Global architecture of the EEA data information system

InfoRAC visit 2017

Jan Bliki, Copenhagen 19-21 April 2017

Producing European datasets

Countries deliveries

33 countries
XML, SHP, GML, KML, CSV, Access, etc...

European datasets

80+ European Datasets

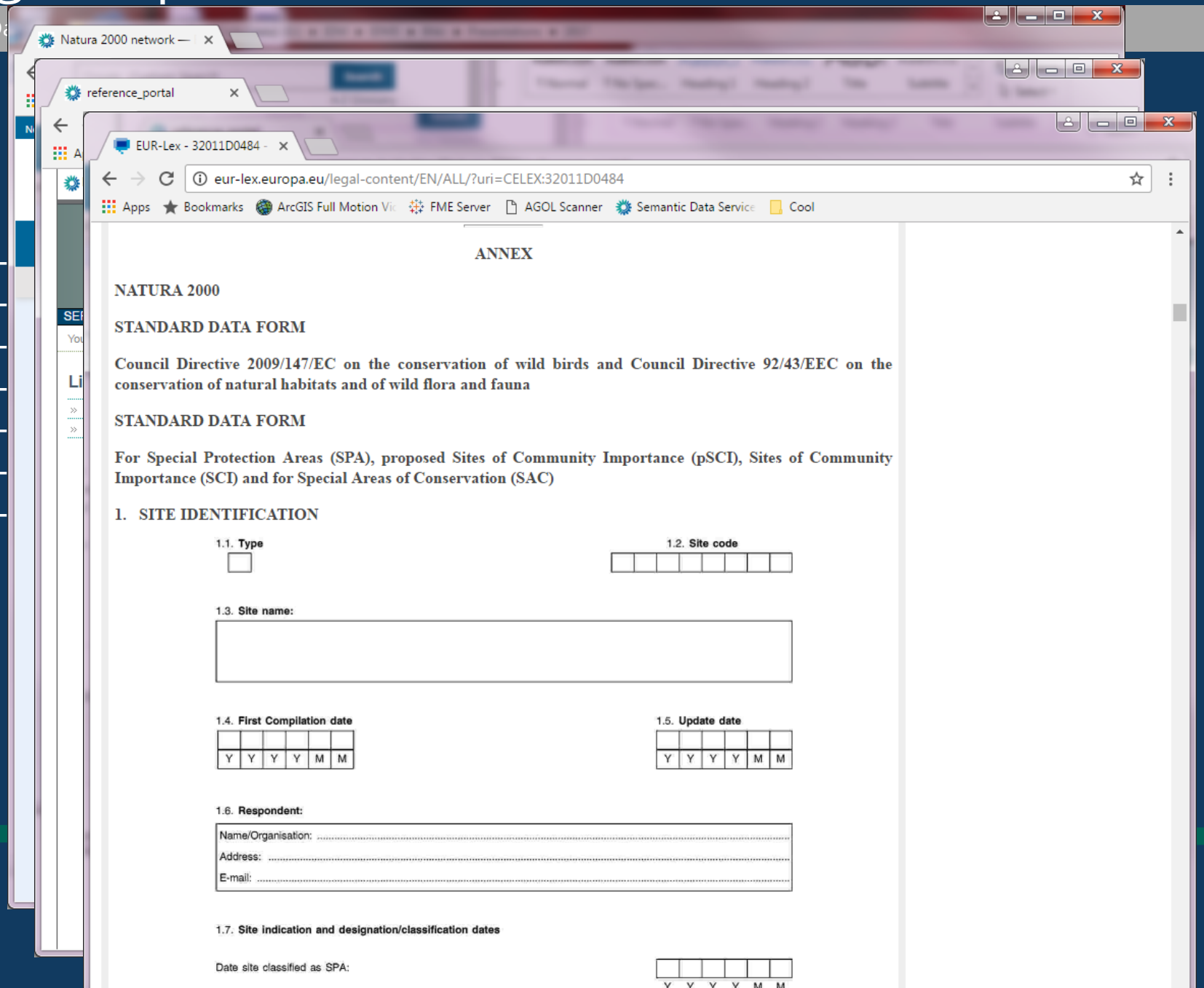
European products

Downloads,Web sites,applications

Producing European datasets

Data planning

Data flow preparation

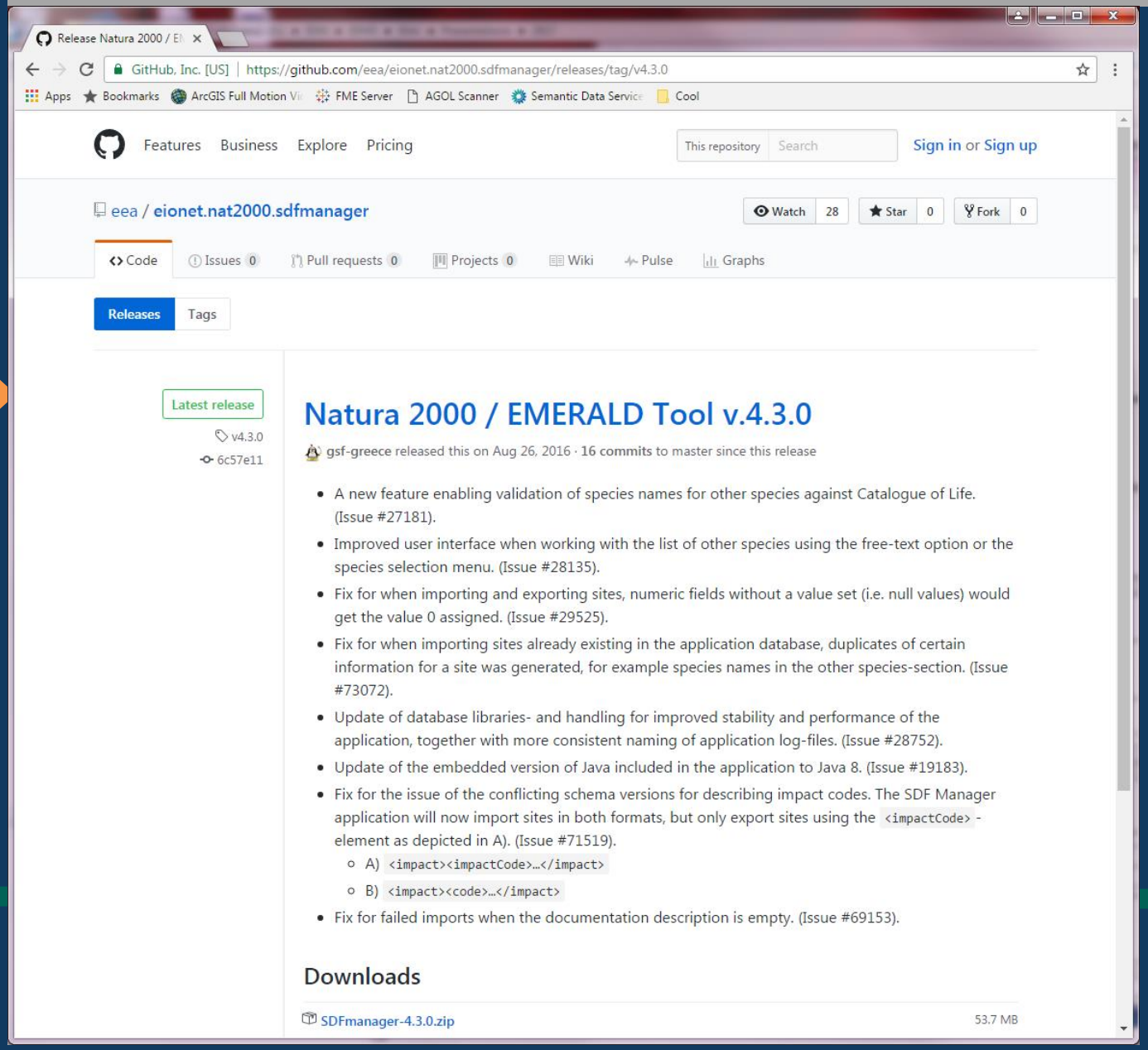


Producing European datasets

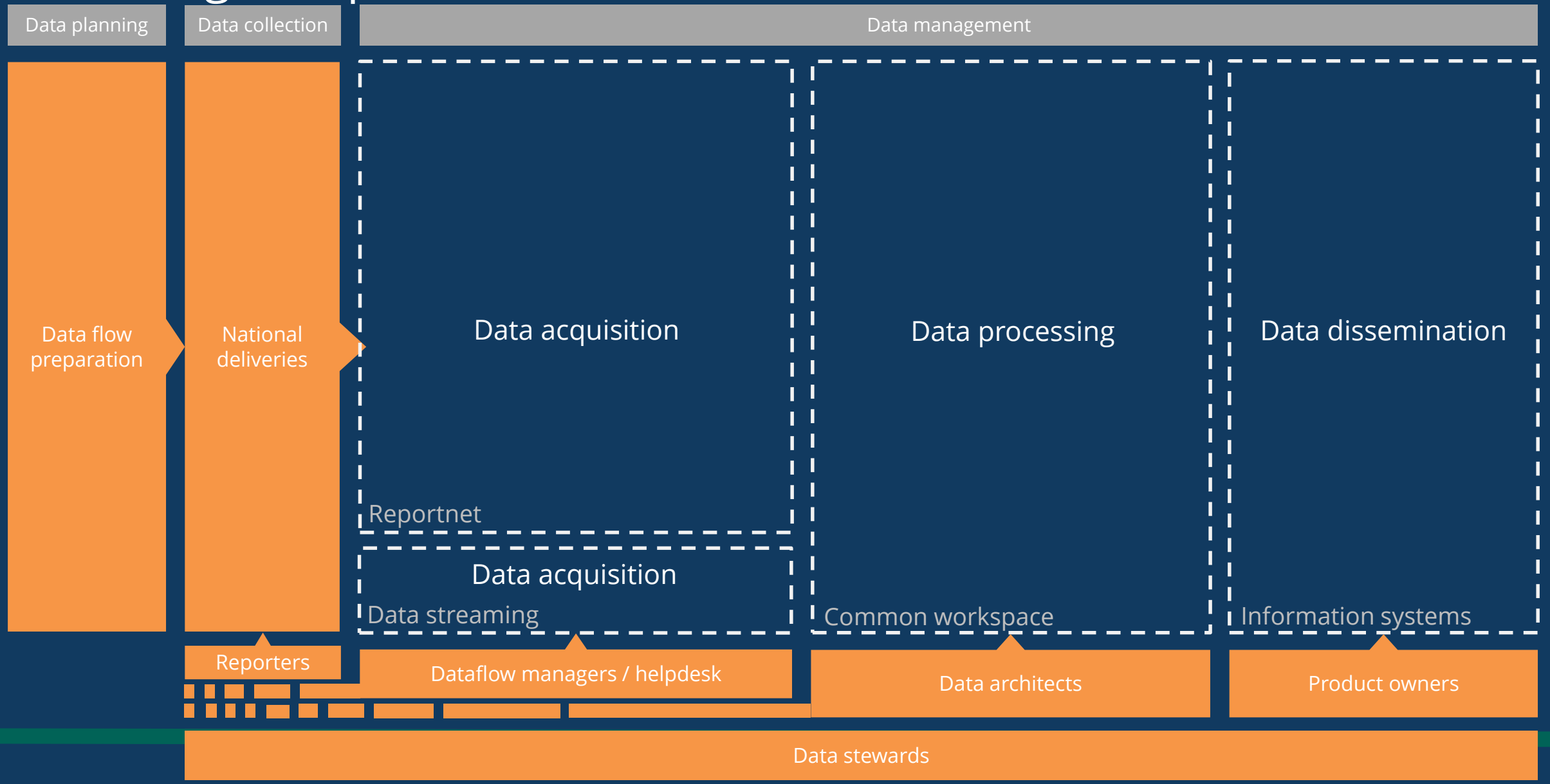
Data planning | Data collection | Data management

Data flow preparation

National deliveries

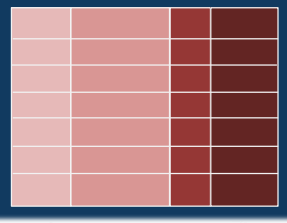


Producing European datasets

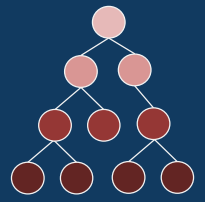


Examples of data structures

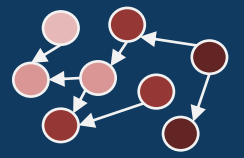
Relational

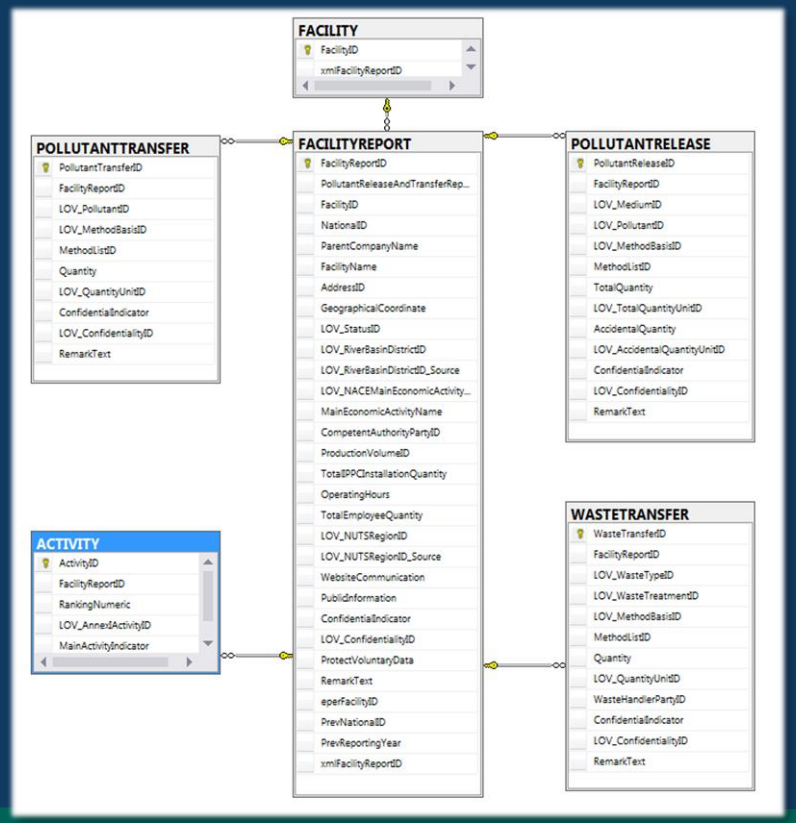


Document

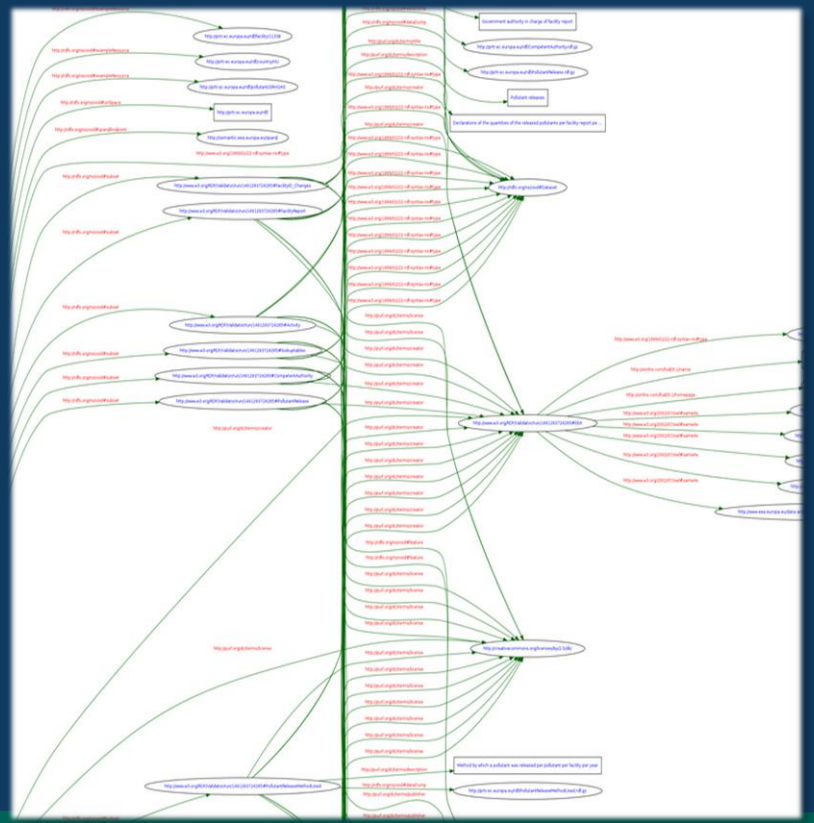


Graph






```
<!-- Created by EPERXMLCreator 0.1 13-04-2011 16:14:19 -->
<report xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="eper.xsd">
  <SubmissionDate>2010-03-01</SubmissionDate>
  <Country>BE</Country>
  <ReportYear>2004</ReportYear>
  <Facility>
    <NationalID>Bxl 04</NationalID>
    <ReportedYear>2004</ReportedYear>
    <ParentCompanyName>SOLVAY</ParentCompanyName>
    <FacilityName>PEPTISYNTHA</FacilityName>
    <Address>Rue de Ransbeek 310</Address>
    <PostCode>1120</PostCode>
    <GeographicCoordinateSystem>WGS84</GeographicCoordinateSystem>
    <Longitude>4.40302</Longitude>
    <Latitude>50.90434</Latitude>
    <NaceCode>24.14</NaceCode>
    <ProductionVolume/>
    <RegulatoryBodies/>
    <ContactName>Valérie Stoop</ContactName>
    <PhoneNumber>+32 2 775 75 39</PhoneNumber>
    <FaxNumber>+ 32 2 775 77 72</FaxNumber>
    <Email>vst@ibgebim.be</Email>
    <Activity code="4.5">
      <MainActivity>1</MainActivity>
      <NosePCode>107.03</NosePCode>
    </Activity>
    <Emission>
      <EmissionType>1</EmissionType>
      <PollutantName>Dichloromethane (DCM)</PollutantName>
      <Method>M</Method>
      <EmissionValue>3700</EmissionValue>
    </Emission>
  </Facility>
  <Facility>
    <NationalID>Bxl 06</NationalID>
    <ReportedYear>2004</ReportedYear>
    <ParentCompanyName>SIOMAB - Agence Régionale pour la Propreté</ParentCompanyName>
    <FacilityName>SIOMAB - Agence Régionale pour la Propreté</FacilityName>
    <Address>Quai Léon Monnoyer 8</Address>
    <PostCode>1000</PostCode>
    <GeographicCoordinateSystem>WGS84</GeographicCoordinateSystem>
    <Longitude>4.38088</Longitude>
    <Latitude>50.88414</Latitude>
    <NaceCode>90.00</NaceCode>
    <ProductionVolume/>
    <RegulatoryBodies/>
    <ContactName>Valérie Stoop</ContactName>
    <PhoneNumber>+32 2 775 75 39</PhoneNumber>
    <FaxNumber>+ 32 2 775 77 72</FaxNumber>
    <Email>vst@ibgebim.be</Email>
```

Producing European datasets

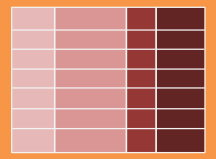
Data planning

Data collection

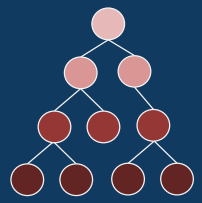
Data management

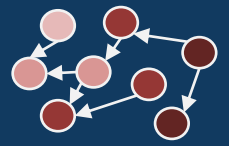
Data flow preparation

National deliveries



Data acquisition

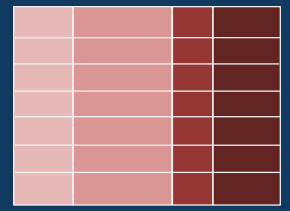


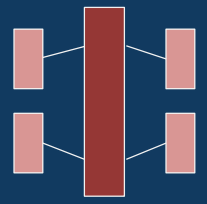


Reportnet

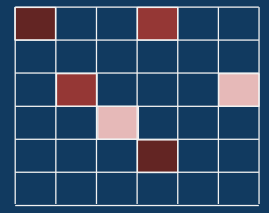
Data acquisition

Data streaming

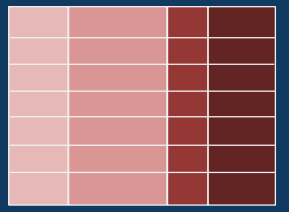




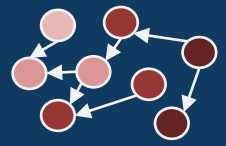
Data processing



Common workspace



Data dissemination



Information systems

Reporters

Dataflow managers / helpdesk

Data architects

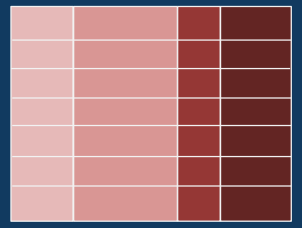
Product owners

Data stewards

European Environment Agency

Database types in EEA

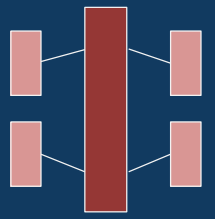
Relational



- Data is usually stored in row by row manner (row store)
- Standardized query language (SQL)
- Data model defined before you add data
- Joins merge data from multiple tables
- Results as tables

Pros: Mature ACID transactions with fine-grain security controls
Cons: Requires up front data modeling, does not scale well.
Example: Microsoft SQL Server, PostgreSQL, MySQL, SQLLite

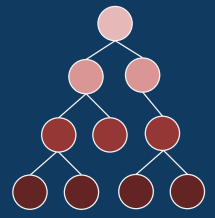
Analytical



- Based on “Star” schema with central fact table for each event.
- Optimized for analysis of read-analysis of historical data
- Common query language easy to implement.

Pros: Fast queries for large data
Cons: Not optimized for transactions and updates.
Example: Microsoft Analysis Server, OLAP, Tableau

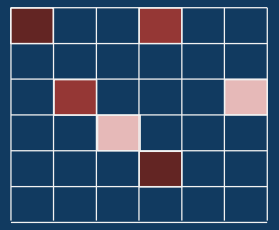
Document



- Data stored in nested hierarchies (XML, JSON)
- Logical data remains stored together as a unit
- Any item in the document can be queried
- Roughly equivalent to the programming concept of an object

Pros: No object-relational mapping layer, ideal for search
Cons: Complex to implement, strong contrast to relational databases
Example: Microsoft DocumentDB, Elastic search

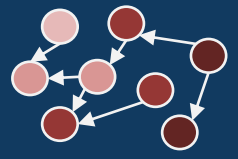
Grid - based



- Data stored in cells with a spatial position
- Transfers spatial data into tabular form
- Relates data that has only spatial relationships

Pros: Scalable and performing
Cons: Needs a spatial reference and from the same spatial scale.

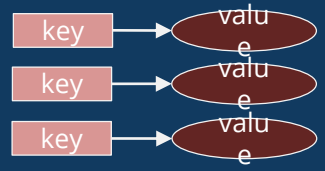
Graph



- Data stored in a series of nodes and properties
- Queries are really graph traversals
- Ideal when relationships between data is key. (ex. Social networks)

Pros: Fast network search, works with public linked data sets
Cons: Poor scalability when graphs don’t fit into RAM, specialized query language.
Example: Virtuoso

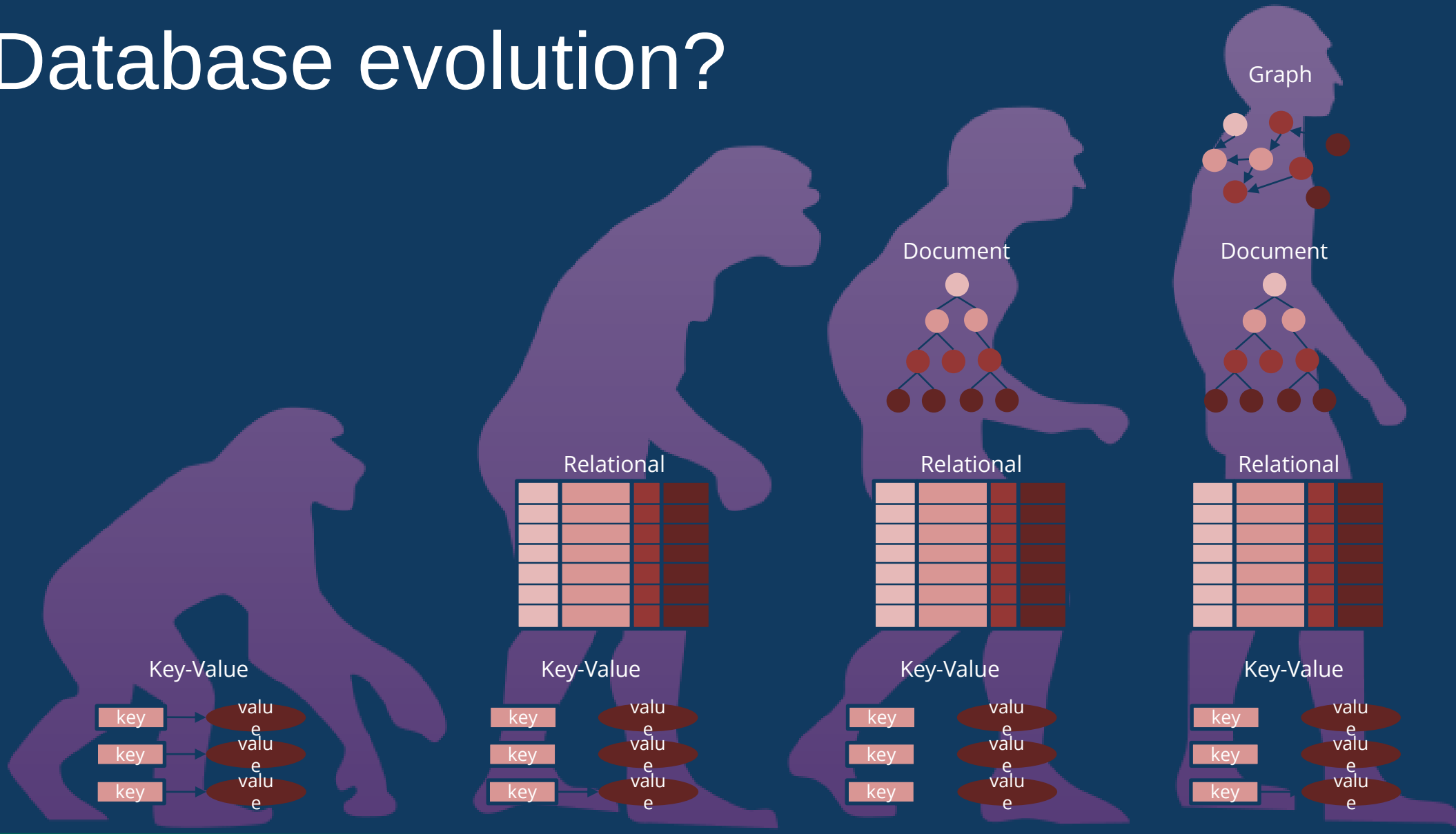
Key-Value

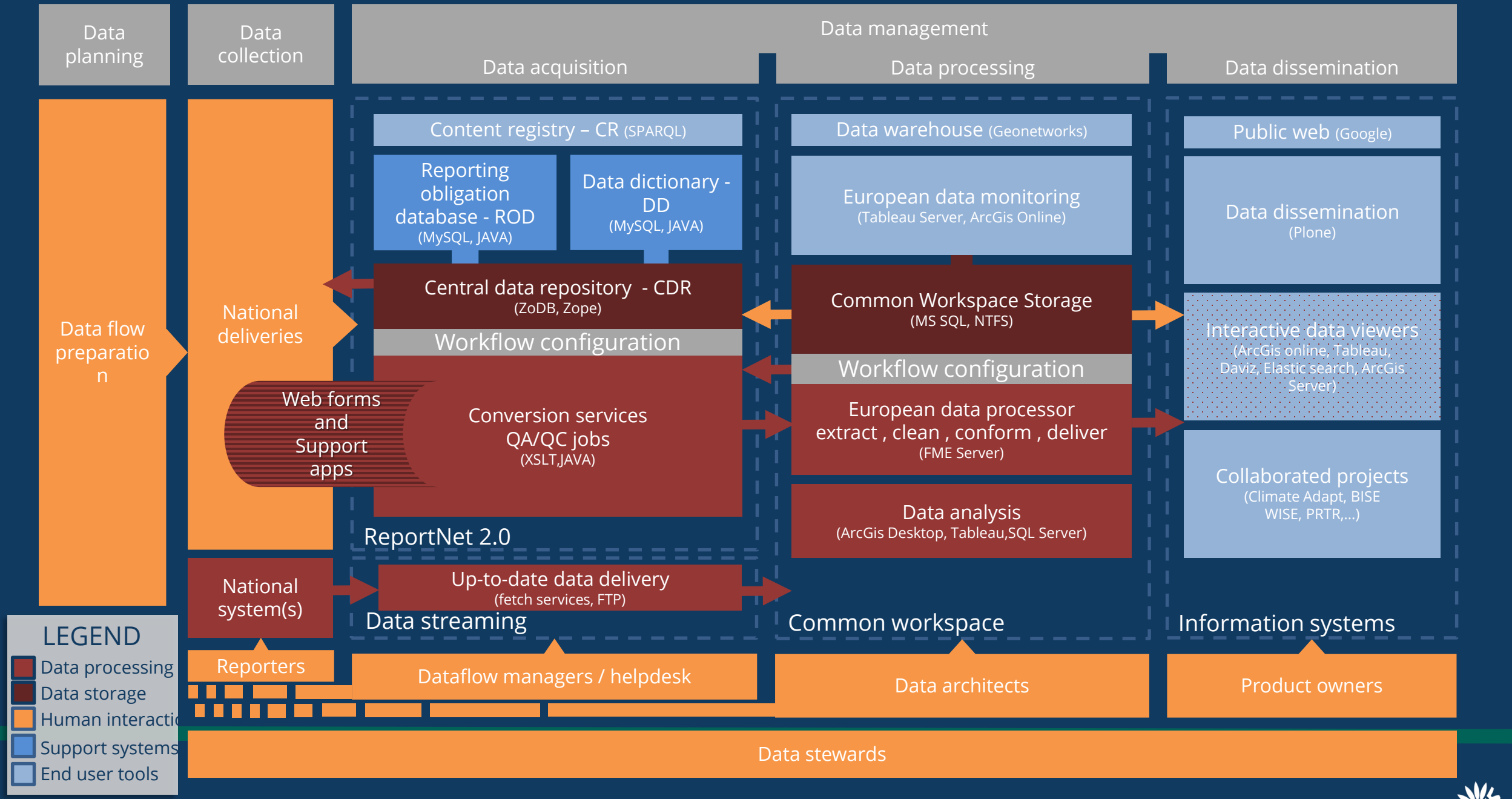


- Keys used to access opaque blobs of data
- Values can contain any type of data (images,video,files)

Pros: Scalable, simple API (put,get,delete)
Cons: No way to query based on the content of the value
Example: Azure Blob-storage

Database evolution?





European Environment Agency


QUESTIONS ?