Regional Course on Statistical Business Registers

Session 8: SBR Maturity Model (continued)

- Dimension 6: IT environment
- Dimension 7: Interoperability

UNITED NATIONS SIAP Statistical Institute for Asia and the Pacific

United Nations Statistics Division

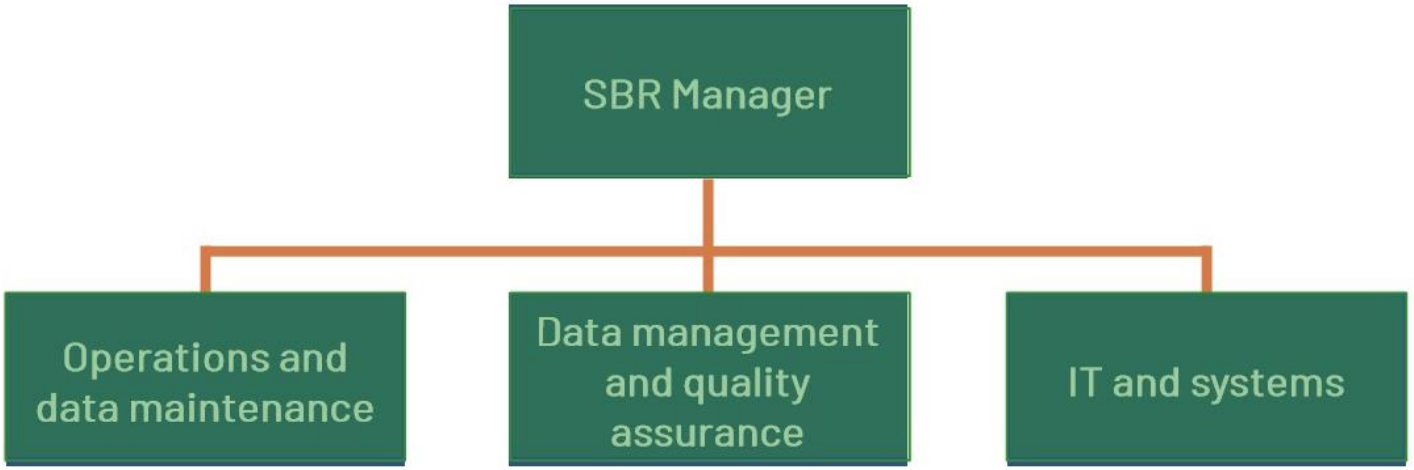
Overview

Sessions 6 & 7 covered the first 5 dimensions of the SBR Maturity Model

↓

This session will cover the final two dimensions, on the **IT environment** and **interoperability**

Dimension 6: IT Environment



IT and systems

- Ensuring that the systems and software needed for extracting data are properly programmed and optimized.
- Maintaining, and possibly advances or further developing, the software that enables SBR staff to easily access and update the SBR content.
- Administering the database processes and tables that make up the SBR, and ensuring that the tables are accessible and available for production of the outputs.

Dimension 6: IT Environment

It's important to keep in mind several aspects in the development of an IT infrastructure for SBRs so that it is in line with the resources and priorities of the NSO.

Effective partnerships with internal and external stakeholders are crucial.

Implementation of the SBR must focus on achieving its primary role.

Seek to minimize unnecessary conceptual and technical complexities.

Dimension 6: IT Environment

Goal

Develop a system that fits within the NSO's infrastructure and is compatible with other systems, i.e. administrative data acquisition systems and business survey collection systems

Project management

- Significant undertaking: should be managed as a project, using project management methodology, if available
- Effort should be appropriate to size, complexity and risk
- Other common PM methodologies like PRINCE2 (PRojects IN Controlled Environments) and PMI/PMBOK (Project Management Body of Knowledge)

Dimension 6: IT Environment

Software development methodology

Agile implementation:

- Demonstrating value to shareholders quickly
- Managing change and risk continuously
- Frequent evaluation and realignment of scope, cost, schedule, quality

2 phases of IT development:

- Acquiring initial db infrastructure
- Then programming and process development

Follow organizational IT standards

If one is not already implemented, should adopt an appropriate one

Dimension 6: IT Environment

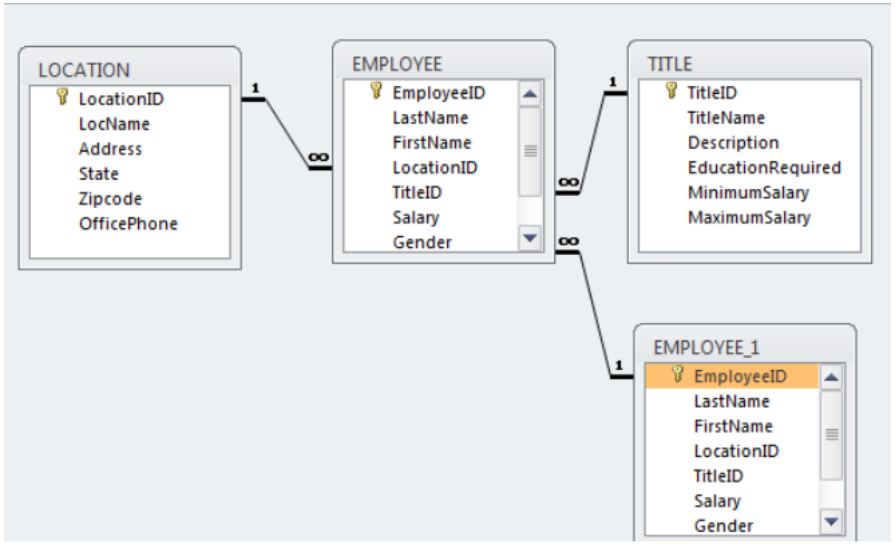
Database management systems

- Security
- Scalability
- Concurrent users
- Flexible and standardized query language

Relational database management systems (RDBMS)

Additionally:

- Referential and data integrity
- Efficient storage and performance
- Modularity



Dimension 6: IT Environment

Database options

- Low cost: Microsoft Access (no concurrent users, limited capacity)
- Others: Oracle, Microsoft SQL Server, MySQL, MariaDB and PostgreSQL

Technologies should be:

- Accessible, staff able to be trained
- Long shelf life
- Able to integrate external systems, administrative data especially, through APIs
- Use clear definitions and concepts

Dimension 6: IT Environment

Other elements

- Should be a place (within or external to db) for storing and accessing data on: frames, snapshots, reporting burden, etc.

- In primary role of providing survey frames, db should include data on collection process: units sampled, how and where data about units will be collected and results of collection efforts -> inputs into “respondent burden module” for use across agency/statistical office, including efforts to mitigate that burden

Dimension 6: IT Environment

Other elements

Unique Identifiers

Essential, will discuss more in the session on the Global Initiative on Unique IDs

Data retention policy

- Don't just change the information in the system
- Instead add information about the change, including the time when change occurred (or was implemented), and maintain the old value
- Can help in reconstructing the register and maintaining the historical register

Documentation

User manual/wiki/technical documentation

Dimension 6: IT Environment

ADB and Statistics Norway-developed systems

Adaptable, NSOs maintain ownership of the system and code

Statistics Norway: STATBUS Statbus.org + github.com/statisticsnorway/SBR

- Runs on Microsoft's IIS (Internet Information Server) and is written in ASP.NET
- Choice of underlying database – Microsoft SQL Server, PostgreSQL, MySQL
- Data can be uploaded in CSV or XML format

- ✓ Includes data quality checks, system user management
- ✓ Different languages, different classification systems
- ✓ Define, construct, maintain statistical units
- ✓ Classify units by sector/activity
- ✓ Produce survey frames and register-based statistics
- ✓ Upload from different sources
- ✓ Multiple ID fields including computer generated; up to 3 languages

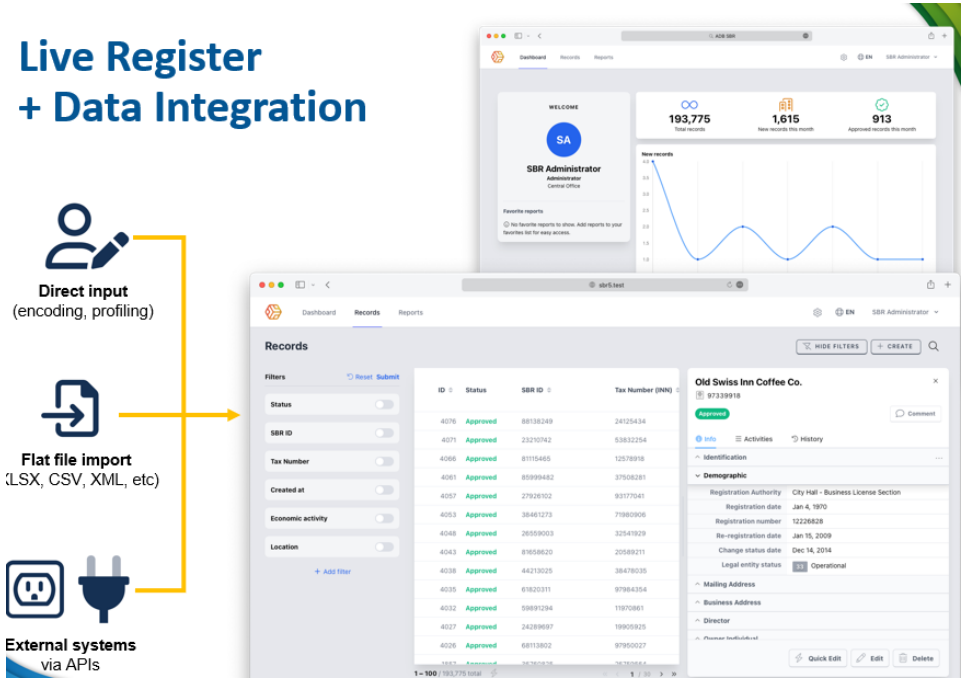
Dimension 6: IT Environment

ADB and Statistics Norway-developed systems

ADB SBR System

Many of the same features as STATBUS:

- Web-based application
- User accounts
- Create frames/snapshots, historical data
- API integration



These systems are not “plug-and-play”!

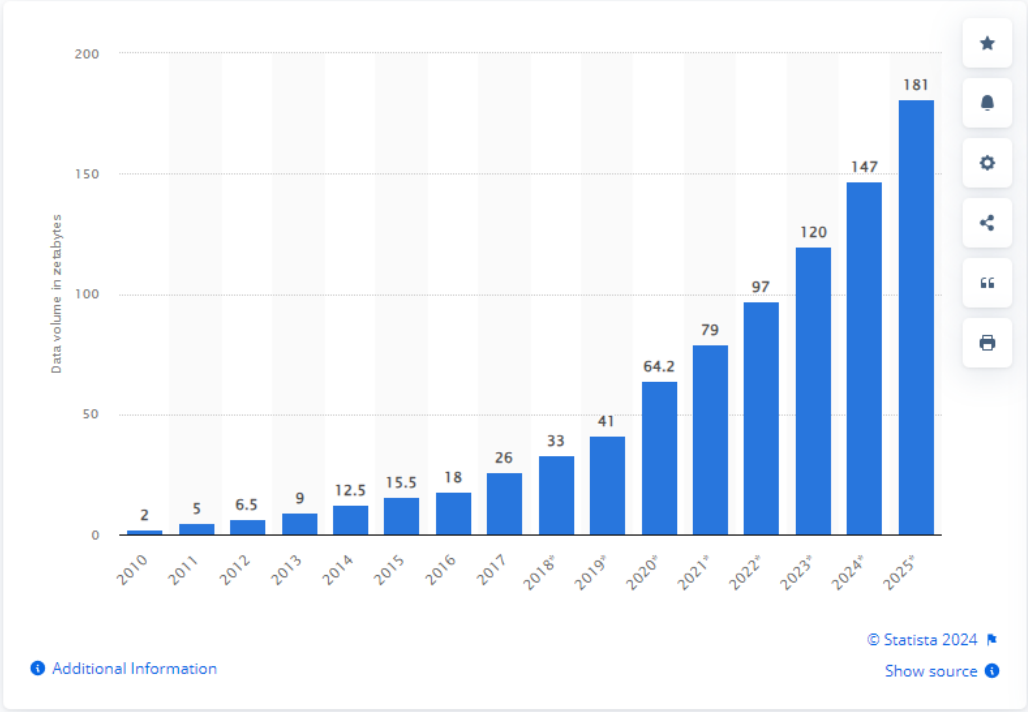
- Require collaboration with these organizations
- Internal planning and management and maintenance are still needed

Dimension 6: IT Environment

Big Data, Artificial intelligence (AI) and Machine Learning (ML)

Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025

(in zettabytes)



DOWNLOAD

PDF XLS PNG PPT

Sources

- Show sources information
- Show publisher information
- Use Ask Statista Research Service

Release date

June 2021

Region

Worldwide

Survey time period

2010 to 2020

Supplementary notes

* The data was taken from various publications released over several years: Forecast for the years 2018 and 2019 as of 2018; Forecast for 2020 as of May 2021; Forecast for 2021 to 2025 as of March

UNITED NATIONS SIAP Statistical Institute for Asia and the Pacific

United Nations Statistics Division

Dimension 6: IT Environment

Big Data, Artificial intelligence (AI) and Machine Learning (ML)

- Many web scraping tools available online
- Companies providing data -> now also providing web scraping services
- Can be used to help identify web stores/online retail
- AI can help write code for scraping, etc.
- Language recognition is a major component

Dimension 6: IT Environment

Big Data, Artificial intelligence (AI) and Machine Learning (ML)

Singapore example

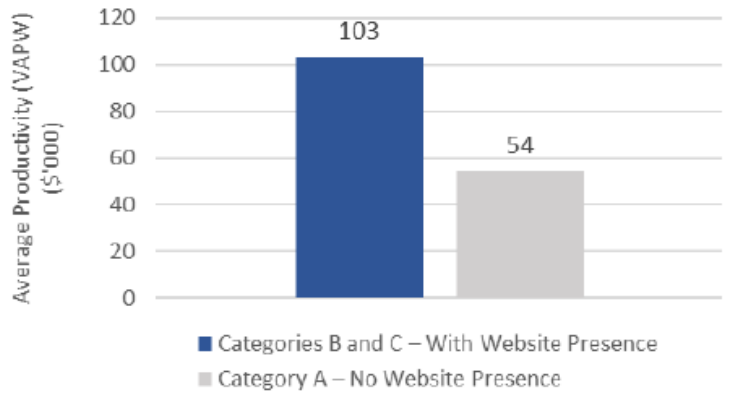
PERFORMANCE OF CLASSIFIERS EXPLORED

Classifier	Test Set Accuracy
Random Forest	79%
Gradient Boosting Machine	77%
Voting Classifier	77%
Logistic Regression	72%
Neural Network	71%
AdaBoost	70%
Support Vector Machine	68%
Naïve Bayes	57%

FEATURE IMPORTANCE OF SELECTED WORDS

Feature Word	Feature Importance
Shop	0.044
Cart	0.041
Price	0.027
Facebook	0.021

AVERAGE PRODUCTIVITY OF ENTERPRISES WITH AND WITHOUT CORPORATE WEBSITES, 2020



https://www.singstat.gov.sg/-/media/files/publications/economy/ssn222-pg1-4.ashx

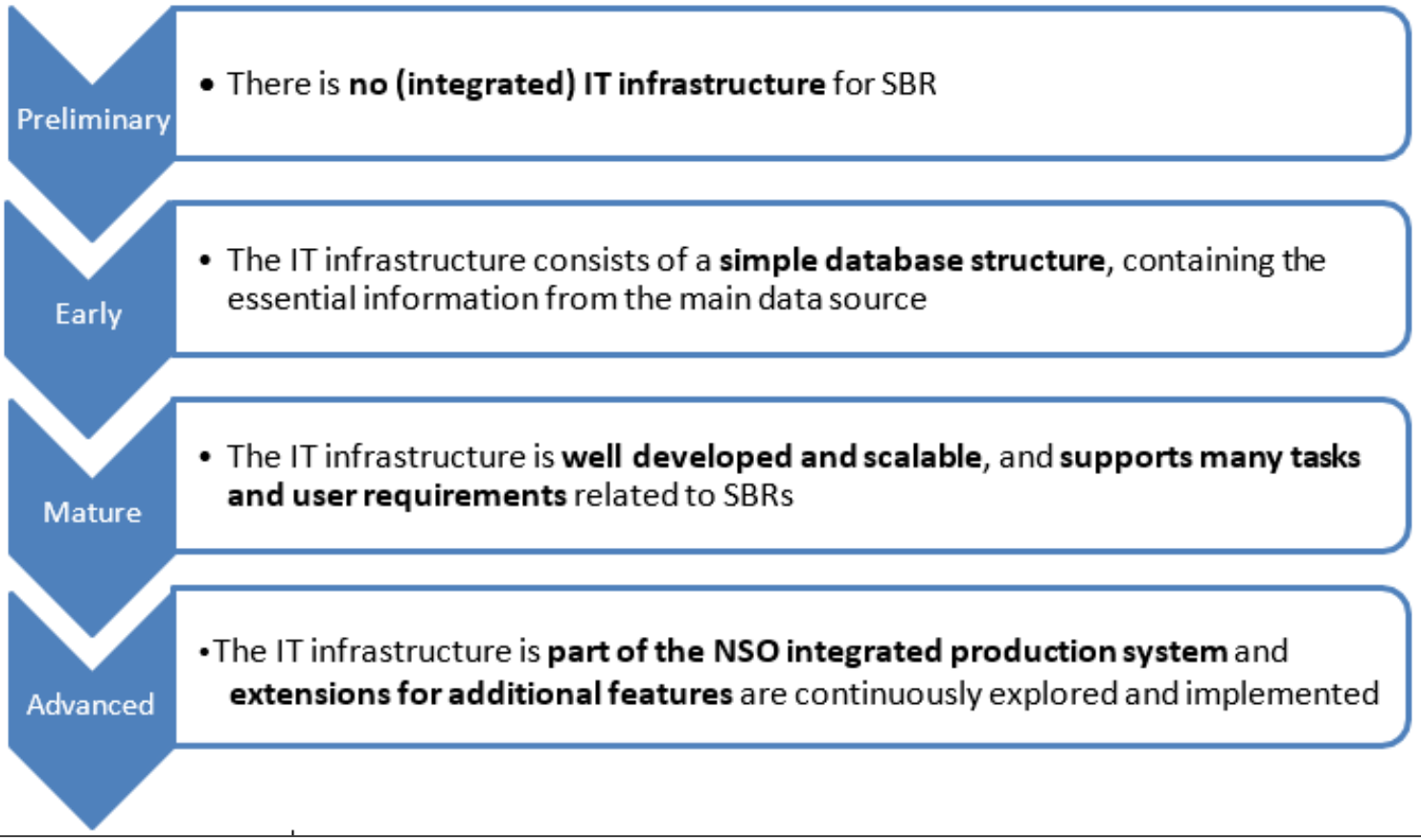
UNITED NATIONS SIAP Statistical Institute for Asia and the Pacific

United Nations Statistics Division

Dimension 6: IT Environment

Figure 8

Summary of stages of maturity for Dimension 6: IT environment



Dimension 6: IT Environment

Preliminary →

- No integrated insfrastructure
- Records kept manually, i.e. Excel
- Basic maintenance strategy

Early →

- Simple db structure with essential data
- May have other applications, but not integrated

Mature →

- Managed like a project, maintained, dedicated staff
- Modular, scalable, supports historical register, unique IDs

Advanced

- Part of integrated system, continuously improved
- New technologies: big data, web scraping, portals for businesses to enter info

Dimension 6: IT Environment

Questions:

1. Does your SBR use a relational database?
2. Have you had positive/negative experience with specific types of software?
3. Have you had any experience with ADB/STATBUS products?

Dimension 7: Interoperability

Interoperability

The ability of computer systems or software to exchange and make use of information in a coordinated manner

In SBRs

The ability of an SBR to communicate and exchange standardized data with other registers, be they domestic, regional, or global

Data can be easily re-used and processed in different applications, allowing different information systems to work together

Dimension 7: Interoperability

Interoperability

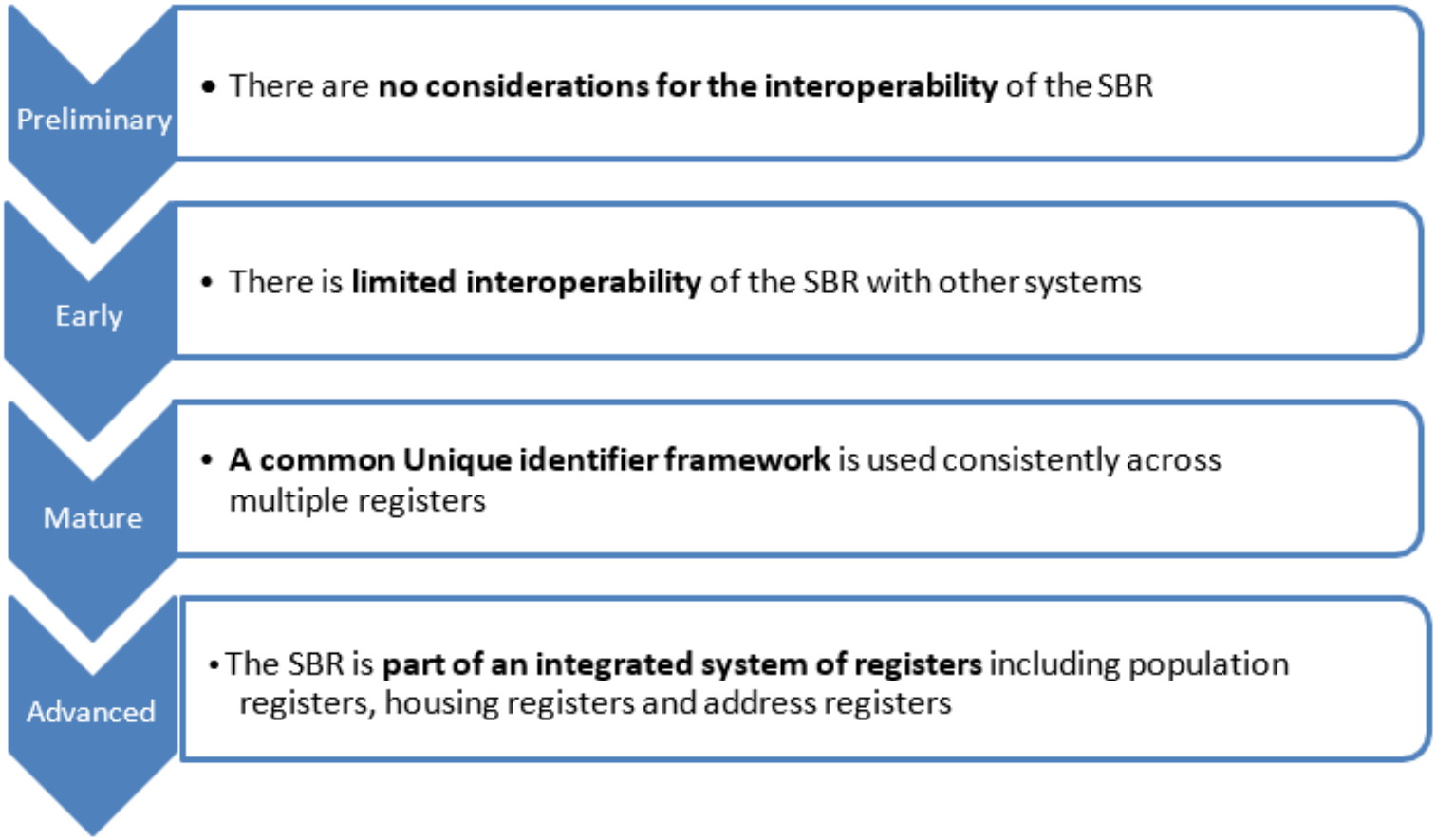
- Related to IT infrastructure but important enough to be separately assessed
- Increases efficiency

Requires

- A unique ID
- Standardized language/classification/characteristics

Dimension 7: Interoperability

Figure 9
Summary of stages of maturity for Dimension 7: Interoperability



Dimension 7: Interoperability

Preliminary →

- Matching done manually
- No considerations for interoperability

Early →

- Some automated routines
- Some interoperability with administrative sources

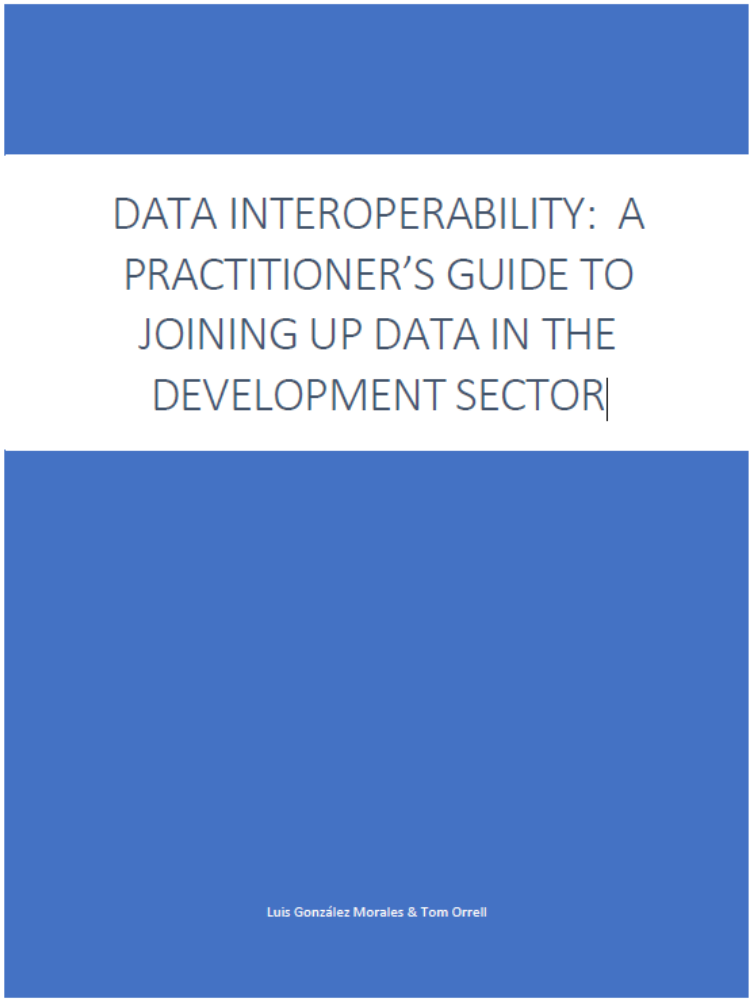
Mature →

- Common unique ID to link administrative & statistical registers
- Compatability with international standards

Advanced

- Microdata linking implemented
- Interoperable with international sources using global identifiers
- Integration with population, housing, other registers

Dimension 7: Interoperability



Available at this link

Dimension 7: Interoperability

Questions:

1. Does your SBR meet the minimum requirements, namely:

   Use of a unique identifier +

   A standardized language/classification/characteristics?

2. Which, if any, other registers can be linked with your SBR? Could this be done in a more efficient way?

3. Has your office implemented micro data linking?

Dimension 7: Interoperability

Thank you!