



OTTAWA 2023

64TH WORLD STATISTICS CONGRESS



WELCOME.



**OTTAWA 2023**

64TH WORLD STATISTICS CONGRESS



**IPS0313**  
**Metadata-driven Data Gate**  
**at Statistics Estonia**

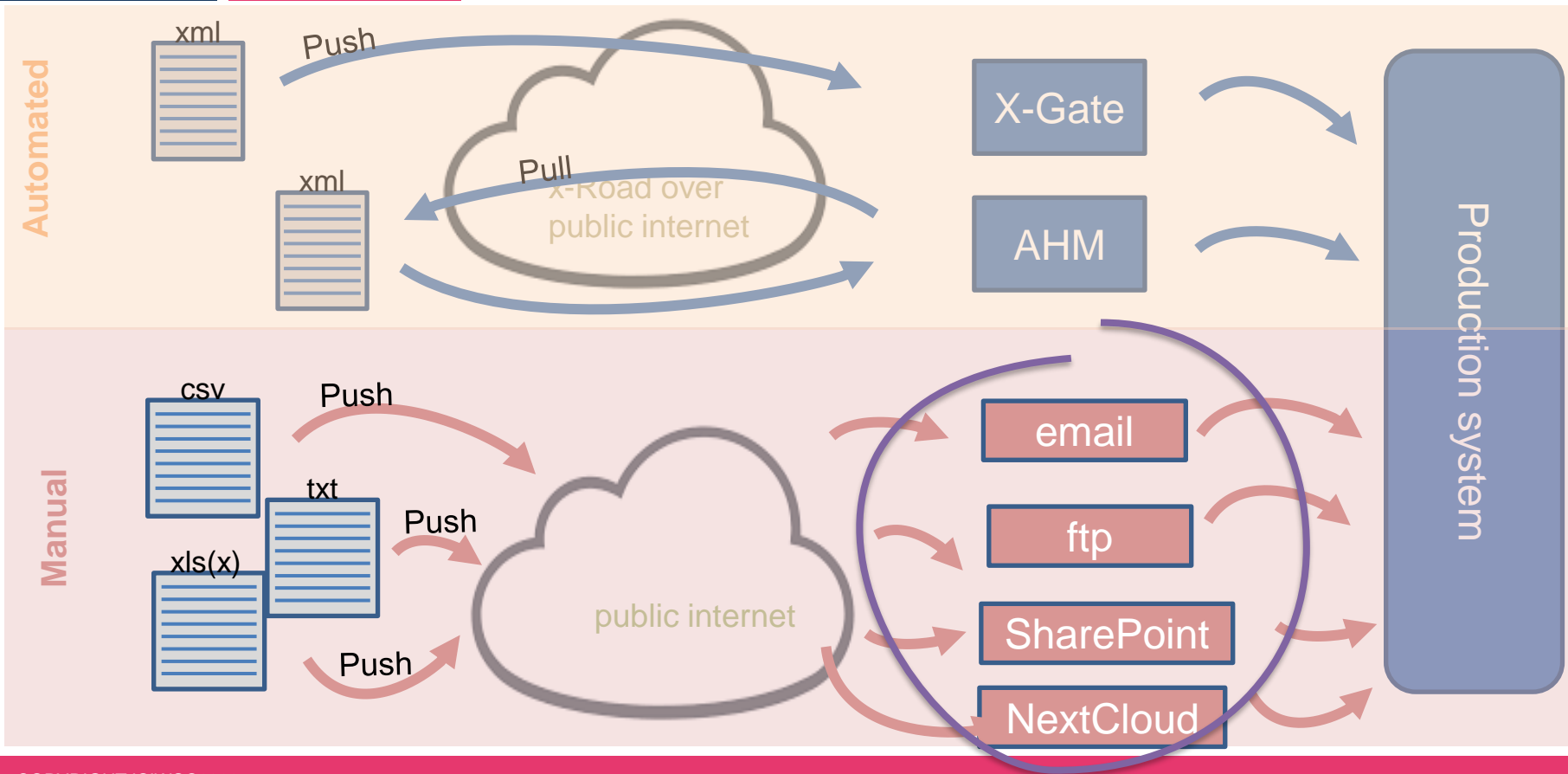
---

Mrs Kaia Kulla  
Statistics Estonia  
Thursday 20 July, 10:00AM – 12:00AM

1. Current data capture methods
2. Goal - minimizing manual effort in receiving data
3. Metadata for Data Gate, but not only
4. DDI comes in
5. Benefits
6. Use case

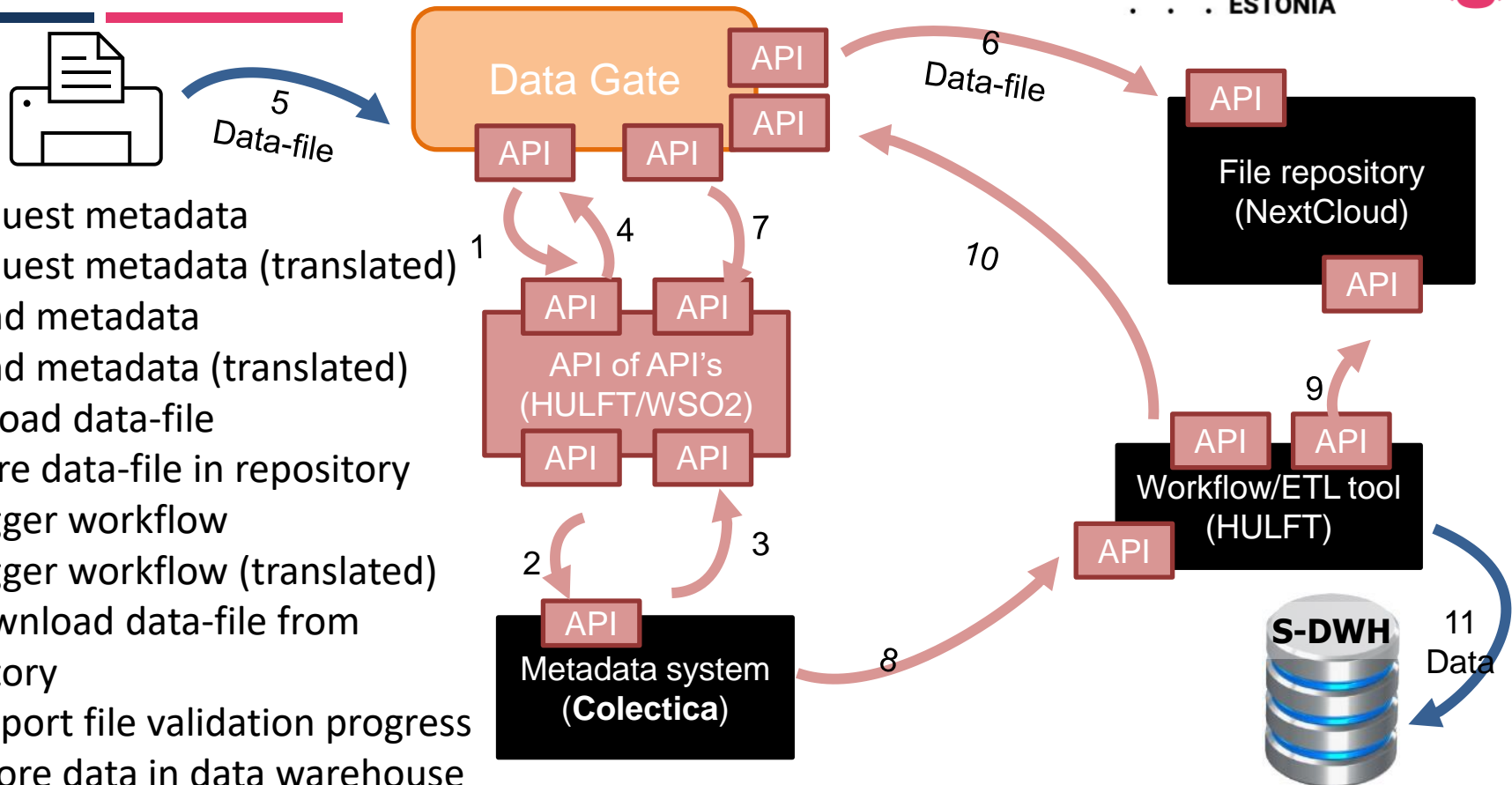
# Current data capture methods

# Current Data Capture Methods



# Goal - minimizing manual effort in receiving data

# New solution – Data Gate

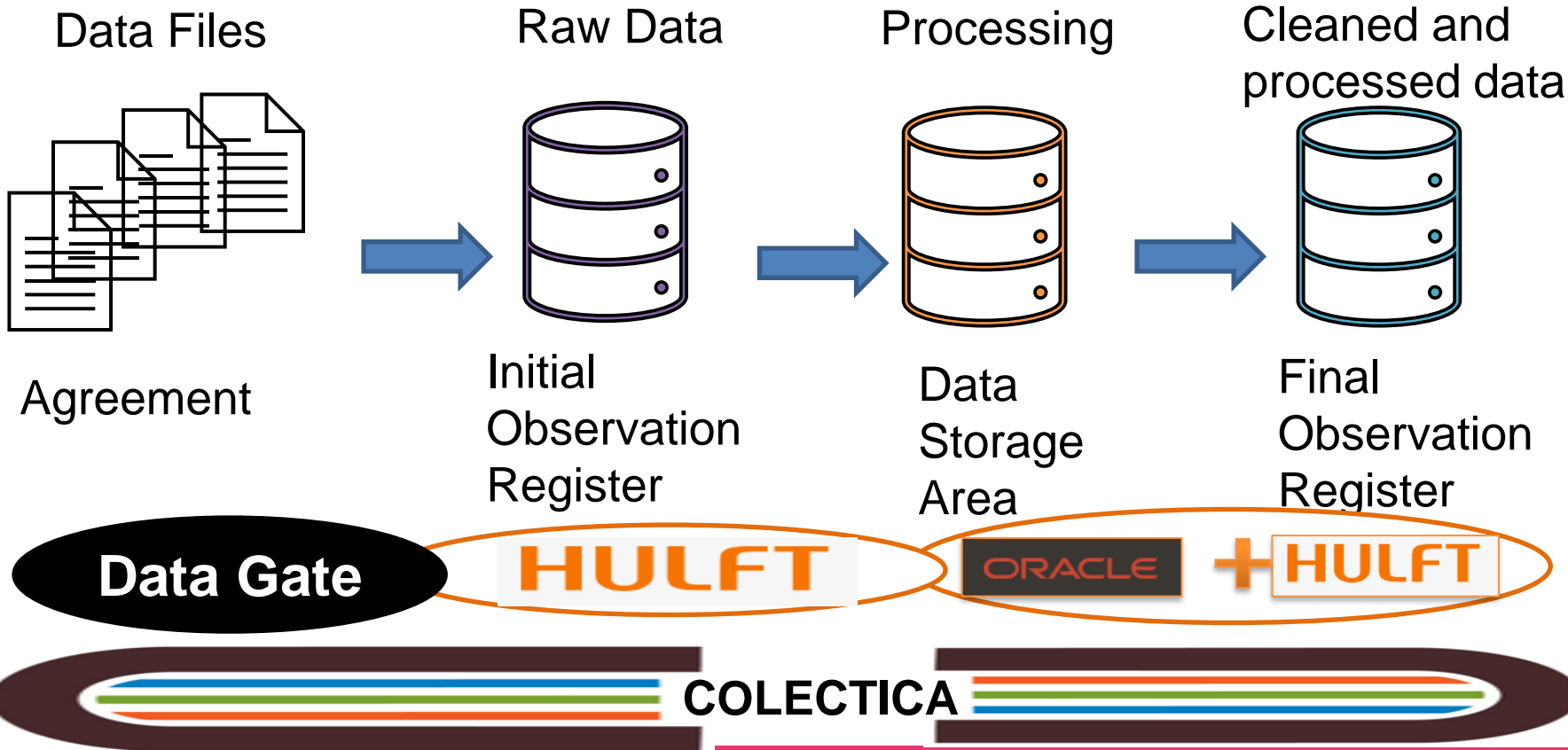


- 1 – request metadata
- 2 – request metadata (translated)
- 3 – send metadata
- 4 – send metadata (translated)
- 5 – upload data-file
- 6 – store data-file in repository
- 7 – trigger workflow
- 8 – trigger workflow (translated)
- 9 – download data-file from repository
- 10 – report file validation progress
- 11 – store data in data warehouse

# Metadata for Data Gate, but not only



# 9 Data capture steps to consider



# 10 Metadata for Data Gate – what we need?






- Obligations -> originate from agreement
  - Information about data provider, incl. information about administrative register
  - Datasets
    - Variables, types, descriptions
    - Code Lists and Classifications
    - Description of dataset (files: type, separator, )
  - Data submitting date(s), frequency
  - Data collection dates
  - Periods covered by the data
- But also, information about
  - Processing
  - Storage
  - Data Lineage
  - ???

# Users to consider











- Several applications
  - Denodo
  - Data editing application
  - Data collection applications
  - Environment for scientists (in development)
  - Etc.
- Programs (R-language, Oracle)
- Physical users
  - Analysts (inside)
  - Scientists (outside)
  - Students (outside)
  - Etc.

# DDI comes in

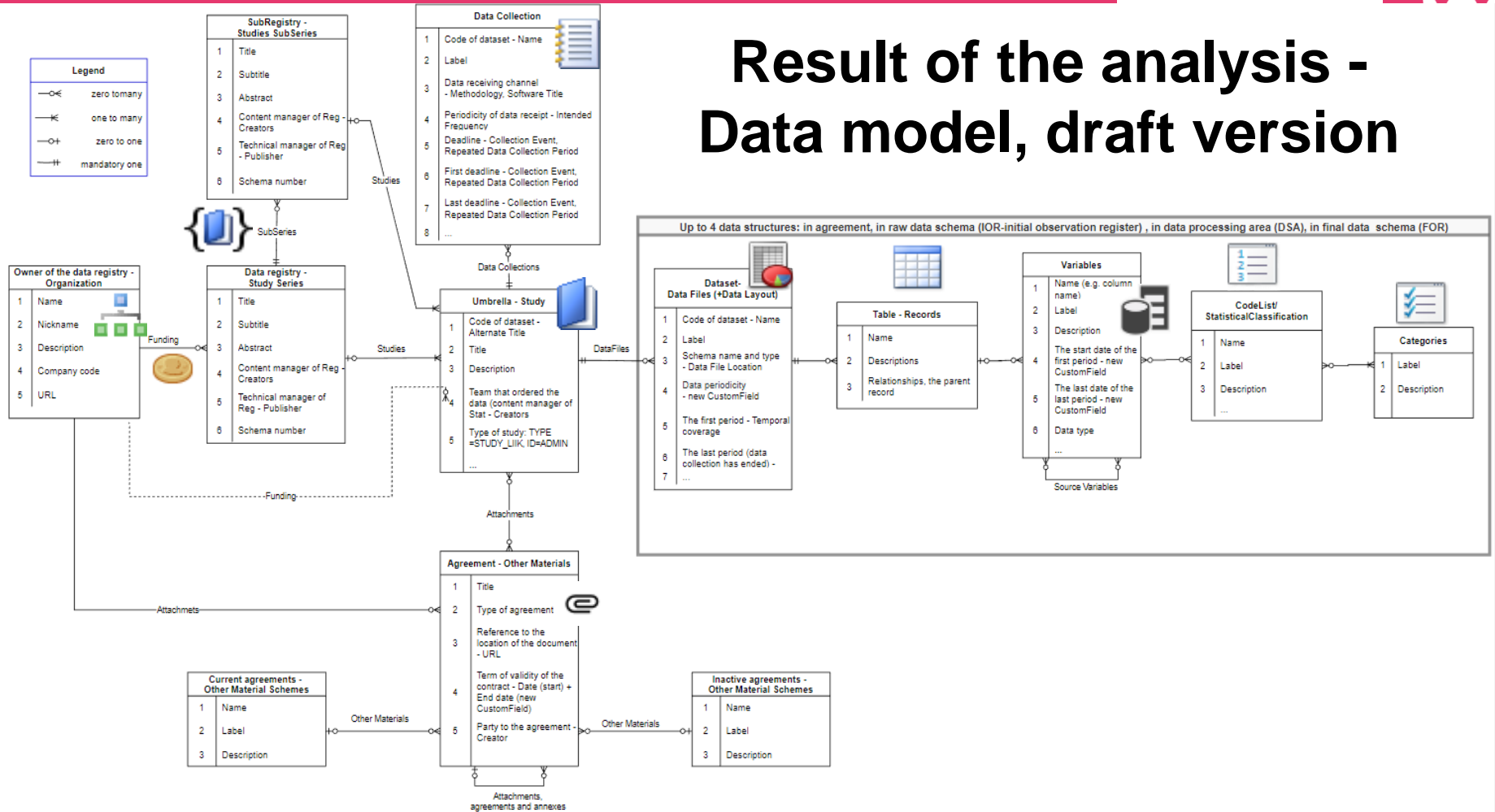
# Where to start?

- Collect existing information from everywhere (existing data, Confluence, Excel files, emails, Word documents) in one place -> **Colectica** 
- What is missing, what would simplify everyday work? 
- Use of metadata as much as possible in the data capture, data storage and data processing processes
- Discover DDI model and features of Colectica 
  - Models behind Colectica and our current metadata system are different
  - DDI is much more detailed compared to our current model (Statistical activities, Variables, Code Lists, for processing: Data Sets, Tables, Variables)
  - Lack of attributes or nests for information that we already have
  - DDI Controlled Vocabularies don't match always with our need

# Result of the analysis - per DDI item

- DDI Items involved:
  - Series -> Data Registry or Information System 
    - Subseries -> Sub-registry, if that exists
  - Study -> kind of umbrella for data capture 
  - Data Files (Data Layout) -> Dataset 
    - Records -> Tables within the Dataset
  - Organization -> Owner of the Data Registry 
  - Other Material -> Agreement 
- Meaning remains the same as a meaning of DDI item
  - Data Collection 
  - Variables 
  - Code Lists and Categories  
  - Classifications and Classification Items 

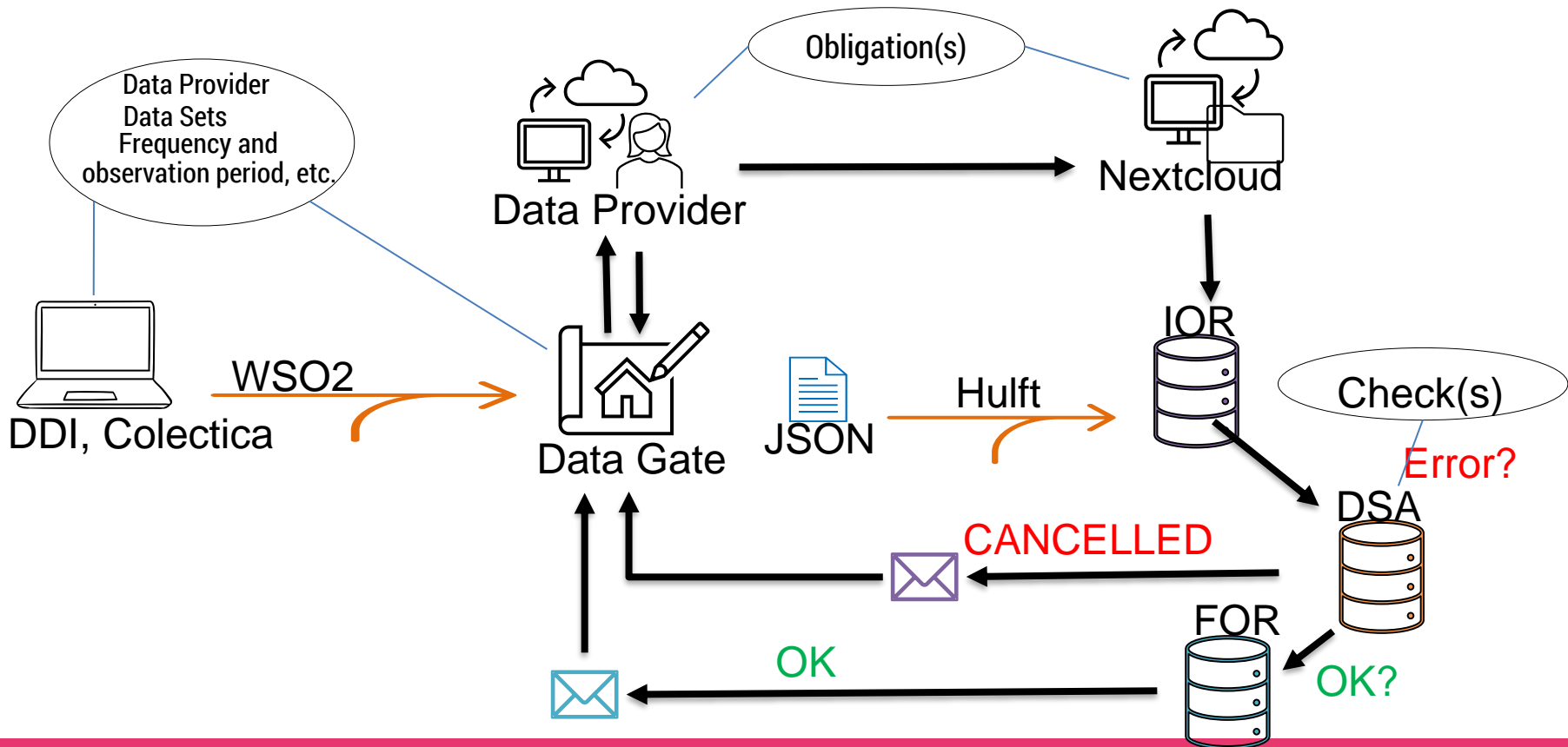
# Result of the analysis - Data model, draft version



# Benefits



# Simplified process and counterparts



# Benefits, what we (will) get



For data capture purposes, in Data Gateway

- Automatic processes for data capture (at least from SE side)
- All information about agreements, metadata based on these agreements and obligations are in one place
- Calendar (schedule) for data capture – until now this was lacking
- Monitoring panel – which data in which state, status etc.
- Validation of captured data according to the agreement

Also, ...

- All metadata, that we have and located everywhere or that we missed, are stored now or will be stored in one place
- Reuse of metadata
- Overview of existing administrative registry data
- Implementing DDI stimulates the thinking of what else could be possible to document
  
- In the future:
  - Automatization of pseudonymization – document in metadata, what data should pseudonymize

# Use case

# Data Provider Portal – List of Obligations

STATISTICS



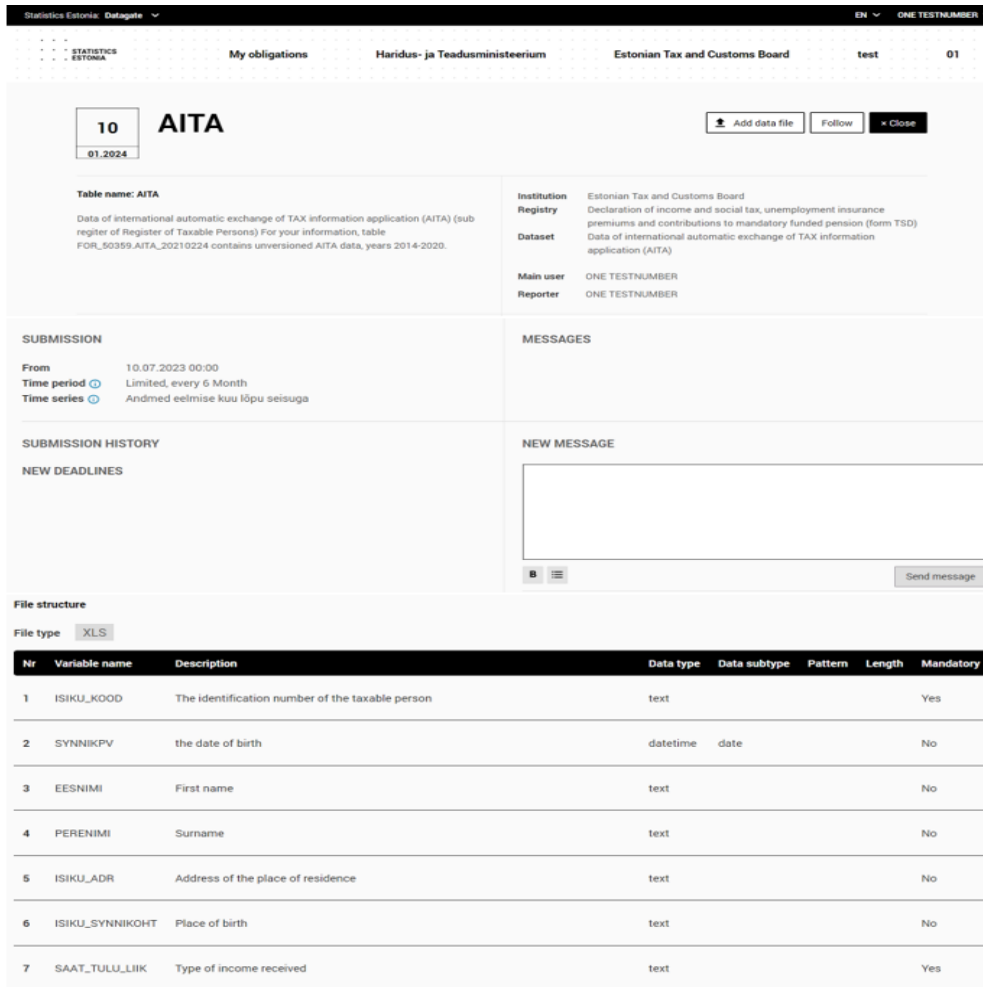
All obligations are listed with submission dates and statuses

The screenshot shows the 'Estonian Tax and Customs Board' portal. At the top, there are navigation links for 'My obligations', 'Haridus- ja Teadusministeerium', 'Estonian Tax and Customs Board', 'test', and 'Uus Asutus OÜ'. The main heading is 'Estonian Tax and Customs Board' with a 'Reporters' button. Below this is a 'Filters' section with input fields for 'From' (10.07.2023), 'To' (11.07.2024), 'Obligation/dataset name' (Enter name), and 'Status' (All). There is also a 'Reporter' dropdown set to 'ONE TESTNUMBER', buttons for 'Following' and 'All deadlines', and a 'Show all faulty & missed' toggle. A calendar for June 2023 is visible on the right. The main content area displays a table of obligations:

Submission Date	Obligation Name	Reporter	Status
01.2024	Eeklefinantsimata andmestiku (leelaadimine) (ENG) Silt saad ühes laadides andmestiku (ENG)	ONE TESTNUMBER	Add data file
10 01.2024	AITA Data of international automatic exchange of TAX information application (AITA) (sub register of Register of Taxable Persons) For your information, table FOR_50359.AITA_20210224 contains unversioned AITA data, years 2014-2020. <a href="#">More</a>	ONE TESTNUMBER	MESSAGES
10 01.2024	AITA Data of international automatic exchange of TAX information application (AITA) (sub register of Register of Taxable Persons) For your information, table FOR_50359.AITA_20210224 contains unversioned AITA data, years 2014-2020. <a href="#">More</a>	ONE TESTNUMBER	MESSAGES
10 07.2024	AITA Data of international automatic exchange of TAX information application (AITA) (sub register of Register of Taxable Persons) For your information, table FOR_50359.AITA_20210224 contains unversioned AITA data, years 2014-2020. <a href="#">More</a>	ONE TESTNUMBER	MESSAGES

# Data Provider Porta

- Overview on one obligation



The screenshot displays the 'Data Provider Porta' interface for the obligation 'AITA'. The top navigation bar includes 'Statistics Estonia', 'My obligations', 'Haridus- ja Teadusministeerium', 'Estonian Tax and Customs Board', and 'test 01'. The main content area features a header with '10 AITA' and a date '01.2024'. Below this, there are buttons for 'Add data file', 'Follow', and 'Close'. The interface is divided into several sections: 'Table name: AITA' with a description of the data; 'SUBMISSION' with details on the time period and series; 'SUBMISSION HISTORY' and 'NEW DEADLINES'; 'MESSAGES' and 'NEW MESSAGE' with a text input field and a 'Send message' button; and 'File structure' which includes a table of variables.

**Table name: AITA**  
Data of international automatic exchange of TAX information application (AITA) (sub register of Register of Taxable Persons) For your information, table FOR\_50359.AITA\_20210224 contains unversioned AITA data, years 2014-2020.

**Institution** Estonian Tax and Customs Board  
**Registry** Declaration of income and social tax, unemployment insurance premiums and contributions to mandatory funded pension (form TSD)  
**Dataset** Data of international automatic exchange of TAX information application (AITA)  
**Main user** ONE TESTNUMBER  
**Reporter** ONE TESTNUMBER

**SUBMISSION**  
**From** 10.07.2023 00:00  
**Time period** Limited, every 6 Month  
**Time series** Andmed eelmise kuu lõpu seisuga

**SUBMISSION HISTORY**  
**NEW DEADLINES**

**MESSAGES**

**NEW MESSAGE**

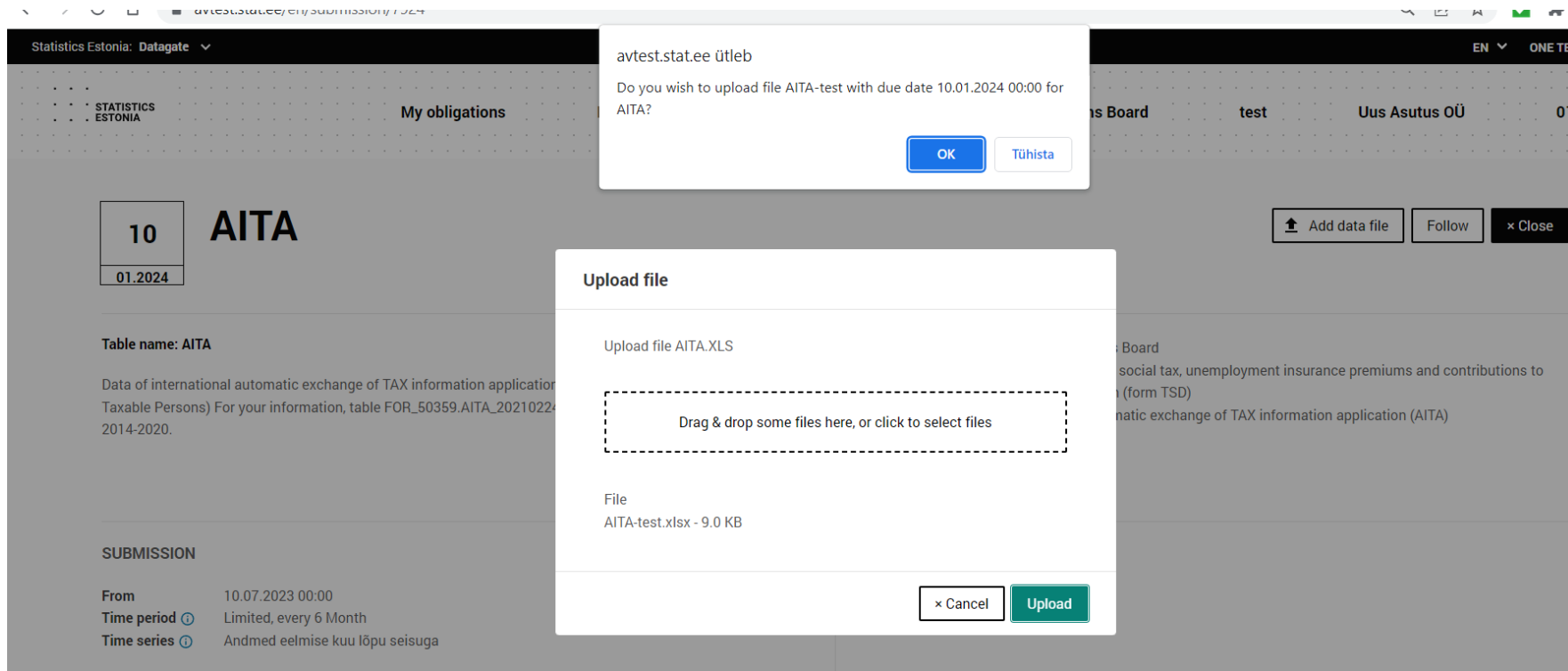
**File structure**  
File type: XLS

Nr	Variable name	Description	Data type	Data subtype	Pattern	Length	Mandatory
1	ISIKU_KOOD	The identification number of the taxable person	text				Yes
2	SYNNIKPV	the date of birth	datetime	date			No
3	EESNIMI	First name	text				No
4	PERENIMI	Surname	text				No
5	ISIKU_ADR	Address of the place of residence	text				No
6	ISIKU_SYNNIKOHT	Place of birth	text				No
7	SAAT_TULU_LIIK	Type of income received	text				Yes



# Data Provider Portal

- Submitting agreed data file




The screenshot displays the 'Statistics Estonia: Datagate' interface. The main content area shows 'My obligations' for table 'AITA' with a submission date of '10 01.2024'. The table name is 'AITA' and the description is 'Data of international automatic exchange of TAX information application (Taxable Persons) For your information, table FOR\_50359.AITA\_20210224-2014-2020.' The submission details are as follows:

SUBMISSION	
From	10.07.2023 00:00
Time period	Limited, every 6 Month
Time series	Andmed eelmise kuu lõpu seisuga

Two modal dialogs are overlaid on the page:

- Confirmation Dialog:** Titled 'avtest.stat.ee ütleb', it asks 'Do you wish to upload file AITA-test with due date 10.01.2024 00:00 for AITA?' and has 'OK' and 'Tühista' buttons.
- Upload Dialog:** Titled 'Upload file', it shows 'Upload file AITA.XLS' and a dashed box for file selection with the text 'Drag & drop some files here, or click to select files'. Below the box, it lists 'File AITA-test.xlsx - 9.0 KB' and has 'Cancel' and 'Upload' buttons.

## Submitted data – data has arrived

 <p><b>Eeldefineerimata andmestiku üleslaadimine (ENG)</b> Siit saad üles laadida andmestiku (ENG)</p>	<p><a href="#">Add data file</a></p> <p>Reporter: ONE TESTNUMBER</p>	
<p><b>10</b> 01.2024 ARRIVED</p> <p><b>AITA</b> Data of international automatic exchange of TAX information application (AITA) (sub regiter of Register of Taxable Persons) For your information, table FOR_50359.AITA_20210224 contains unversioned AITA data, years 2014-2020.</p> <p><a href="#">More</a></p>	<p><a href="#">Add data file</a></p> <p>Institution: Estonian Tax and Customs Board</p>	<p><b>MESSAGES</b></p> <p>Data File is submitted, but it is not final yet.</p>
<p><b>10</b> 01.2024</p> <p><b>AITA</b> Data of international automatic exchange of TAX information application (AITA) (sub regiter of Register of Taxable Persons) For your information, table FOR_50359.AITA_20210224 contains unversioned AITA data, years 2014-2020.</p> <p><a href="#">More</a></p>	<p><a href="#">Add data file</a></p> <p>Institution: Estonian Tax and Customs Board</p>	<p><b>MESSAGES</b></p>

# SE portal – data arrival monitoring

## Monitooring

Kalender Tabel

### Today's load



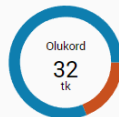
50%  
2 tk laekunud

50%  
2 tk täna saabumas

0%  
0 tk laekumata

0%  
0 tk tagasi lükatud

### Overview of received files



19%  
6 tk töös

81%  
26 tk saabunud

0%  
0 tk tagasi lükatud

0%  
0 tk korras

0%  
0 tk vigased

### Last month's receipts statistics



100%  
14 tk õigeaegselt

0%  
0 tk hilinenud

Välista igapäevased  Näita kõiki vigaseid ja laekumata

Kõik

Minu Jälgimisel Kõik

aita  Alates: 26.06.2023 Kuni: 11.08.2024

Otsi asutuse, andmestiku, sisendfaili, esitaja, kliendihalduri või sisuhalduri nime, FOR tabeli nime või numbr järgi

Kohustuse ID	Kohustus/sisendfail	Andmestik	Esitaja	Tähtaeg	Olek
248	AITA	Data of international automatic exchange of TAX information application (AITA)	ONE TESTNUMBER	10.01.2024 00:00	SAABUNUD
248	AITA	Data of international automatic exchange of TAX information application (AITA)	ONE TESTNUMBER	10.01.2024 00:00	SAABUMAS
248	AITA	Data of international automatic exchange of TAX information application (AITA)	ONE TESTNUMBER	10.07.2024 00:00	SAABUMAS
248	AITA	Data of international automatic exchange of TAX information application (AITA)	ONE TESTNUMBER	10.07.2024 00:00	SAABUMAS



# SE portal – detailed overview of arrivals

STATISTICS  
ESTONIA



p\_tt1

KORRAS

Märgi lõpetatuks

Lisa andmefail

× Sulge

ORKESTREERIMINE

Alusta

Lõpeta

Alusta uuesti

Korda samm

Jäta viiga vahele

Töövoog 29.12.2022 12:01 - 29.12.2022 13:22

Faali struktuuri kontroll

Korras

Faali sisu kontroll

Korras

IOR-i laadimine

Korras

Andmete töötluskontroll

Korras

FOR-i viimine

Lõpetatud

KIRJEID

Sisendfaalis

IORis 4

FORis 4

TÖÖVOO KULG

Algus 29.12.2022 12:01

Lõpp 29.12.2022 13:22

Kestvus 1 tundi

KOHUSTUSE ANDMED

Kohustus [p\\_tt1\\_ID 214](#)  
Asutus test  
Register provv\_tt  
Andmestik provv\_tt\_ID 63  
FOR 66666, provv\_tt1  
Andmeesitaja SIRLI NIIBO-TAMM  
Töövoog Proovivoog  
Esitamine Tähtajatu, vahega 1 Kuu  
Ajaline kaetus 29.12.2022 12:00, ID 6803  
Tähtaeg  
Automaatedastus Ei  
Seotud lingid

Lisa link

KEHTIV

TEATED

TÄHTAJA MUUDATUSED

Muuda tähtaega

TEGEVUSTE LOGI

Faali üleslaadimine 29.12.2022 13:21

Tauno Tammi: Skaneerimise staatus: OK Üleslaadimise staatus: OK HULFFI saatmise staatus: OK

> Faali üleslaadimine 29.12.2022 12:01

UUS TEADE

# SE portal – describing of obligations

## New obligation

Status  
Kirjeldamisel | Loobu | Salvesta

### ORGANIZATION AND DATASET

**Organization\***  
Vali või sisesta uus

**Dataset\***  
Vali või sisesta uus

**Register\***  
Vali või sisesta uus

**Input file\***  
Vali või sisesta uus  Critical

### SUBMISSION

**Time series\***  
Vali ajaline kaetus

**From\***  
Palun vali kuupäev

**To\***  
Palun vali kuupäev

Unlimit  Limited  One  
Frequency period e In

### PERSONS

**Client manager**  
Määra kliendihaldur

**Content manager**  
Määra sisuhaldur

### FILE STRUCTURE

**File format\***  
 txt  csv  xls(x)  xml  json

**Column separator\***  
Punkt, koma vm

**String attribute**

**Decimal separator\***  
Punkt, koma vm

Nr	Variable name	Description	Data type	Data subtype	Pattern	Length
1	Muutuja nimi	Kirjeldus	Formaat	Alamformaat	Formaadi kuju	Pikkus

Status  
Kirjeldamisel | Loobu | Salvesta



OTTAWA 2023

64TH WORLD STATISTICS CONGRESS



THANK YOU.

