

# ECHA case study – **How to manage data in a regulatory environment**

June 2021 - ECDC consultation on  
digital technologies in public health

*Athanasios Demiridis and  
Blanca Martinez de Aragon*  
European Chemicals Agency



# ECHA's Goals

The European Chemicals Agency (ECHA) is the EU regulatory agency that works for the **safe use of chemicals**.

Its vision is to be the centre of knowledge on the sustainable management of chemicals for the **benefit of citizens and the environment**



Environment



Health



Innovation &  
competitiveness

# ECHA's Goals

Data is key to achieve these goals!



Environment

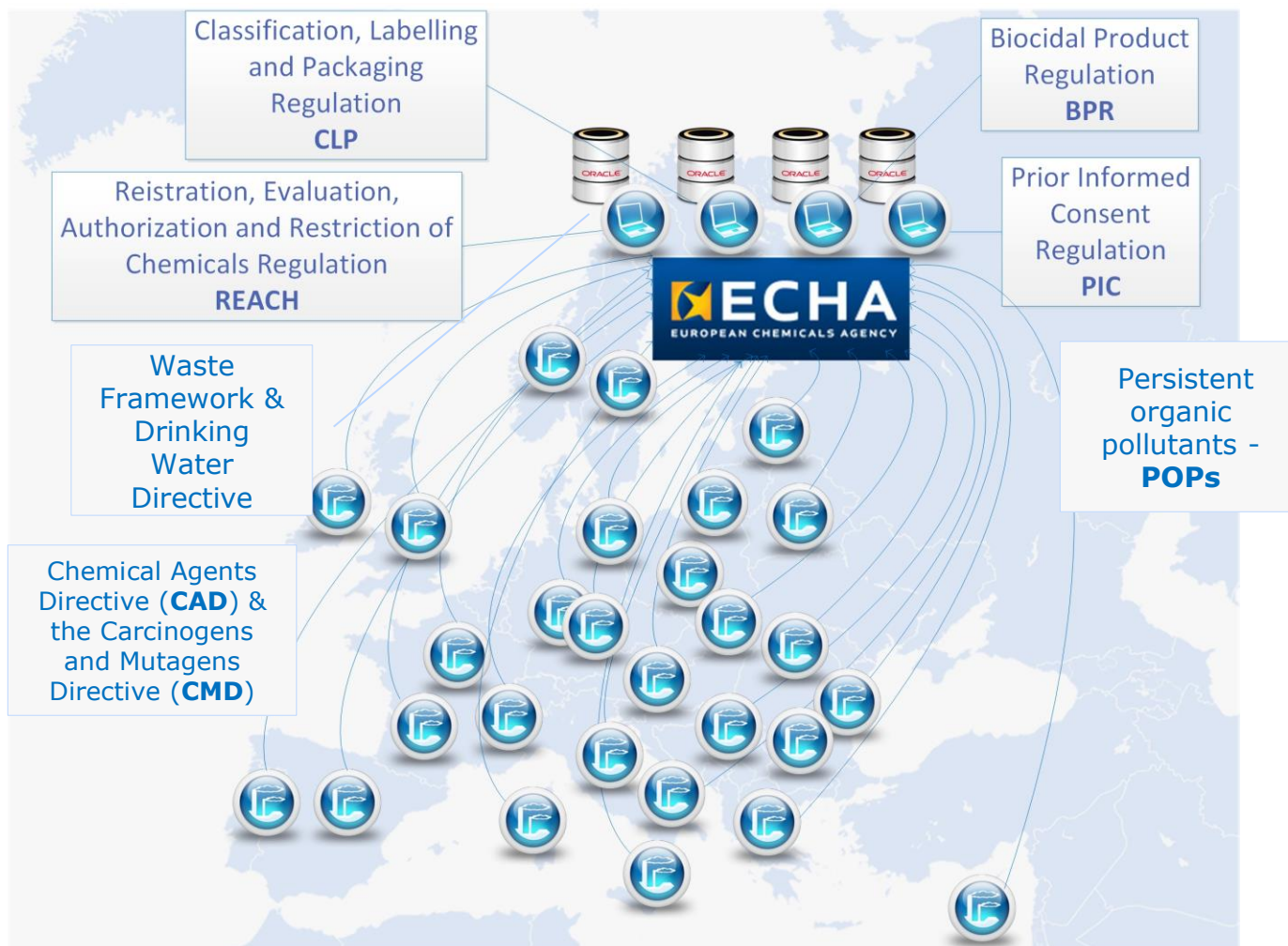















Health



Innovation &  
competitiveness

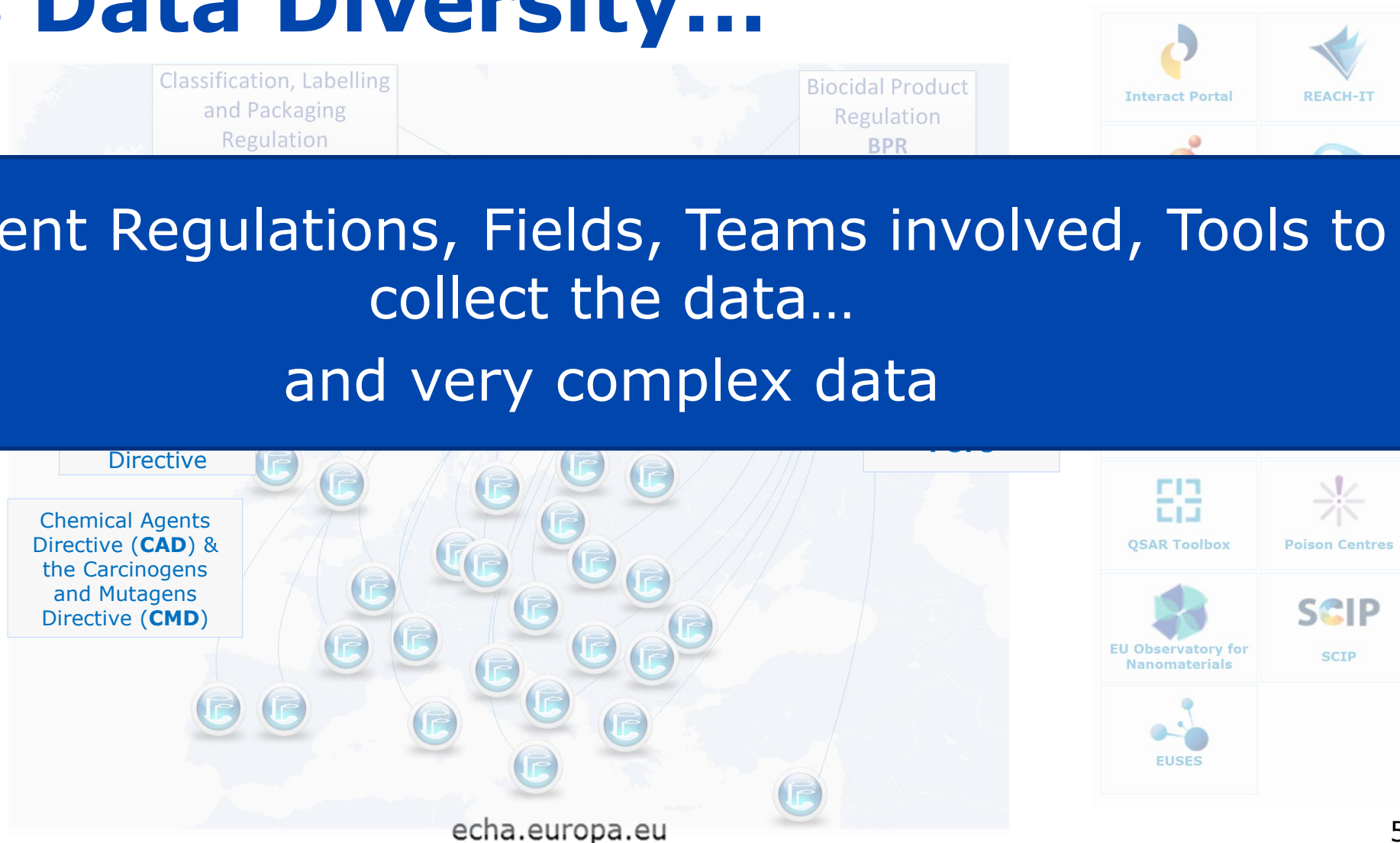
# ECHA's Data Diversity...



 Interact Portal	 REACH-IT
 IUCLID 6	 CHESAR
 R4BP 3	 SPC Editor
 ePIC	 ECHA Cloud Services
 QSAR Toolbox	 Poison Centres
 EU Observatory for Nanomaterials	 SCIP
 EUSES	

# ECHA's Data Diversity...

Different Regulations, Fields, Teams involved, Tools to collect the data...  
and very complex data



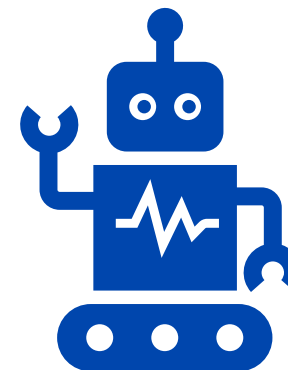
## How do we manage all the data and make sure we have a common understanding?



Data Format



Enterprise Data  
Model (EDM)

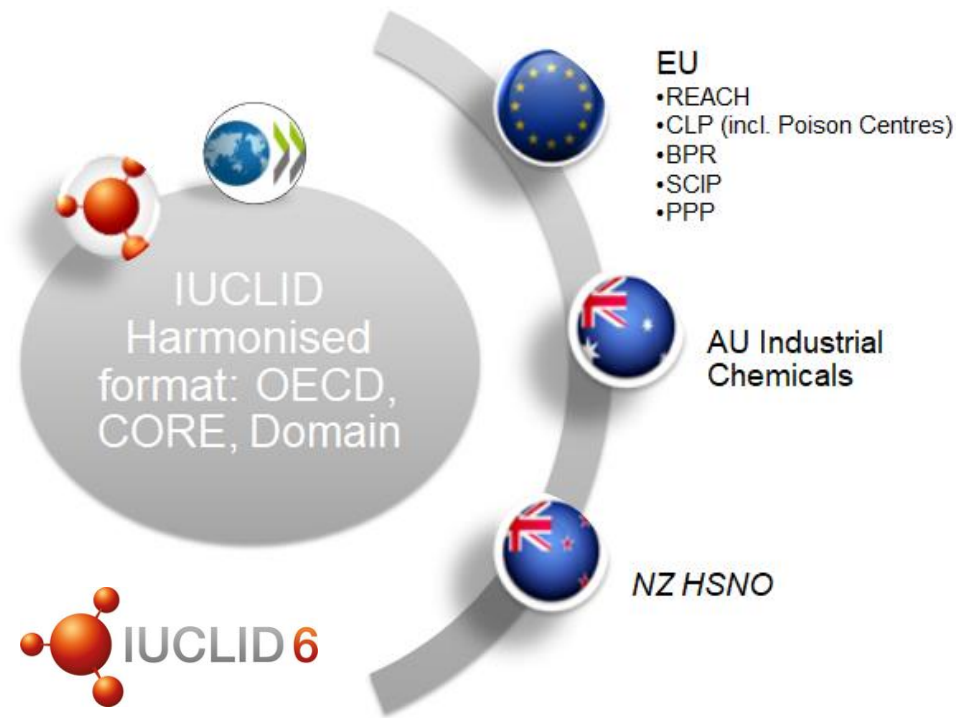


Technologies



# Data Format

- **IUCLID** is the International Uniform Chemical Information Database maintained by ECHA and the OECD
- A **Common Harmonized Format**, multilingual
- Facilitates to the industry from different Member States the **data collection and data sharing** (*and to ECHA!*)
- It is **updated** once a year in collaboration with the user community



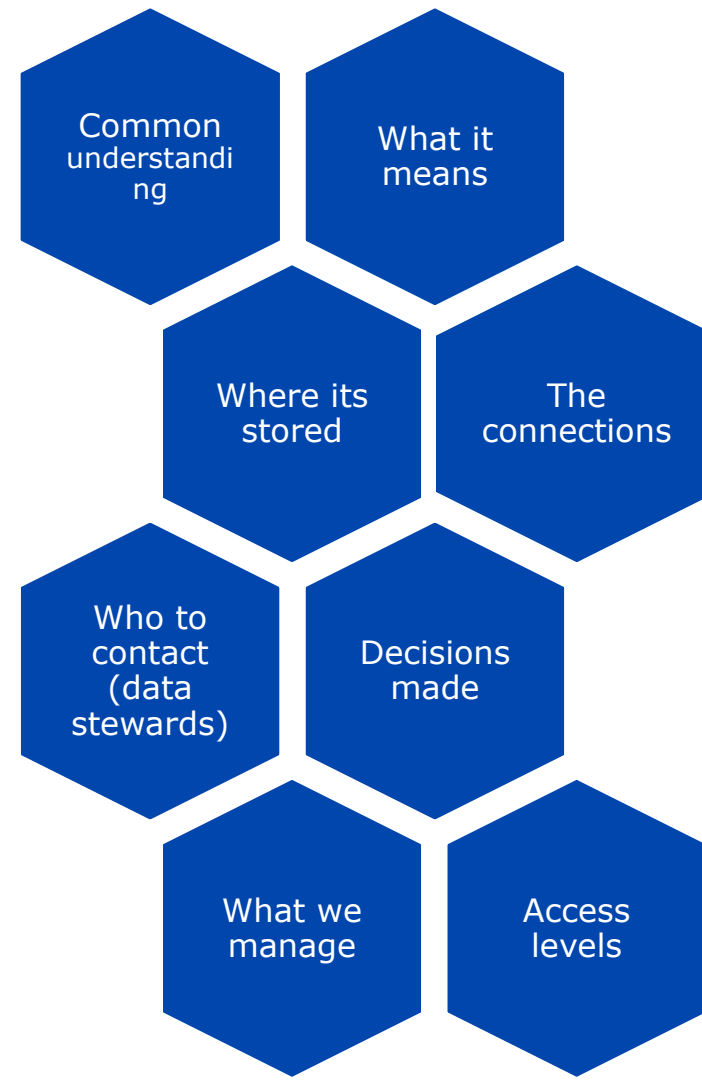
Not all the incoming data has this format



# The Enterprise Data Model

Creates a **shared understanding** across the organisation of:

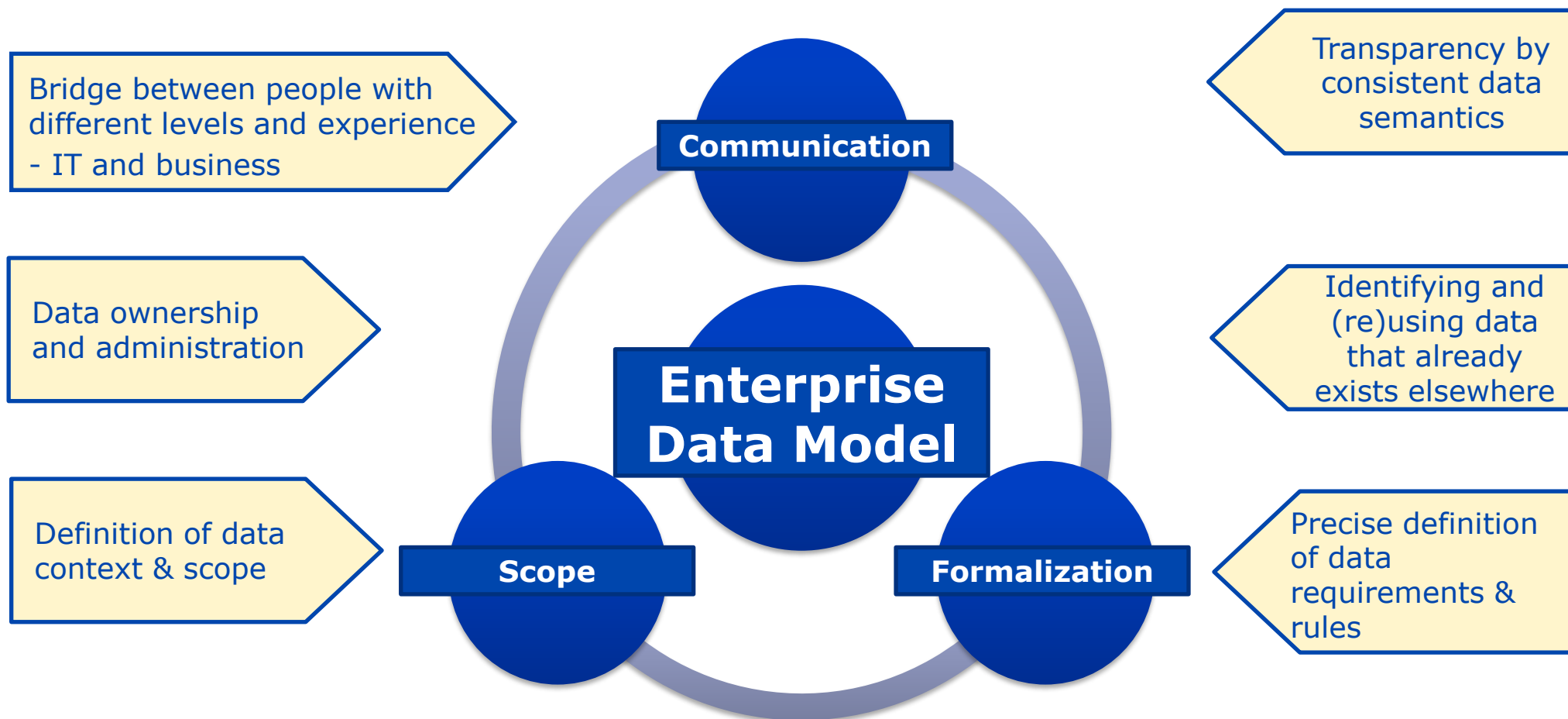
1. the **elements** of importance to ECHA's business (e.g. *chemical substances, file/submissions..*)
2. the **facts about these element** that the business needs to keep record of (e.g. *substance identifiers, level of completeness..*)

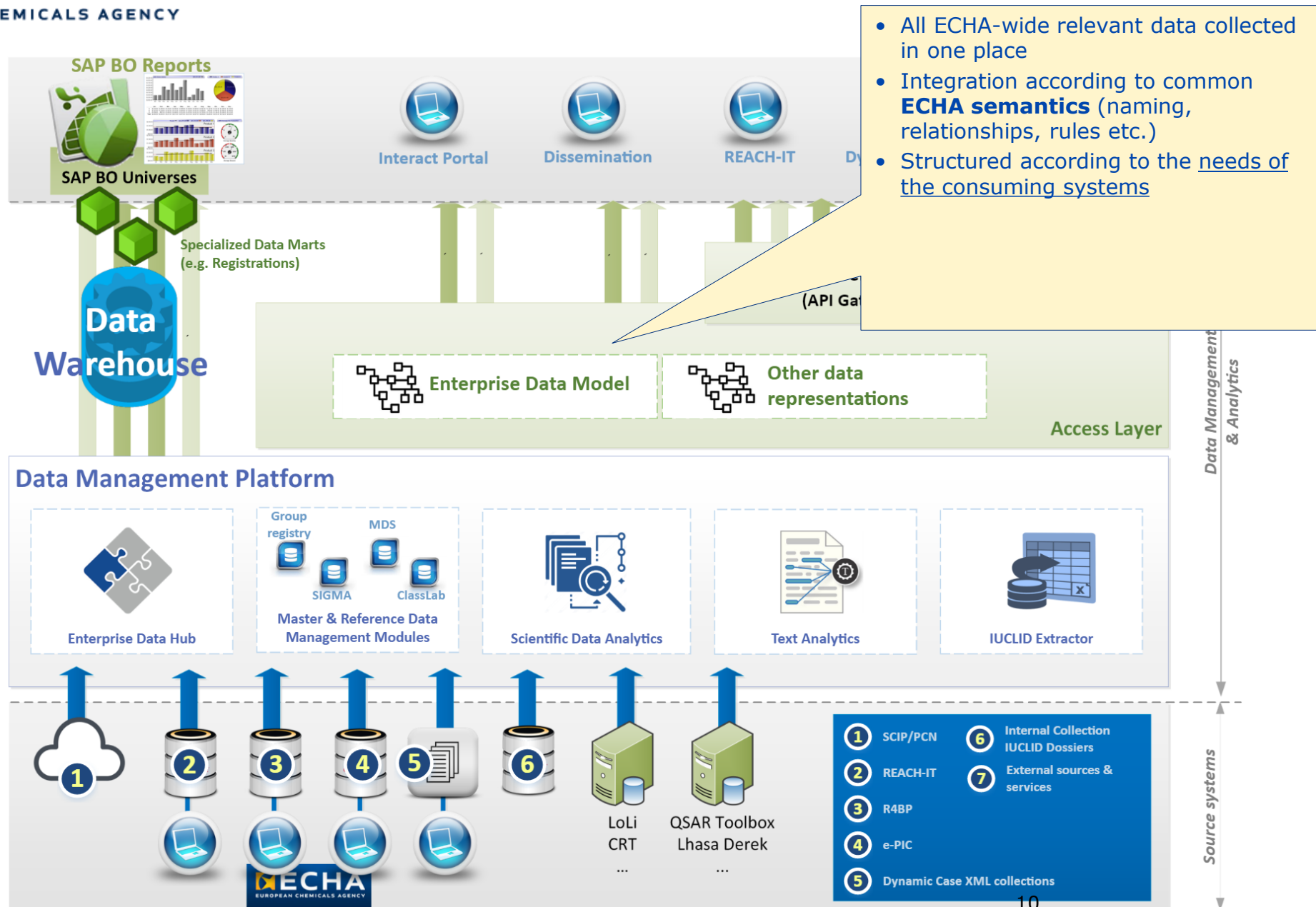


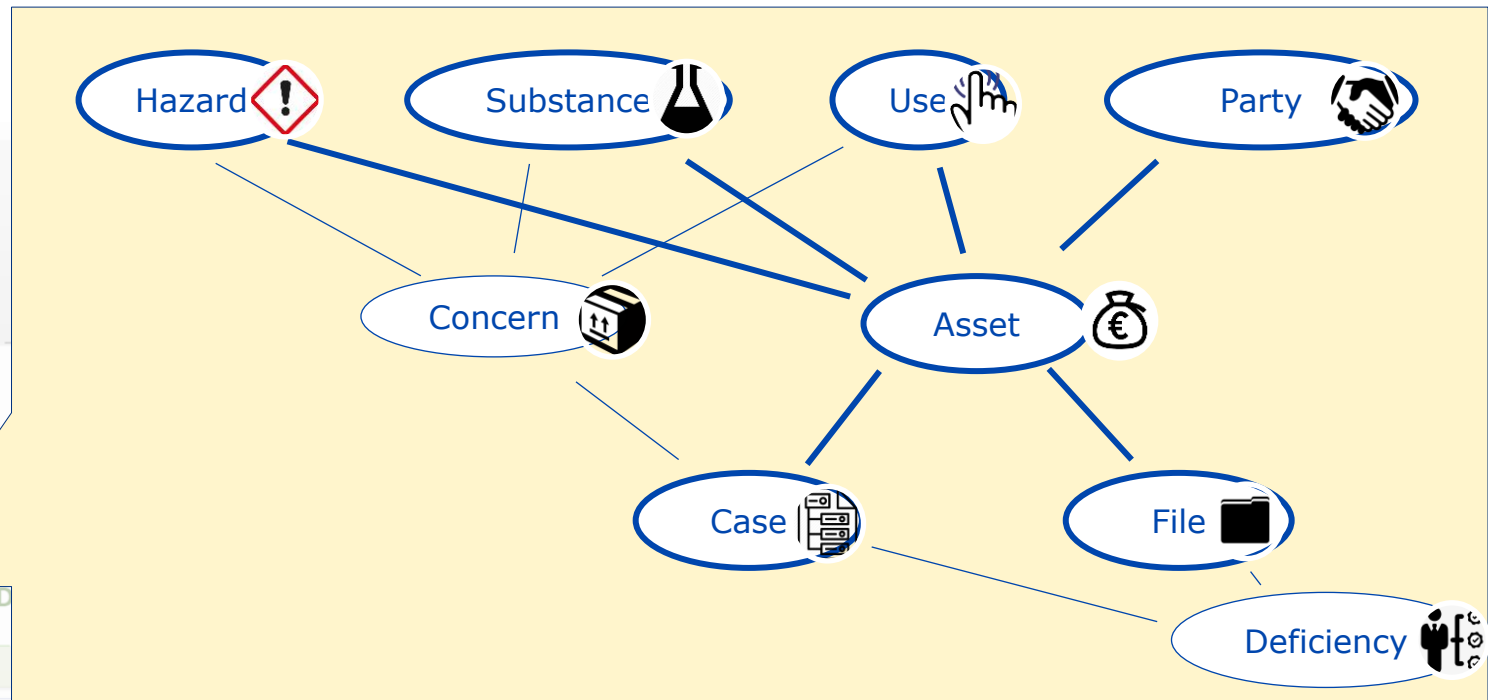
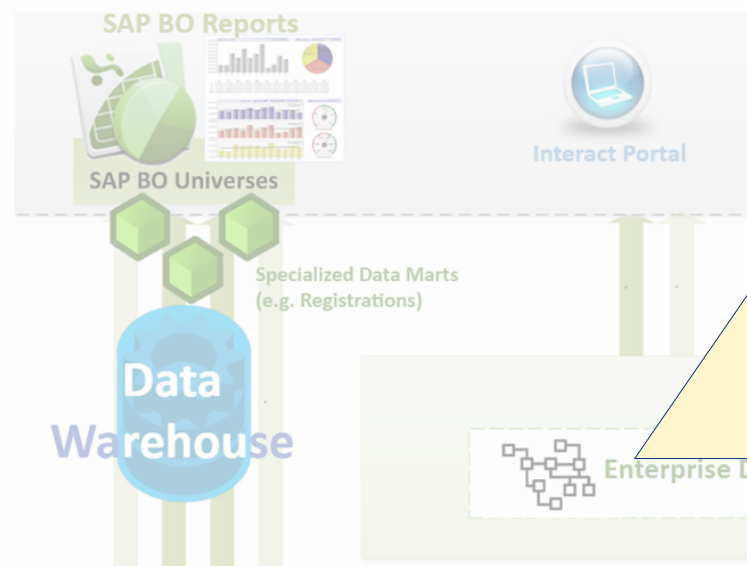




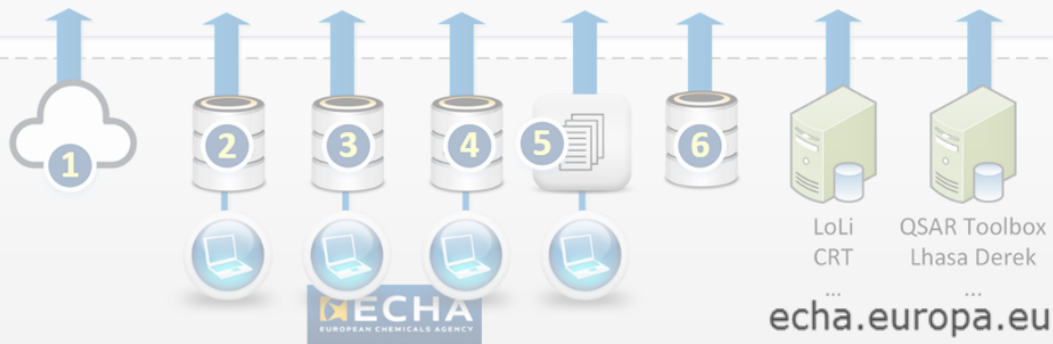
# Why is it important?



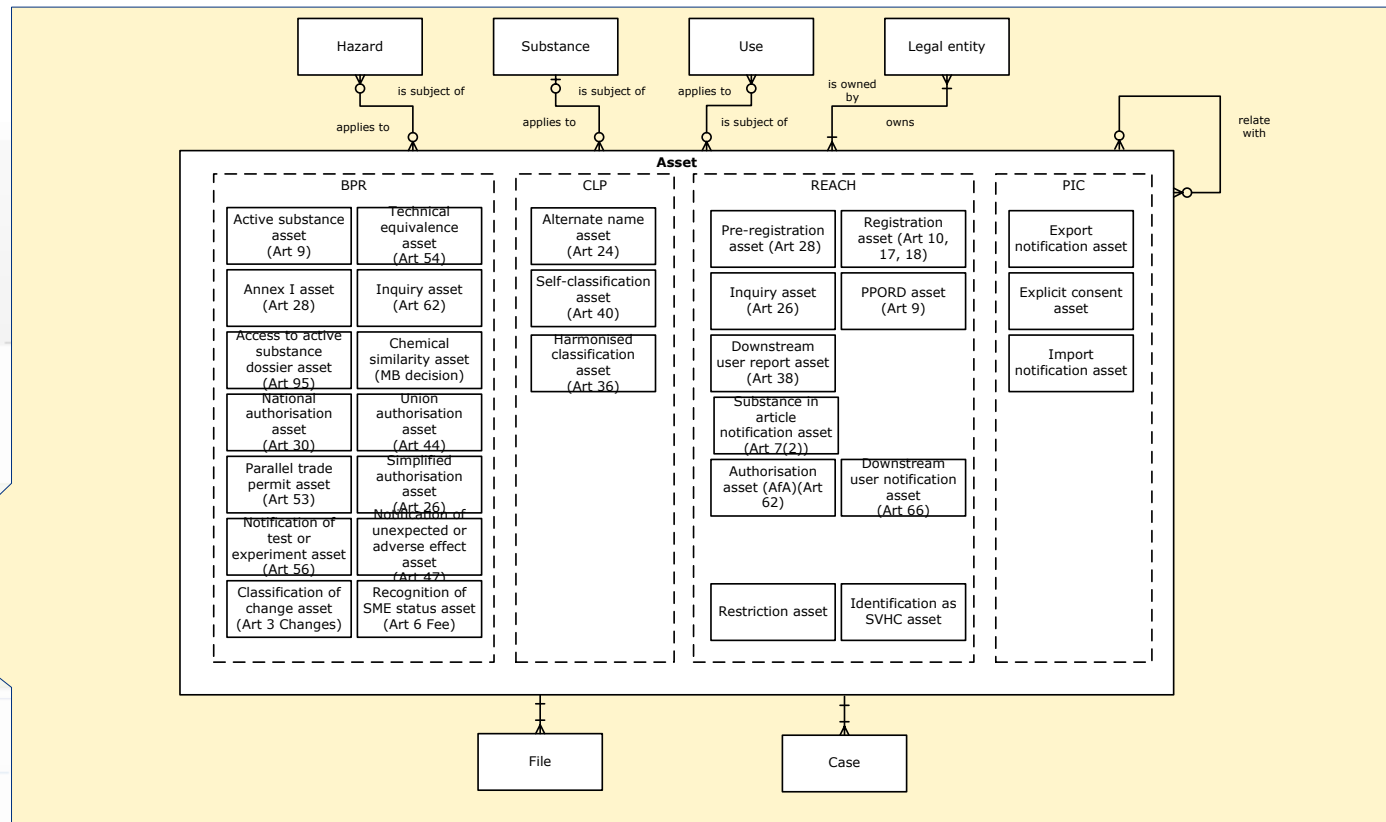
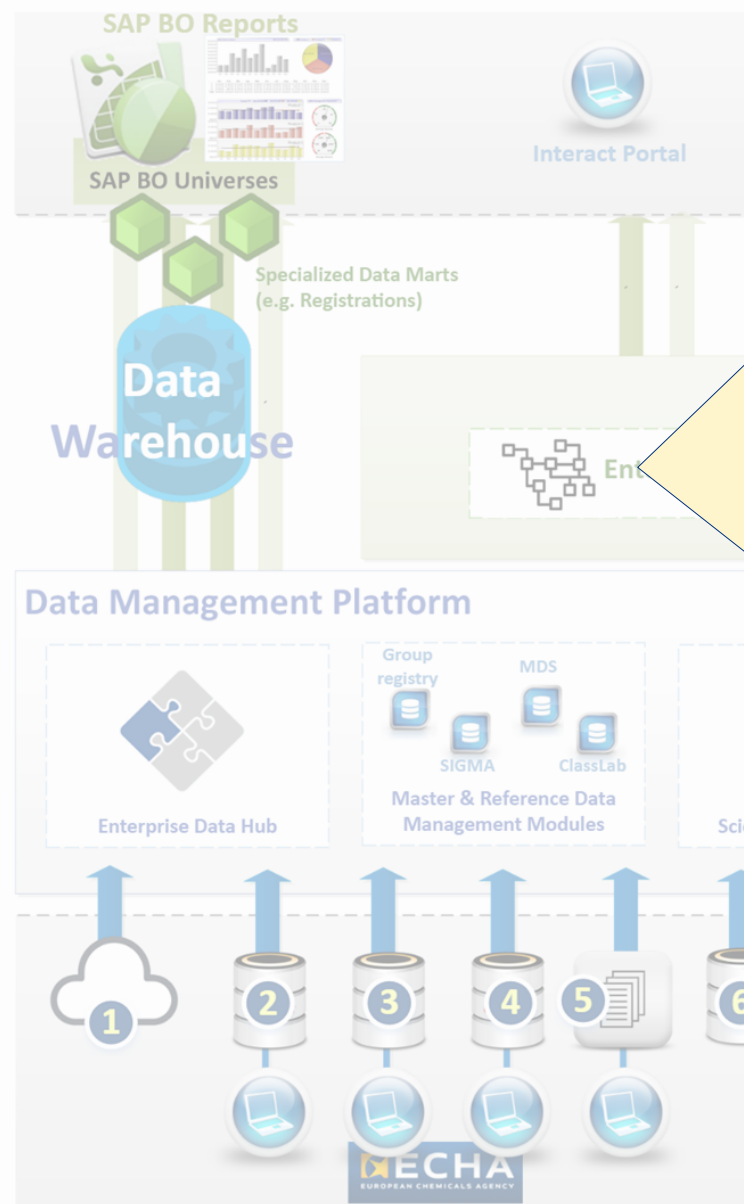




### Data Management Platform

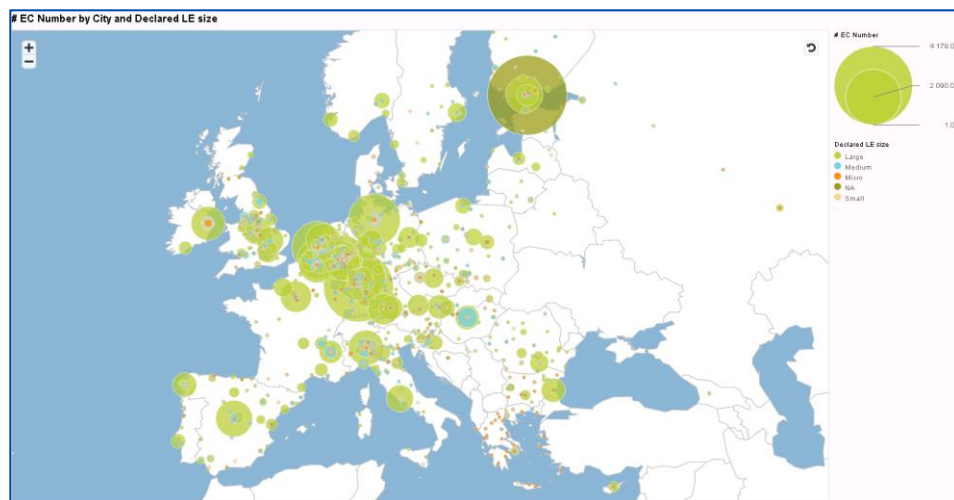


- |                                |                                       |
|--------------------------------|---------------------------------------|
| 1 SCIP/PCN                     | 6 Internal Collection IUCLID Dossiers |
| 2 REACH-IT                     | 7 External sources & services         |
| 3 R4BP                         |                                       |
| 4 e-PIC                        |                                       |
| 5 Dynamic Case XML collections |                                       |



- 1 SCIP/PCN
- 2 REACH-IT
- 3 R4BP
- 4 e-PIC
- 5 Dynamic Case XML collections
- 6 Internal Collection IUCLID Dossiers
- 7 External sources & services

# Allowing to bring the information together...



Monitoring ClassLab  
& REACH-IT

### 4,4'-isopropylidenediphenol

Bisphenol A; BPA

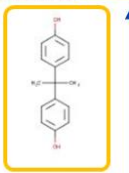
Regulatory process names: 2 Translated names: 29 CAS names: 1 IUPAC names: 21 Trade names: 15 Other identifiers: 8

#### Substance identity

EC / List no.: 201-245-8

CAS no.: 80-05-7

Mol. formula: C<sub>15</sub>H<sub>16</sub>O<sub>2</sub>



#### Hazard classification & labelling

**Danger!** According to the **harmonised classification and labelling** (ATP09) approved by the European Union, this substance may damage fertility, causes serious eye damage, may cause an allergic skin reaction and may cause respiratory irritation.

**Additionally**, the classification provided by companies to ECHA in **REACH registrations** identifies that this substance may damage fertility or the unborn child and is toxic to aquatic life with long lasting effects.

#### Properties of concern

- R** Toxic to Reproduction
- Ss** Skin sensitising
- ED** Endocrine Disrupting

#### Important to know

- Substance included in the **Community Rolling Action Plan (CoRAP)**.
- Substance of very high concern (SVHC) and included in the **candidate list** for authorisation.
- Some uses of this substance are restricted under **Annex XVII** of REACH.

#### How to use it safely

ECHA has no data from registration dossiers on the precautionary measures and guidance on safe use for this substance.

#### About this substance

This substance is manufactured and/or imported in the European Economic Area in 1 - 10 tonnes per year.

This substance is used in articles, by professional workers (widespread uses), in formulation or re-packing, at industrial sites and in manufacturing.

#### Consumer Uses

ECHA has no public registered data indicating whether or in which chemical products the substance might be used. ECHA has no public registered data on the routes by which this substance is most likely to be released to the environment.

#### Article service life

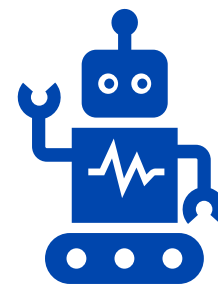
Other release to the environment of this substance is likely to occur from: outdoor use in long-life materials with low release rate (e.g. metal, wooden and plastic construction and building materials) and indoor use in long-life materials with low release rate (e.g. flooring, furniture, toys, construction materials, curtains, foot-wear, leather products, paper and cardboard products, electronic equipment). This substance can be found in complex articles, with no release intended: vehicles and machinery, mechanical ...

about INFOCARD - Last updated: 21/08/2019

Monitoring all source systems  
for occurrences (RML/EDH)

Monitoring Dynamic Case

**We keep up with rapidly  
changing requirements  
using New Technologies**





# Example: SCIP



## Duty holders

**Any supplier of articles containing a substance of very high concern (SVHC)** on the Candidate List above 0.1% w/w



**As from 5 January 2021**

Must submit SCIP notification to ECHA



EU producers & Importers

EU *assemblers*

EU distributors



## SCIP

SCIP is the database for information on **S**ubstances of **C**oncern **I**n articles as such or in complex objects (**P**roducts) established under the **W**aste **F**ramework **D**irective (**W**FD).

The SCIP database makes the information **publicly available**, in particular to waste operators and consumers.

# SCIP: Main Challenges



## **How to handle the huge volume of submissions?**

Estimated at 20 million dossiers in the first year. Some days more than 500.000 were submitted. Previous architecture could not cope with it.



## **How to ensure and improve the Quality of Data, Duplication, Integrity, Consistency?**



## **How to reuse existing components? Making sure that the solution is interoperable and scalable**

# SCIP: The Solution

- ✓ **Harmonised data format – IUCLID:** using a common vocabulary to have consistency, a common understanding and simplify data management.
- ✓ **Lambda architecture** (*Batch layer, Serving layer and Speed layer*): Allows to process huge amounts of data and accessing it with low latency.
- ✓ **Data Validation rules:** Custom validation mechanisms to identify anomalies and process only correct dossiers. **Automation** to avoid manual work.
- ✓ **Monitoring and stabilisation** of the SCIP data integration/refresh process to become even more resilient to data volume increases.
- ✓ **Private Cloud:** Higher performance than on premise and better for scalability

# SCIP: The Solution

- ✓ **Harmonised data format – IUCLID:** using a common vocabulary to have consistency, a common understanding and simplify data management.
- ✓ **Lambda architecture** (*Batch layer, Serving layer and Speed layer*): Allows to process huge amounts of data and accessing it with low latency.
- ✓ **Data Validation rules:** Custom validation mechanisms to identify anomalies and process only correct dossiers. Automation to avoid manual work.
- ✓ **Monitoring and stabilisation** of the SCIP data integration/refresh process to become even more resilient to data volume increases.
- ✓ **Private Cloud:** Higher performance than on premise and better for scalability

Designing the architecture to be scalable, prepared for future growth

# Next Steps



Study on the use of Public Cloud to further reduce overhead.



Expand the Data Interoperability across institutions.



Enhance the data analysis (e.g. natural language processing, identify trends, predictions..)



Further dissemination of the information including visualizations.



Define and agree on the archiving strategy



# Thank you!

[Athanasios.DEMIRIS@echa.europa.eu](mailto:Athanasios.DEMIRIS@echa.europa.eu)

[Blanca.MARTINEZDEARAGON@echa.europa.eu](mailto:Blanca.MARTINEZDEARAGON@echa.europa.eu)

