III Congress of Medical Statisticians is Changing World of Official Health Statistics: Subjectivity and Globalism

Moving from Health Data Statistics to Health Data Analytics

Zlatan Sabic, Ph.D.
The World Bank Group
zsabic@worldbank.org

Moscow, September 23, 2021

WORLD BANK GROUP

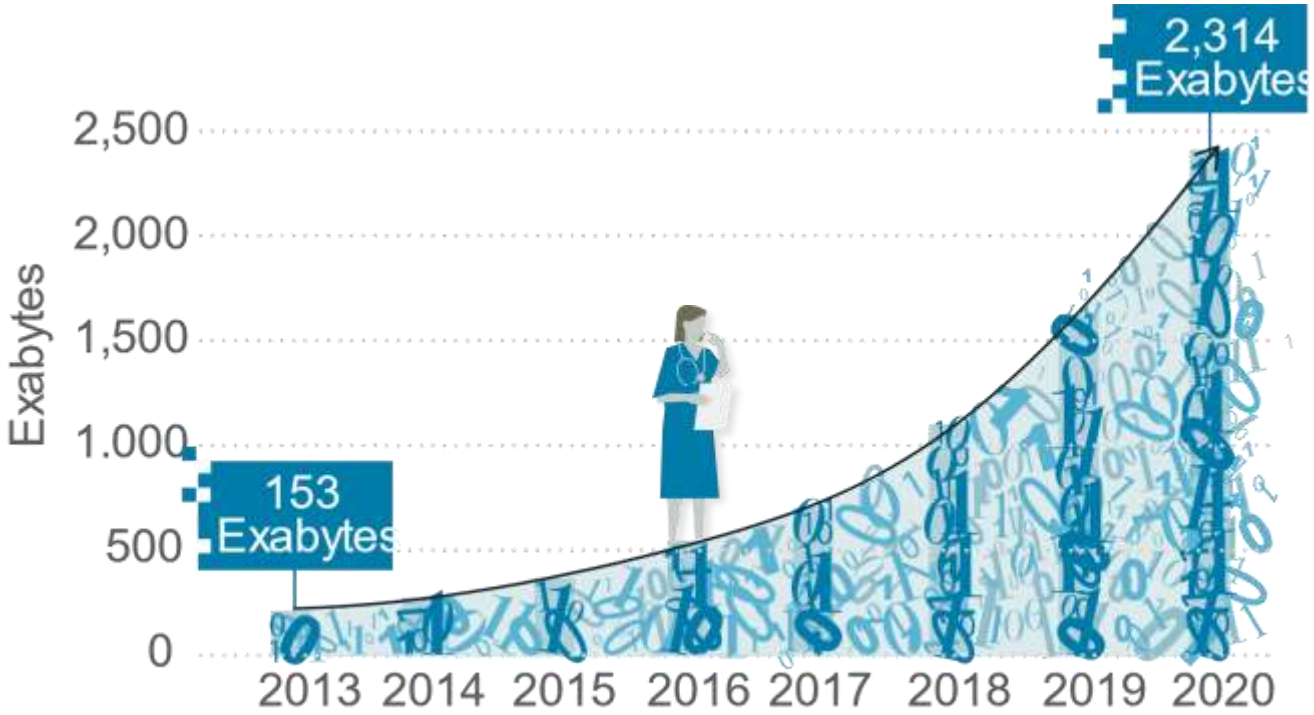
Health Systems Generate **Vast Amounts of New, Rapidly Changing Data**

By 2020, volume of health data in the world will exceed **2.3 million trillion megabytes**

2.5 trillion
megabytes added daily

30% of all global data are health data

80% of health data are unstructured



Source: Data Science and Digital Innovation: A Global Perspective, Marelize Görgens and David Wilson | The World Bank, 2019

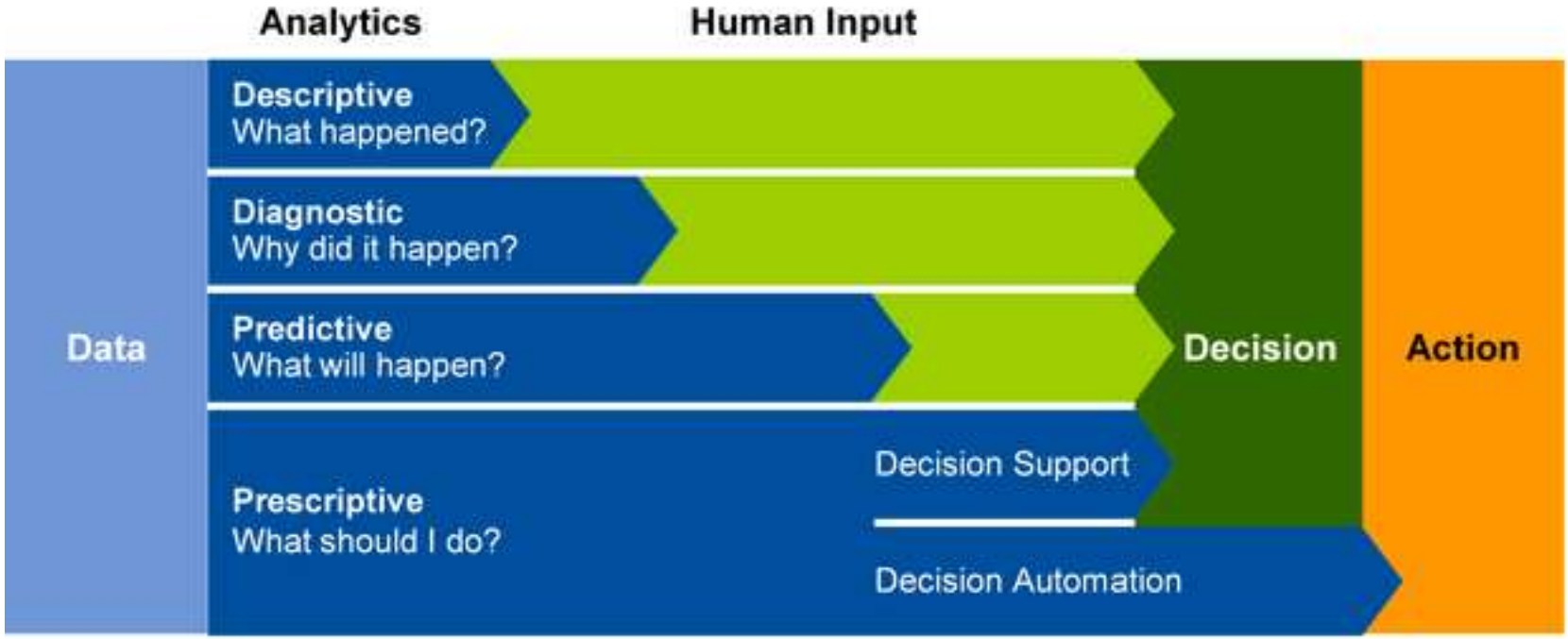
WORLD BANK GROUP

Health Data Statistics
It's all about reports and indicators?

- **WHO 100 core indicators** (*https://www.who.int/healthinfo/indicators/2018/en/*)
- **OECD framework**
- **EU core health indicators**
- **Basic healthcare system statistics/public health**
- **Public expenditure reviews**
- **Country specific (Australia, UK, Netherlands, USA, Canada, …)**
- **Strategy specific: Health Key Performance Indicators (KPIs)**
- **…**

A COVID-19 lesson: We need more dynamic and flexible system(s) of health data reporting and utilization

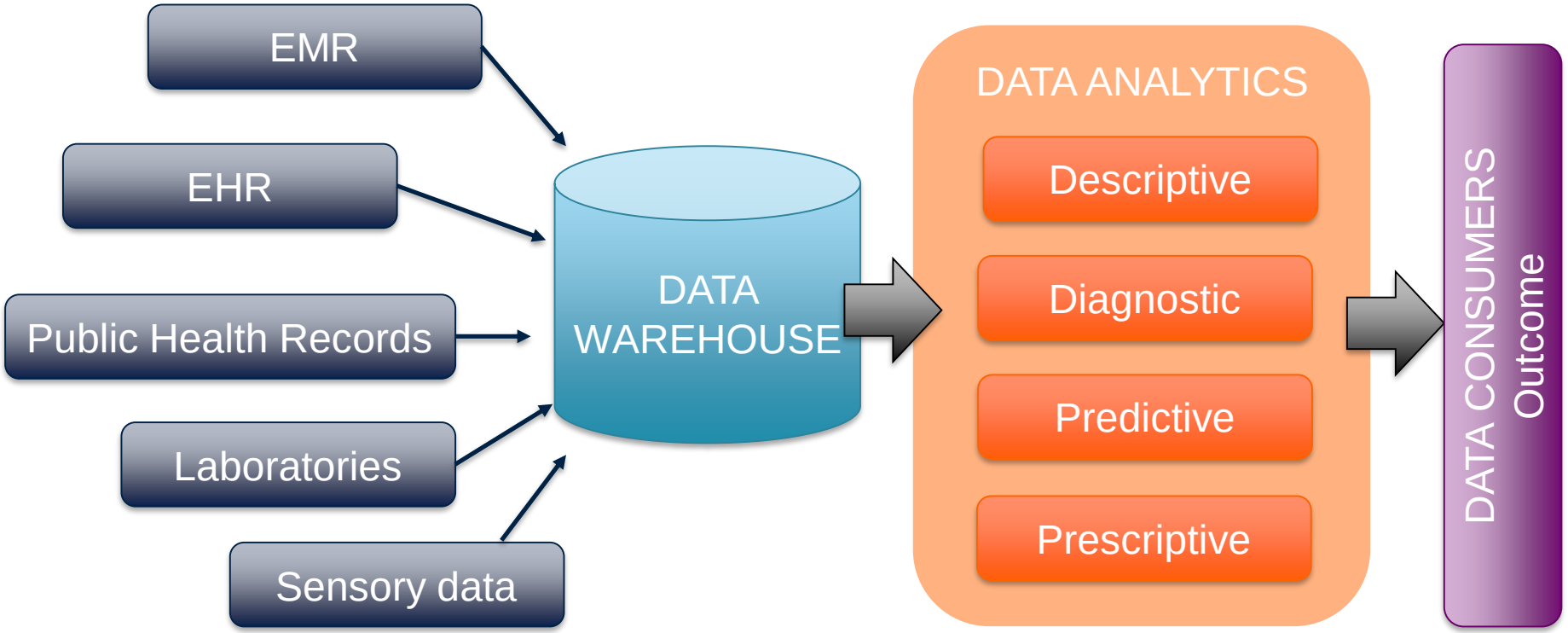
Moving from Statistics to Analytics



This requires tools but also new approach to data governance – what are data sources, quality of data, cross-use of data sets, dynamic vs. static data sets, dynamic access to data sets, reports and indicators, …

Source: Gartner (2014), Newsroom: Gartner Says Advanced Analytics Is a Top Business Priority, https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority

Health data – processing/tools



Tools are available, e.g.

IBM Watson Health – IBM DataProbe tool

- Iowa Medicaid Enterprise (IME) needed help fighting fraud, waste and abuse in its new managed care program.
- IME utilized the flexible analytic capabilities of IBM® DataProbe® software to run specialized algorithms that mined for the highest priority vulnerabilities.
- The team built a database by integrating and standardizing six years of fee-for-service and managed care data history, including claims (paid and denied), eligibility, provider files and several reference files.
- The team created a library of custom predictive analytics to identify complex schemes as well as simple billing mistakes.
- As a result, **over USD 41.5 million was recovered during a 2- year period**.

Source: https://www.ibm.com/case-studies/iowa-medicaid-enterprise-watson-health

Tools are available, e.g.

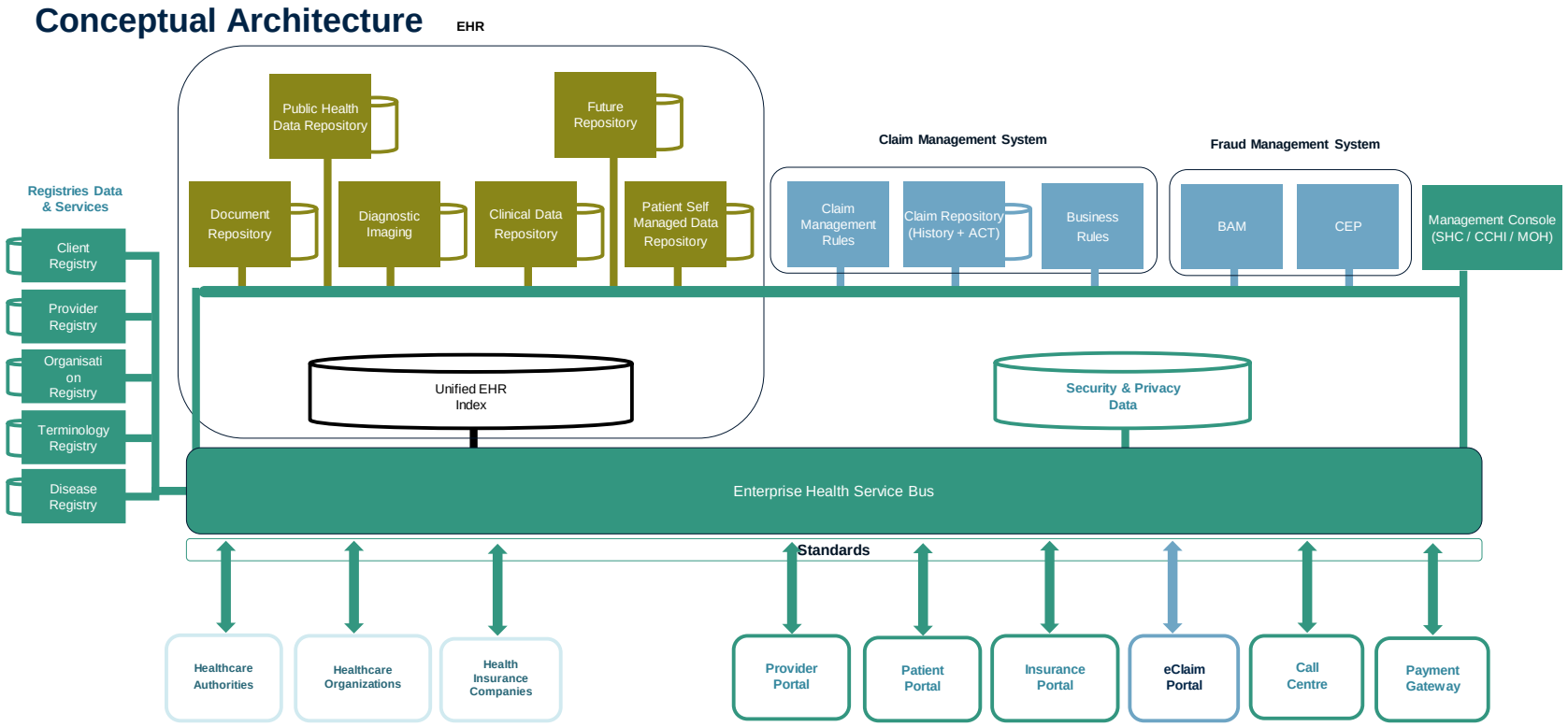
AyasdiIA for Healthcare

- Clinical Variation Management (what's going on in hospital; best care practices)
- Population Risk Stratification
  - uses machine intelligence to discover nuanced subpopulations automatically,
  - predicts future risk trajectories and drivers of risk, and
  - inform the most effective interventions for delivering the best outcomes
- Advanced Analytics for Fraud, Waste, and Abuse Detection
  - Identify new patterns of aberrant behavior
  - Validate and improve existing detection models
  - Prioritize fraud leads
  - Improve detection in the pre-payment cycle

Source: AyasdiIA for Healthcare: Intelligent Application Suite, https://www.ayasdi.com/wp-content/uploads/2018/12/Ayasdi-for-Healthcare-06.03.20.pdf

Tools are available (technology level), e.g.

Saudi eHealth Exchange (SeHE)



Source: eHealth Standard-based Interoperability framework and policy, Saudi Health Council, 2019.

However, there are challenges

- **Accuracy.** Reporting of patient data into EMRs or EHRs is not entirely accurate yet.
- **Unified format.** Huge volume of data that is not easy to capture with traditional EHR format. No perfect data organization by healthcare providers. A need to codify. Medical coding systems like International Classification of Diseases (ICD) code are useful but have their own limitations.
- **Cleaning.** The data needs to cleansed to ensure the accuracy, correctness, consistency, relevancy, and purity after acquisition. Use of machine-learning techniques to reduce time and expenses.
- **Meta-data.** To have complete, accurate, and up-to-date metadata regarding all the stored data.
- **Querying.** In absence of proper interoperability between datasets the query tools may not access an entire repository of data. Medical coding systems like ICD, SNOMED-CT, or LOINC help.
- **Data sharing; Security; Visualization; Image pre-processing; Storage;**

Source: Dash, S., Shakyawar, S.K., Sharma, M. *et al.* Big data in healthcare: management, analysis and future prospects. *J Big Data* **6**, 54 (2019). https://doi.org/10.1186/s40537-019-0217-0

To move from data statistics to data analytics

- **Introduce dynamic analytical data sets**
- **Ensure the alignment of health data analytics with business objectives**
- **Aim for balanced data governance**

- Invest in data sources
- Respect general data privacy and security regulations and standards
- Aim for advanced data analytics, but respect maturity levels
- Track data utilization

Dynamic Analytical Data Sets
- the core of new approach -

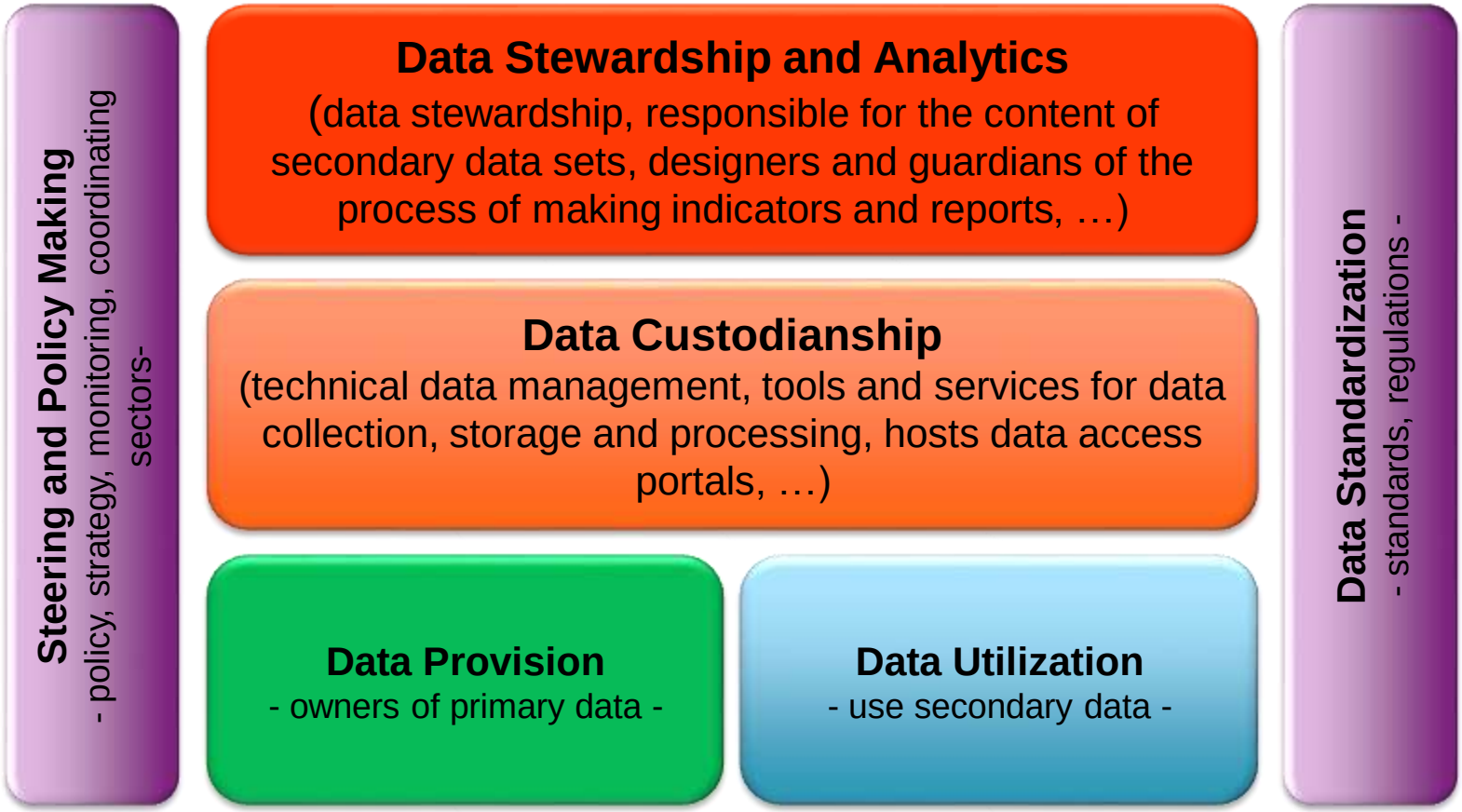
analytical data set is a group of individual but de-personalized data elements collected according to specified rules of particular domain/priority topic or subdomain

- automatic feed and de-personalization from primary data sets
- available to data consumers in raw form (not summarized into indicators, or pre-processed) – can have different meanings in different contexts
- Independent, transparent and freely available for the use by authorized and certified users
- catalogs of analytical data sets are defined dynamically

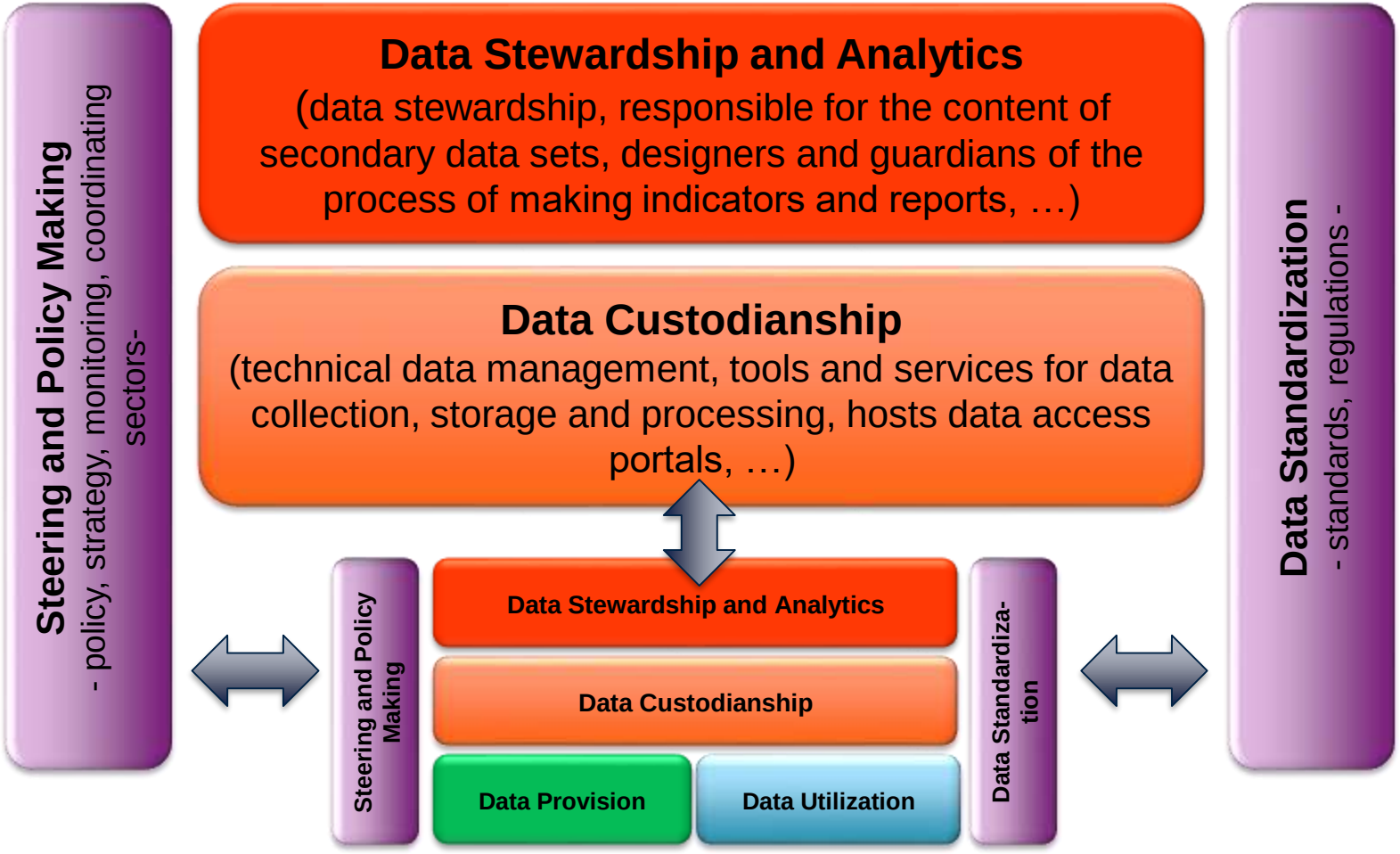
Align health data analytics with business objectives - the core of new approach -

- Do not collect data for the sake of collecting data, or “because we can”
- Health data should be collected and processed with the clear purpose of achieving the business objectives of stakeholders
- The data analytics model should be flexible (allowing for both central and distributed data analytics) and adaptable to the data consumers' needs
- Align dynamically metadata and catalogues to the business purpose:
  - Analytical data sets
  - Indicators
  - reports

Balanced Health Data Governance
- the core of new approach -



REPLICATED HEALTH DATA ANALYTICS GOVERNANCE MODEL TO SUB-NATIONAL LEVELS



III Congress of Medical Statisticians is Changing World of Official Health Statistics: Subjectivity and Globalism

THANK YOU!

Moving from Health Data Statistics to Health Data Analytics

Zlatan Sabic, Ph.D.
The World Bank Group
zsabic@worldbank.org

Moscow, September 23, 2021

WORLD BANK GROUP