

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

D

Data flow
preparation

ANNEX

NATURA 2000

STANDARD DATA FORM

Council Directive 2009/147/EC on the conservation of wild birds and Council Directive 92/43/EEC on the conservation of natural habitats and of wild flora and fauna

STANDARD DATA FORM

For Special Protection Areas (SPA), proposed Sites of Community Importance (pSCI), Sites of Community Importance (SCI) and for Special Areas of Conservation (SAC)

1. SITE IDENTIFICATION

1.1. Type

1.2. Site code

--	--	--	--	--	--	--	--

1.3. Site name:

1.4. First Compilation date

Y	Y	Y	Y	M	M

1.5. Update date

Y	Y	Y	Y	M	M

1.6. Respondent:

Name/Organisation:

Address:

E-mail:

1.7. Site indication and designation/classification dates

Date site classified as SPA:

Y	Y	Y	Y	M	M



Producing European datasets

Data planning

Data collection

Data management

Data flow
preparation

National
deliveries

The screenshot shows a GitHub repository page for 'Natura 2000 / EMERALD Tool v.4.3.0'. The page is viewed in a browser window with the URL <https://github.com/eea/eionet.nat2000.sdfmanager/releases/tag/v4.3.0>. The repository is owned by 'eea' and has 28 watchers, 0 stars, and 0 forks. The 'Releases' tab is selected, showing the latest release 'v4.3.0' (commit 6c57e11) released on August 26, 2016. The release notes list several updates and fixes, including validation of species names, UI improvements, and database updates. A 'Downloads' section at the bottom shows the file 'SDFmanager-4.3.0.zip' with a size of 53.7 MB.

Release Natura 2000 / EMERALD Tool v.4.3.0

ee / eionet.nat2000.sdfmanager

Code Issues 0 Pull requests 0 Projects 0 Wiki Pulse Graphs

Releases Tags

Latest release

v4.3.0
6c57e11

gsf-greece released this on Aug 26, 2016 · 16 commits to master since this release

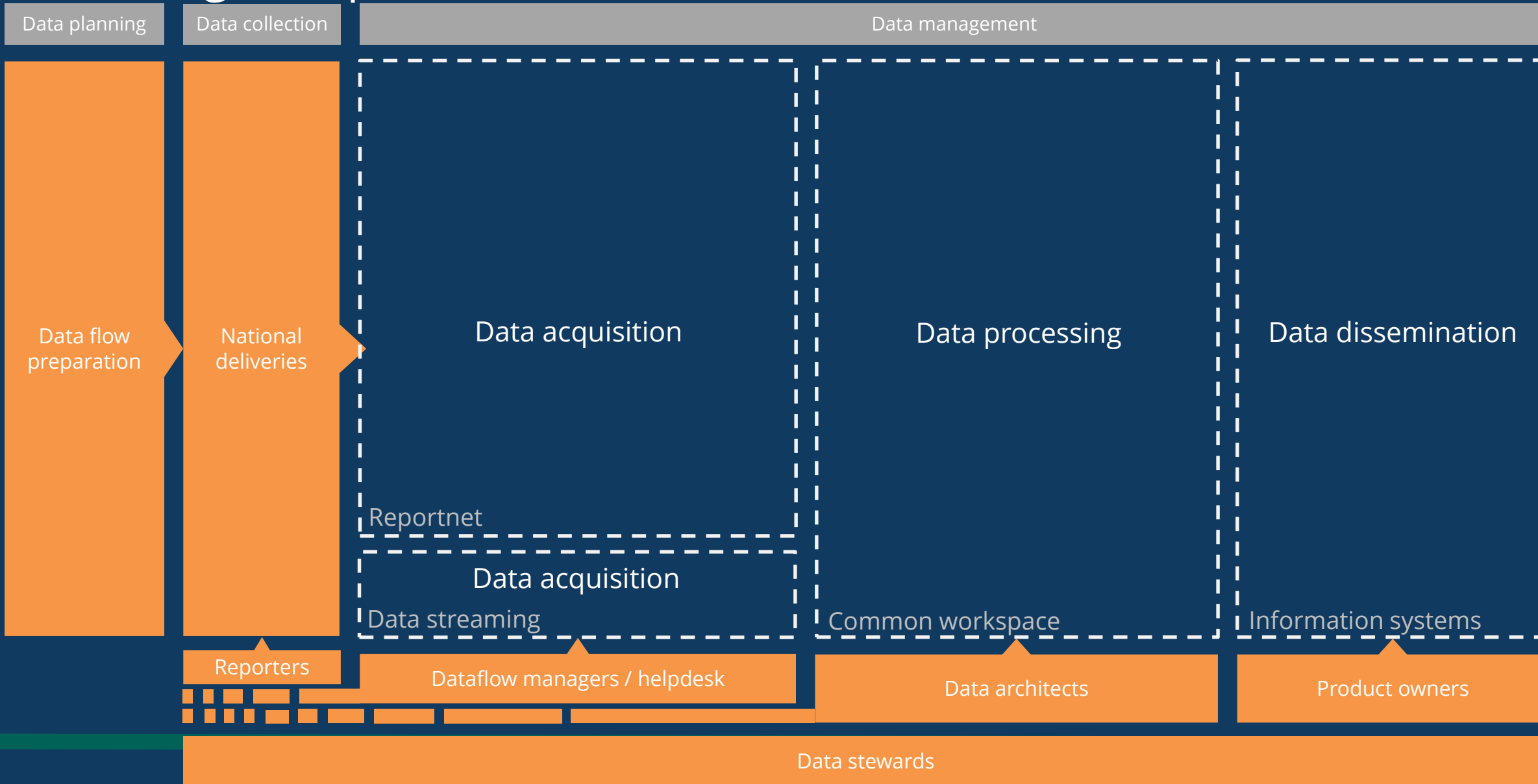
- A new feature enabling validation of species names for other species against Catalogue of Life. (Issue #27181).
- Improved user interface when working with the list of other species using the free-text option or the species selection menu. (Issue #28135).
- Fix for when importing and exporting sites, numeric fields without a value set (i.e. null values) would get the value 0 assigned. (Issue #29525).
- Fix for when importing sites already existing in the application database, duplicates of certain information for a site was generated, for example species names in the other species-section. (Issue #73072).
- Update of database libraries- and handling for improved stability and performance of the application, together with more consistent naming of application log-files. (Issue #28752).
- Update of the embedded version of Java included in the application to Java 8. (Issue #19183).
- Fix for the issue of the conflicting schema versions for describing impact codes. The SDF Manager application will now import sites in both formats, but only export sites using the `<impactCode>` - element as depicted in A). (Issue #71519).
 - A) `<impact><impactCode>...</impact>`
 - B) `<impact><code>...</impact>`
- Fix for failed imports when the documentation description is empty. (Issue #69153).

Downloads

SDFmanager-4.3.0.zip 53.7 MB

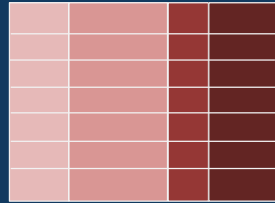


Producing European datasets

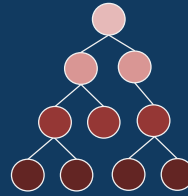


Examples of data structures

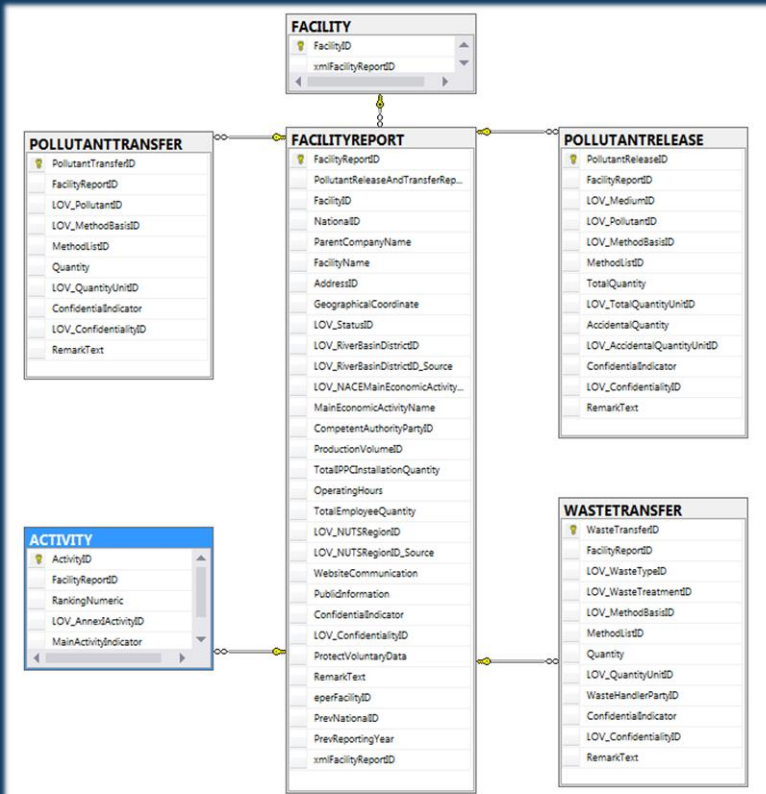
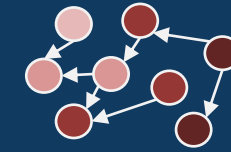
Relational



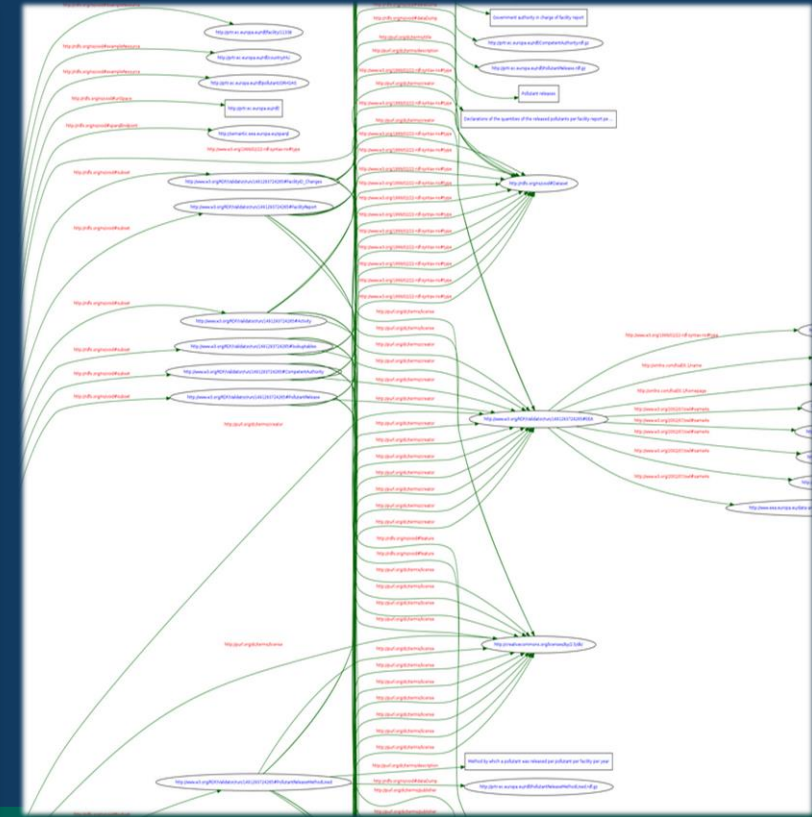
Document



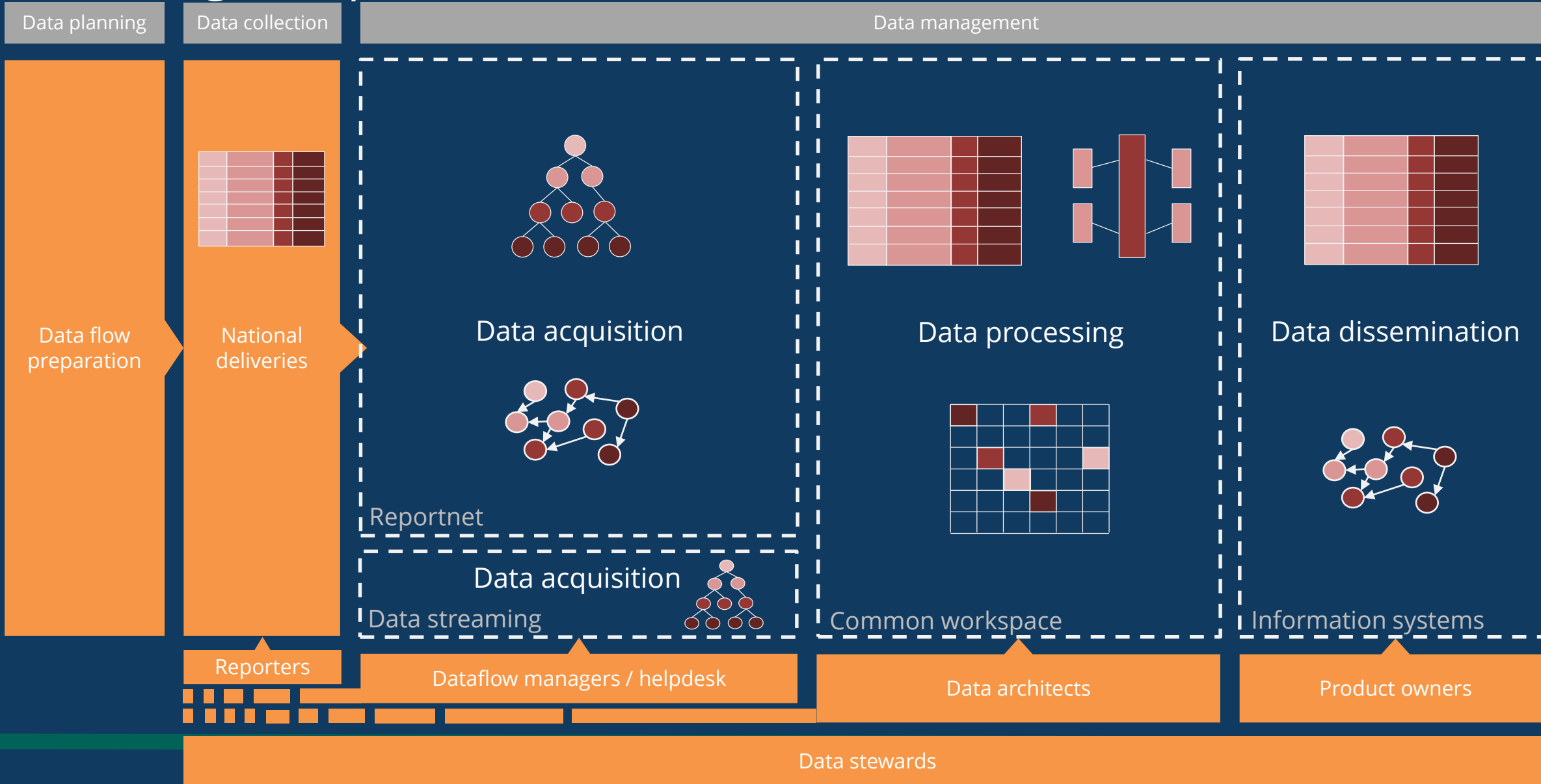
Graph



```
<!-- 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>Bx1 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 codes="4.5">
      <MainActivity>1</MainActivity>
      <NosePCCode>107.03</NosePCCode>
    </Activity>
    <Emission>
      <EmissionType>1</EmissionType>
      <PollutantName>Dichloromethane (DCM)</PollutantName>
      <Method>H</Method>
      <EmissionValue>3700</EmissionValue>
    </Emission>
  </Facility>
  <Facility>
    <NationalID>Bx1 06</NationalID>
    <ReportedYear>2004</ReportedYear>
    <ParentCompanyName>SIOHAB - Agence Régionale pour la Propreté</ParentCompanyName>
    <FacilityName>SIOHAB - 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>
  </Facility>
</report>
```

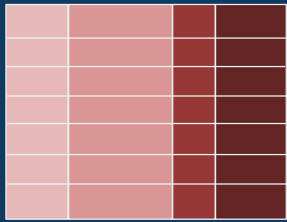


Producing European datasets



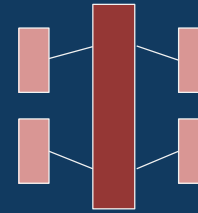
Database types in EEA

Relational



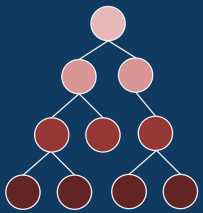
- Data is usually stored in row by row manner (row store)
 - Standardized query language (SQL)
 - Data model defined before you add data
 - Joins merge data from multiple tables
 - Results as tables
- Pros: Mature ACID transactions with fine-grain security controls
Cons: Requires up front data modeling, does not scale well.
Example: Microsoft SQL Server, PostgreSQL, MySQL, SQLite

Analytical



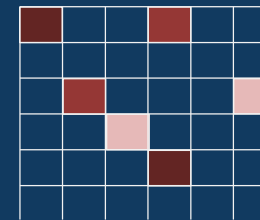
- Based on "Star" schema with central fact table for each event.
 - Optimized for analysis of read-analysis of historical data
 - Common query language easy to implement.
- Pros: Fast queries for large data
Cons: Not optimized for transactions and updates.
Example: Microsoft Analysis Server, OLAP, Tableau

Document



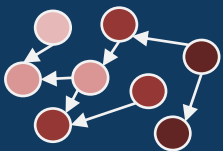
- Data stored in nested hierarchies (XML, JSON)
 - Logical data remains stored together as a unit
 - Any item in the document can be queried
 - Roughly equivalent to the programming concept of an object
- Pros: No object-relational mapping layer, ideal for search
Cons: Complex to implement, strong contrast to relational databases
Example: Microsoft DocumentDB, Elastic search

Grid - based



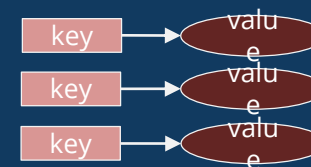
- Data stored in cells with a spatial position
 - Transfers spatial data into tabular form
 - Relates data that has only spatial relationships
- Pros: Scalable and performing
Cons: Needs a spatial reference and from the same spatial scale.

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

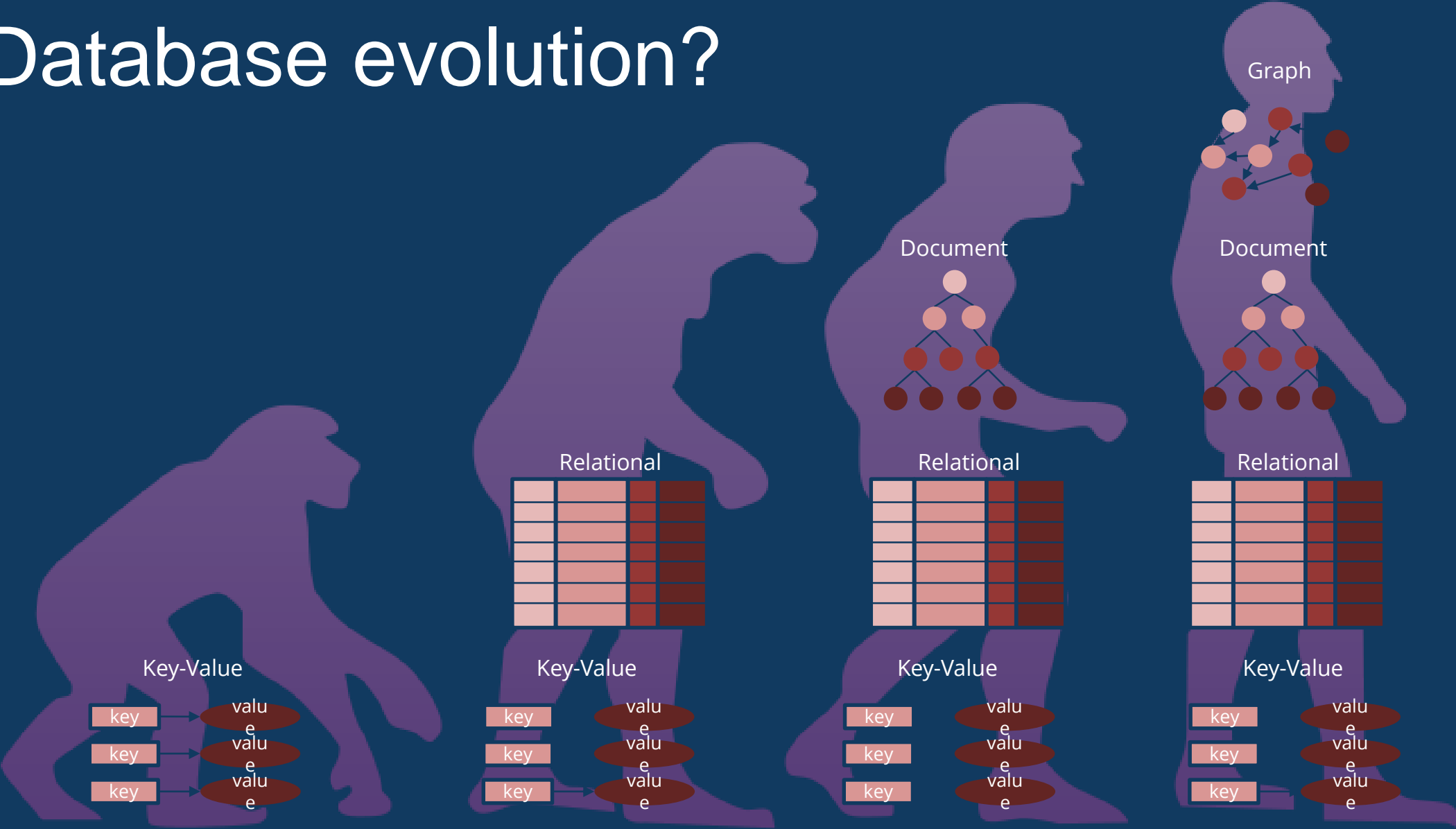
Key-Value

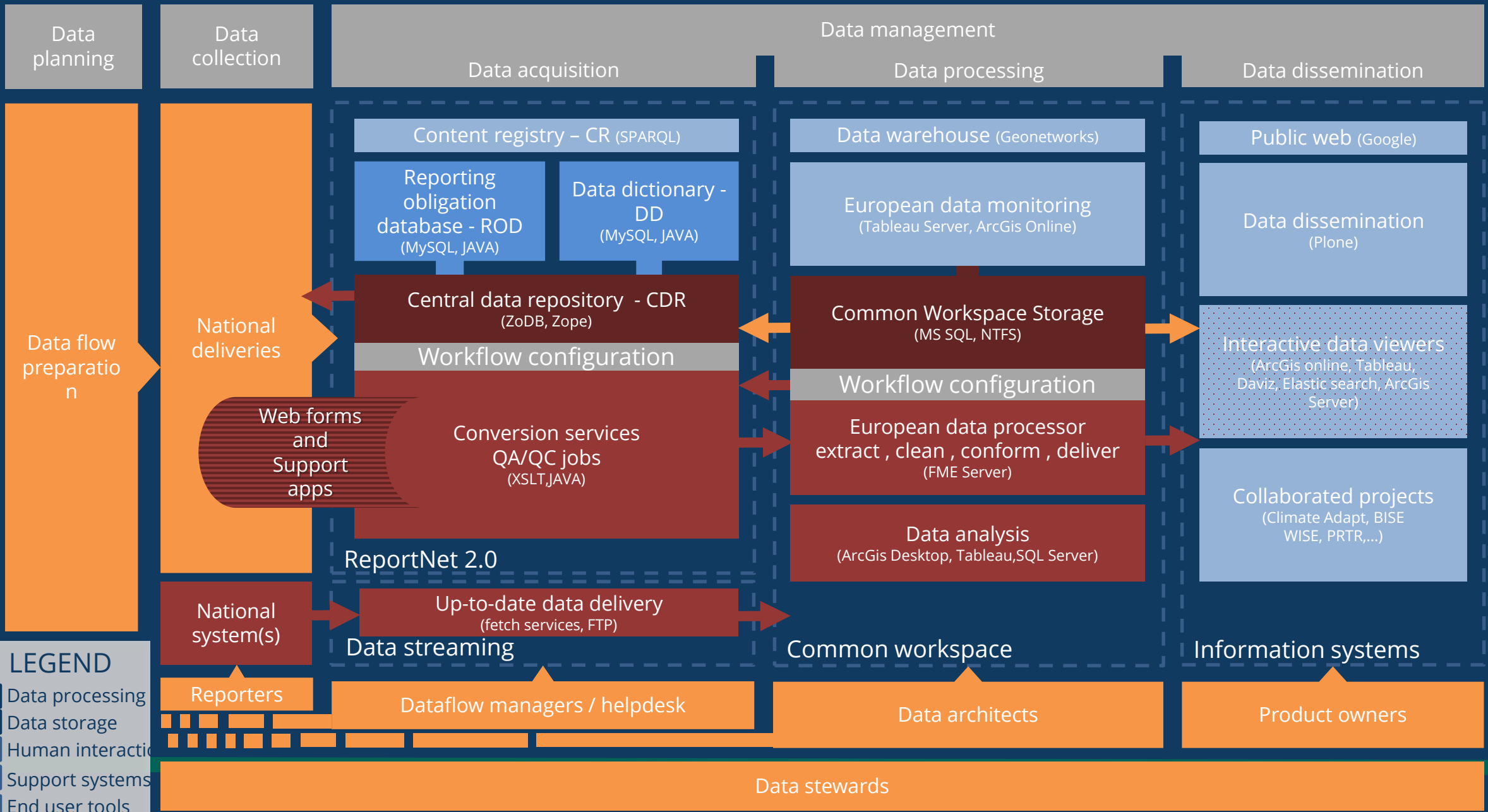


- Keys used to access opaque blobs of data
 - Values can contain any type of data (images, video, files)
- Pros: Scalable, simple API (put, get, delete)
Cons: No way to query based on the content of the value
Example: Azure Blob-storage



Database evolution?







QUESTIONS ?

Jan Bliki, Copenhagen 19-21 April 2017