A live multi-source statistical (Business) Register

How administrative data integration in
Statistics Netherlands
is implemented
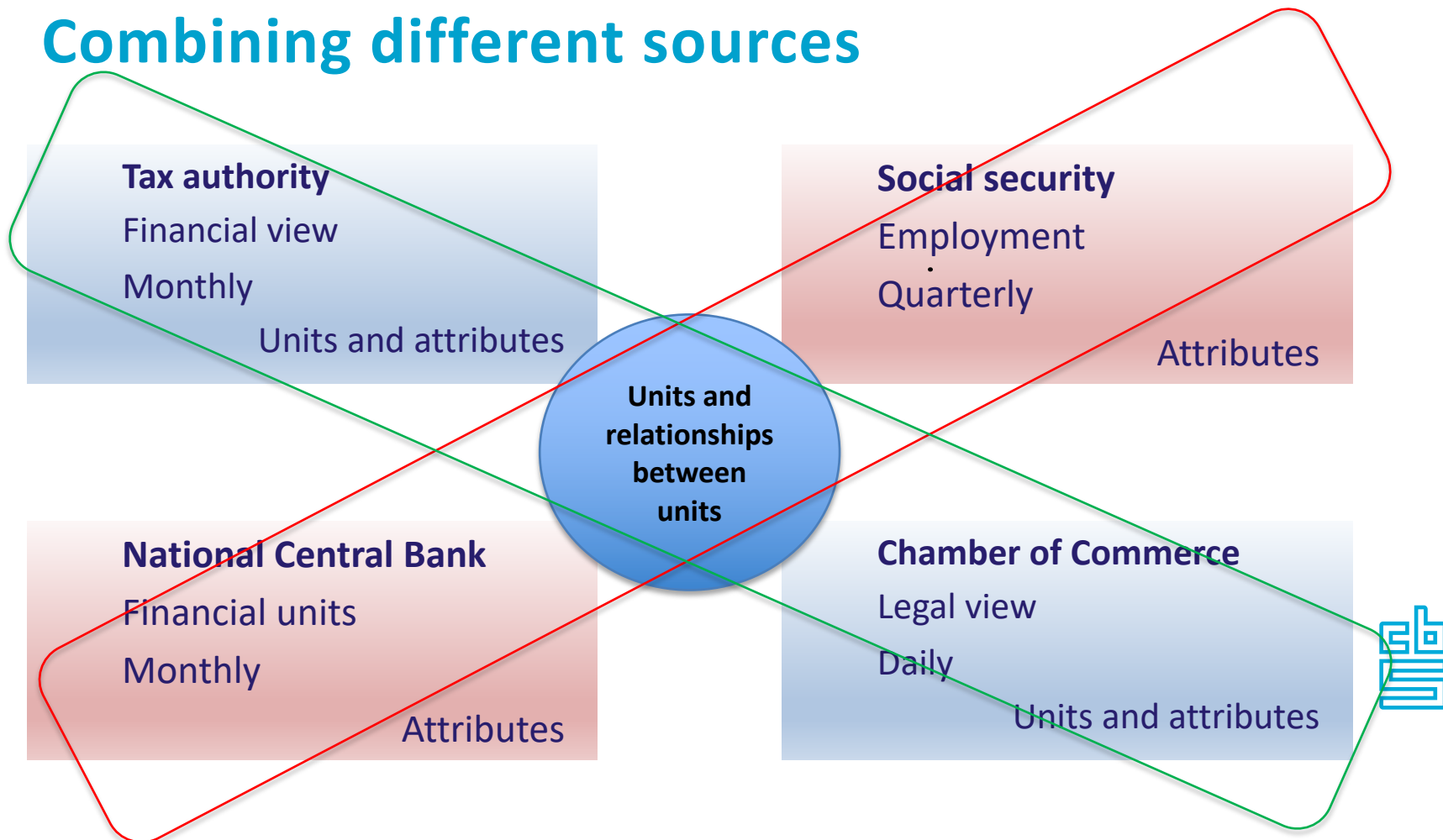
Topics

1. Combining administrative sources
2. Challenges in integration
3. SN input and output area
4. Metadata (management)
5. Future challenges

Combining different sources

Tax authority

Financial view

Monthly

Units and attributes

Social security

Employment

Quarterly

Attributes

Units and relationships between units

National Central Bank

Financial units

Monthly

Attributes

Chamber of Commerce

Legal view

Daily

Units and attributes

Challenges in integrating data from different sources

Different target populations

Collaborate with sources
- Understand the (target) population
- In depth (business) analysis
- Be informed on changes

Quality of data(delivery)

Quality checks
- Which corrections/additional information is needed?
- Monitor data delivery
- Report derogations and request clarification
- Be informed on updates

Different timing aspects

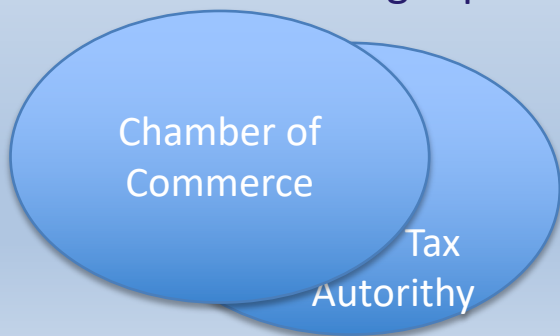**Timelags and priority**

Different technical representation

Transform data to comparable technical format

Integration of units and attributes

Units

- Two sources
- CoC is main source.
- Tax Authority only for CoC undercoverage
- For a small population different timing aspects

Chamber of Commerce

Tax Autorithy

Attributes

Multiple sources

Time lag & Priority

Source	Priority	Timelag
Profiler	1	12
CoC	2	∞
TA	3	-

NACE changed!

CoC	6420	1-2-2018

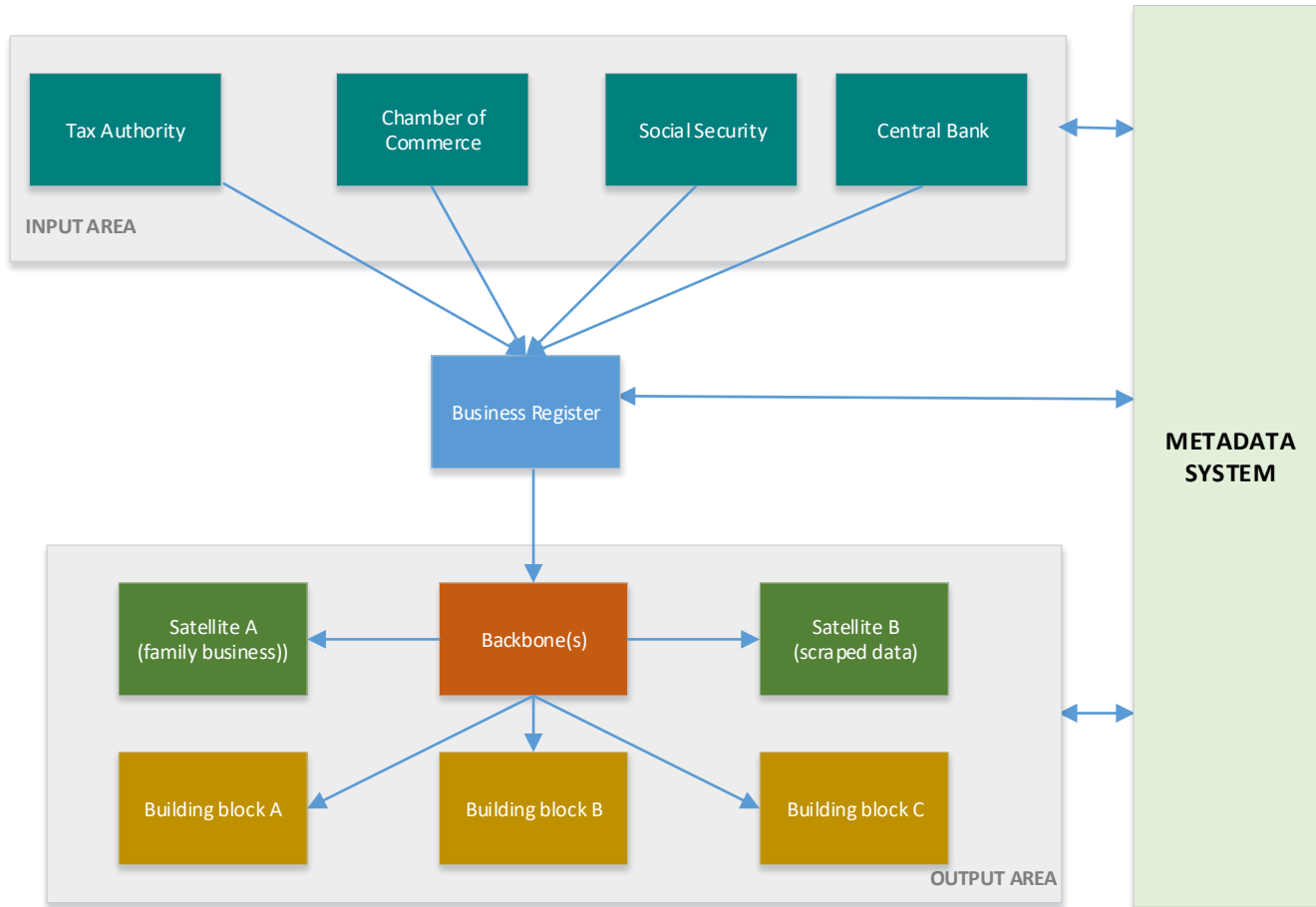| TA	7010	1-2-2019

Pr	7010	1-3-2019

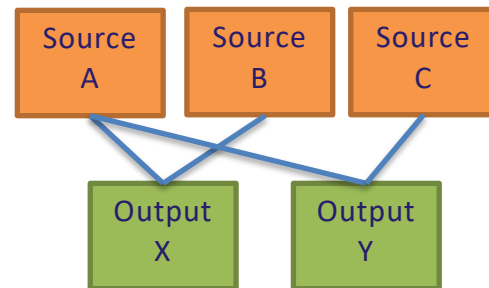| 7010	1-3-2019

Overview combining sources

Meta data system

1. Describe what sources you have
   - Comparable
   - Menu of data for (new) users
2. Link adm. sources to output
3. Sources change
   - Impact of changes. Know where to adapt (in input, throughput and output)

Tool for loading metadata received from adm sources

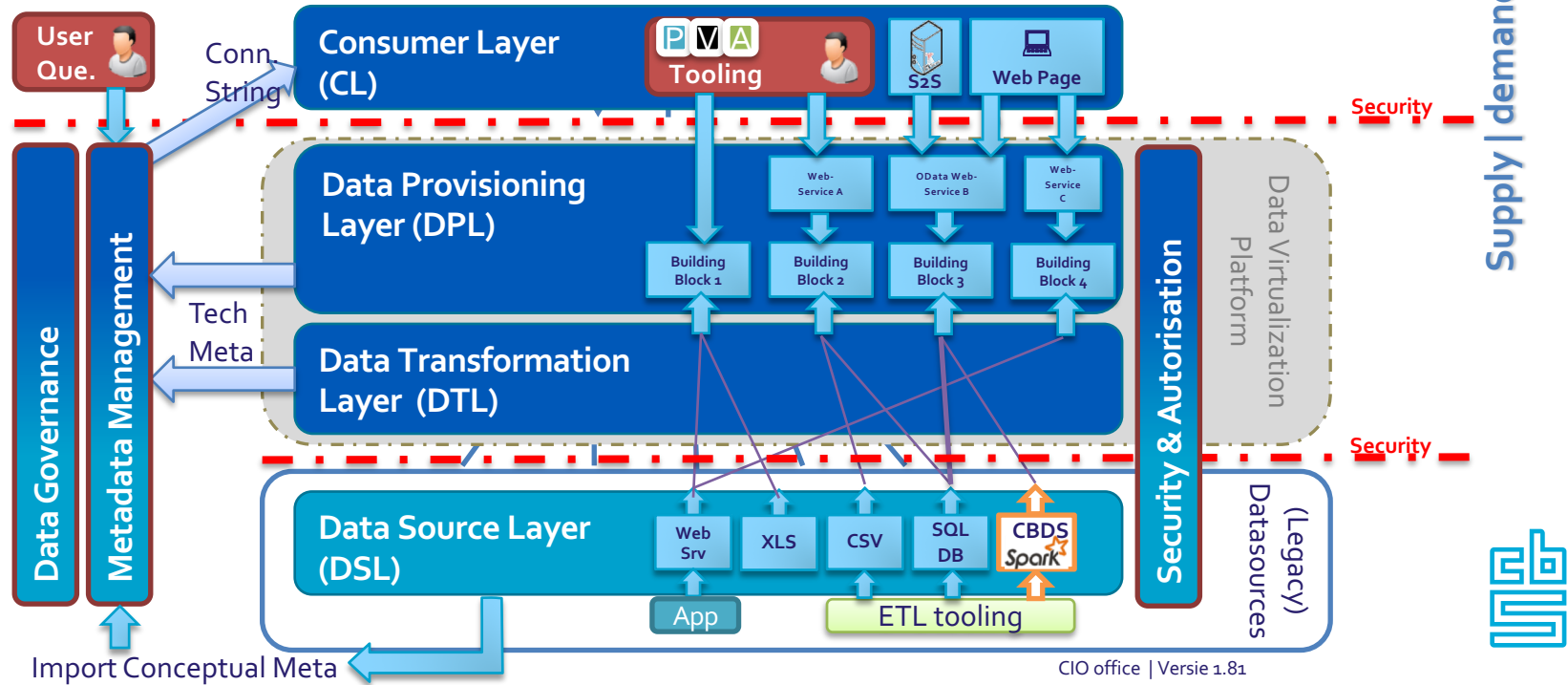| Source A	Source B	Source C

Output X	Output Y

Future challenges

1. Uniforming all adminstrative data processes

2. More flexible in using new administrative sources and integrating the new information

3. Webscraping

4. Using big data in the BR

Data Architecture

Data Architecture diagram — CIO office | Versie 1.81

Key elements:
- User Que. → Conn String
- Consumer Layer (CL): Tooling (P V A), S2S, Web Page
- Data Provisioning Layer (DPL): Web-Service A, OData Web-Service B, Web-Service C; Building Block 1, Building Block 2, Building Block 3, Building Block 4
- Data Transformation Layer (DTL)
- Data Source Layer (DSL): Web Srv, XLS, CSV, SQL DB, CBDS Spark; App; ETL tooling
- Data Governance
- Metadata Management (Tech Meta)
- Security & Autorisation
- Data Virtualization Platform
- (Legacy) Datasources
- Supply | demand
- Security
- Import Conceptual Meta
- Existing / New
- P = Data Prep, V = Data Visualization, A = Data Analytics

Questions?

- b.coenen@cbs.nl