

Data Modernization Using Advanced Technologies

November 15, 2022



ازدهار البلدان كرامة الإنسان



الأمم المتحدة

الاسكوا
ESCWA

Foundational base



ازدهار البلدان كرامة الإنسان



SG Data Strategy
ESCWA Data Strategy
ESCWA Digital Strategy
ESCWA Innovation Strategy

Initiatives:



ازدهار البلدان كرامة الإنسان



Manara (Regional Knowledge Hub)

Scattered portals

Scattered statistical databases

Big data experiments

GIS use cases

Measurement

SDGs support

ESCWA Data Eco System

Reform:



ازدهار البلدان كرامة الإنسان



Clustering
Projectization
Internal Synergy
Data infrastructure reform

Value:



ازدهار البلدان كرامة الإنسان



External Synergy
Tool-based work (portals, visuals, etc)
Surgical policy support
Data democratization



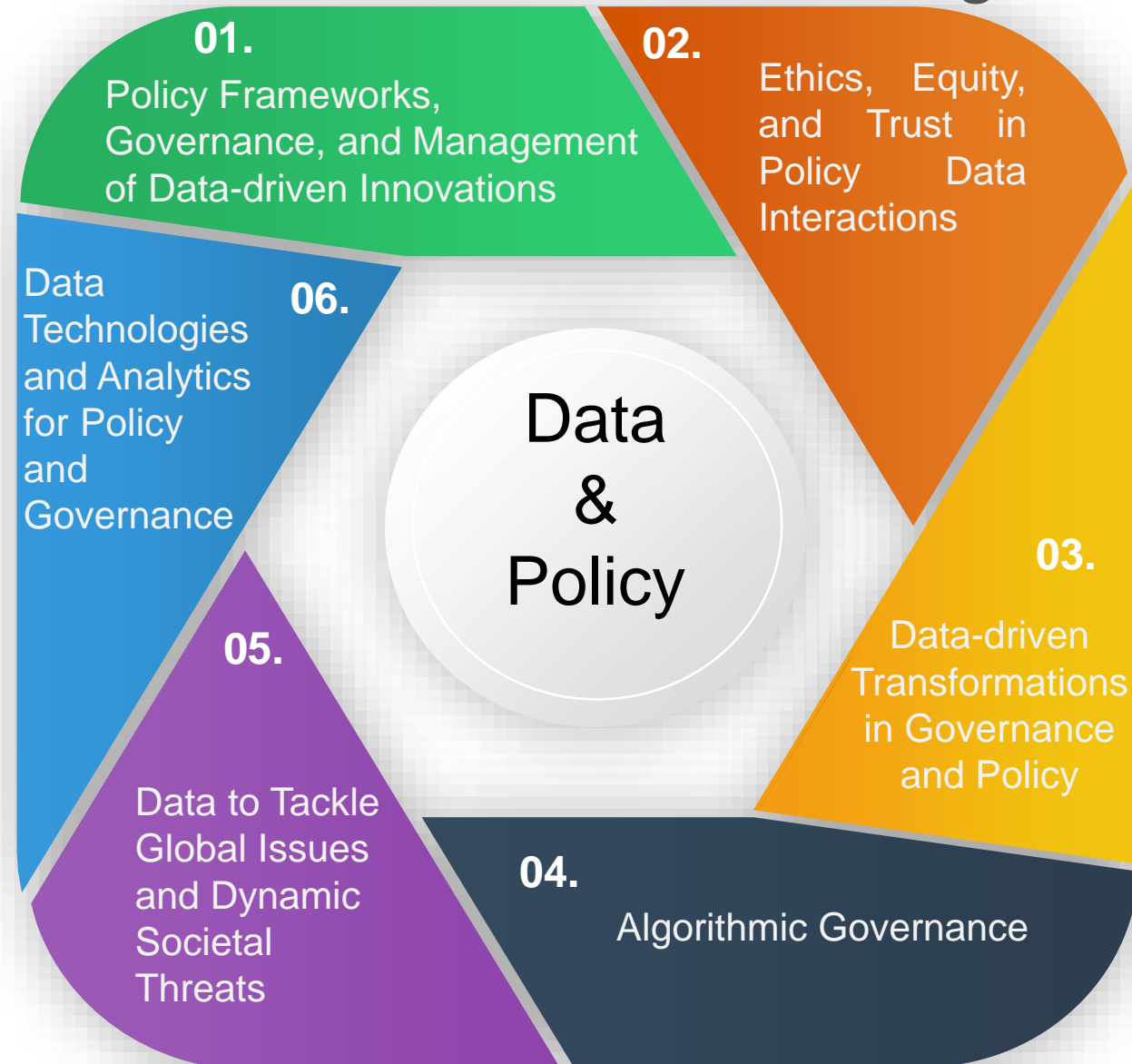
ازدهار البلدان كرامة الإنسان



Data Modernization through a Policy Lens:

6 pillars at the synergy of data & policy

6 Pillars of Data Modernization through a Policy Lens:



Ethics, Equity, and Trust in Policy Data Interactions

(1 of 2)

Why ID matters

63 million people or **15 percent** of the population in the Arab region lack official proof of identity - typically members of the poorest and most vulnerable groups

1 in 2 women in low-income countries does not have an ID, limiting their access to critical services and participation in political and economic life

Source: World Bank, "ID4D Global Dataset", 2018

Why Digital ID matters

161 countries have ID systems using digital technologies

3 to 13 percent of GDP value equivalent could be unlocked through implementing digital ID by 2030

50 to 70 percent of the full economic potential could be unlocked by **basic digital ID alone** in emerging economies

100 uses cases of digital ID systems identified in seven countries

Source: McKinsey Global Institute, "Digital identification: A key to inclusive growth", 2019

Ethics, Equity, and Trust in Policy Data Interactions

(2 of 2)

Types of Digital Identities

Type	Identities
Individuals	National ID (NID)
	Civil Registration
	Social Security Number
	Personal Tax ID
	Drivers License Number
	Voting Registration Record
	Hospital/School/Job ID
Financial	IBAN/SWIFT (Bank IDs)
	Bank Account Number
	Blockchain Wallets
Ideas/Arts	Trademark/Patent ID
	NFTs

Type	Identities
Businesses	Business Registration
	Business Tax ID
	Import Export ID
	Vendor ID
IT Equipment	Device ID
	Phone ID (IMEI)
	Sim Card ID (ICCID)
	Phone Number
Virtual	IP Address
	Social Media Account
	Email Account
	Avatars (Metaverse, VR)

Data-driven Transformations in Governance and Policy

(1 of 3)

Research on policy indicates that *policy modeling* has witnessed innovations in parallel to the increased availability of data sources and tools, and currently includes 12 areas:

Domestic and international trade	Energy, communications, infrastructure, and transportation
Environmental and natural resources management	Fiscal and government spending
Institutional, regulation, and negotiation	Labor, employment, and population
Monetary, banking, and investment	Production and consumption
Technological and R&D	Welfare and social
Economic growth and development	Miscellaneous policy modeling

Source: Ruiz Estrada MA (2011) Policy modeling: Definition, classification and evaluation. *Journal of Policy Modeling* 33: 523–536. Available at: <https://doi.org/10.1016/j.jpolmod.2011.02.003>



ازدهار البلدان كرامة الإنسان



Data Technologies and Analytics for Policy and Governance

The example of the ESCWA Data Ecosystem

What would it take to build a successful data ecosystem

First things first...

KEY ISSUES TO ADDRESS

(simplified)

Vision

- Why are we interested in data?
- What are we trying to achieve?
- Is the goal improvement of current practices or radical transformation?

Use cases

- What are the main macro use cases?
- How much value do we expect from each?
- What are their key constraints and requirements?

Analytics

- Do we have a team with the necessary profiles?
- Have we implemented the appropriate analytics functions?
- Do we have a thorough analytics process?

Data governance

- Have we defined and measured data quality indicators?
- Do we have in place a data management organization?
- Have we launched basic data hygiene actions?

Data infrastructure

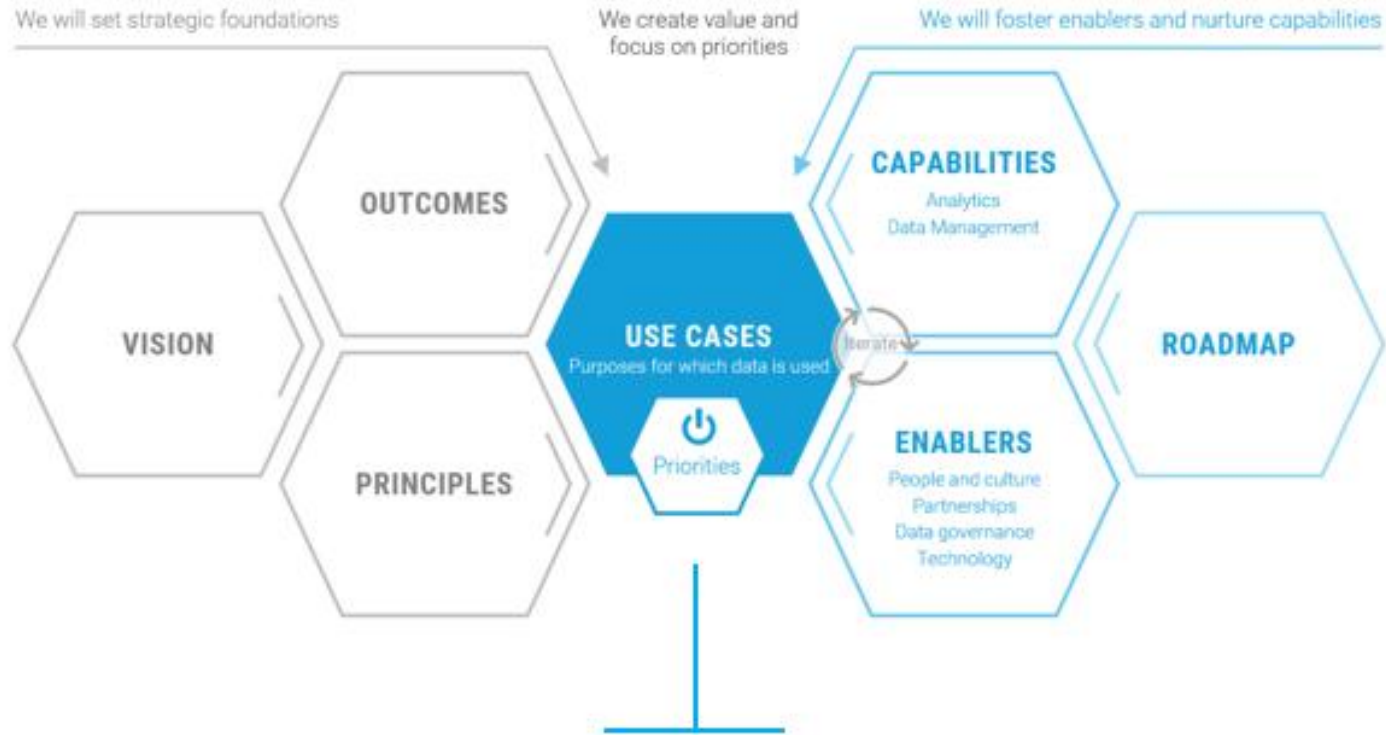
- Do we have an infrastructure suitable to our vision and to future use cases?
- Are we using the proper technologies?
- Do we have the appropriate operating model?



ازدهار البلدان كرامة الإنسان



A problem-driven approach to building enablers and capabilities



Strategy for the Development of a Data Ecosystem

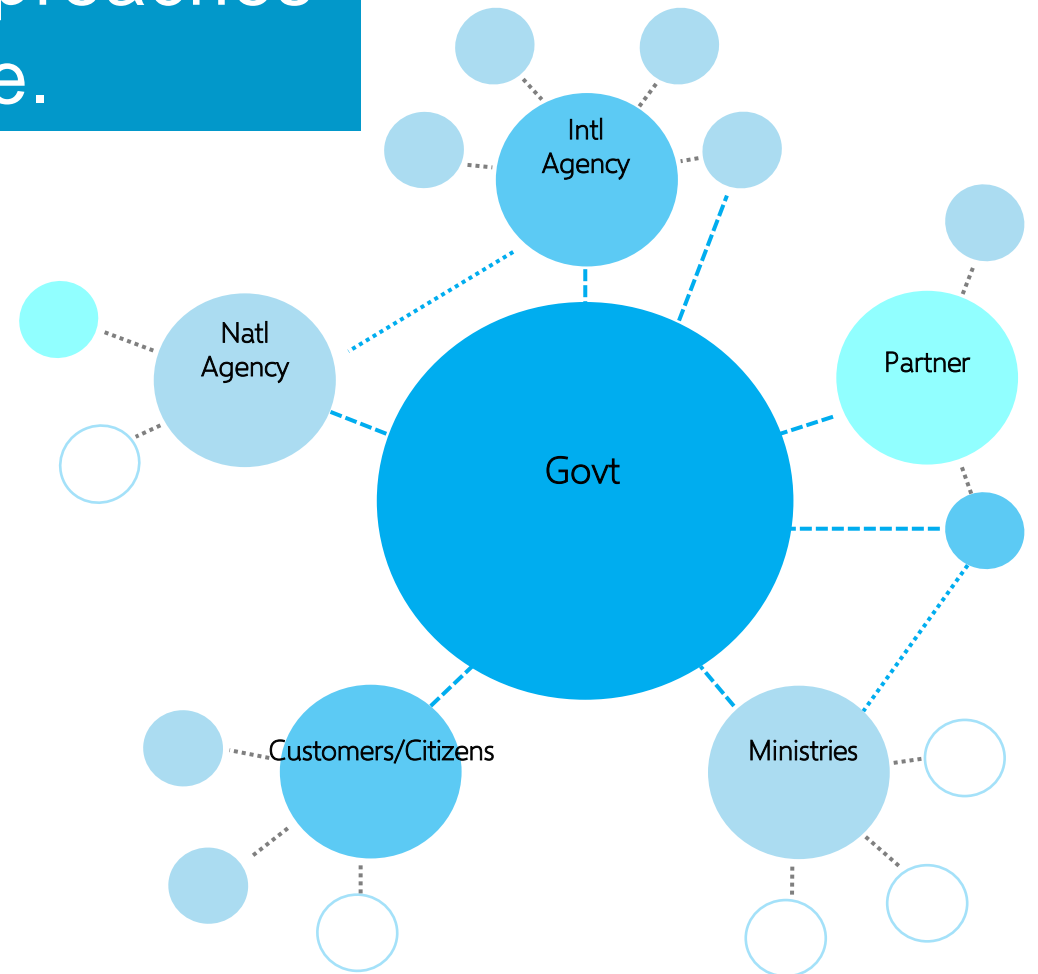
The strategy orients the strategic foundation and enabling environment to focus on use cases and priorities that add value for stakeholders.

MODERN DATA INTEGRATION ARCHITECTURES

A broad spectrum of data integration approaches exist in the current data landscape.

- **A Federated Architecture**
To achieve a holistic and integrated approach to the use of the facilities of the coordinated sharing organizations and governments today, modern data architectures must be explored.

▪ **Provides visibility** to all participating national, regional, and international organizations (where appropriate).
Given the interconnected nature of all our institutions and organizations today, a federated architecture can be used to enhance and support the mission and utility of data.
▪ **Allows different entities to link their data hubs** activities and lead to more successful adoption of modern data technologies.



CRITICAL SUCCESS FACTORS

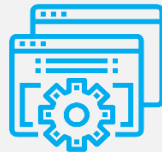
Governments and organizations are challenged to leverage and process large amounts of data to provide better services, improve efficiencies and effectiveness of their operations and provide interactive user experiences targeted at different audiences. To do this, any data and AI strategy framework must:

**RESPOND TO
CLIENTS/CITIZENS
NEEDS**



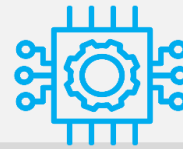
Serve as the main entry point to authoritative business/government data/services, and the corresponding metadata, from the organization itself and other relevant organizations/suppliers/partners required for the effective delivery of services

**PROVIDE A MODERN
ARCHITECTURE AND
TOOLS**



For the integration, presentation, analysis, visualization, communication, use and re-use of data.

**BE BASED ON A
TECHNOLOGY THAT
SERVES USERS' AND
PRODUCERS' NEEDS**



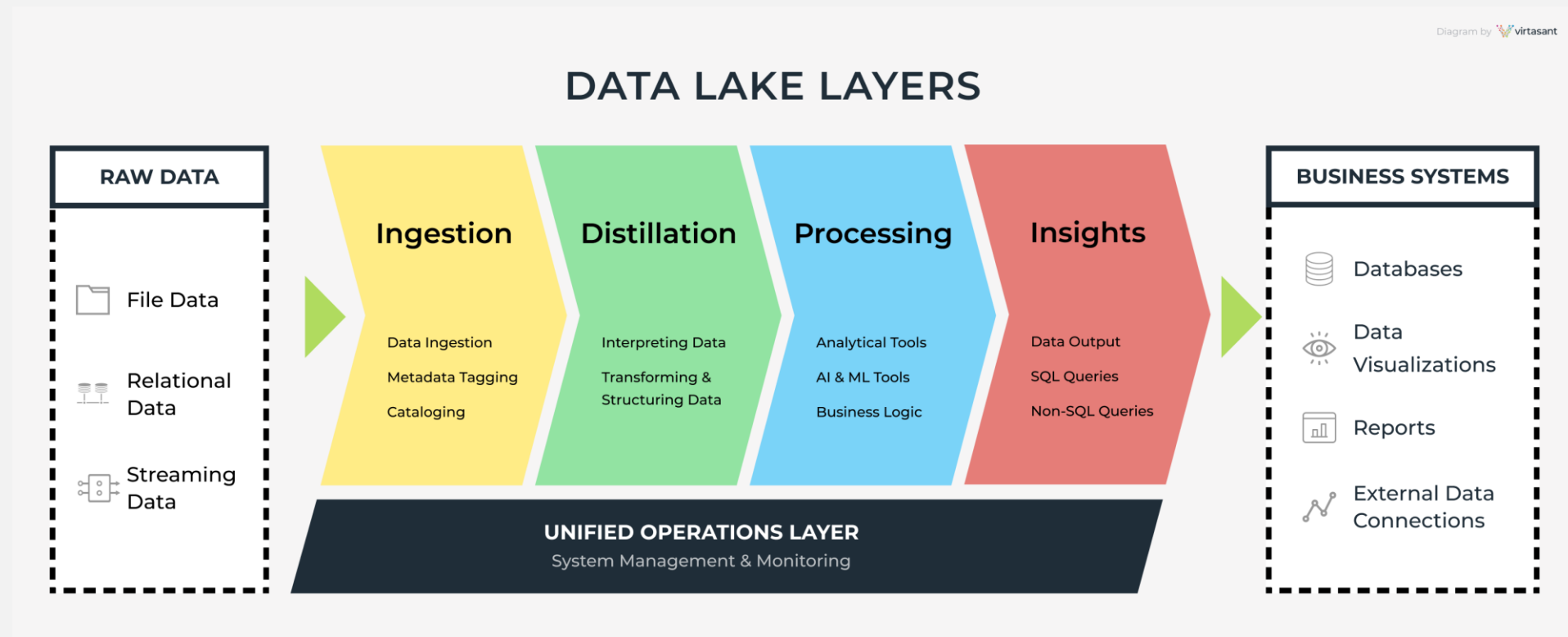
Through modern architecture and tools for the integration, presentation, analysis, visualization communication, use and reuse of data, and metadata.

**ENSURE THE
COMMON USE OF
STANDARDS**

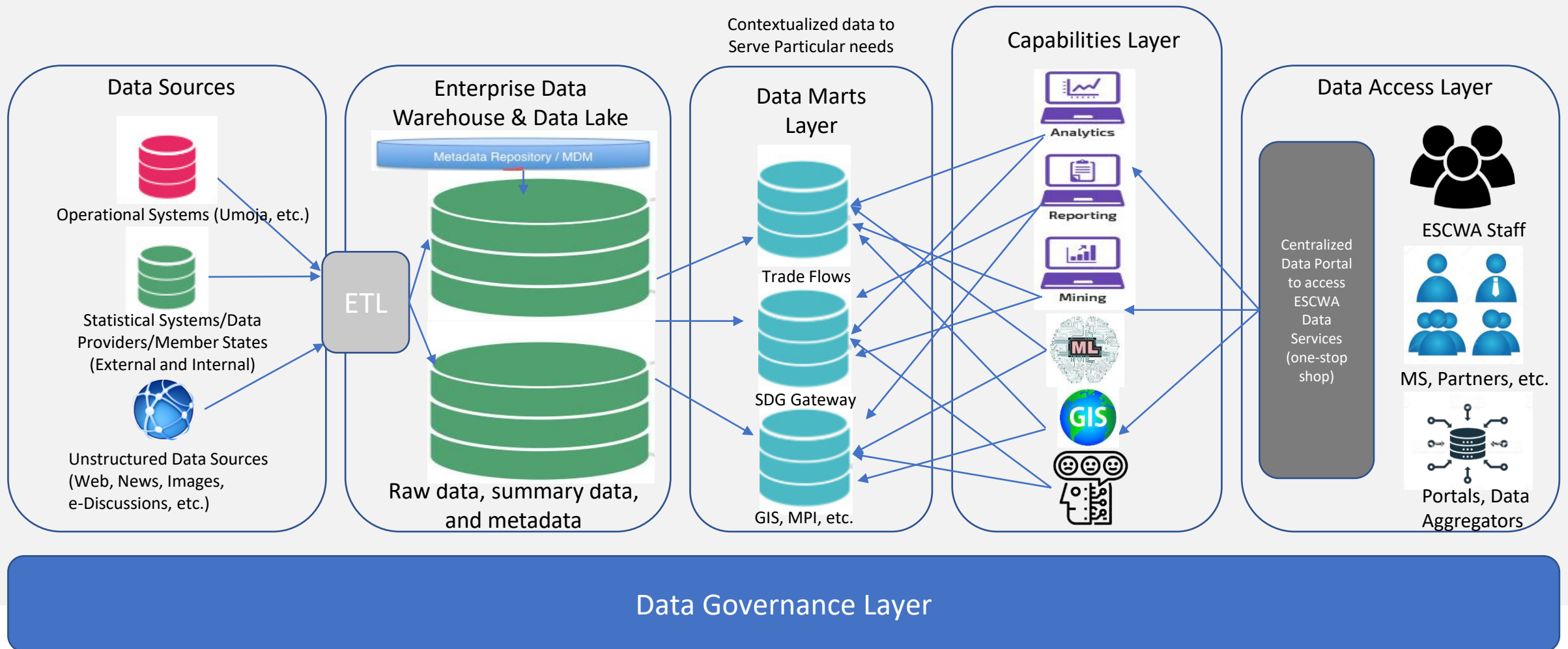


For data exchange including metadata standards, concepts, and classifications to facilitate interoperability, traceability, integration, and usability.

High-level data operations



High-level reference architecture for the ESCWA Data Ecosystem





ازدهار البلدان كرامة الإنسان



Demo of the ESCWA Data Ecosystem

Use Cases

Global & ESCWA

Using Twitter data to analyze public perception on the nature of the

At Regional Big Data & Knowledge Hub - ESCWA

and Reports Using Regional Labour Market

Use Twitter data to provide an

Collect a variety of data from the public reaction to

that the data localising to provide a

platform papers, datasets, the regional data news, super

through a single interface. This can include

show for policy makers and employers

different data variables

<https://skillsmonitor.unescwa.org>

<https://openknowledge.worldbank.org/handle/46288/22656>

MANARA

Hub -

MANARA



Parsing text to create new digital data

```
>>> import layoutparser as lp
>>> model =
lp.Detectron2LayoutModel('lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config')
>>> layout = model.detect(image) # You need to load the
image somewhere else, e.g., image = cv2.imread(...)
>>> lp.draw_box(image, layout, ...) # With extra
configurations
```



Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

over union (IOU) level [0.50:0.95], on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

Table 4 is a core metric designed for the COCO competition. It is

- Region Types
- Text
 - Title
 - Figure
 - Table

Paper with Complex Layouts

```
>>> import layoutparser as lp
>>> model =
lp.Detectron2LayoutModel('lp://PrimaLayout/mask_rcnn_R_50_FPN_3x/config')
>>> layout = model.detect(image) # You need to load the
image somewhere else, e.g., image = cv2.imread(...)
>>> lp.draw_box(image, layout, ...) # With extra
configurations
```



- Region Types
- TextRegion
 - ImageRegion
 - TableRegion

Magazine Scans & Websites

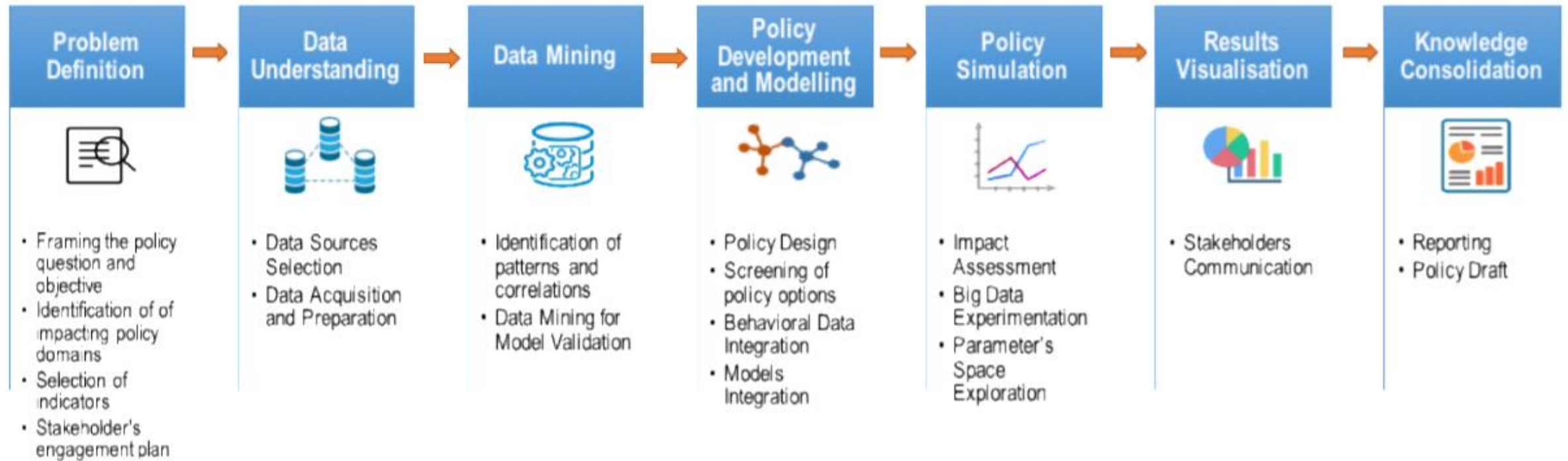
```
>>> import layoutparser as lp
>>> model =
lp.Detectron2LayoutModel('lp://HJDataset/faster_rcnn_R_50_FPN_3x/config')
>>> layout = model.detect(image) # You need to load the
image somewhere else, e.g., image = cv2.imread(...)
>>> lp.draw_box(image, layout, ...) # With extra
configurations
```



- Region Types
- Text Region
 - Title Region
 - Title
 - Subtitle

Historical Documents

Sample of Practical Application of AI & Data Driven Policy Making Framework



The ESCWA Regional Data Ecosystem: Open Data and Capabilities for the Region

Aggregates various
data sources

Provides powerful
data analysis
capabilities using AI
and statistical means

The ML lab: offering
costly tools for free to
governments,
students, researchers,
policymakers, etc..

Technical and Policy Advisory
Services, and more...
All for FREE...



ازدهار البلدان كرامة الإنسان



الأمم المتحدة

الإسكوا
ESCWA

Thank you!