EFSA IUCLID DATA REUSE VISION

- EFSA has received **IUCLID dossiers for PPP** applications for about 4 years.

- The costs of the migration to IUCLID have been visible for all parties involved (industry, MSs, EFSA), while benefits not fully exploited yet.

- The systematic (re)use of the structured data in IUCLID through dashboards could be a tangible benefit for everyone leading to faster and more consistent risk assessments.

- EFSA has launched a project last year to ingest/extract selected data from EFSA Agency IUCLID instance hosted by ECHA and build **dashboards** with these data.

- Two tools were identified to extract IUCLID data (with a long-term preference for the latter):

  - the ECHA data extractor (**DE**).

  - a new data product based on **API** to be developed by EFSA.

- EFSA has managed to extract data and test these tools since this summer, and today we present the first outcomes

API = Application Programming Interface

ECHA DATA EXTRACTOR

- As short-term solution while building the data product, EFSA is testing ECHA Data Extractor since August 2025 to extract production data
  - Mature tool – ECHA has been developing it for years
  - Its use does not require programming skills
  - Designed for systematic extraction of data across dossiers including attachments

- The first tests on **EFSA "metabolites"** use case has highlighted some limitations:
  - cannot deal with different contexts in a single extraction (we have 4 contexts in our database)
  - complex output (several files that need to be merged afterwards)
  - not designed for the extraction of the full dossier content (it requires manual selection of all existing sections and sub-sections under the Data Extractor table of content)

- Since the data product is already allowing us to extract with API the data that we need, and we have the technical expertise to run the API calls within the team, we are considering to focus "all" our resources on the data product already now.
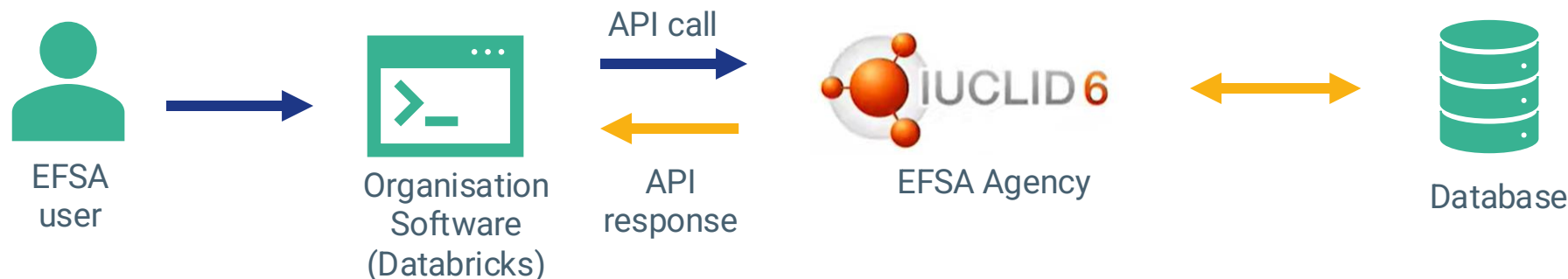
PPP DATA PRODUCT

Steps needed to prepare dashboards with IUCLID PPP data:

1. Retrieve data from EFSA Agency (data ingestion)

2. Organise the data (semantic layer and data marts)

3. Prepare the dashboards (data visualisation)

PPP DATA PRODUCT: 1 DATA INGESTION WITH API CALLS

- The work begins by launching API calls to EFSA Agency instance to retrieve the data.
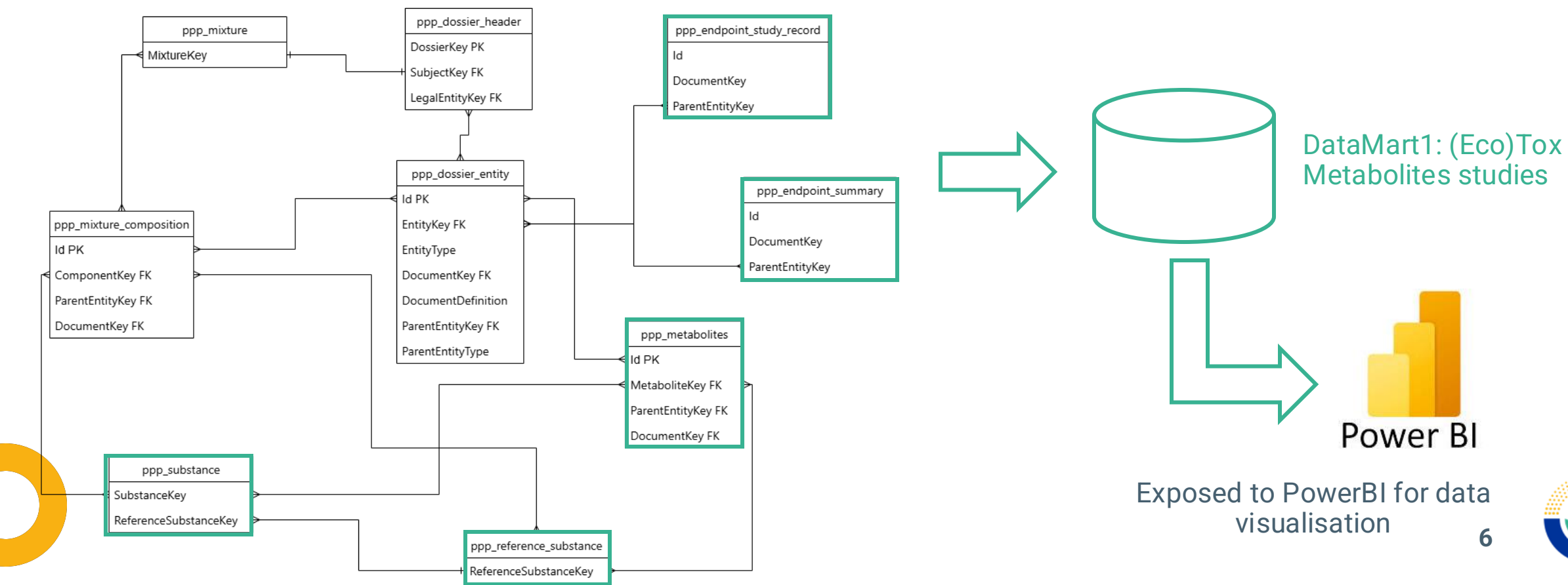

- Available API calls are documented by ECHA (not all of them are available to EFSA by default in the Gateway)

- EFSA has to:

  - Explore ECHA's documentation.

  - Test API calls and ask for addition of new ones to ECHA when needed (i.e. whitelisting process).

  - Implement API calls in an efficient ETL (Extract, Transform, Load) pipeline.

- We are currently exploring ways to optimize API calls performance:

  - **Sequential approach** is time consuming (n. of requests is a multiple of the entities/documents contained in dossier)

  - **Parallel approach** under exploration by EFSA to reduce data extraction time

5
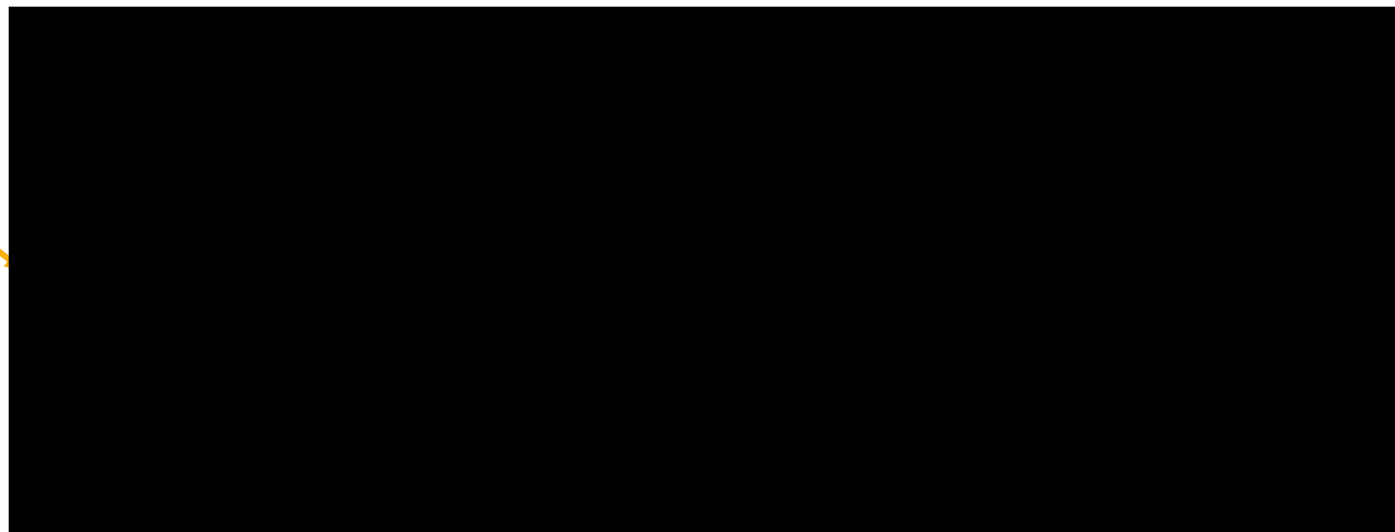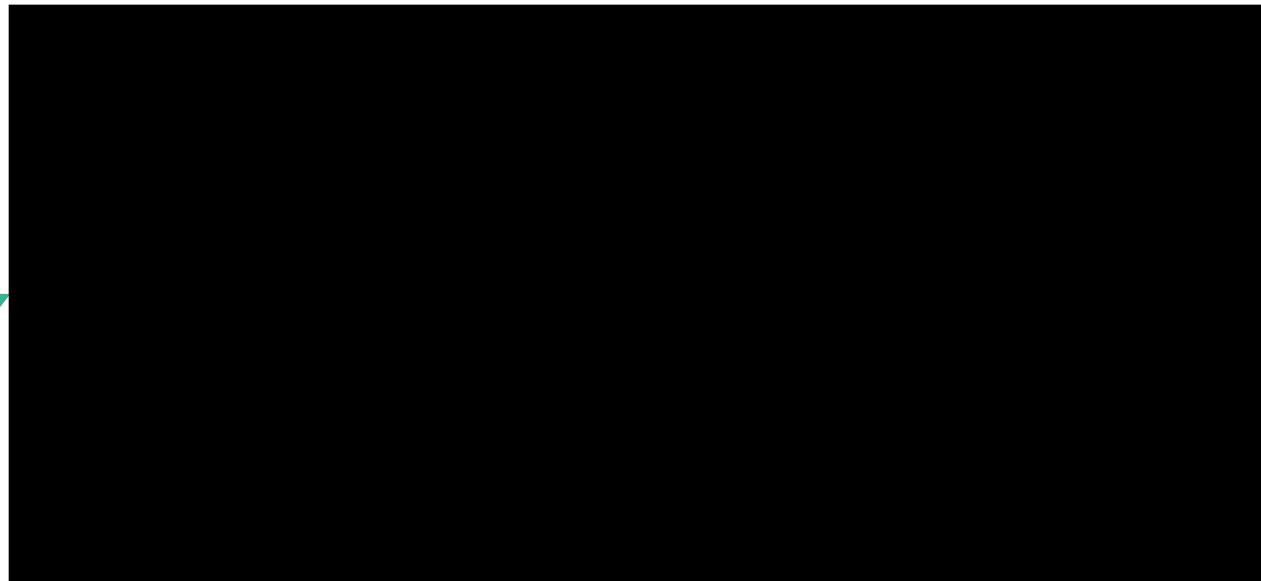
PPP DATA PRODUCT: 2 SEMANTIC LAYER AND DATA MARTS

Data retrieved is being organized to be reused for different use cases.

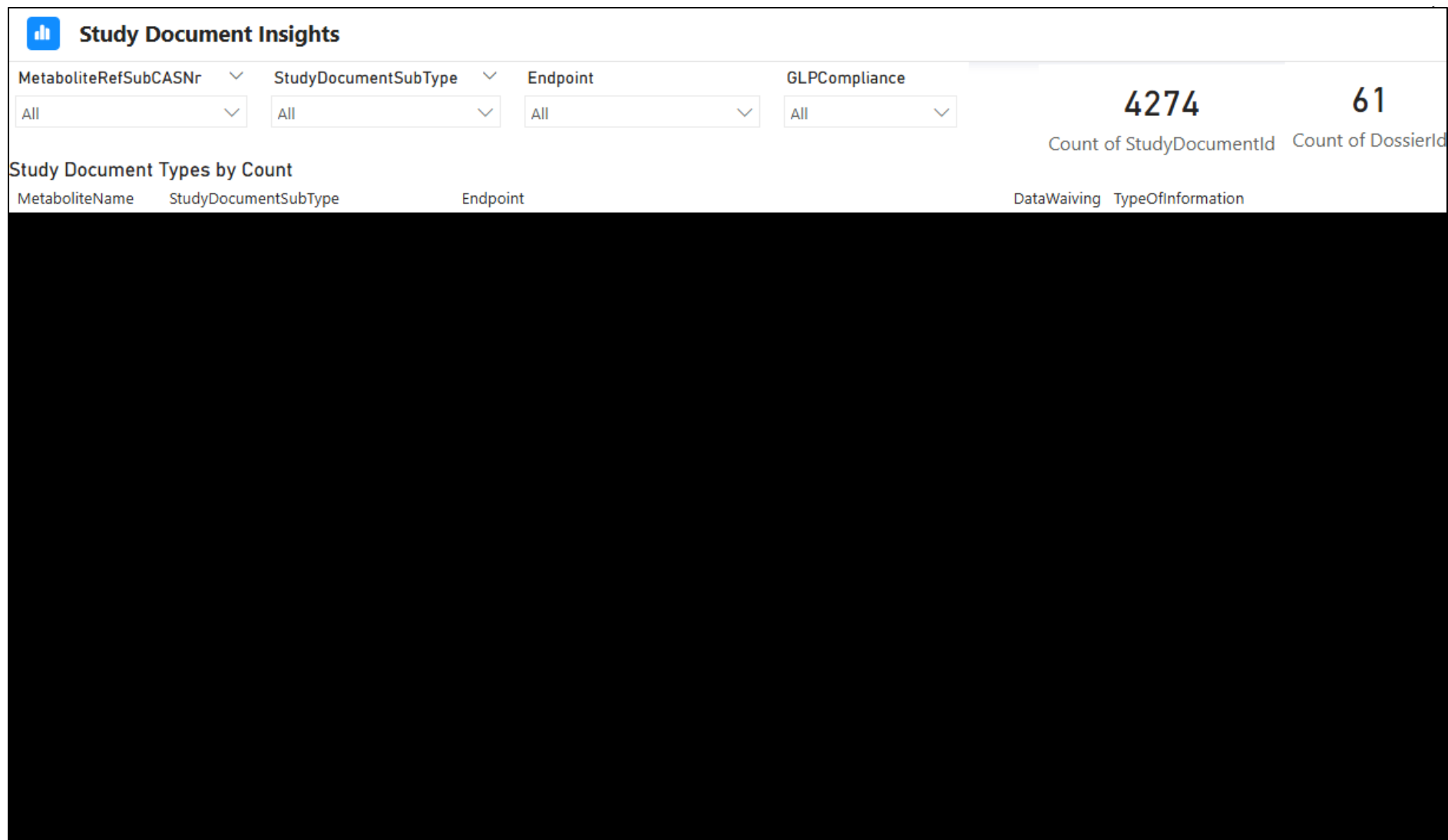- Different tables by entity/document type

DataMart1: (Eco)Tox Metabolites studies

Exposed to PowerBI for data visualisation

6

PPP DATA PRODUCT: 2 SEMANTIC LAYER AND DATA MARTS

dossieranalytics
- ppp_dossier_document
- ppp_dossier_entity
- ppp_dossier_header
- ppp_endpoint_study_record
- ppp_endpoint_summary
- ppp_metabolites
- ppp_mixture
- ppp_mixture_composition
- ppp_reference_substance
- ppp_substance

PPP DATA PRODUCT: 3 DATA VISUALISATION

Power BI

Study Document Insights

MetaboliteRefSubCASNr	StudyDocumentSubType	Endpoint	GLPCompliance	4274	61
All	All	All	All	Count of StudyDocumentId	Count of DossierId

Study Document Types by Count

MetaboliteName	StudyDocumentSubType	Endpoint	DataWaiving	TypeOfInformation

LONG-TERM PLAN - FULL INGESTION OF IUCLID PPP DATA

- EFSA is working with ECHA on a technical solution that would allow EFSA to ingest the entire dossiers from EFSA Agency IUCLID database (all data from all IUCLID fields and all attachments from all PPP dossiers)

- After this step, EFSA will make available this data and attachments in SharePoint providing:

  - Easier access to attachments, which will be organised in document libraries

  - Possibility to search across attachments and fields and across dossiers

  - Customised advanced searches and filters

  - Security features (access, handling of confidential data and collaboration possibilities)

- A similar solution has just been put in place for all the other food domains

- Timeline: 2027, + more detailed planning ongoing

ADVANCED SEARCH IN SHAREPOINT

- Search Box
- Search results
- Filter

DORA   Panels ⌄   Audit trail   FAQ   Data Protection Notice

Dossier Overview for Risk Assessment   Home   + Create site   ☆ Not following   ⬈ Share ⌄   ⋯

Dossier Search

🔍 Enter your search terms…

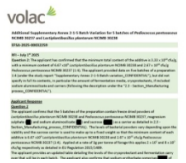**FEED-2023-21053**

Document filters

Type ⌄

Sensitivity ⌄

Section ⌄

Status ⌄

File type ⌄

CONCLUSIONS

NEXT STEPS AND TIMELINE

- By end of 2025
  - Finalization of the semantic layer and first dashboard prototype on metabolites.

- From 2026
  - Start with the **involvement of Member States** in:
    - testing and enhancing the dashboard prototype
    - define requirements for additional dashboards linked to use cases

  - Start the work for the full ingestion of PPP IUCLID dossiers to provide access to dossier data (including attachments) via Sharepoint.