



# Data harmonisation processes in HUMBOLDT

Dr. Christine Giger

HUMBOLDT Scientific Manager

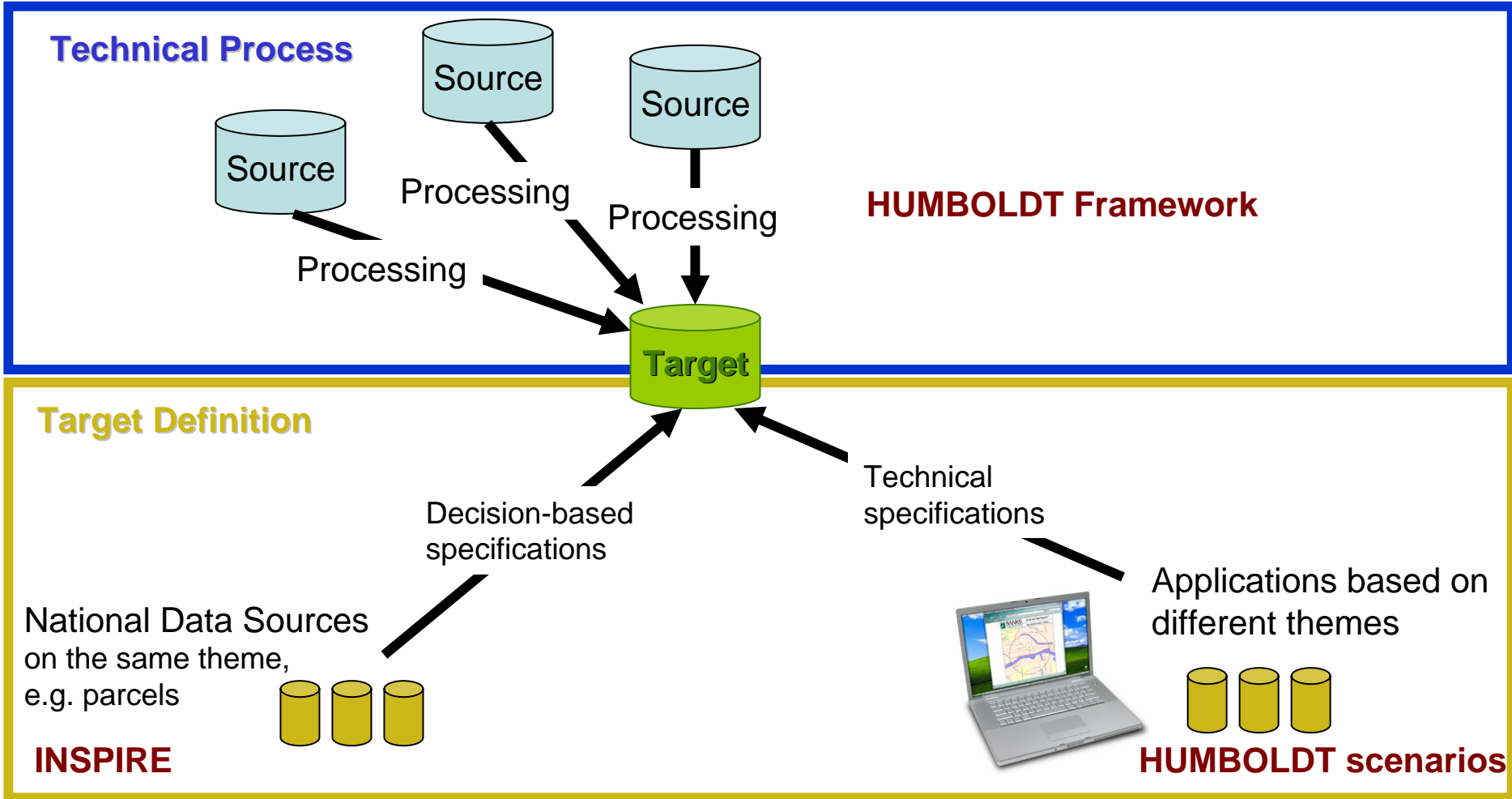
## Overview

- ▣ Data harmonisation in HUMBOLDT
- ▣ Results of the state of the art analysis
- ▣ Current results
- ▣ Next steps



# Data Harmonisation

 Data harmonisation processes



## What is data harmonisation?

### **INSPIRE**

“process of developing a common set of **data product specifications** in a way that allows the provision of access to **spatial data** through **spatial data services** in a representation that allows for combining it with other harmonised data in a coherent way.”

NOTE: This includes agreements about coordinate reference systems, classification systems, application schemas, etc.

### **HUMBOLDT**

“creating the possibility to **combine data from heterogeneous sources** into **integrated, consistent and unambiguous information products**, in a way that is of **no concern to the end-user**” (A 3.5-D1)



# Results of the state of the art analysis

254 documents, which were considered for SoA

	tools	user groups and needs	software architecture	data harmonisation and management
Number of investigated documents	48	25	52	41
INSPIRE state of play reports	30	0	28	30
<b>Documents per Activity</b>	<b>78</b>	<b>25</b>	<b>80</b>	<b>71</b>

## ■ User groups, their needs and requirements

### Approach for the first phase:

#### ■ Literature

#### ■ Goals:

- Characterisation of user groups and roles as well as application domains
- Basic harmonisation requirements

### Approach for the second phase:

#### ■ Questionnaires and interviews

#### ■ Goals:

- detailed information from the Scenario users
- more information on users from further application domains
- Detailed information on uses cases and working solutions
- better insight into interoperability and harmonisation requirements

## Results of the second phase

The top five requirements are:

- ▣ Problems with **different data formats** → provide interoperable access to heterogeneous data sources
- ▣ Problems with **different data models** → provide solutions for data model harmonisation
- ▣ Problems due to **missing / inconsistent / outdated metadata** → provide solutions to search for and possibly capture / publish metadata
- ▣ Problems with the **meaning of objects, i.e. semantics** → provide solutions like application domain dictionaries and thesauri
- ▣ Problems with **different coordinate reference systems** → provide Coordinate Transformation Service(s)

## Harmonisation approaches and best practices

- European projects as well as national projects and initiatives were investigated
- Focus on working solutions for SDI and their evaluation

## Data harmonisation components

### Data harmonisation components (INSPIRE, RISE)

(A) INSPIRE Principles	(B) Terminology	(C) Reference model
(D) Rules for application Schemas and feature catalogues	(E) Spatial and temporal aspects	(F) Multi-lingual text and cultural adaptability
(G) Coordinate referencing and units model	(H) Object referencing modelling	(I) Data translation model/guidelines
(J) Portrayal model	(K) Identifier Management	(L) Registers and registries
(M) Metadata	(N) Maintenance	(O) Quality
(P) Data Transfer	(Q) Consistency between data	(R) Multiple representations
(S) Data capturing	(T) Conformance	

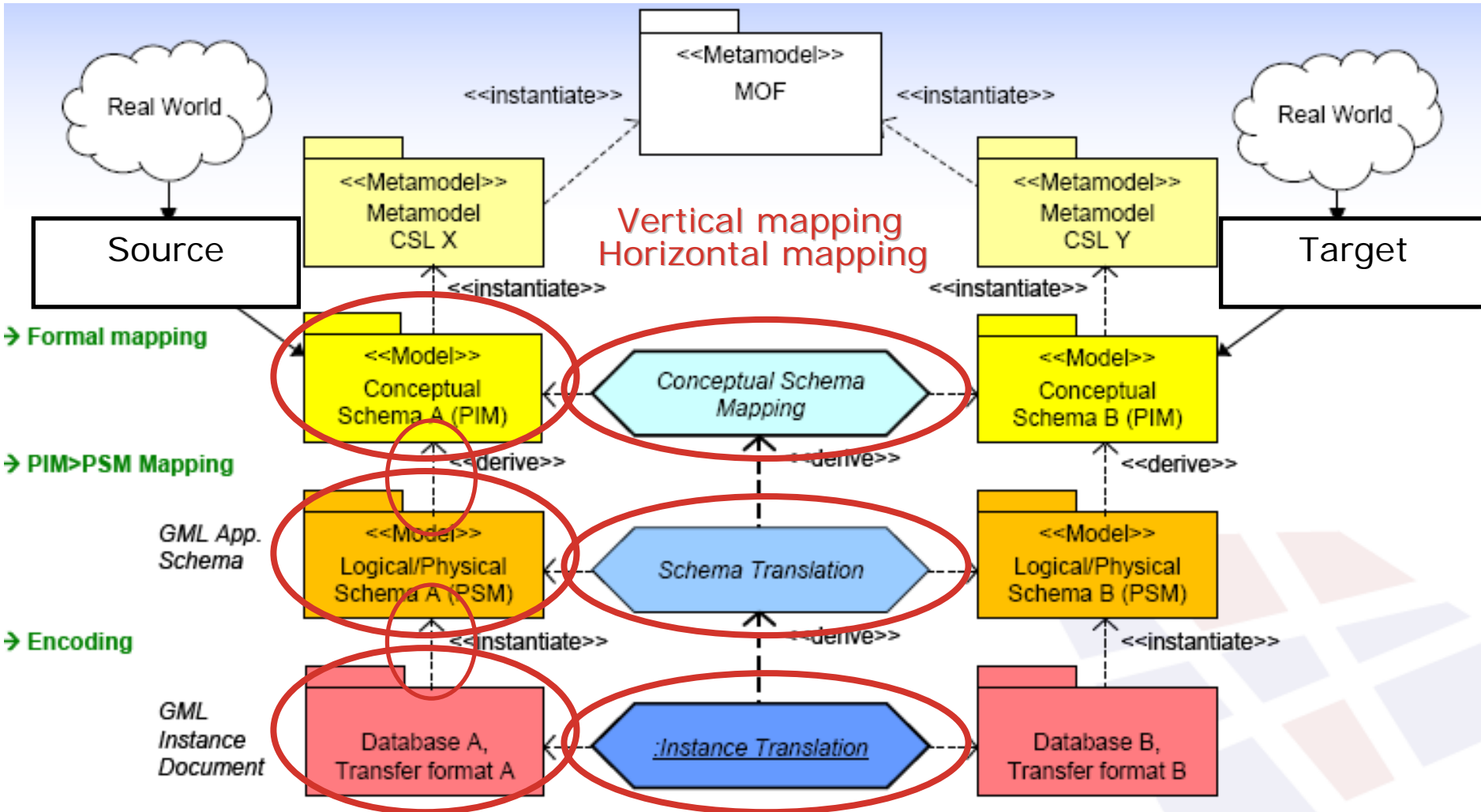
HUMBOLDT

+ computational models (process models), their constraints and parameters as 21st component

Obviously basic: **data model harmonisation**  
(schema mapping)

OMG (Object Management Group):

- ▣ Architecture for open distributed systems (e.g. used by OGC)
- ▣ Basis for state-of-the-art, mainstream IT: Model-Driven Approach (MDA)
  - PIM: Platform Independent Model
  - PSM: Platform Specific Model
  - **UML: Unified Modeling Language**
  - MOF: Meta Object Facility (basis of UML and other specialised description languages)
  - **MDA used as basis for interface standardization in ISO/TC211, CEN/TC287, OGC**



## Tools to support the harmonisation process

169 tools have been evaluated:

- ▣ 15% data modelling, schema translation
- ▣ 30% data and metadata discovery and management
- ▣ 55% publication, access and visualisation

Fully developed solutions as well as prototypes or demonstrator applications have been considered

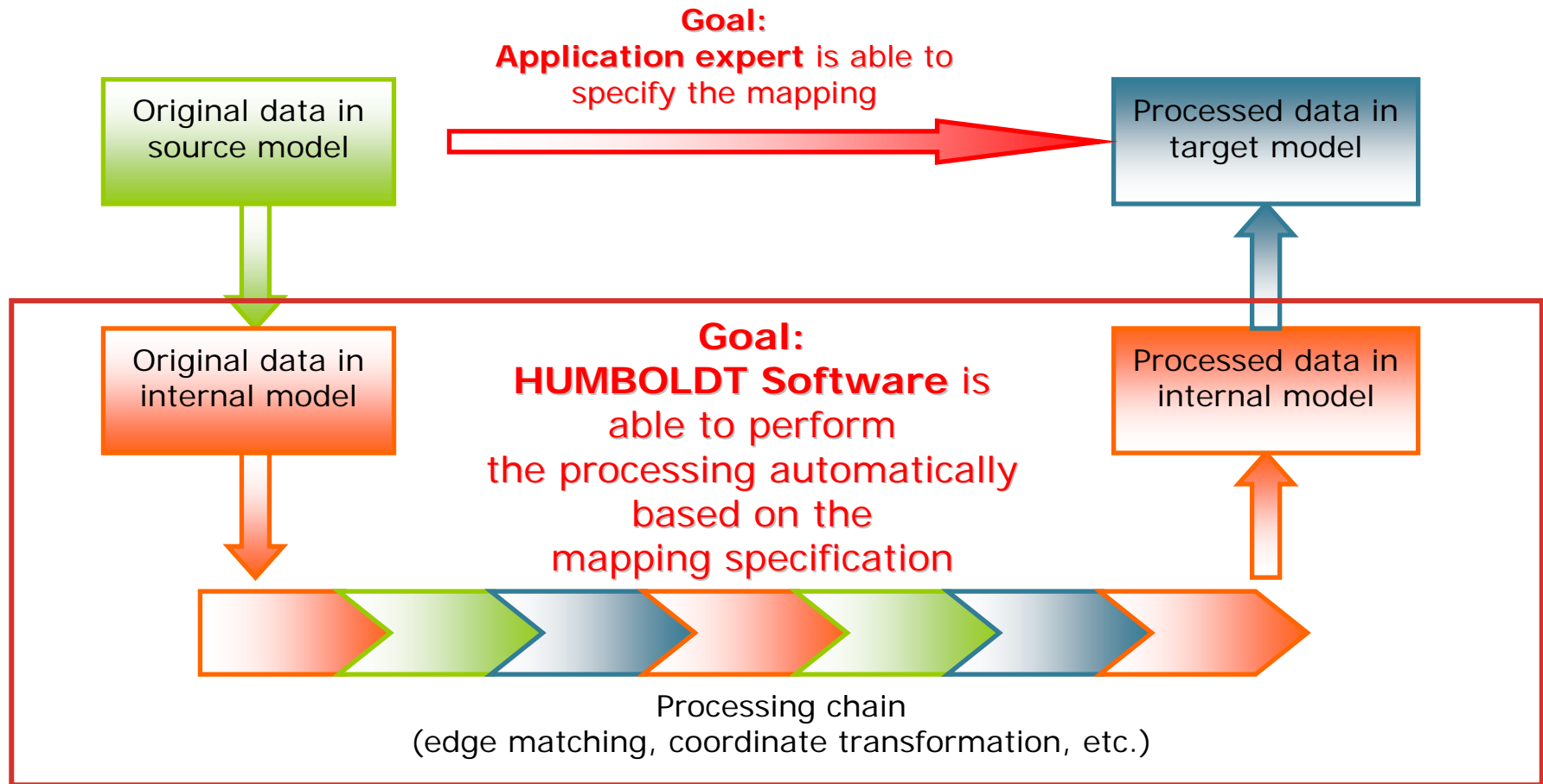
## Results of the State of the Art Analysis: Selection of open issues

- Existing tools do not support all **mappings and steps in a harmonisation workflow**
- Existing tools do not work **service-oriented** or cannot be integrated easily in a service-oriented architecture or framework
- Existing tools do not fully support testing whether a **model or data set is semantically consistent and correct**
- Lack of tools and methods to help in the process of **capturing the 'domain knowledge' and terminology** in an application domain (e.g. Humboldt scenario)



# Current results

## Technical processing of data



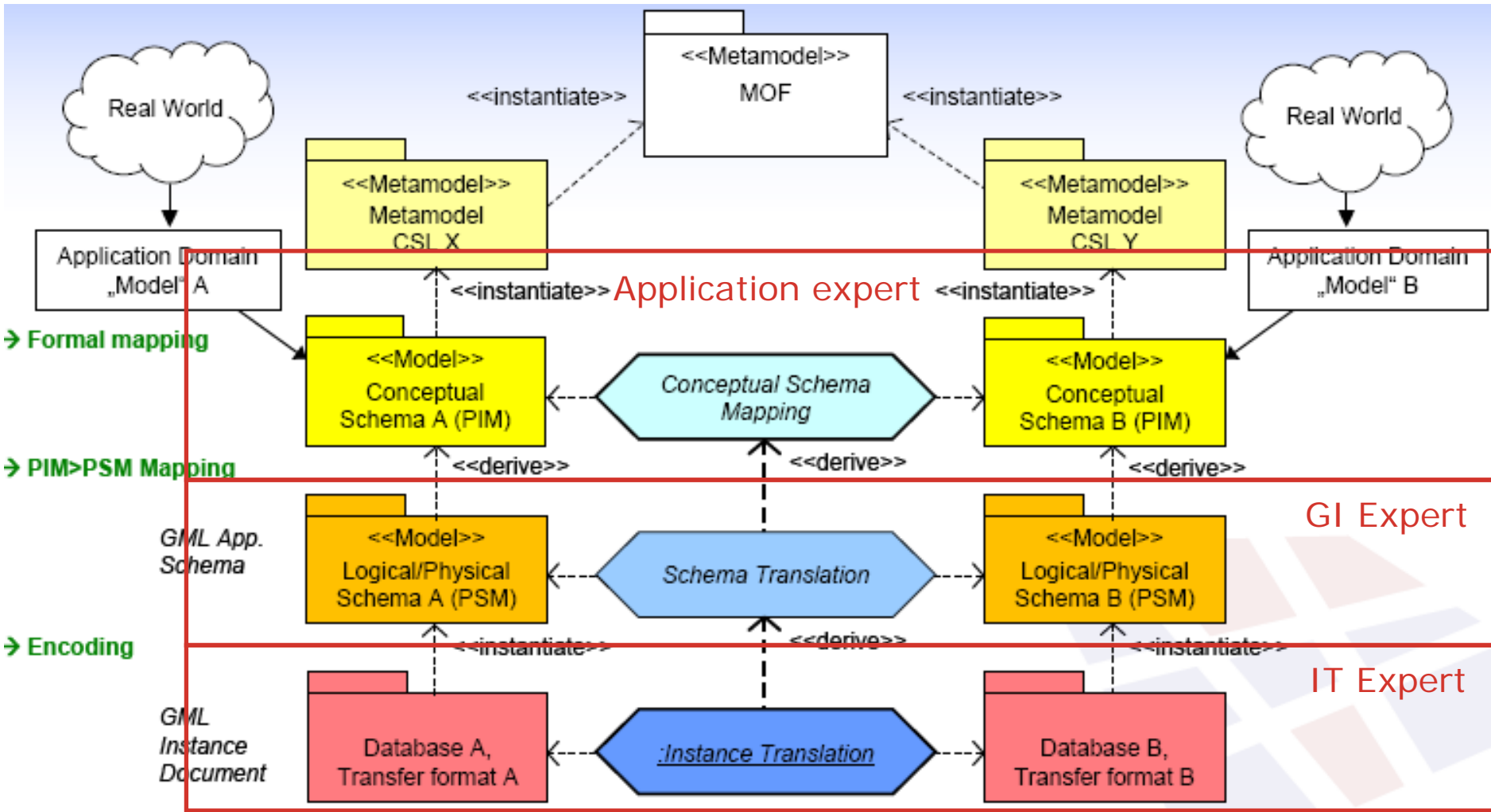
## ■ Prerequisites for implementing this process

Data specification containing the source and target model specification:

- Humans are able to easily specify, read and understand it
- Computers are able to process it

Mapping specification between source and target model

- Humans are able to easily specify, read and understand it
- Computers are able to process it



## ■ Prerequisites for implementing this process

Tools have to involve different categories of experts in different ways:

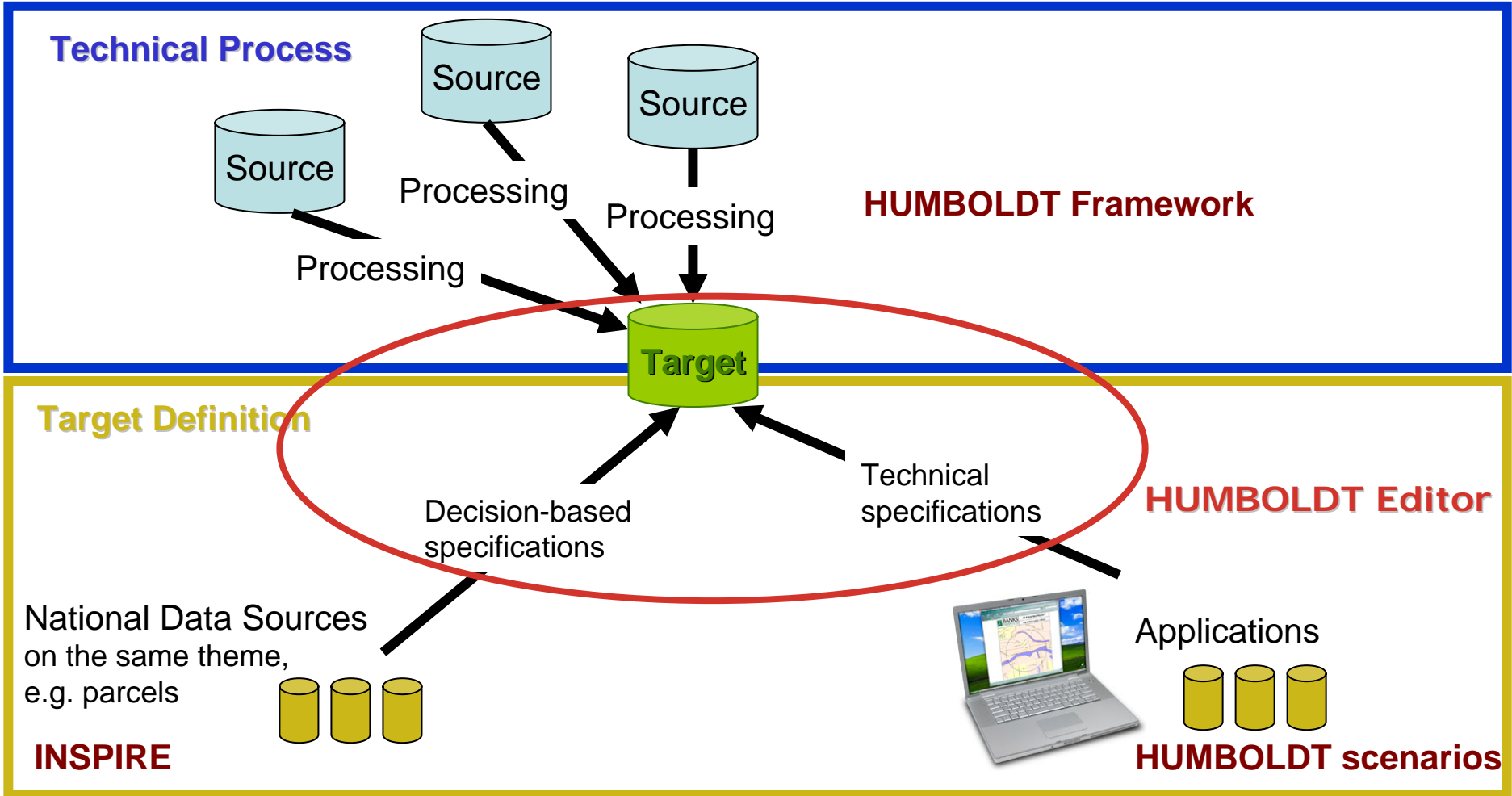
- **Application experts** who have to **express their knowledge** in the data and mapping specifications of the **horizontal mapping** between two models on the platform-independent (conceptual) level
- **GI experts** who have to **implement their GI knowledge** in the **vertical mapping** of platform-independent to platform-dependent model, so that this can be carried out automatically based on the specification of the application experts
- **IT experts** who have to **implement their IT knowledge** in the encoding/decoding processes (**vertical mapping**) in order to carry out these mappings automatically based on the results of the previous steps

## ▣ Lessons learned so far

- ▣ The HUMBOLDT framework has to **support the process of data specification** in order to obtain human-readable and computer-processable representations
- ▣ This will also support the **target definition process**, which is a “non-technical”, decision-based harmonisation process in the HUMBOLDT scenarios (and in INSPIRE)

Intermediate result: the **HUMBOLDT editor**

■ Data harmonisation processes



## The HUMBOLDT editor

Is an easy-to-use editor for the specification of data descriptions

- ▣ Descriptions are human-readable and computer-processable

Is less than a “usual” UML editor:

- ▣ does not offer any functionality, that is not directly related spatial data specifications

Is more than a “usual” UML editor:

- ▣ It offers additional features, e.g.
  - Relevant for the application experts (users) and developers:
    - a **basic data model** containing basic spatial data types,
  - Relevant for the GI and IT experts (HUMBOLDT developers):
    - **Support of vertical mapping: serialisation** possible to any transfer standard or other representation (e.g. XML, **GML**, **ISO19131**, ...)
    - **Support of horizontal mapping: model-to-model translation** options based on MDA, e.g. direct export to an internal model without previous serialisation



## Next steps

## The HUMBOLDT editor

- ▣ Is currently tested in an alpha version in the HUMBOLDT scenarios
- ▣ Will soon be available in a beta version for user@humboldt registered users  
([http://www.esdi-humboldt.eu/get\\_involved/user.html](http://www.esdi-humboldt.eu/get_involved/user.html))
- ▣ Feedback appreciated!

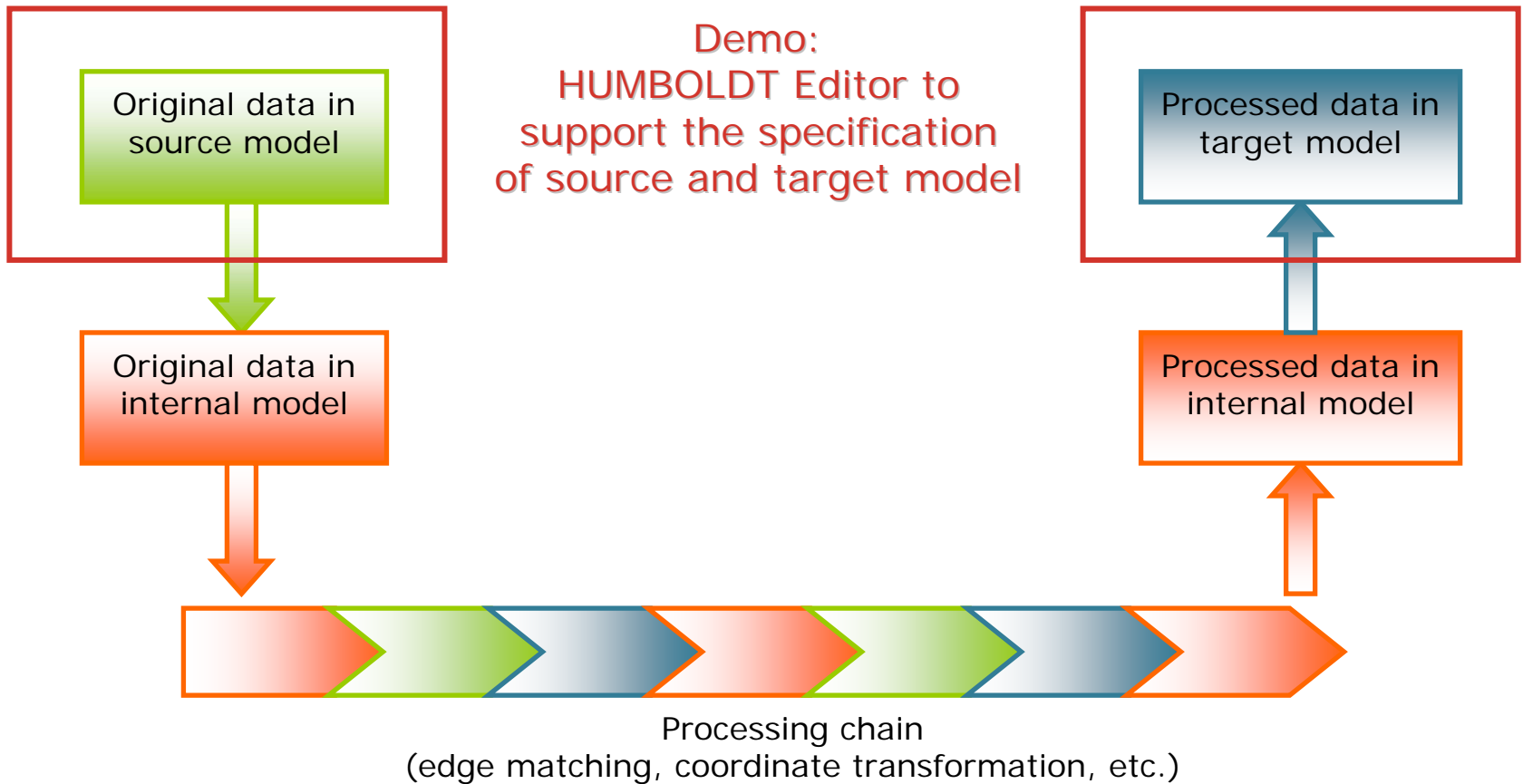
## The HUMBOLDT framework

### Framework version 1 is currently under development

- After extensive testing of existing tools new approaches are developed
- Further tools for horizontal and vertical mapping of data representations will be available
- The current specification is tested and validated against the end user requirements in the HUMBOLDT scenarios
- For one of the scenarios (HS-Protected Areas) a demo of the HUMBOLDT framework will be available soon

### Please register for our Newsletter!

## Technical processing of data




## Example

### Creation of a simple hydrographic model

#### Waterbodies

- Rivers
- Lakes

#### attached network of sensors



Thank you for your attention!  
Any questions?

Christine Giger

cgiger@ethz.ch

www.esdi-humboldt.eu