors and fosters the principle of reciprocity

Statistics Poland

Pro-active and value creation role of Data Stewards in B2G report

Strategy and partnerships

Internal engagement

Achieving scale

leader or team of organisation well-versed in collaborative and responsible usage of data assets to provide public value

Communication

Data preparation and distribution

Statistics Poland

Pro-active and value creation role of Data Stewards in B2G report

Act

Protect

Data steward

Collaborate

Data stewardship

- help define and implement data definitions, shared metadata, standardized, controlled dictionaries and standards;

- set data quality guidelines that face requirements for what is considered to be good data quality;

- consistent management of data resources throughout the entire life cycle in order to preserve their quality, integrity, and consistency and avoid redundancies and integrity failures;

- facilitate re-use and provide access to data resources (for internal or external purposes);

- maintain high-quality metadata;

Statistics Poland

Data stewardship

- cooperate with others engaged in creating, collecting, accessing, using, sharing, maintaining data;

- collect needs and feedback on the quality of data resources for which they are responsible, ensuring that data is fit for purpose;

- be aware of the data protection policies, intellectual property, and information security, ensuring data is protected, and security procedures are enforced,

- organize and contribute to communication and promotional activities to increase awareness and use of data resources for which they are responsible;

- build, support, and share knowledge; help users understand the data better, recommend improvements;

Data stewardships and official statistics

Statistics Poland

Challenges

- Exploding Data Ecosystems

- Skyrocketing Number of Actors

- Overwhelming Complications of Complexity

- Lacking Skills, Competencies, Resources

- Insatiable Demand for Insight

Statistics Poland

Current activities

- UN HLG-PCCB

- UN Data Stewardship in Official Statistics Group - Workstreams

- UNECE Group on Data Data Stewardship in Official Statistics

- Eurostat (Partnership Group)

Statistics Poland

Current activities of statistical offices

- Coordination of activities for the development, production and dissemination of official statistics

- Access to administrative data for statistical purposes

- Being consulted in case of new information systems or change in information systems

- Document other statistics and assess their quality

Statistics Poland

Data stewardship vs current legal state of play

- Specific role of European Official Statistics stemming from their legal mandate

  - enshrined in the TFUE (art. 338)

  - developed by 223/2009 regulation and

  - national provisions related to official statistics

Statistics Poland

Data stewardship... whaaat?

What	How	Why
Data ownership Data custody Data management Data stewardship B2G data sharing	Role of the official statistics in a broader, technological progress-driven data ecosystem	Goal: Improve the use of data in society by making government and other data available to all users in an open and effective way

Pillars

Governance
Partnerships and collaboration
Access
Methods

Statistics Poland

Strategic questions on the roles of NSIs

- What will be the role of national statistical offices and official statistics in society?

- What are the traditional functions national statistical offices have to maintain? How the new role might change values and traditions?

- How national statistical offices will respond to competition from other organisations and what strategic partnerships are needed?

- How to maintain high trust in delivering quick insights to changes in society?

Source: 67th CES plenary session (26-28 June 2019)
The role of national statistical systems in the new data ecosystem (https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_25-1906075E.pdf)
The updated paper: Implementation of the new role of national statistical offices at the time of expanded possibilities https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2020/ECE_CES_2020_10-2005282_E.pdf

Statistics Poland

Leading role in data governance
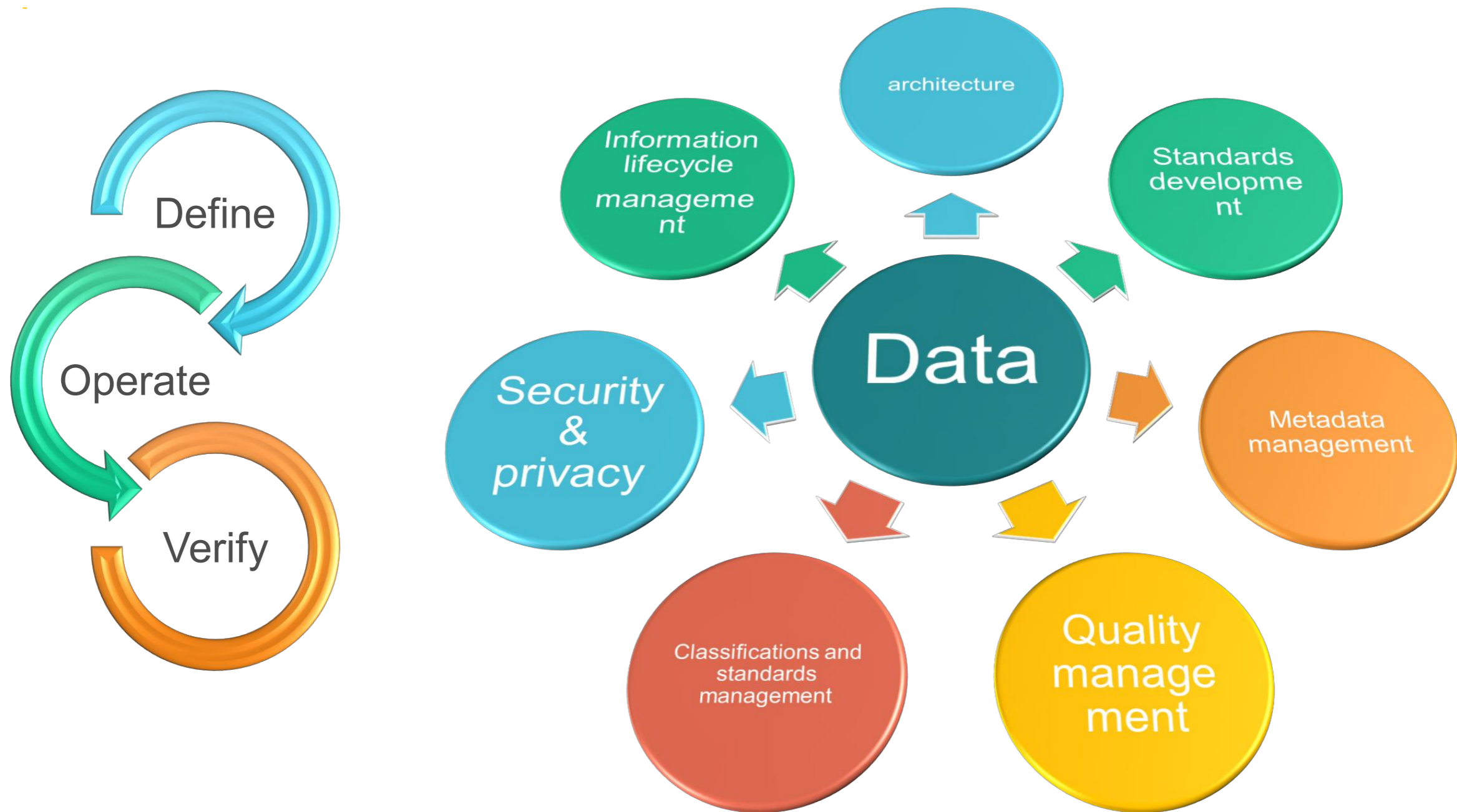
Data Stewardship activities
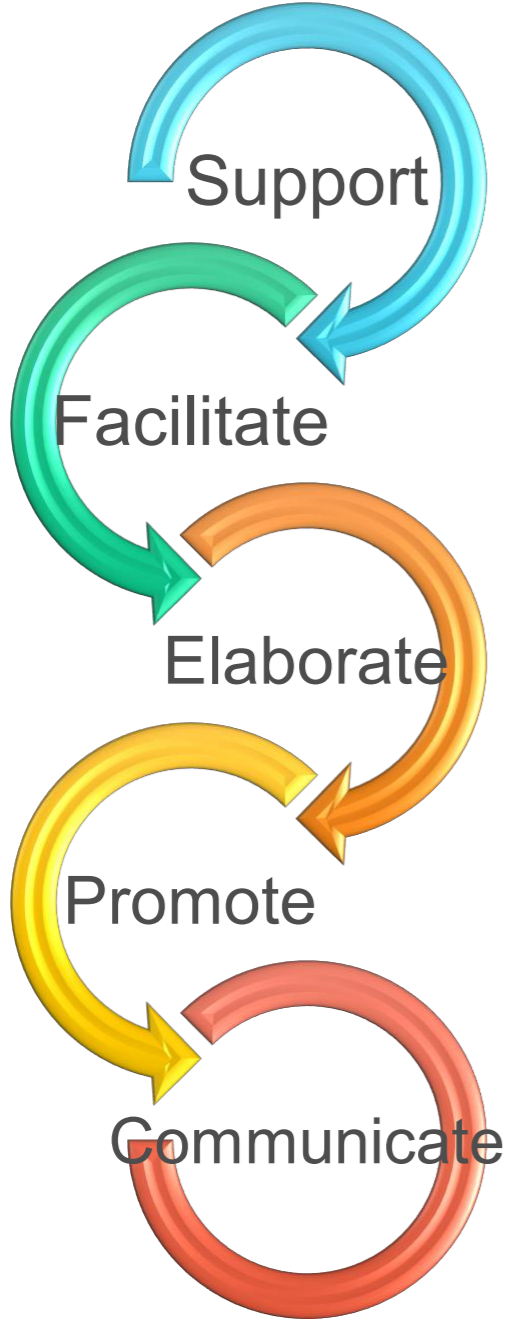
Data stewardship in official statistics

- sample and measure

- foster analysis

- provide insight

- open the data

- be in-between

- facilitate sharing

- enable innovation

- unleash the power of data

- inform society

Statistics Poland

Statistics Poland

Statistics Poland

Coordination role vs data stewardship

- 223/2009 regulation

  - Art. 5.1

    - The national statistical authority designated by each Member State as the body having the responsibility for coordinating all activities at national level for the development, production and dissemination of European statistics (...)

  - Art. 6.2 (f)

    - (the heads of NSIs shall)coordinate the statistical activities of all national authorities that are responsible for the development, production and dissemination of European statistics, as set out in Article 5(1);

Statistics Poland

Polish attempts to address new needs in law

- The revision of the the Law of 29 June 1995 on Official Statistics was made in the field of:

    - adding a definition of the term "non-public information systems",

    - entitling official statistics to collect personal data from non-public information systems,

    - introducing a guarantee that the provision of personal data to public statistics services by providers of publicly available telecommunications services does not constitute a violation of the confidentiality of communications on telecommunications networks.

Statistics Poland

A practical addendum –
A new model of collaboration – the Polish case

- "SMUP" – Public Services Monitoring System

  - a response to needs defined by local government administration

  - Development and implementation of a universal, publicly available on-line, providing integrated, high-quality data, that will allow the optimizations of the process of providing public services

- Partners

  - Ministry of Internal Affairs and Administration

  - Association of Polish Cities

  - Association of Polish Districts (counties, LAU1)

  - Silesian Union of Gminas and Poviats

Statistics Poland

Why such a service?

- Lack of a consistent monitoring system for public services, which would allow every citizen to gain access to information on the availability and quality of delivery of goods and services, for which public authorities are responsible

- The institutions responsible for the delivery as well as the recipients of services have limited possibilities to compare their

  - quantity and quality,

  - availability and financial efficiency

- Don't get mistaken.

  - It is not another thing like we used to do.

  - It's an integrated, coherent, and holistic data governance system for measuring public services served by regional and local authorities (not central government).

Statistics Poland

What can we attain/obtain?

- New model of collaboration with local government administration (being at the same time data provider and data user) will make it possible :

  - to jointly define their needs and design appropriate indicators to measure and monitor public services

  - to discuss how to monitor some public services & to exchange experiences with experts from Statistics Poland (in working groups)

- Benefits:

  - to strengthen and develop collaboration with local government administration

  - to develop or improve statistical products -in response to user needs

  - standardisation and re-use of collected data

  - reduce the local government administration's burden (currently they have to report to various public administration institutions).

  - increase of availability and usefulness of data collected so far in various structures and systems (registers) by different data owners (institutions of public administration)

Statistics Poland

Statistics Poland

Thank you

for your attention!