Economic and Social Commission for Western Asia

Enhancing Capacity Development Using Data Science to fill the SDG indicator gaps

Cluster 4: Statistics, Information Society and Technology

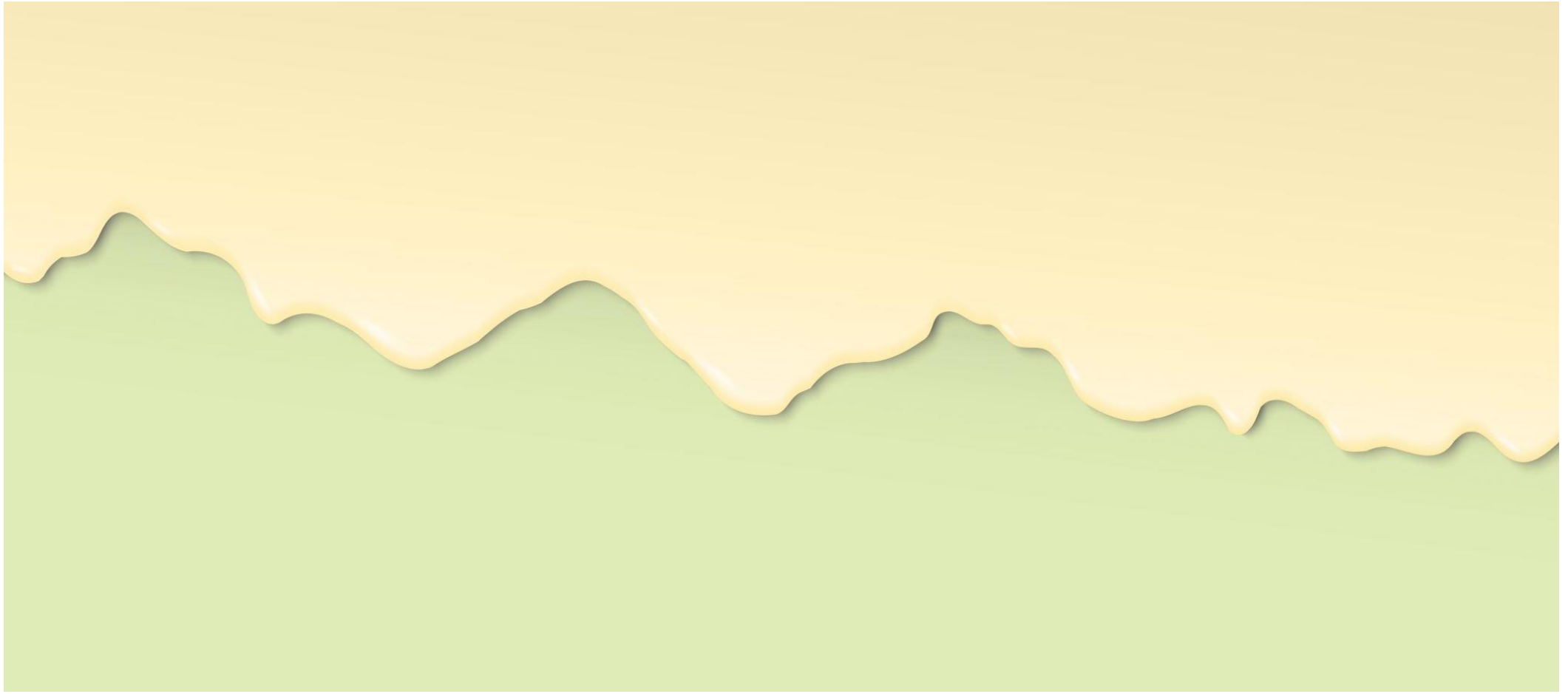
UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

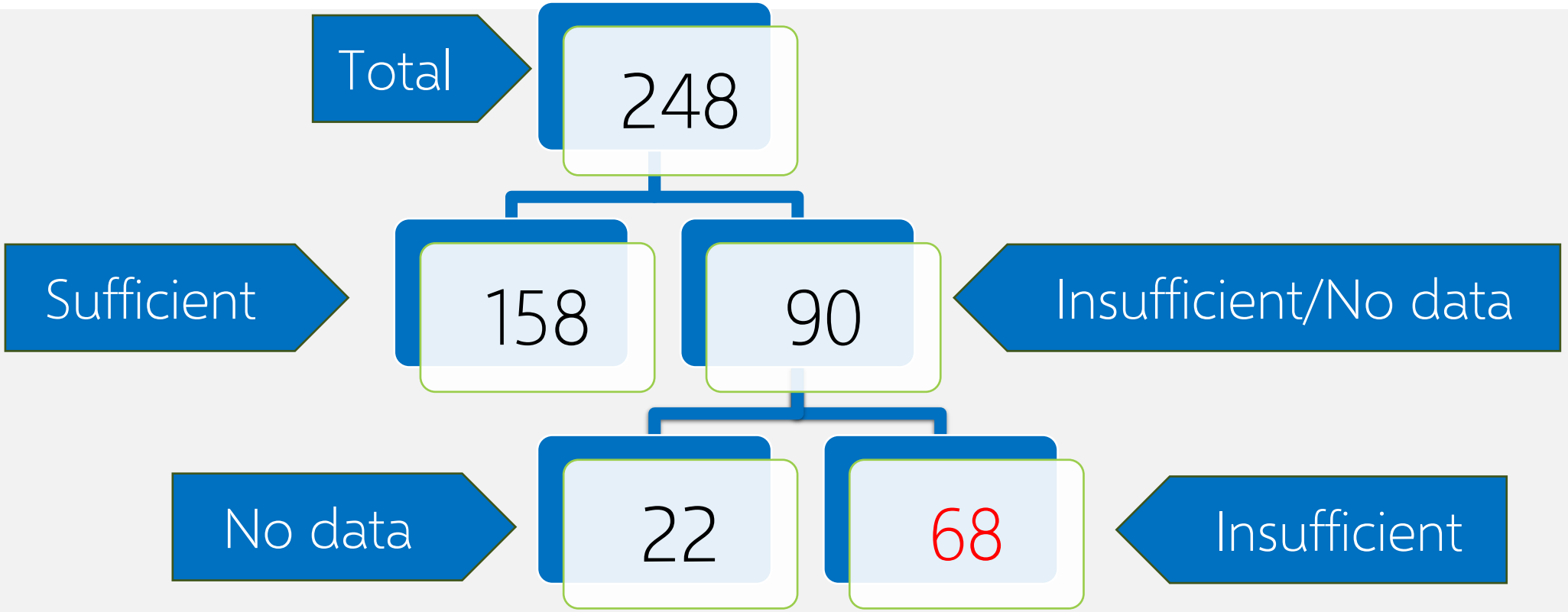
Agenda

1. SDGs indicators availability issues and Data science
2. Geospatial for measuring SDGs indicators: example
3. Data Science for SDGs: methods and example
4. Capacity development and regional cooperation
5. Conclusion Remarks

1. SDGs indicators availability issues

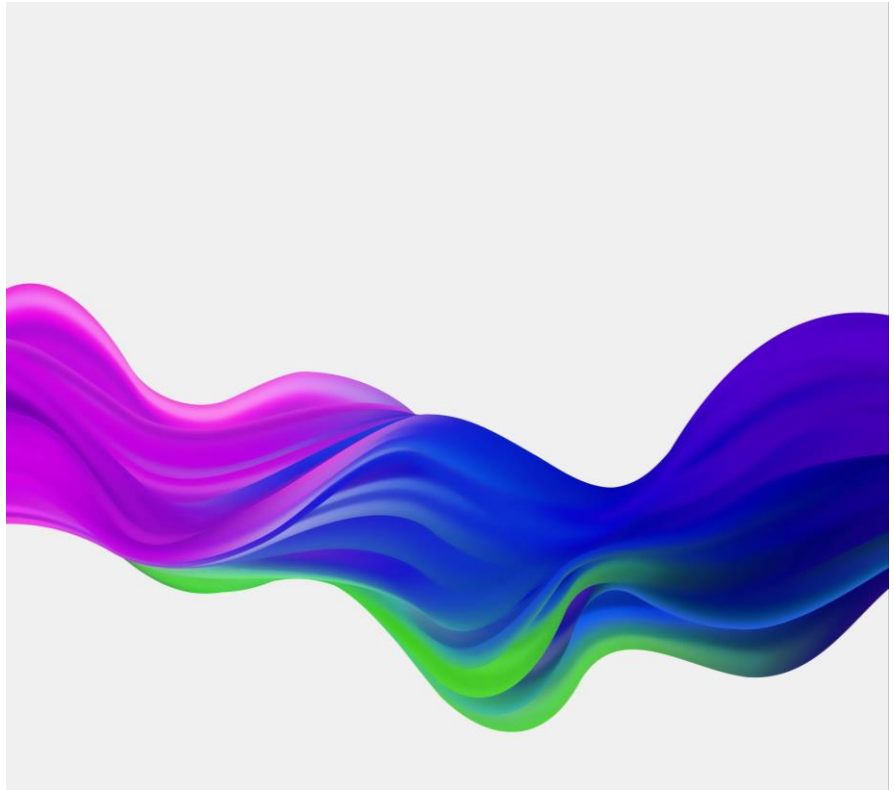


Map of SDG indicators From ESCWA Portal



Assessment of Goal 1

- Large data gap
- Big budgets to fill the gap
- Data science and big data solutions



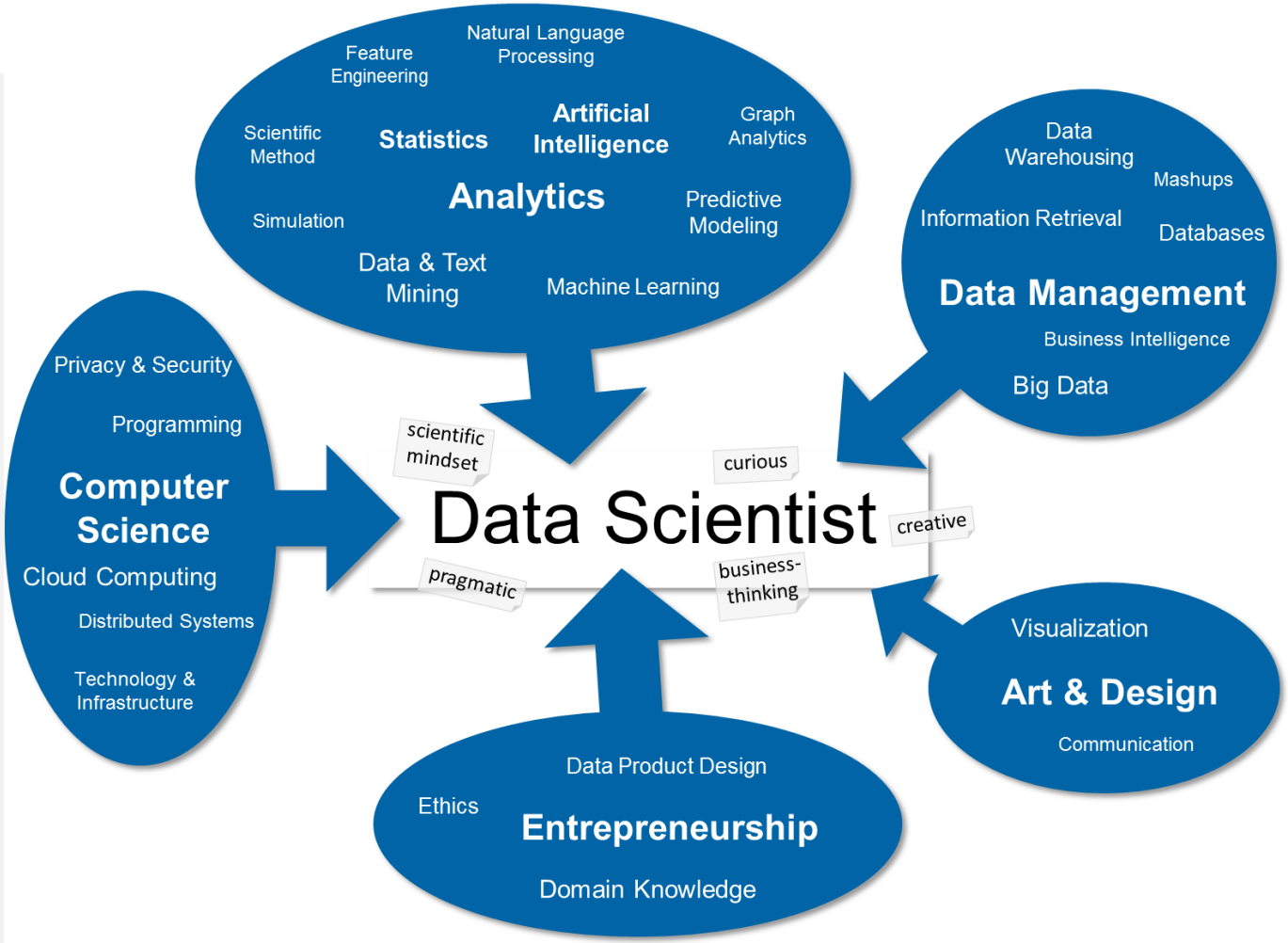
Example: the cost of 1.2.2

Type of Survey	Estimated Budget Range	Factors Affecting Costs
Household Income and Expenditure Surveys (HIES)	Hundreds of thousands to several million dollars	Survey scale, frequency, complexity, and sample size
Census Data	Tens of millions to hundreds of millions of dollars	National population size and complexity
Labor Force Surveys	Several hundred thousand to a few million dollars	Sample size and data collection methods
Social and Demographic Surveys	Hundreds of thousands to a few million dollars	Survey scope, sample size, and complexity
National Poverty Surveys	Several hundred thousand to a few million dollars	Survey complexity, coverage, and data collection methods
Administrative Data	Varies; typically, lower compared to dedicated surveys	Data management and reporting requirements
Data from International Organizations	Minimal direct costs	Analysis and integration into national reports

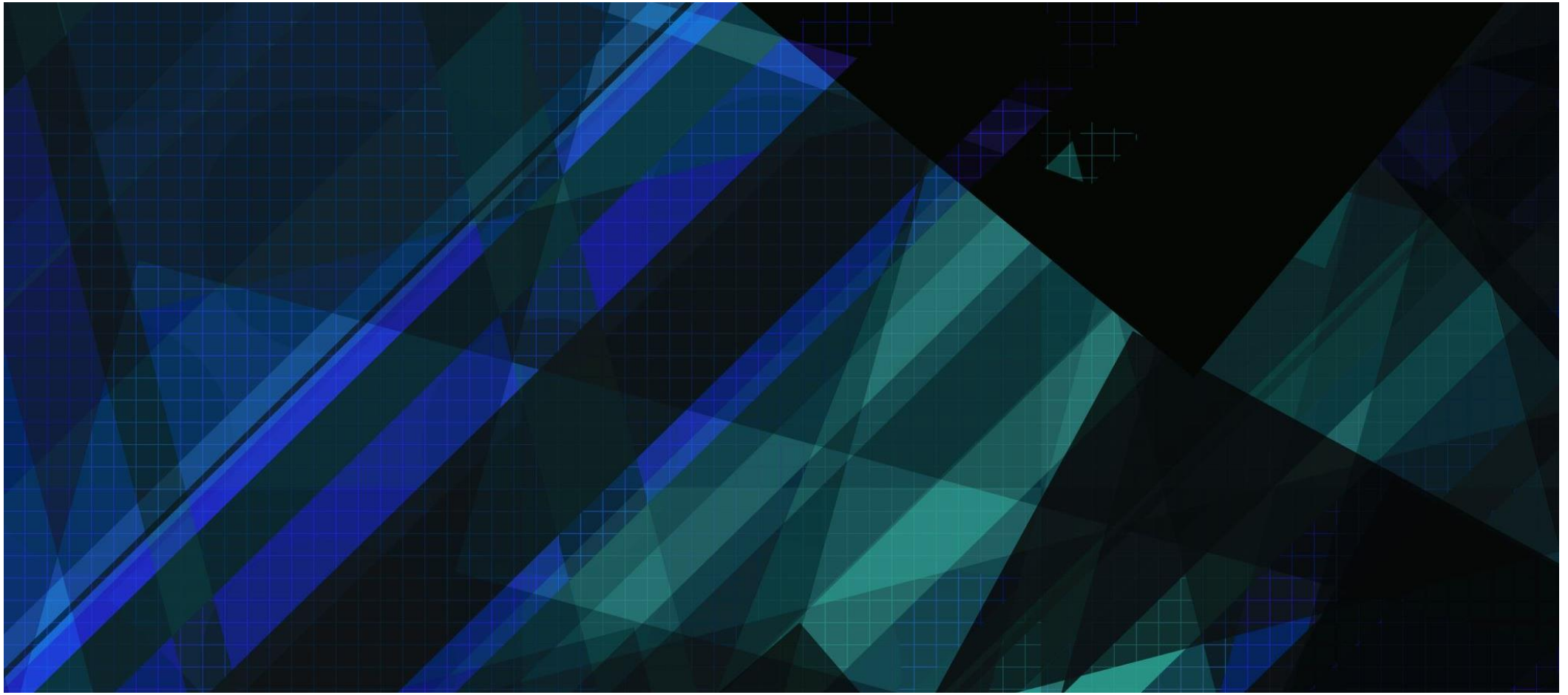
Science and Data science

Science	Data science
is a systematic way of building and organizing knowledge in the form of testable explanations and predictions.	is a branch of science that deals with collecting, cleaning, analyzing and extracting knowledge and insights from data.
studies the natural and physical world through **experiments**, **observations** and **measurements**.	studies and analyzes data using **mathematics**, **statistics**, **computer science** and **machine learning** techniques
both use the **scientific method** to understand phenomena and solve problems	

Data science skills



2. Geospatial for measuring SDGs indicators: example



Geospatial information for measuring SDGs indicartors

Most of the indicators include elements of geography, place, and location.

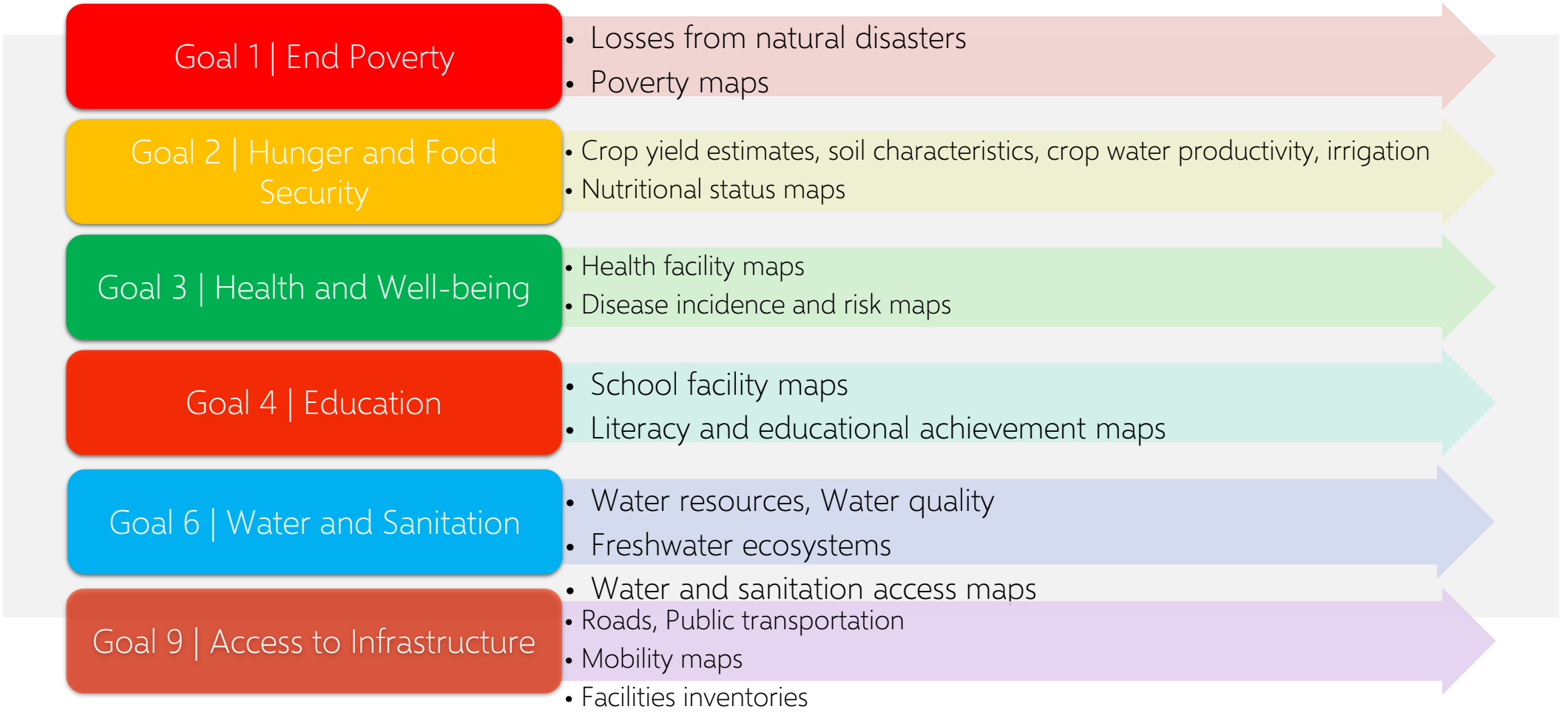
- As data in itself – geospatial data is used directly for the indicator construction (geospatial data = indicator)
  - ✓ **Indicator 15.1.1:** Forest area as a percentage of total land area

- Support statistical data – geospatial data is used in combination with other data to estimate an indicator (geospatial and other data -> indicator)
  - ✓ Indicator 11.3.1 Ratio of land consumption rate to population growth rate

- Enrich statistical data – geospatial data is used to enrich the indicators, although the indicator does not require a geospatial breakdown (analysis, enrichment of the indicators)
  - ✓ Indicator 6.3.2: Percentage of water bodies with good ambient water quality

- Geospatial data can help in communication and gives possibilities for geographical disaggregation of data:
  - 231 Indicators disaggregated by geographic location, urban/rural, region, etc.
  - Administrative data often come with geospatial information (e.g. address, administrative unit, etc.).

Geospatial information for measuring SDGs indicators



Geospatial information for measuring SDGs indicators

Goal 11 | Cities	• Access to public green space • Substandard housing maps
Goal 12 | Sustainable Consumption	• Energy productivity maps • Pollution maps
Goal 13 | Combating Climate Change	• CO_2 emissions • Exposure to extreme storms and droughts
Goal 14 | Marine and coastal ecosystems	• Coastal/Marine protected areas • Harmful algal blooms • Eutrophication
Goal 15 | Terrestrial ecosystems	• Land cover, land degradation, bio-diversity • Protected areas
Goal 16 | Peaceful and inclusive societies	• Maps of political violence • Crime maps • Refugee and IDP movement

Geospatial information for measuring SDGs indicators: example

Indicator 11.3.1 Ratio of land consumption rate to population growth rate

Data required:

- Population size at 2 different years (census/WPP)

- Built up areas (Census/GI)

Steps:

- Extract built up areas for each measurement year:

- Delimit city boundaries for most recent 4 year: (Geospatial Proc

- Compute land consumption rate based on total built up area within boundary
LCR (%) = (**LN Urbt+n/Urbt**)/ (**y**)

-Compute population growth rate **PGR** (%) = **LN Popt+n/Popt** (**y**)

- Compute core indicator: **LCRPGR** = **LCR /PGR**

3. Data Science for SDGs: methods and example

Steps

Data preparation Challenge: The Data Transformation

Data processing challenges: Tool Selection for Dealing with Missing Data

SPSS, Stata, Python, and SAS: Strengths & Limitations

Imputation challenges: Multiple Imputation Methods

Validation of the results challenges: Statistical Test Methods, Comparing Data Distributions, etc

But who will accept and how to adopt

Conclusions

Data preparation Challenge: The Data Transformation

- Understanding the Data
- Data Cleaning
- Scaling and Normalization
- Data transformation from long format to wide format
- “Similar” countries in the same data set
- Data availability

Tool Selection for Dealing with Missing Data

Feature	SPSS	Stata
Imputation Methods	-	+
Data Management	+	-
Estimation and Inference	-	+

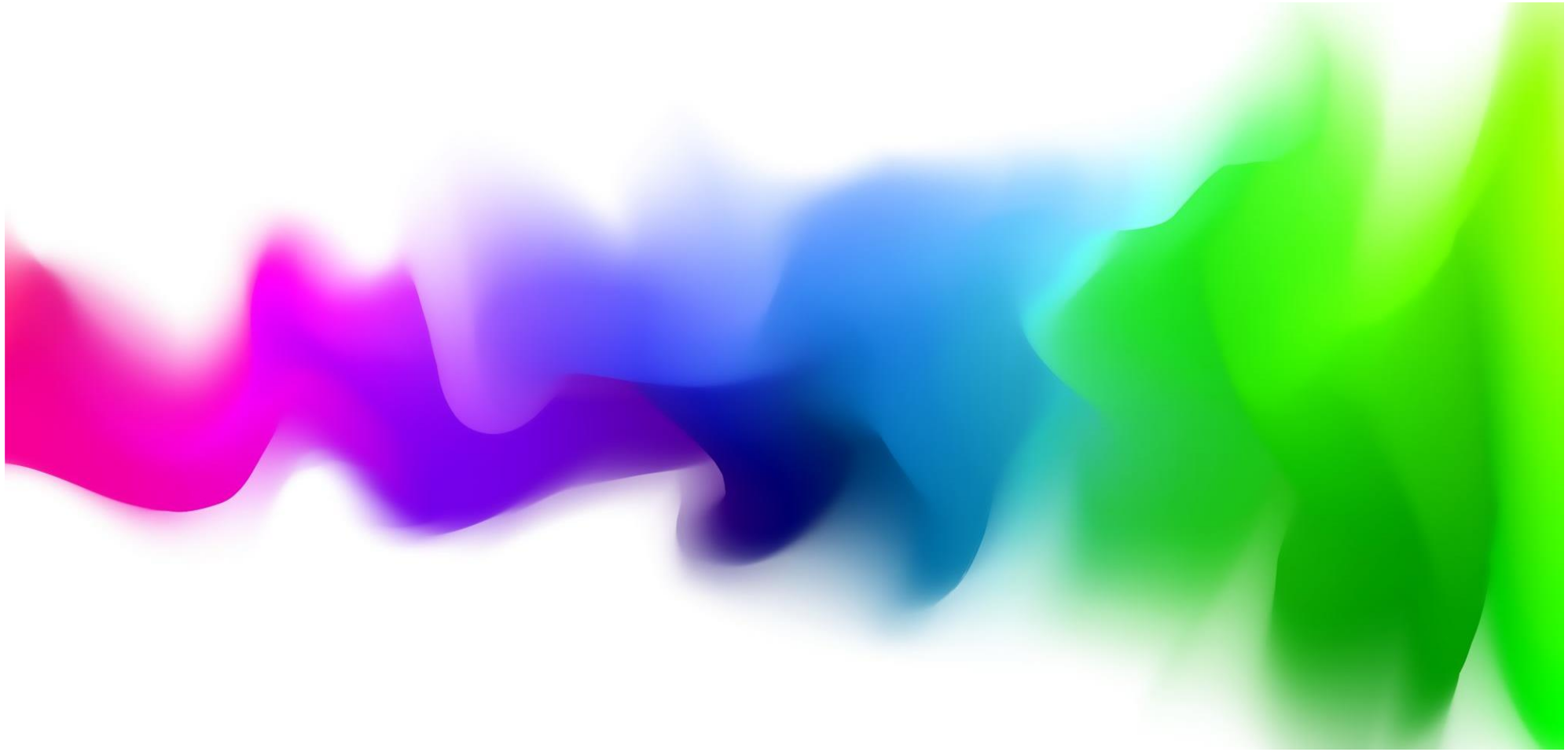
Imputation challenges: Multiple Imputation Methods

Imputation Methods
KNN imputation
EM imputation
Multiple imputation by chained equations (MICE)

Validation of the results challenges

- MCAR test
- Mean test
- Stddev test
- Distribution test
- MSE

4. Capacity development and regional cooperation



Basic concept in big data for official statistics and the SDGs

- Enhance capacity of NSOs in the Arab region
- Provide solid foundation in fundamental concepts of big data for official statistics
- Highlight relevance to SDG implementation and monitoring

Basic concept in big data for official statistics and the SDGs

- ❑ Workshop 1: Basic Concept in Data Science and Big Data for official statistics and the SDGs
- ❑ Workshop 2: Data Science and Big data for official statistics and the SDGs with Python
- ❑ Workshop 3: Data Science and Big data for official statistics and the SDGs: uses cases

5.Conclusion Remarks

- Technological advances are rapidly transforming National Statistical Offices (NSOs).
- ICT including Geospatial Information Technology, is a cross-cutting technology to help monitoring SDGs
- Modernization of practices in the Arab region is necessary.
- Focus is on capacity building tools such as specialized training programs, collaborative data partnerships, and adoption of data science and technology.
- Approaches enhance capacity development of individuals and institutions to collect, analyze, and report data.
- Harnessing data science empowers stakeholders to bridge information gaps.
- This ensures accurate and comprehensive progress monitoring for a sustainable future.

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Thank you