Modern Data & AI Infrastructures to support evidence-based policymaking:

The Case of the UN Arab Region Data & Policy Support Hub

Ahmed Al-Awah, Chief Technology Officer
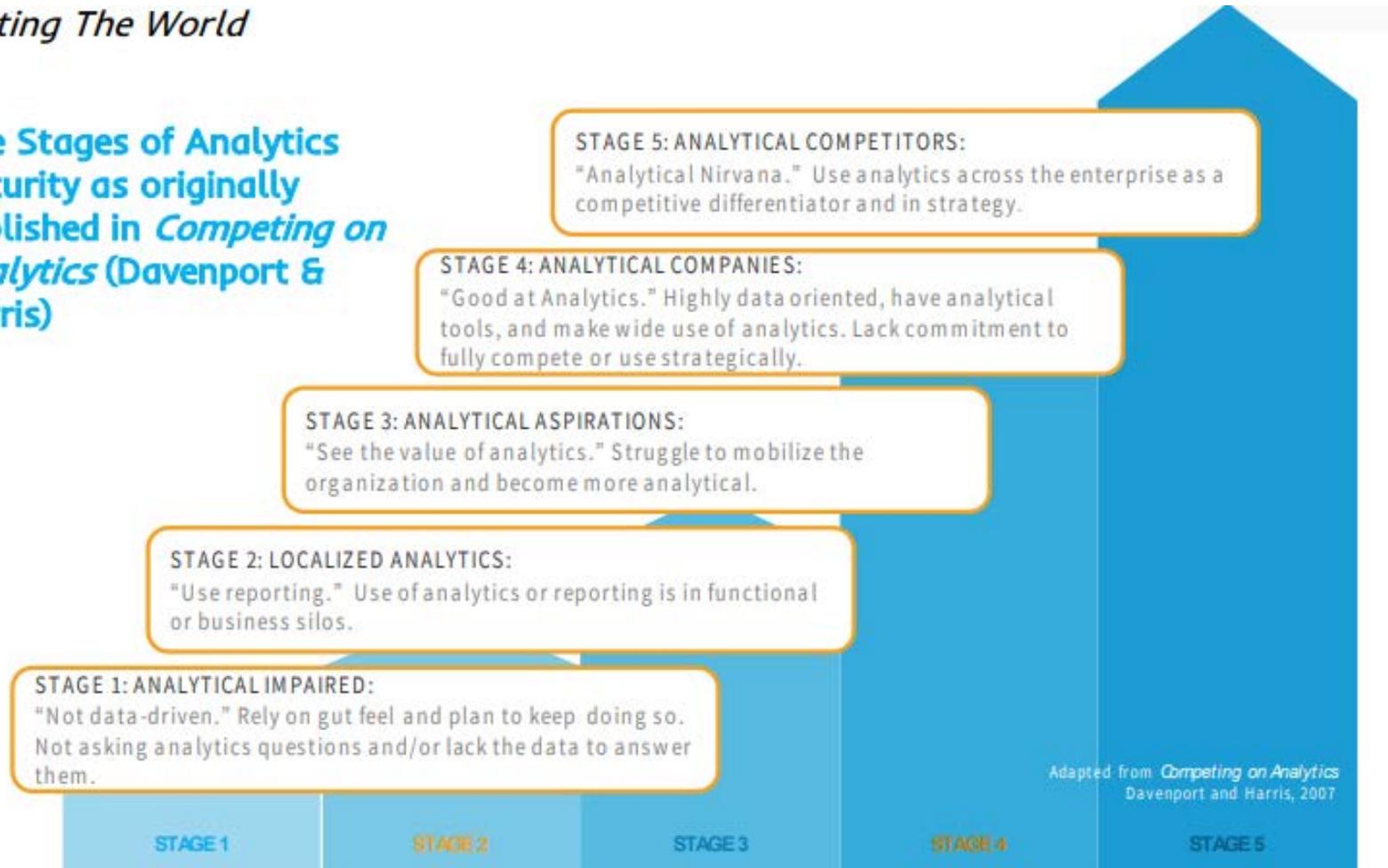*United Nations Economic and Social Commission for Western Asia*
al-awah@un.org

Maturity Assessment

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?
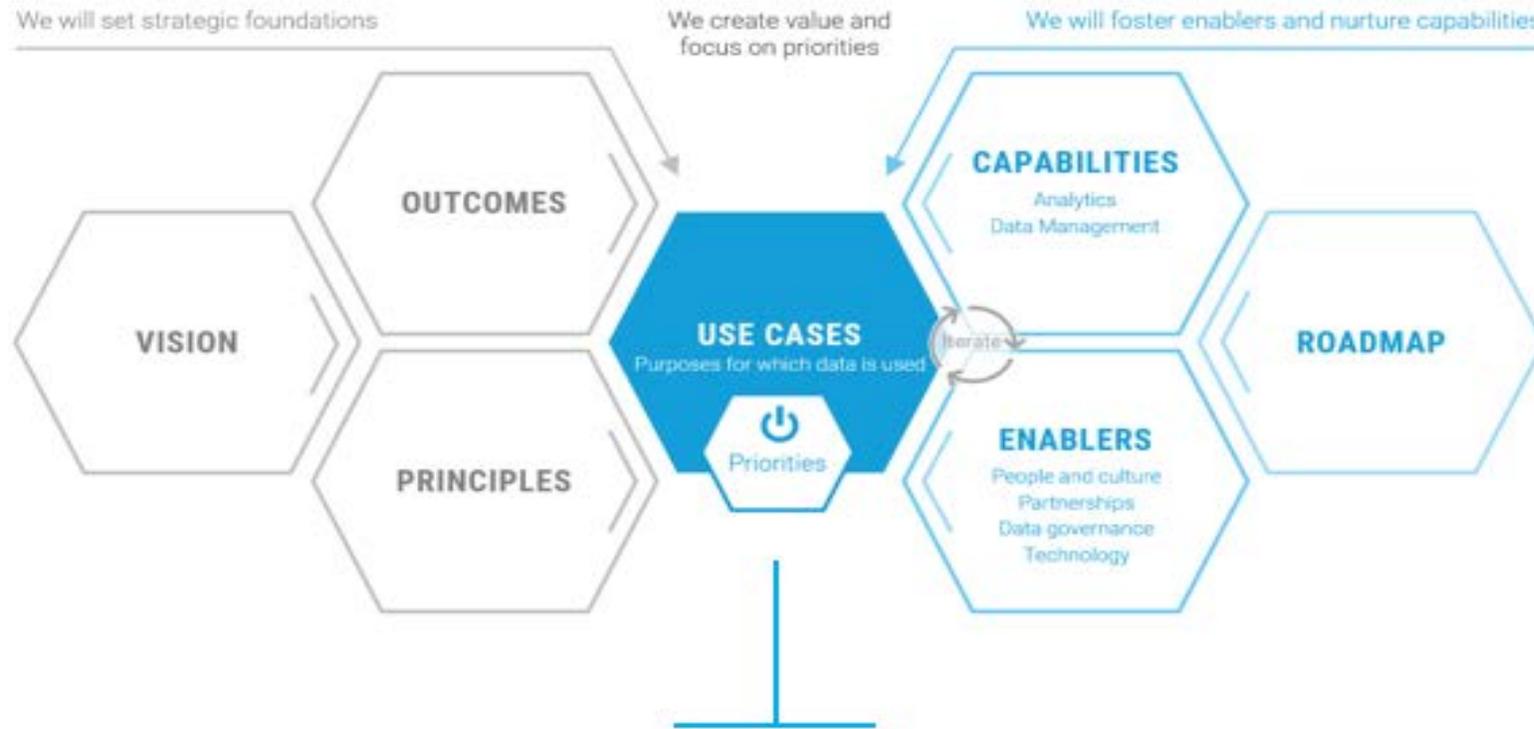- Are we using the proper technologies?
- Do we have the appropriate operating model?

Strategy for Development of a data-driven DSS

A problem-driven approach to building enablers and capabilities

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

Study of Relevant Use Cases

Using Twitter data to analyze public perception of reform in ElSalvador –

This study uses Twitter data to provide a more nuanced understanding of the public reaction to the 2011 reform to the propane gas subsidy in El Salvador. By soliciting a small sample of manually tagged tweets, the study identifies the subject matter and sentiment of all tweets during six one-month periods over three years that concern the subsidy reform. **The paper shows that such an analysis using Twitter data can provide a useful complement to existing household survey data and even potentially replace survey data if none were available.** [...] This study is used as an opportunity to provide methodological guidance for researchers who wish to undertake similar studies, documenting the steps in the analysis pipeline with detail and noting the challenges inherent in obtaining data, classification, and inference.

https://openknowledge.worldbank.org/handle/10986/22656

Forecasting macroeconomic variables:

"**Forecasting macroeconomic variables** is key to developing a view on a country's economic outlook. Most traditional forecasting models rely on fitting data [...] We pursue a new approach to forecasting by employing a number of machine learning algorithms, a method that is data driven, and imposing limited restrictions on the nature of the true relationship between input and output variables. **We apply the Elastic Net, SuperLearner, and Recurring Neural Network algorithms on macro data** of seven, broadly representative, advanced and emerging economies and find that **these algorithms can outperform traditional statistical models, thereby offering a relevant addition to the field of economic forecasting**."

https://www.imf.org/en/Publications/WP/Issues/2018/11/01/An-Algorithmic-Crystal-Ball-Forecasts-based-on-Machine-Learning-46288

**ESCWA
Sample Use Cases:**

**Social Expenditure Efficiency,
Financing for Development,
Using Indices for policymaking, Labor
Market Analysis, and others**

All from a central location ->

Regional Data & Policy Support Hub –

The UN RCP for Arab States Data & Policy Support Hub is a network of interconnected components that work together to capture, process, produce and use data to support decision-making and drive sustainable development in the Arab Region.

The tool brings together data from key data producers in the region (UN, member States, and key international partners), on the basis of sound data governance mechanisms, and contextualizes this data and information in a way that facilitates the creation and sharing of high-quality digital policy support products by leveraging the latest developments in artificial intelligence, geospatial, and big data technologies to provide policymakers in the region with innovative decision-support tools and data analytical capabilities for evidence-based policymaking in the region.

https://data.as-rcp.org
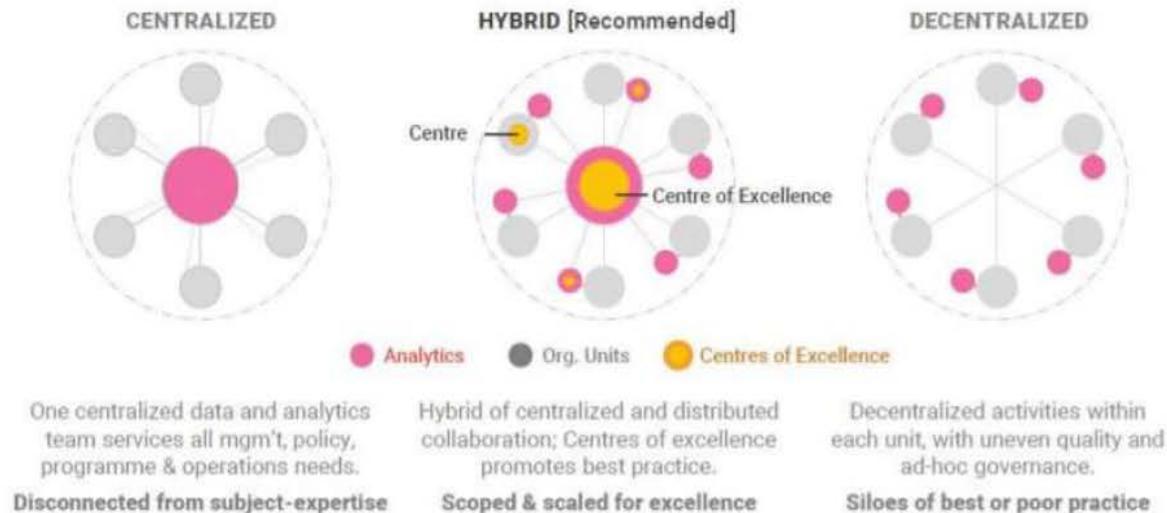
Manara Knowledge Hub -

A Regional Data and Knowledge Hub for contextualizing Publications and Reports for policymaking and the general public using AI

Manara is leveraging different AI technologies to contextualize UN publications, policy papers, datasets, unstructured data sources through a single interface that's serves as a one-stop shop for policymakers and researchers to synthesise different information and data on key development issues in support of evidence-based policymaking.
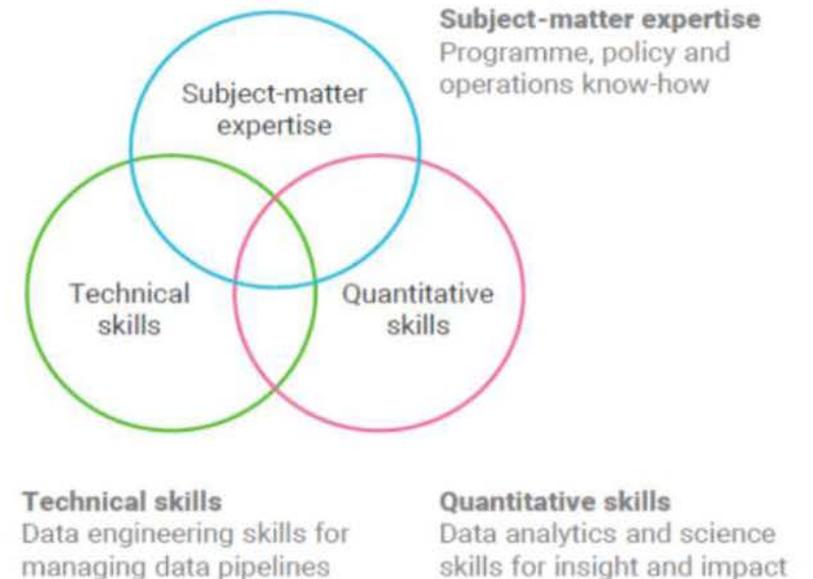
https://manara.as-rcp.org

Developing Use Cases –
Organizational Arrangement Considerations

NURTURE CENTRES OF EXCELLENCE WITH HYBRID MODELS

CENTRALIZED

HYBRID [Recommended]

Centre

Centre of Excellence

DECENTRALIZED

● Analytics   ● Org. Units   ● Centres of Excellence

One centralized data and analytics team services all mgm't, policy, programme & operations needs.

Disconnected from subject-expertise

Hybrid of centralized and distributed collaboration; Centres of excellence promotes best practice.

Scoped & scaled for excellence

Decentralized activities within each unit, with uneven quality and ad-hoc governance.

Siloes of best or poor practice

FORM CROSS-FUNCTIONAL TEAMS

Subject-matter expertise

Technical skills

Quantitative skills

Subject-matter expertise
Programme, policy and operations know-how

Technical skills
Data engineering skills for managing data pipelines

Quantitative skills
Data analytics and science skills for insight and impact

MODERN DATA INTEGRATION ARCHITECTURES

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

To achieve a holistic and integrated approach to the use of the great volume of data required in by Governments today, **modern data architectures must be explored.**

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 in Government operations and activities and lead to more successful adoption of big data technologies.
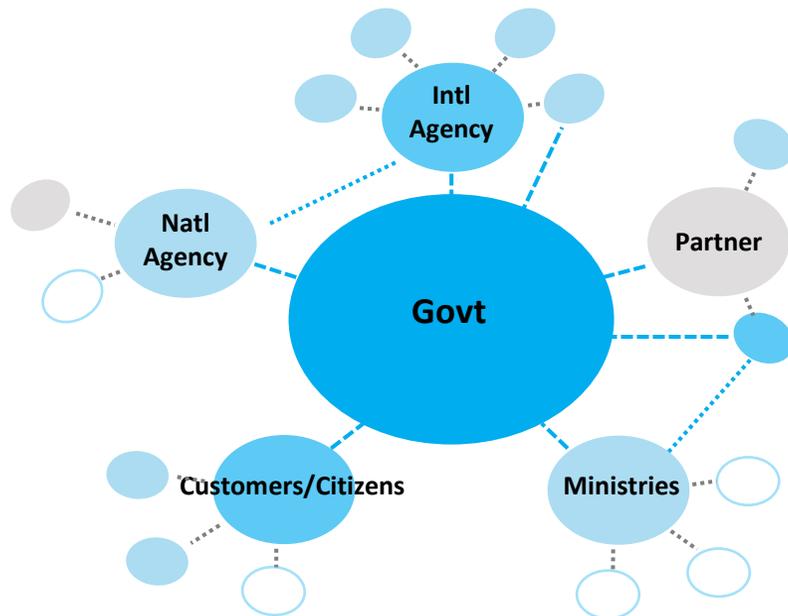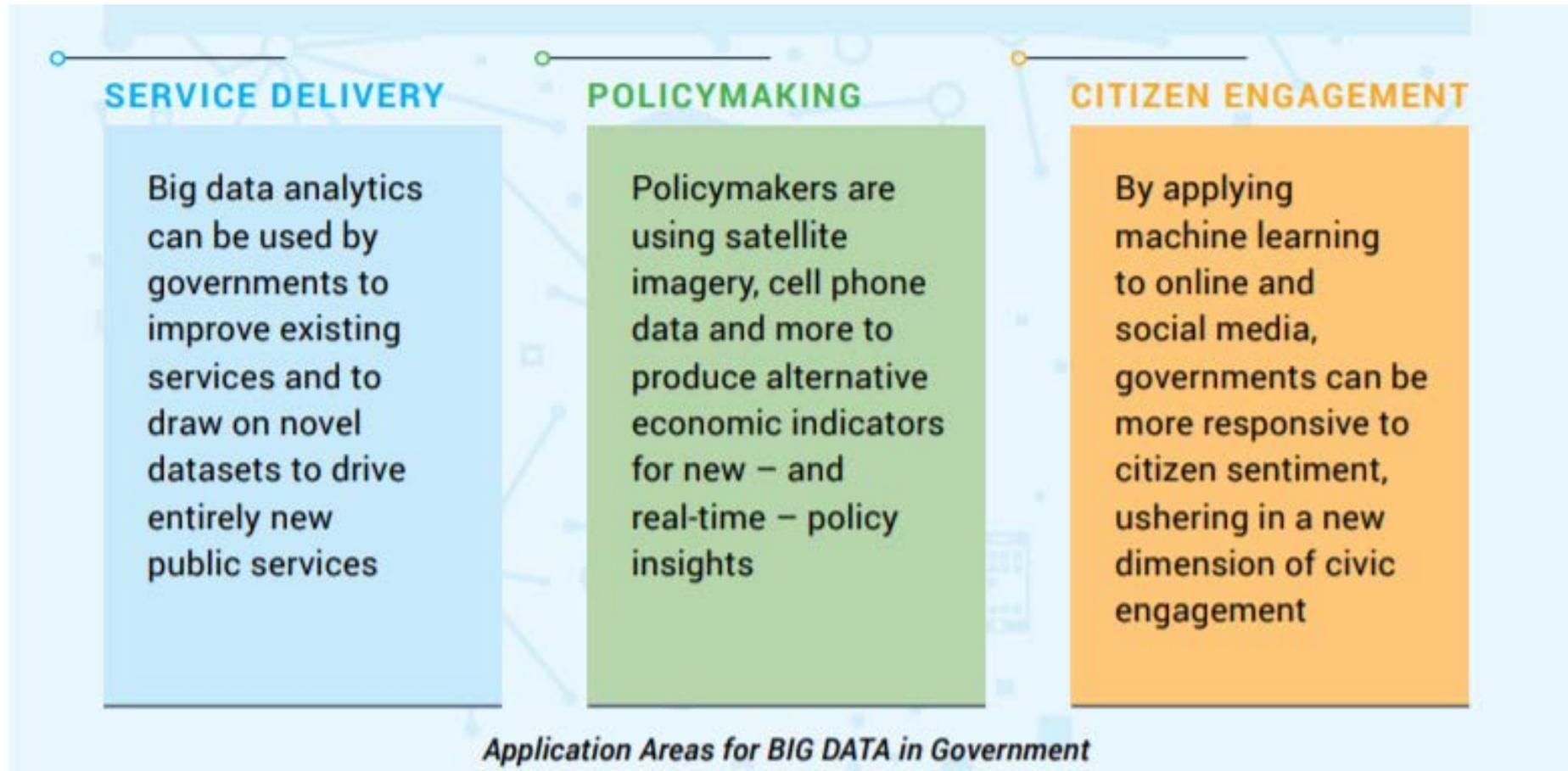

- **A Federated Architecture**

  - **Facilitates the coordinated sharing** and interchange of information among multiple data hubs.

  - **Provides visibility** to all participating national, regional, and international organizations.

  - **Allows different entities to link their data hubs** and share their data more easily with relevant parties.

Leveraging Data Ecosystems for DSS

SERVICE DELIVERY

Big data analytics can be used by governments to improve existing services and to draw on novel datasets to drive entirely new public services

POLICYMAKING

Policymakers are using satellite imagery, cell phone data and more to produce alternative economic indicators for new – and real-time – policy insights

CITIZEN ENGAGEMENT

By applying machine learning to online and social media, governments can be more responsive to citizen sentiment, ushering in a new dimension of civic engagement

Application Areas for BIG DATA in Government

THE CHANGING DATA LANDSCAPE: DATA PRODUCTION AND DATA USE

STRUCTURED DATA　　**UNSTRUCTURED DATA**

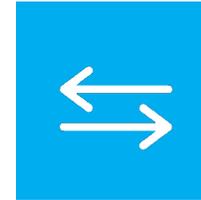**TYPES OF DATA BEING PRODUCED**

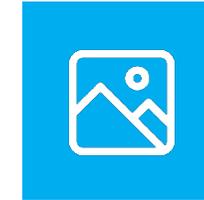| STATISTICAL DATA | ADMIN DATA | TRANSACTIONAL DATA | TEXT & DOCUMENTS | MOBILITY DATA | IMAGES & MEDIA |

CHANGING LANDSCAPE OF DATA USE

IDENTIFY OPPORTUNITIES
to provide services and programs with real-time data

ADDRESS NEEDS
for increased transparency and accountability

INFORM POLICIES
to ensure efficiency and efficacy of development efforts
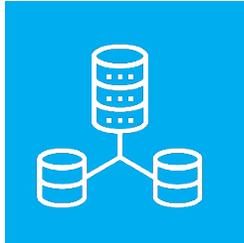
MONITOR AND EVALUATE
to identify and understand impacts

FACILITATE REPORTING
to reduce the burden of sharing critical data

THE CHANGING DATA LANDSCAPE: TECHNICAL FUNCTIONALITIES

INTEGRATE DATABASES
to allow the integration of data across different sources.

VISUALIZE DATA
to facilitate understanding and analysis through improved visualization

DESIGN DASHBOARDS
to enable analysis of inter-related indicators with clear visualization

DISSEMINATE DATA
to increase independent access and use of data

These new architectures are powered by technologies that create more interactive user experiences, enable real-time decision making, and assist with identifying emerging opportunities, requiring an emphasis on more modular, decentralized, service-oriented data sharing and dissemination workflows.

High-Level Overview of Target Policymaking Decision Architecture

Problem Definition	Data Understanding	Data Mining	Policy Development and Modelling	Policy Simulation	Results Visualisation	Knowledge Consolidation
• Framing the policy question and objective • Identification of of impacting policy domains • Selection of indicators • Stakeholder's engagement plan	• Data Sources Selection • Data Acquisition and Preparation	• Identification of patterns and correlations • Data Mining for Model Validation	• Policy Design • Screening of policy options • Behavioral Data Integration • Models Integration	• Impact Assessment • Big Data Experimentation • Parameter's Space Exploration	• Stakeholders Communication	• Reporting • Policy Draft

High-Level Technical Architecture

Data Sources

Operational Systems (Umoja, etc.)

Statistical Systems/Data Providers/Member States (External and Internal)

Unstructured Data Sources (Web, News, Images, e-Discussions, etc.)

ETL

Enterprise Data Warehouse & Data Lake

Metadata Repository / MDM

Raw data, summary data, and metadata

Data Marts Layer

Contextualized data to Serve Particular needs

Trade Flows

SDG Gateway

GIS, MPI, etc.

Capabilities Layer

Analytics

Reporting

Mining

ML

GIS

Data Access Layer

Centralized Data Portal to access ESCWA Data Services (one-stop shop)

ESCWA Staff

MS, Partners, etc.

Portals, Data Aggregators

Data Governance Layer

High-Level Technology Stack
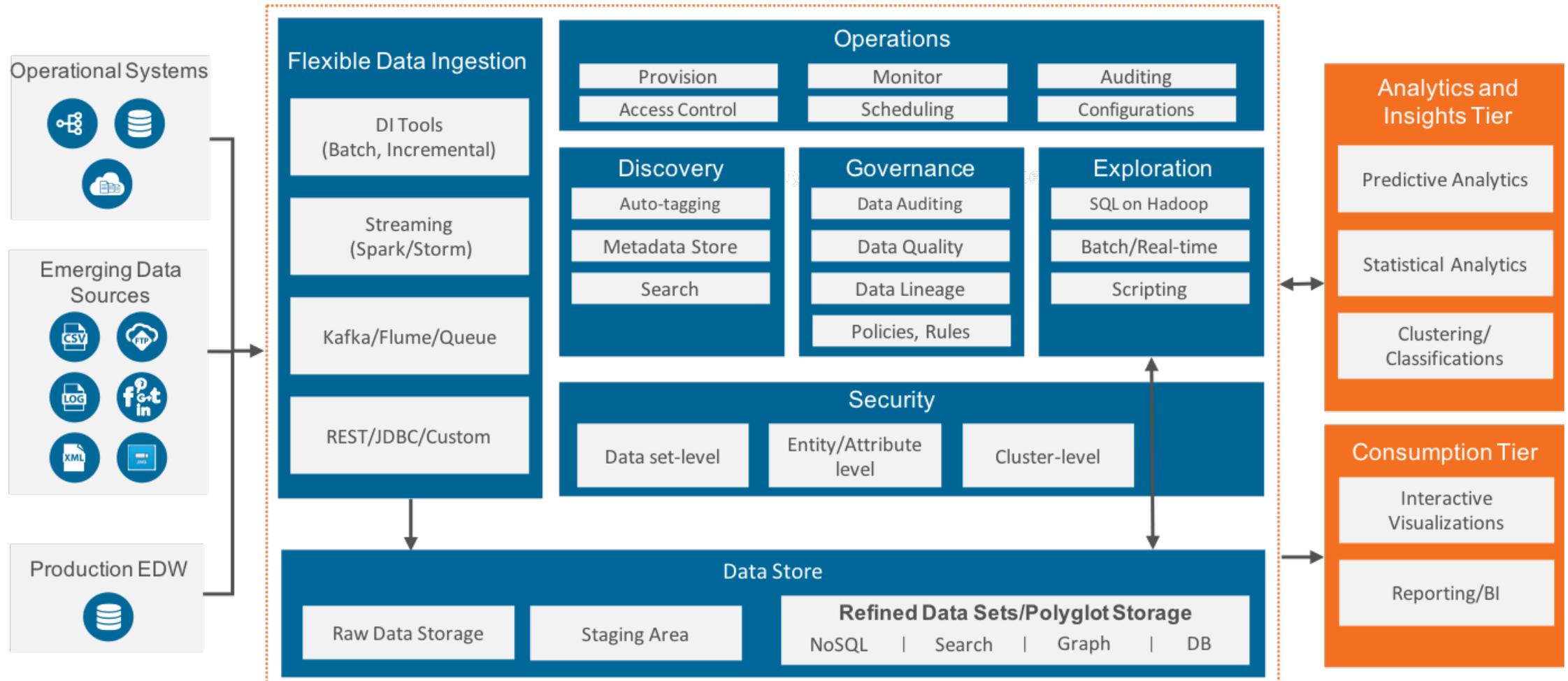
Data Governance Issues

Data Governance Layer

Master Data Management (MDM)
- Data immediacy
- Data completeness
- Data accuracy
- Data availability

Data Life-Cycle Management
- Archive data
- Maintaining data warehouse
- Testing and delivering applications for performance
- Data deletion and disposal

Data Security and Privacy Management
- Sensitive data discovery
- Vulnerability and configuration assessment
- Security policies
- Change auditing
- Activity monitoring
- Auditing and compliance reporting
- Identity and access management
- Protecting data in transit

Demo of the Arab Regional Data & Policy Support Hub

https://data.as-rcp.org

Thanks

Questions?
Ahmed Al-Awah – al-awah@un.org