

Introduction to the new database and harvesting tool by ETC/ATNI

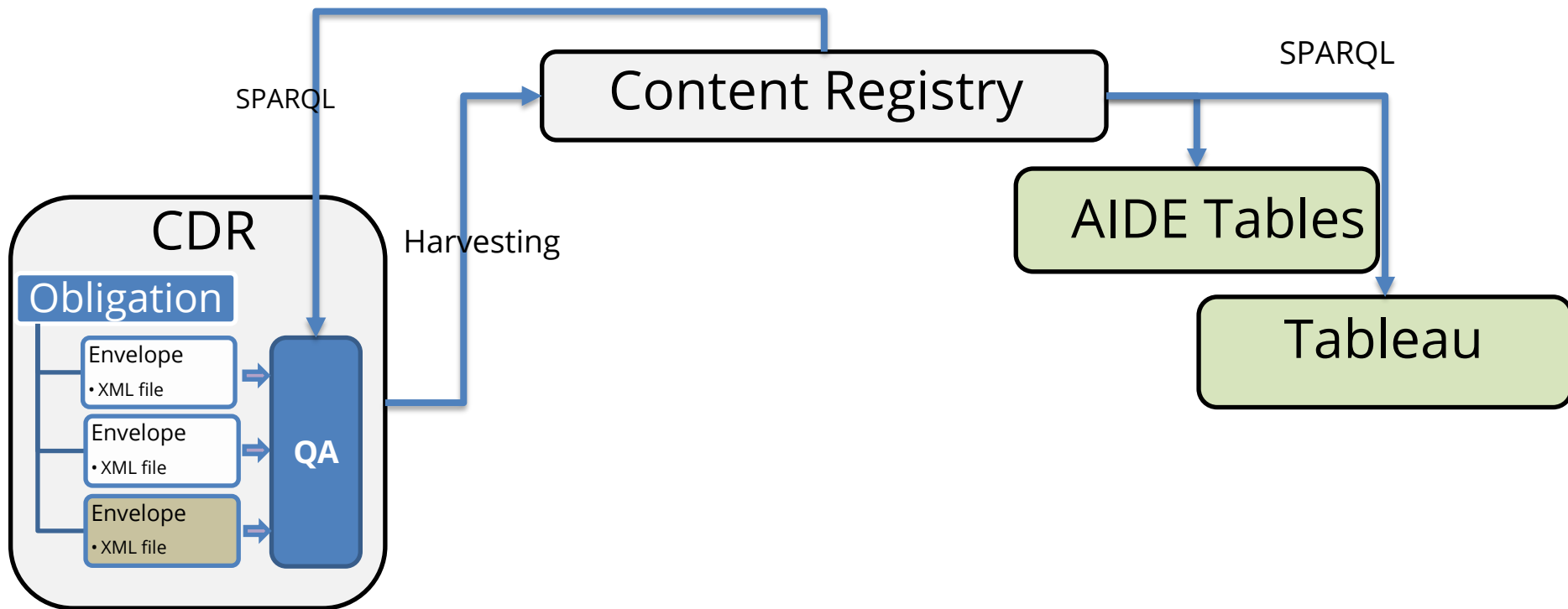
*EEA & ETC-ATNI work
compiled by Luca Liberti*



Presentation Plan

- Current EEA infrastructure
- Database and harvester
- Future Plans

EEA Infrastructure for H-K– Current Status



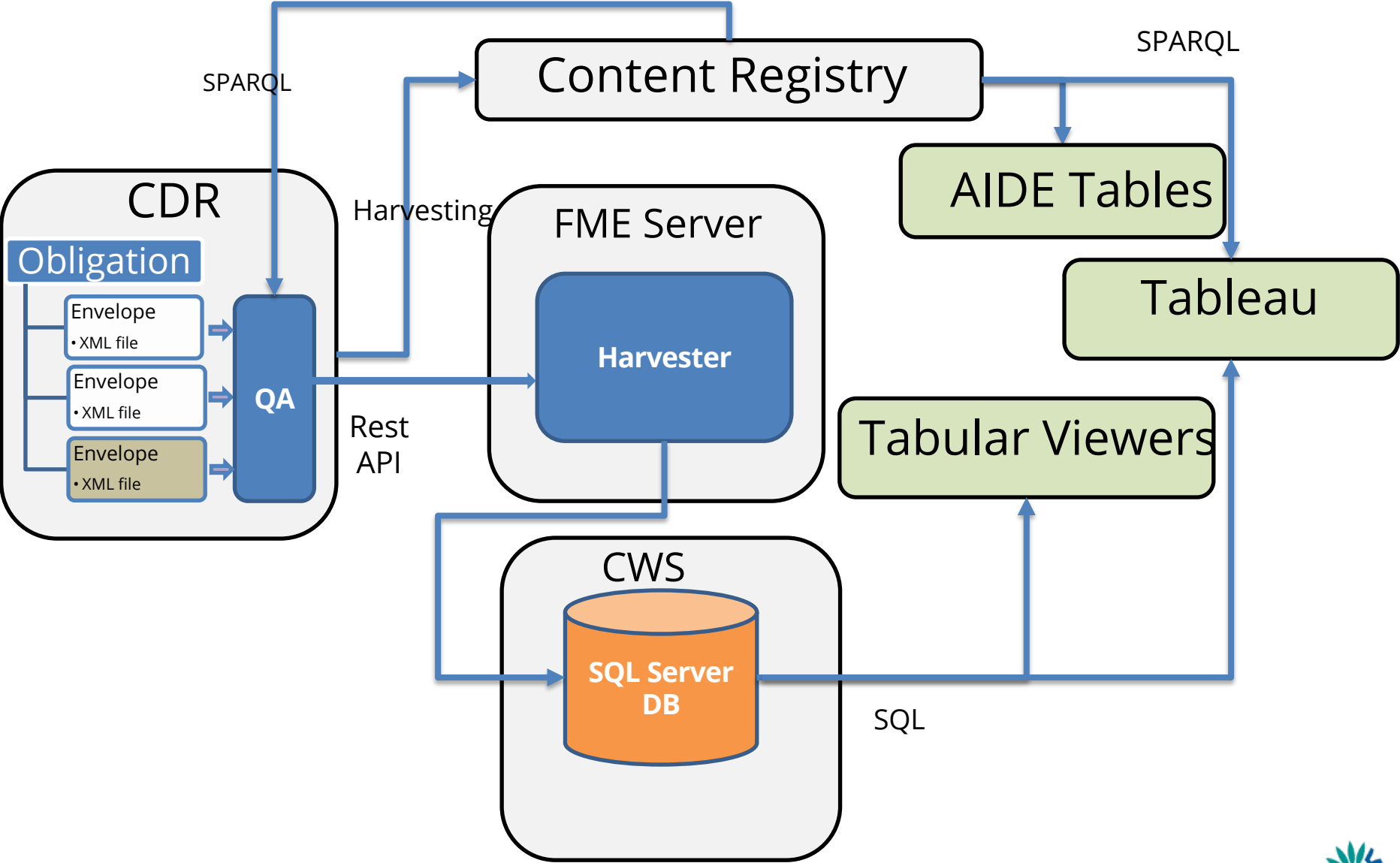
EEA Infrastructure for H-K– Need to improve

- **One of the main issues of H-K dataset is the lack of a database infrastructure**
- **SPARQL has a steep learning curve**
- **CR difficult to interface with other tools**
- **Too few products**

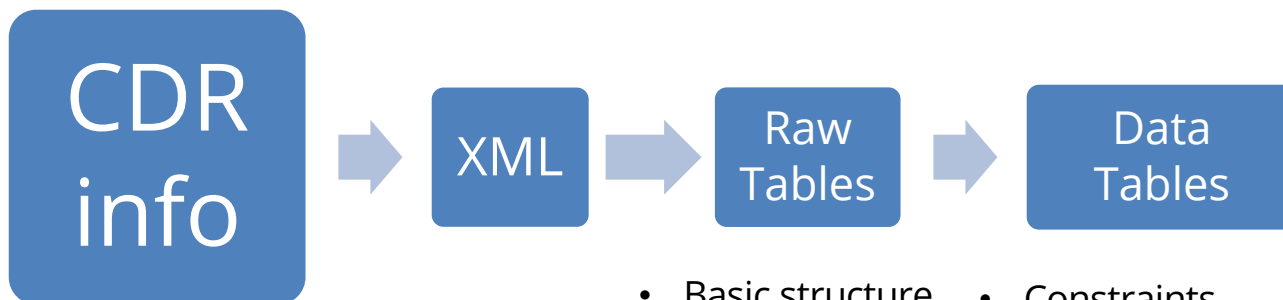
EEA Infrastructure for H-K– NILU Harvester

- **Developed a relational database in SQL Server to store H-K data**
- **Developed a harvester in C# .NET and FME to populate the database with XML data from CDR**
- **Final phase of the project**

EEA Infrastructure for H-K- Improved



Harvester – Harvesting Steps



- Basic structure
- Textual data types
- No relationships
- Constraints (Relationships, Uniqueness,...)
- Data type conversion



Harvester – Database structure

Raw Tables

- Raw import of XML

Data Tables

- Fully qualified data

Dictionaries

- Lookup tables extracted from <http://dd.eionet.europa.eu/vocabularies>

Control Tables

- Harvesting process management tables

Metadata Tables

- Store CDR envelopes metadata information

Harvester – Database design

- Database structure follows the data model
- Each Data model “main” object maps to a Table
- Remove unnecessary object nesting (flatten 1:1 nested elements)
- Table names and field names mostly match the XML tag names
- Integer Autonumber PK in each table to build relationships
- Link to dictionaries are based on shortened codes

<http://dd.eionet.europa.eu/vocabulary/aq/spatialscale/local>

Becomes

local

- Each record points to the original file information

Future Plans

- **Harvest the existing deliveries in the database**
- **Provide feedback on unresolved issues**
- **Develop products on H-K**

Thank You!!