



HUMBOLDT Results on Principles of Data Modeling and Harmonisation

Christine Giger
Scientific Manager
HUMBOLDT Open Day
Lisbon, Portugal, September 28, 2010



“Non-technical” Results of HUMBOLDT

“For Knowledge itself is power”

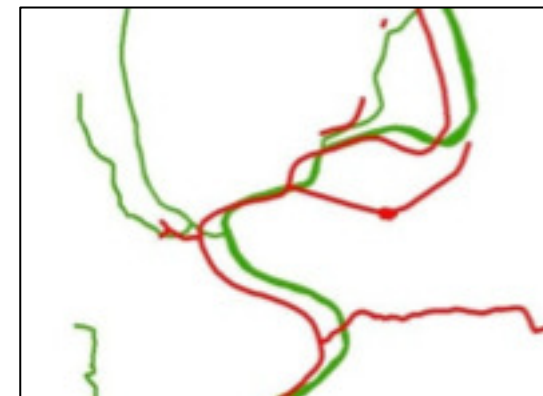
Francis Bacon, 1598

Knowledge and Experiences gathered in HUMBOLDT

- ▣ **Best practices and guidelines** for the process of harmonisation of data models (schema translation)
- .
- ▣ **Testing and evaluation of existing approaches and tools** for data model creation (data specification) and harmonisation

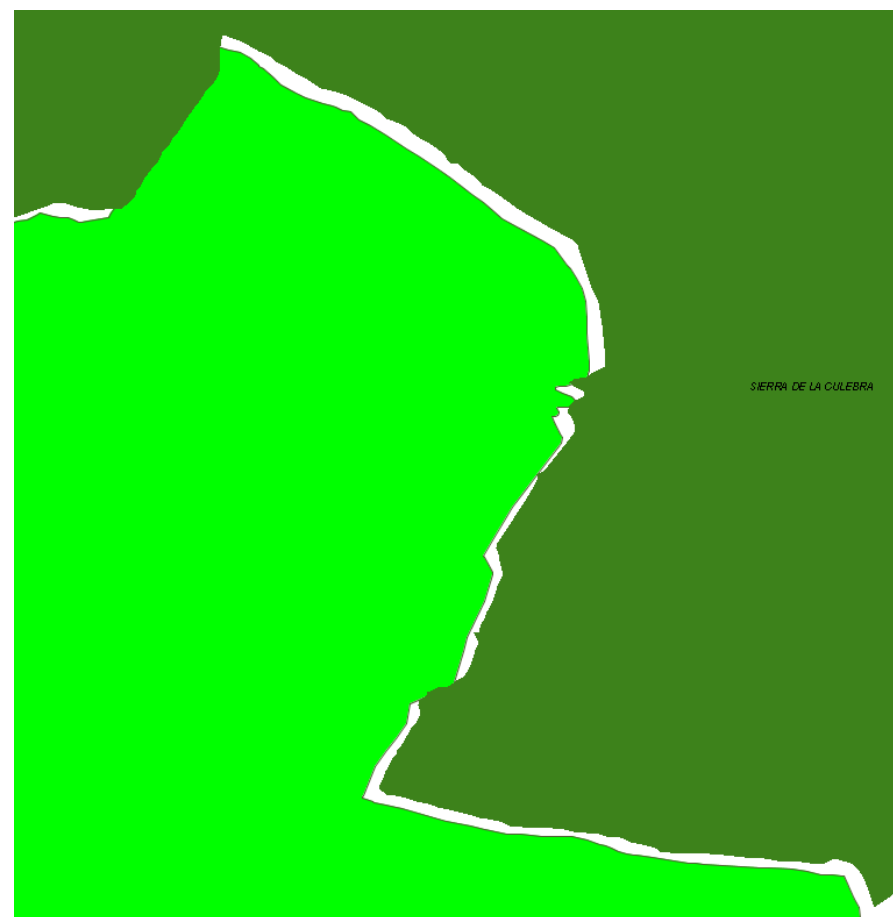
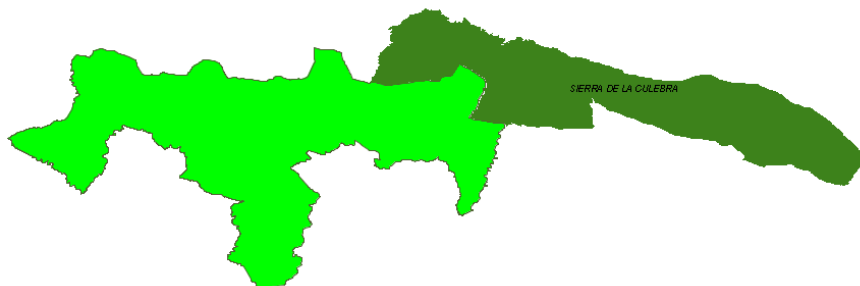


“The Problem”: Geodata Harmonisation Example



■ Data Harmonisation Issues in HUMBOLDT Scenarios

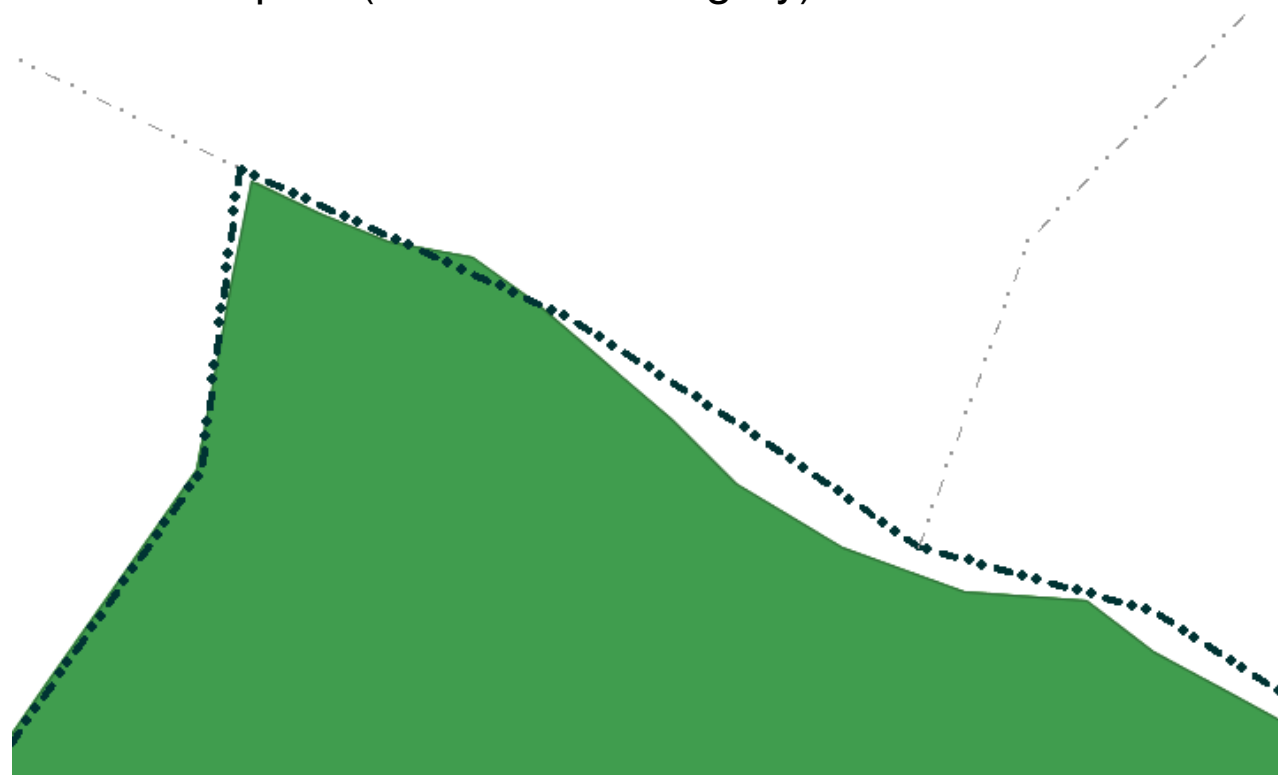
- Spatial (in-)consistency:
A protected area in Portugal
(shapefile light green) and
a protected area in Spain
(wms dark green) that
share a border



■ Data Harmonisation Issues in HUMBOLDT Scenarios

■ Spatial (in-)consistency:

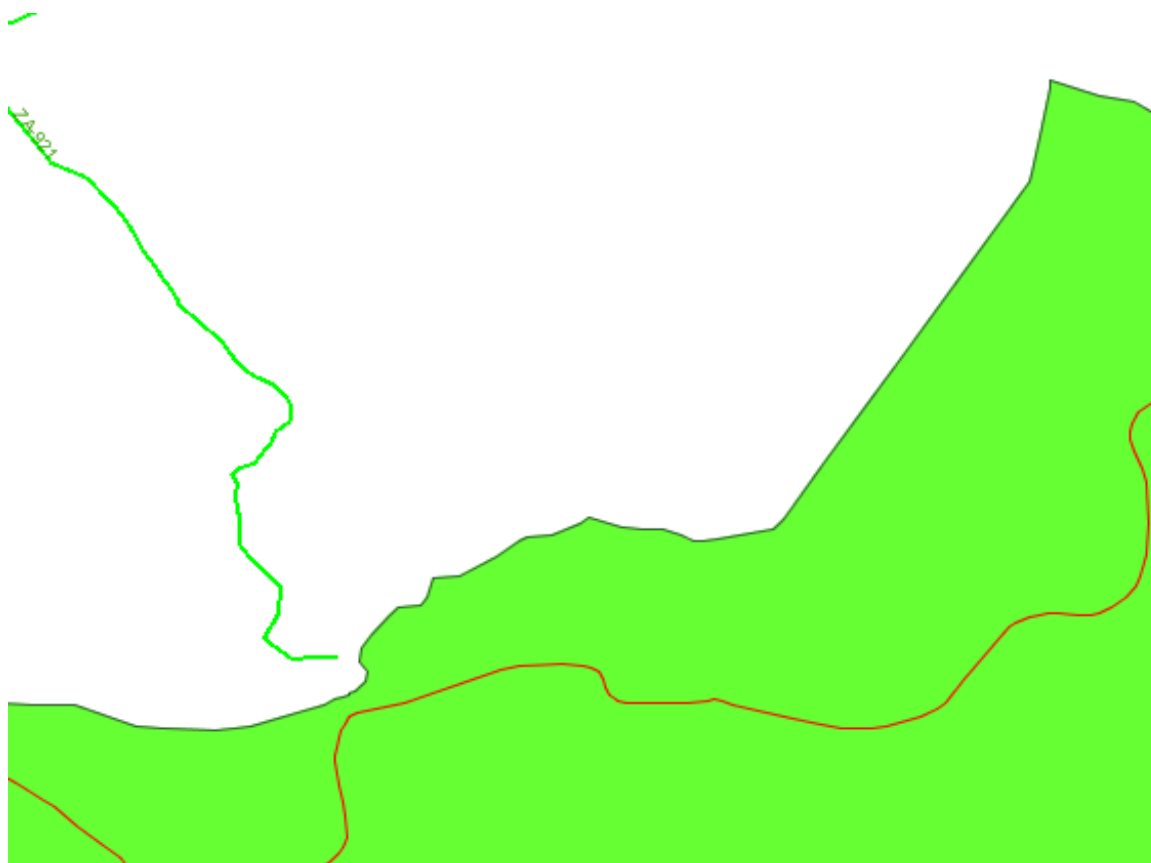
A protected area in Portugal (shapefile dark green) and administrative boundaries in Spain (wms black and grey)



■ Data Harmonisation Issues in HUMBOLDT Scenarios

■ Spatial (in-)consistency:

A road in Portugal
(shapefile red line) and
a road in Spain
(wms green line)

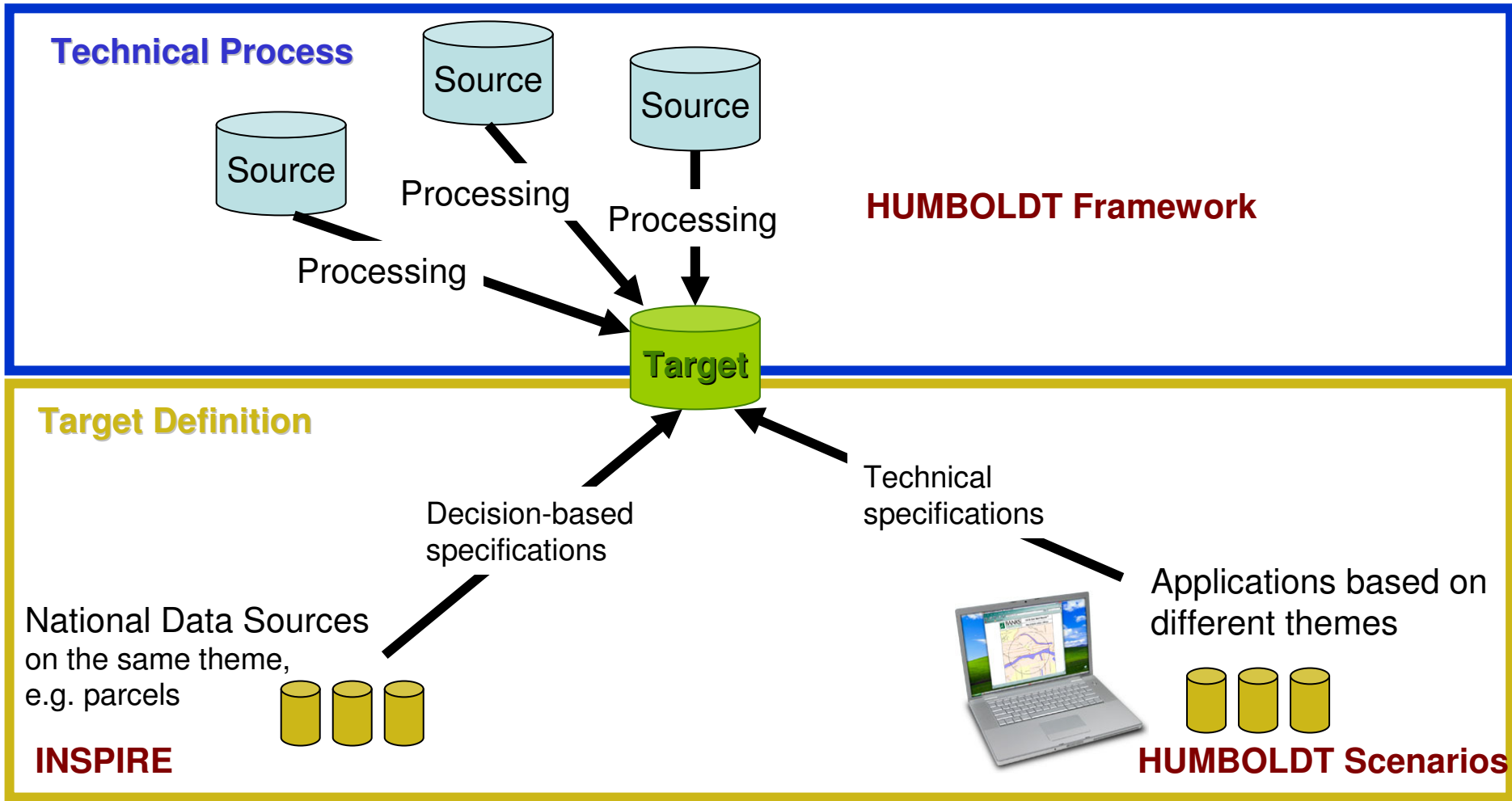




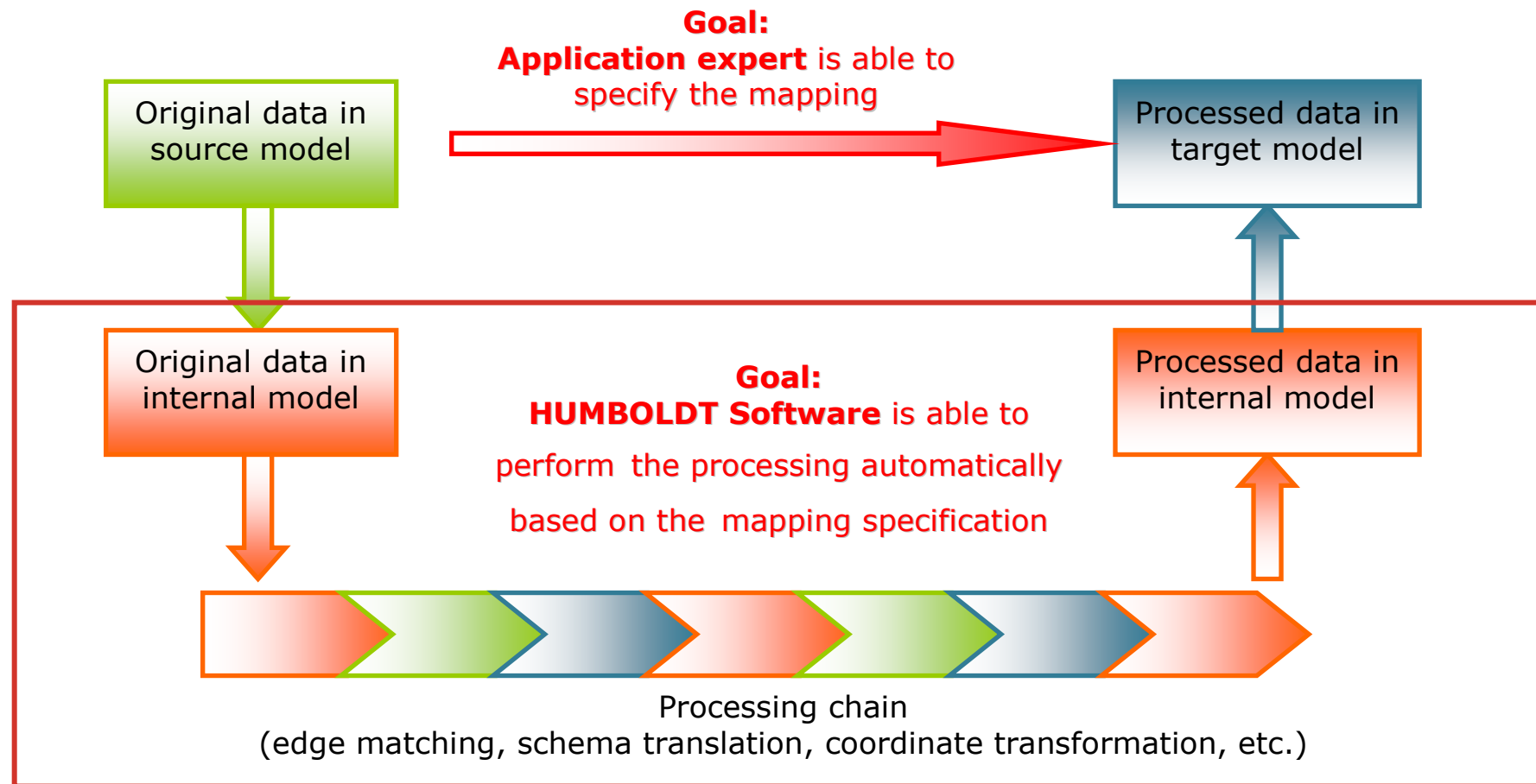
HUMBOLDT's notion of Geodata Harmonisation

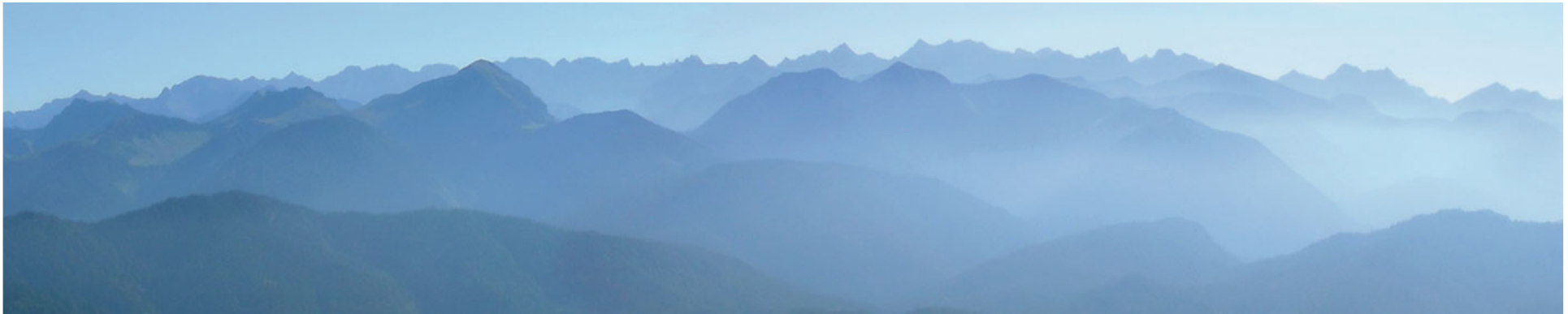


■ Data harmonisation processes



Implication for HUMBOLDT: Technical processing of data

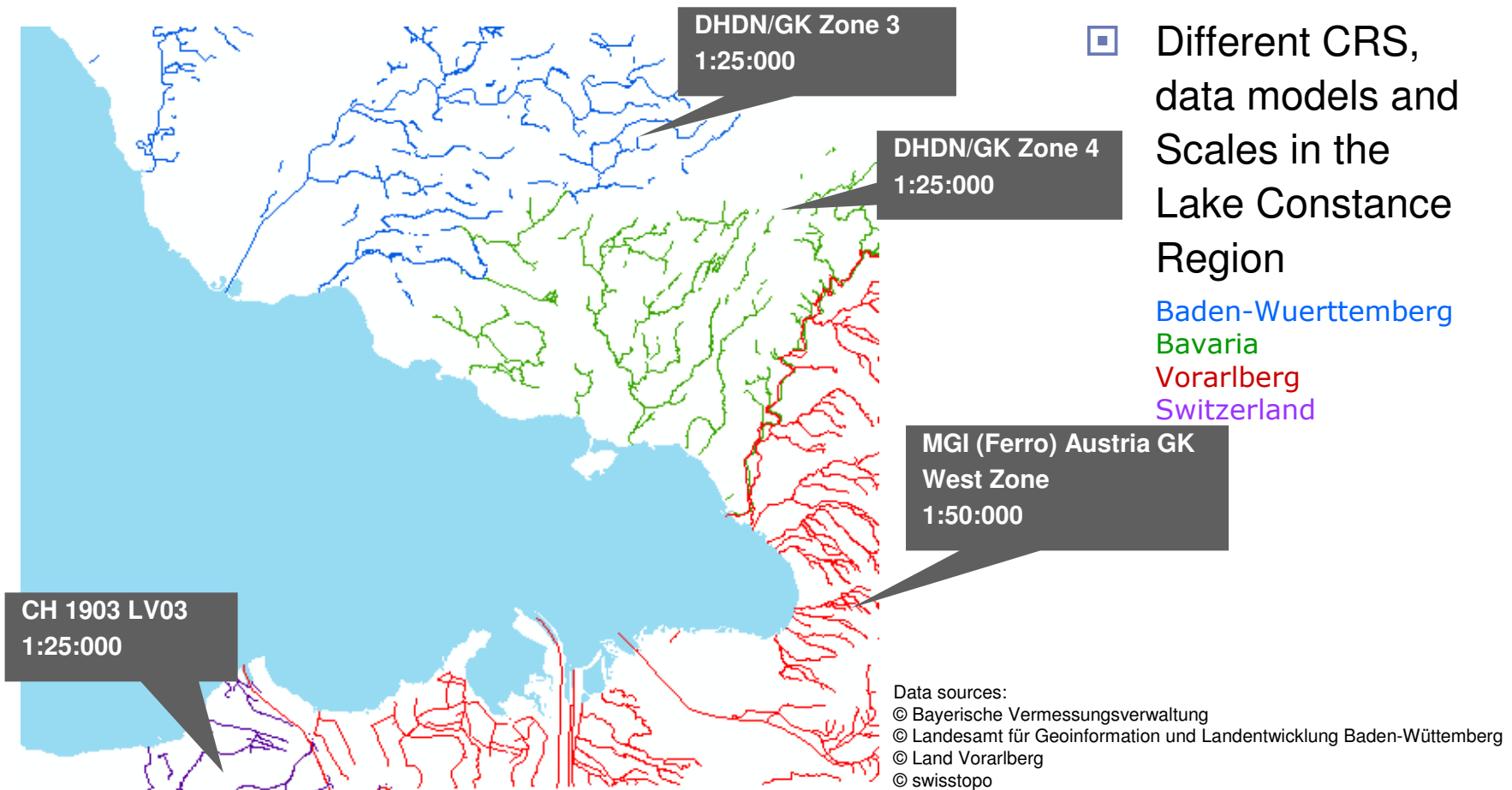




Requirements from different sources

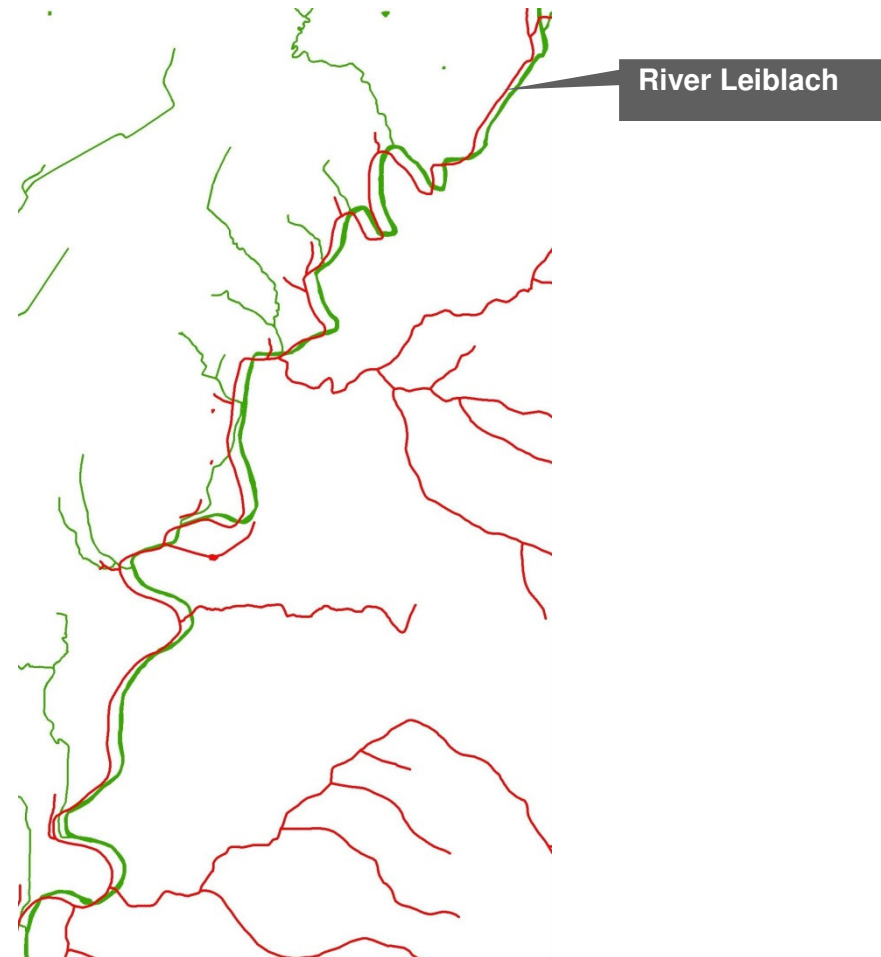


Data Harmonisation Issues in HUMBOLDT Scenarios



Data Harmonisation Issues in HUMBOLDT Scenarios (III)

- ▣ Multiple representation
(detail of image on previous slide)
 - ▣ Partly problem of coordinate transformation
 - ▣ Partly offset correction needed
 - ▣ Partly different geometry



■ Data Harmonisation Issues in HUMBOLDT Scenarios

□ Examples from the Scenarios's requirements list

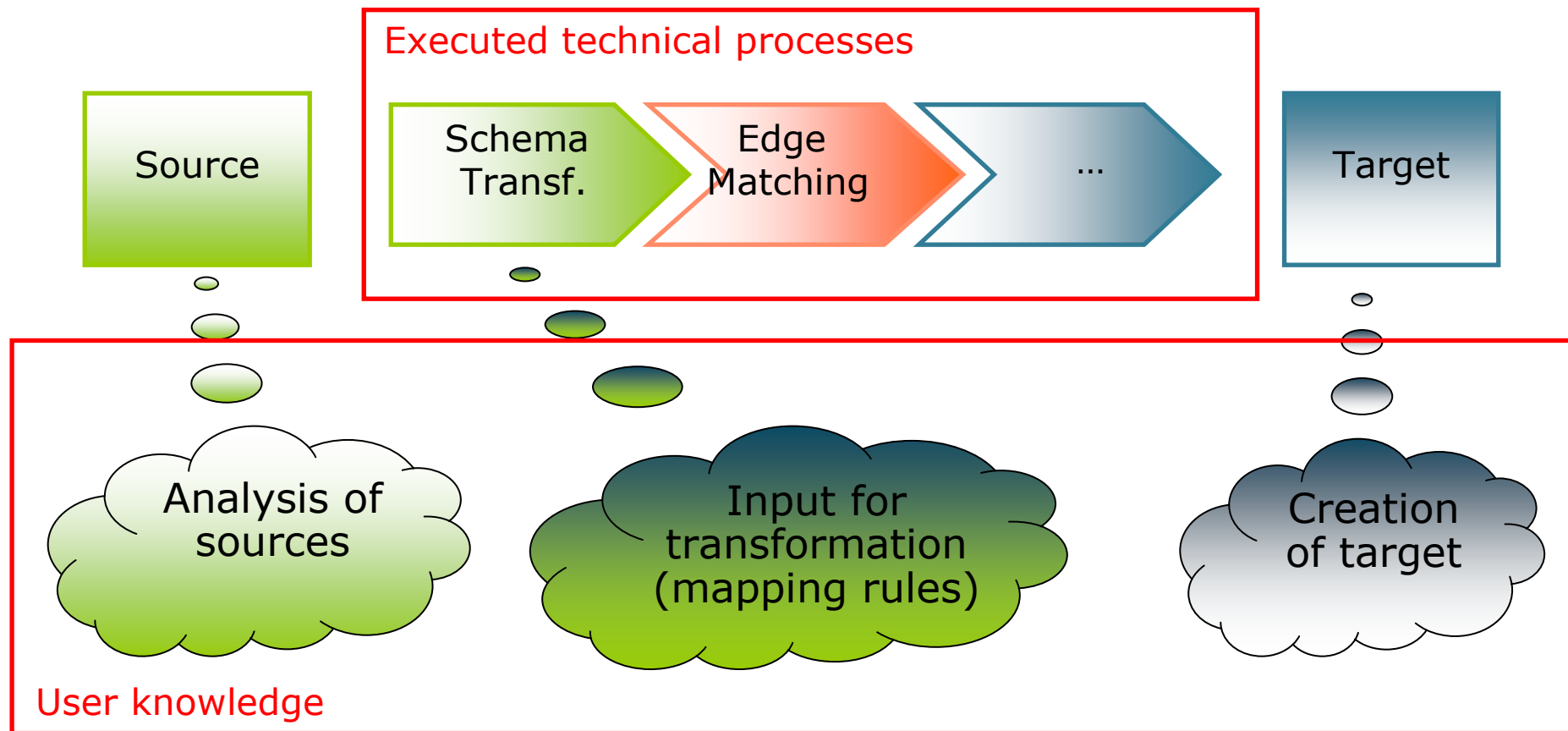
- **geodata formats**
- coordinate reference systems
- **conceptual schemas (data models)**
- **classification schemes**
- scales / resolutions / level-of-detail
- metadata profiles
- natural languages
- **multiple representation of the 'same' spatial objects**
- **spatial consistency issues at the border (edge-matching etc.)**



Introduction to HUMBOLDT Approaches



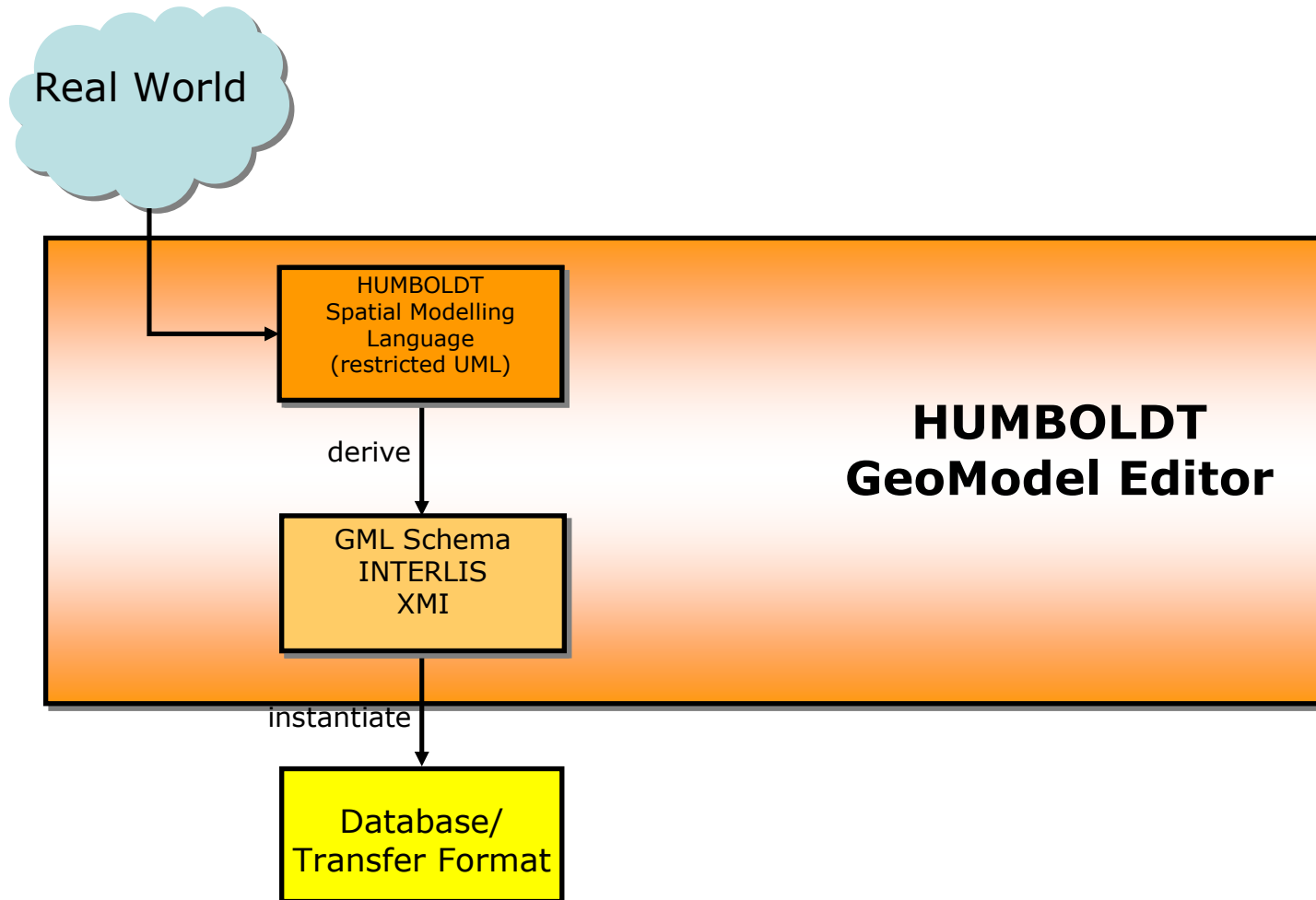
- Functionalities for covering the geodata harmonisation process as a whole

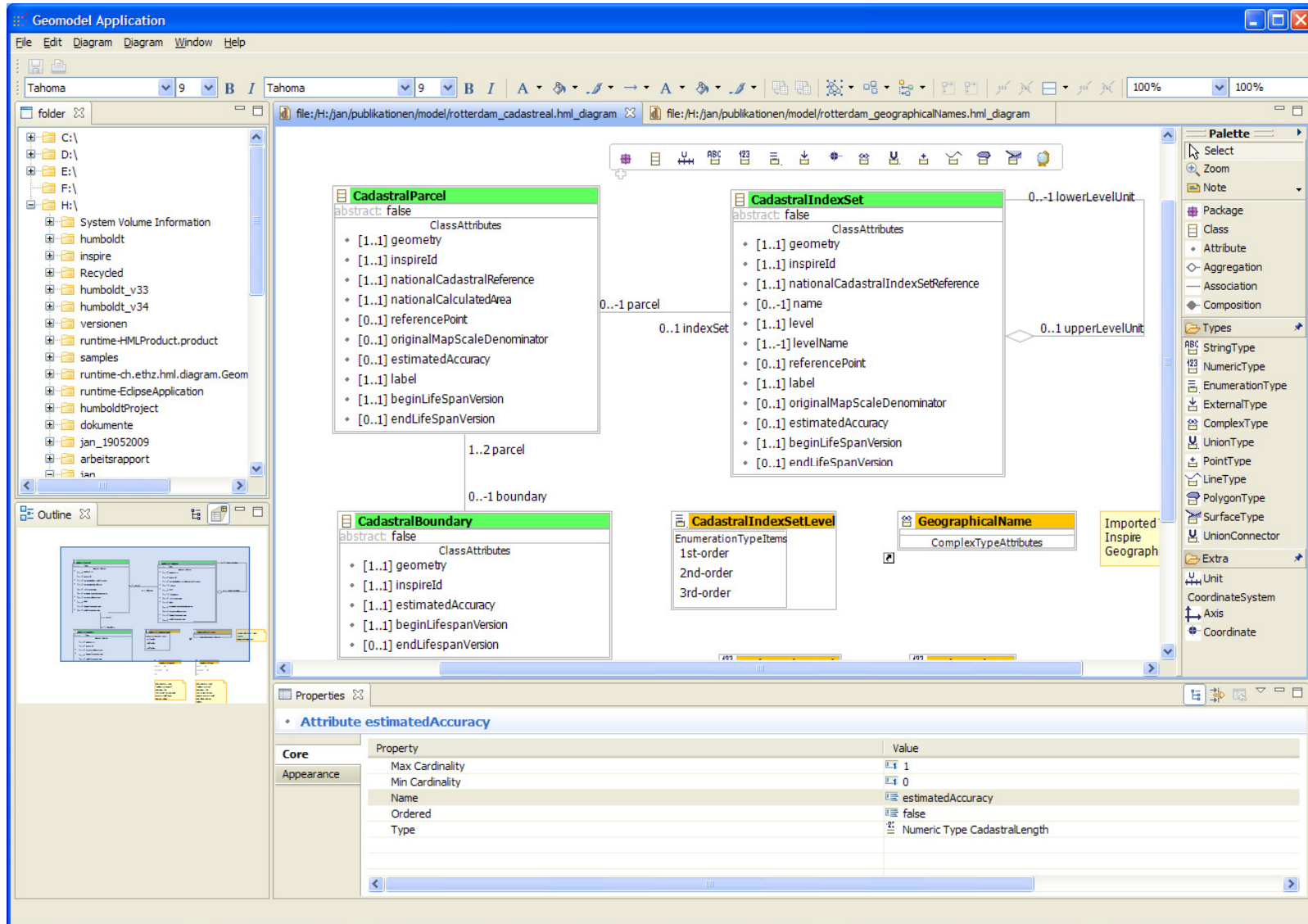


Process Step: Creation of Target

- Defining the information product (target schema, SRS, spatial extent, ...) to which the processed data needs to be transformed







The screenshot shows the Geomodel Application interface. The main workspace displays a UML class diagram with the following classes and relationships:

- CadastralParcel** (abstract: false):
 - ClassAttributes:
 - [1..1] geometry
 - [1..1] inspireId
 - [1..1] nationalCadastralReference
 - [1..1] nationalCalculatedArea
 - [0..1] referencePoint
 - [0..1] originalMapScaleDenominator
 - [0..1] estimatedAccuracy
 - [1..1] label
 - [1..1] beginLifeSpanVersion
 - [0..1] endLifeSpanVersion
- CadastralIndexSet** (abstract: false):
 - ClassAttributes:
 - [1..1] geometry
 - [1..1] inspireId
 - [1..1] nationalCadastralIndexSetReference
 - [0..-1] name
 - [1..1] level
 - [1..-1] levelName
 - [0..1] referencePoint
 - [1..1] label
 - [0..1] originalMapScaleDenominator
 - [0..1] estimatedAccuracy
 - [1..1] beginLifeSpanVersion
 - [0..1] endLifeSpanVersion
- CadastralBoundary** (abstract: false):
 - ClassAttributes:
 - [1..1] geometry
 - [1..1] inspireId
 - [1..1] estimatedAccuracy
 - [1..1] beginLifeSpanVersion
 - [0..1] endLifeSpanVersion
- CadastralIndexSetLevel** (EnumerationTypeItems):
 - 1st-order
 - 2nd-order
 - 3rd-order
- GeographicalName** (ComplexTypeAttributes):
 - Imported Inspire Geograph

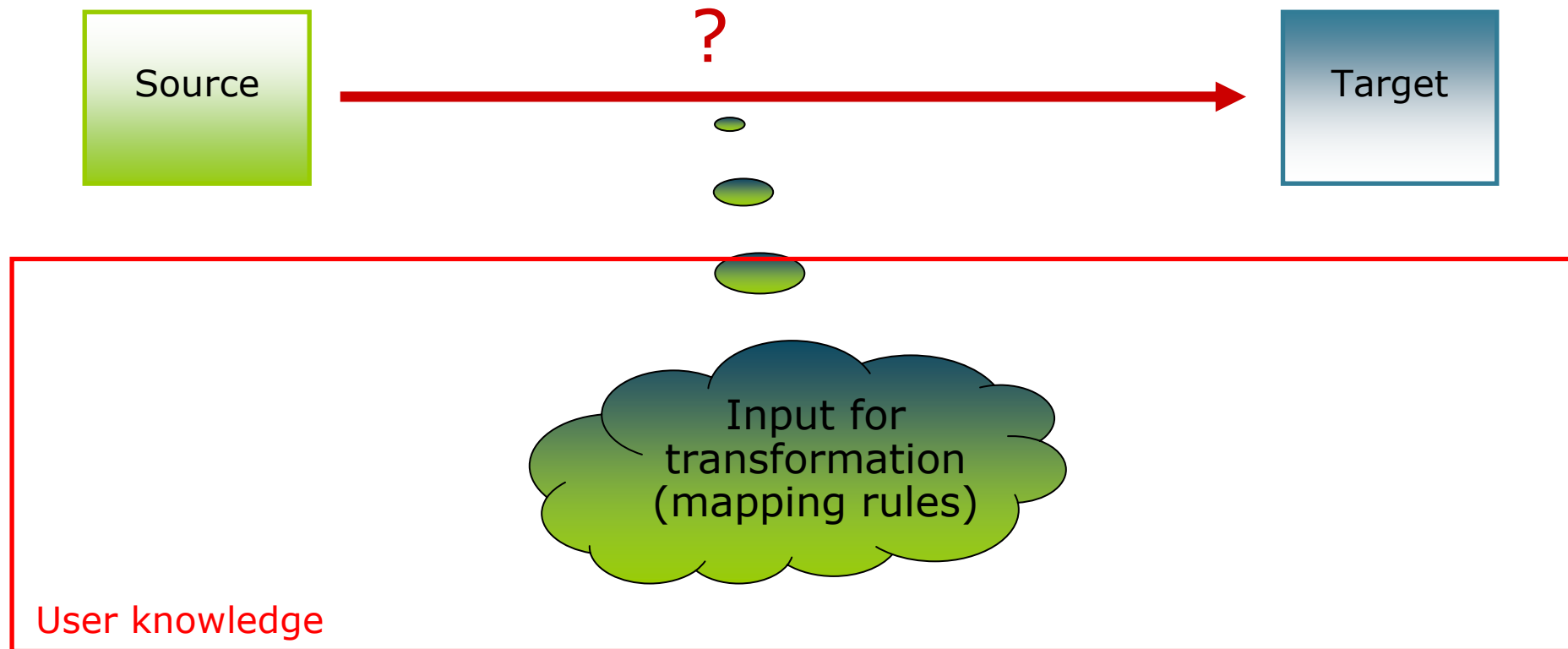
Relationships shown in the diagram:

- CadastralParcel** (0..-1 parcel) is associated with **CadastralIndexSet** (0..1 indexSet).
- CadastralParcel** (1..2 parcel) is associated with **CadastralBoundary** (0..-1 boundary).
- CadastralIndexSet** (0..-1 lowerLevelUnit) is associated with **CadastralIndexSetLevel** (0..1 upperLevelUnit).

The Properties window at the bottom shows details for the attribute **estimatedAccuracy**:

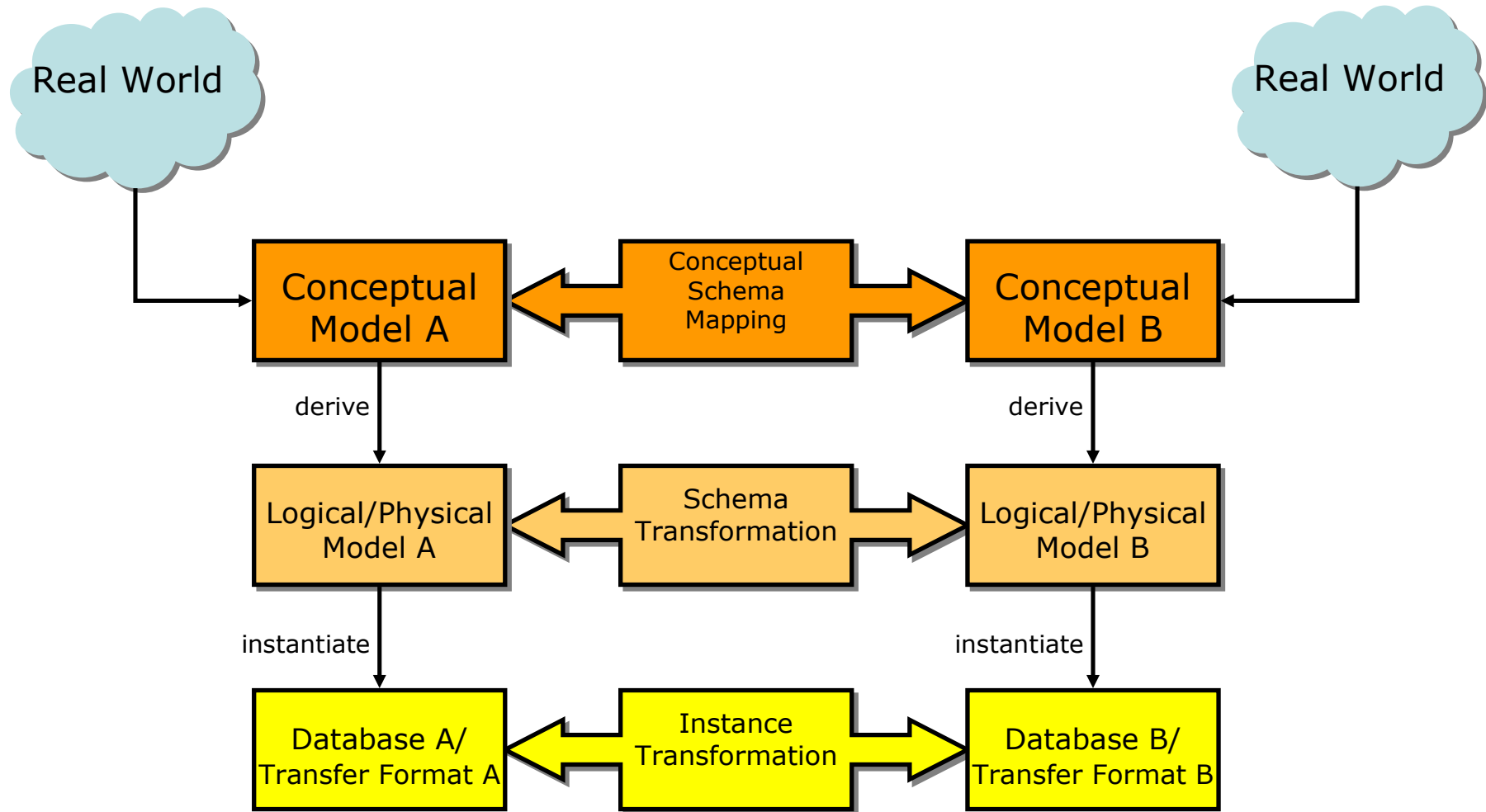
Property	Value
Max Cardinality	1
Min Cardinality	0
Name	estimatedAccuracy
Ordered	false
Type	Numeric Type CadastralLength

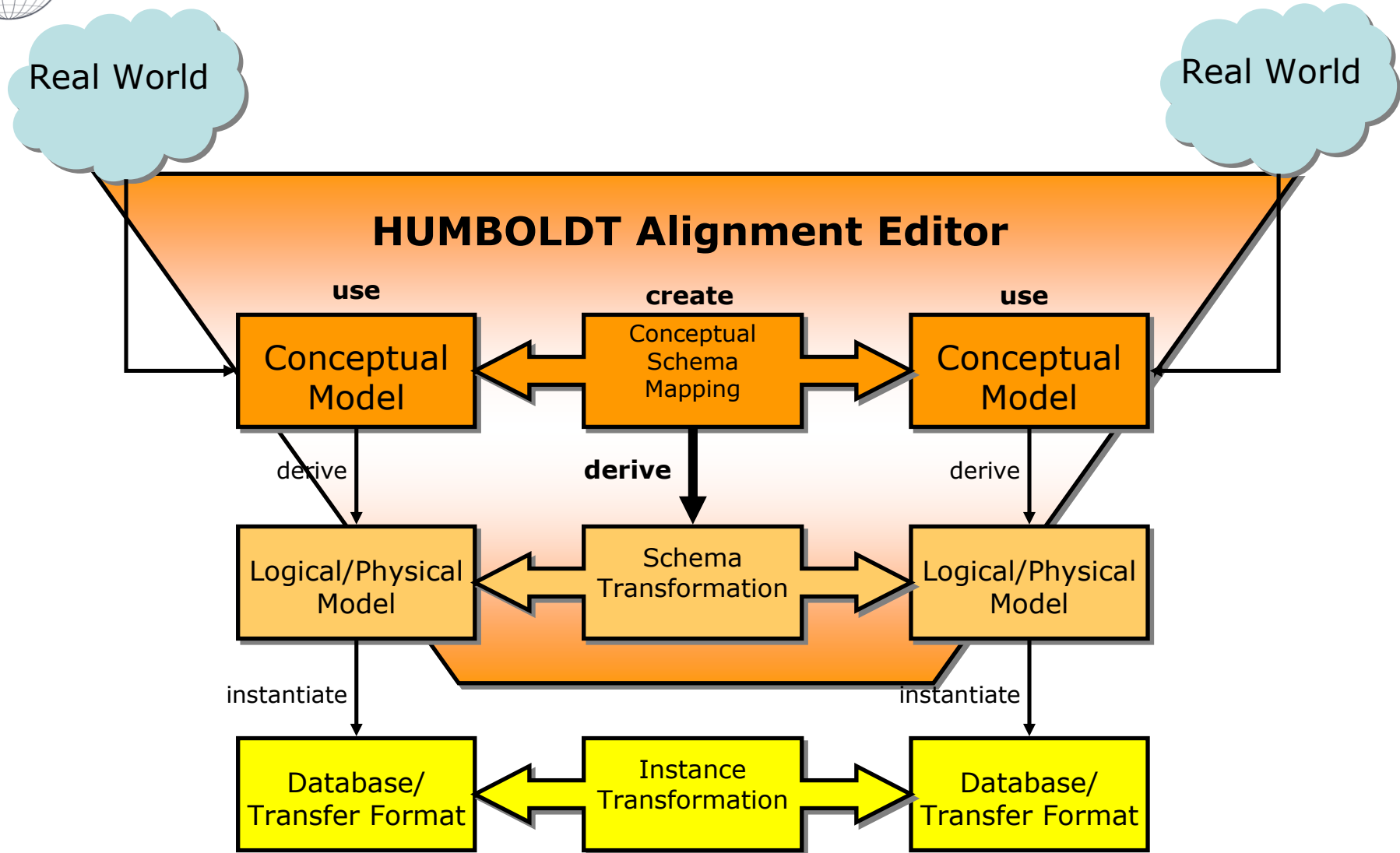
 Process step 2: Schema to Schema Mapping

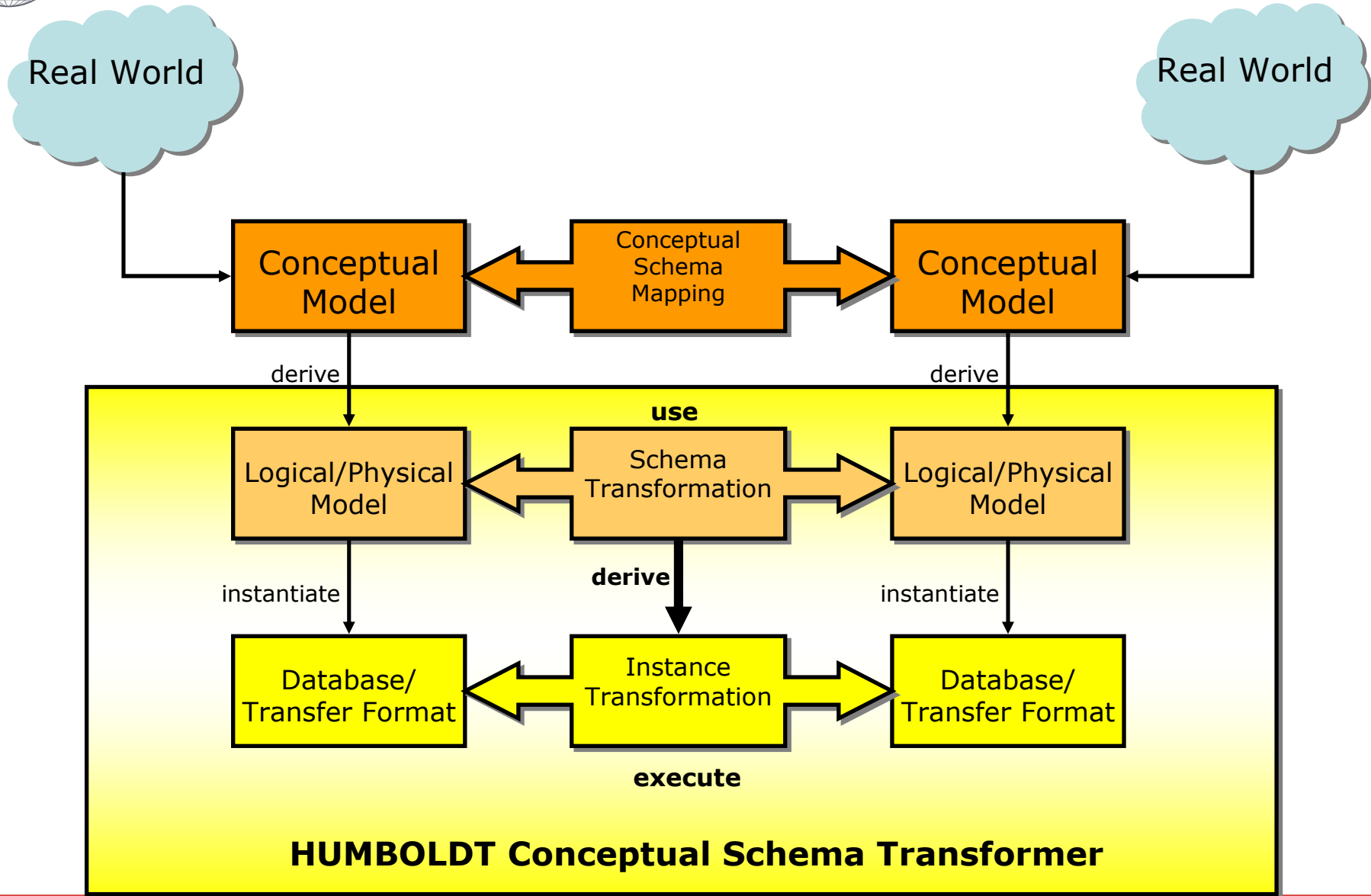


• Necessary Schema Mapping Operations

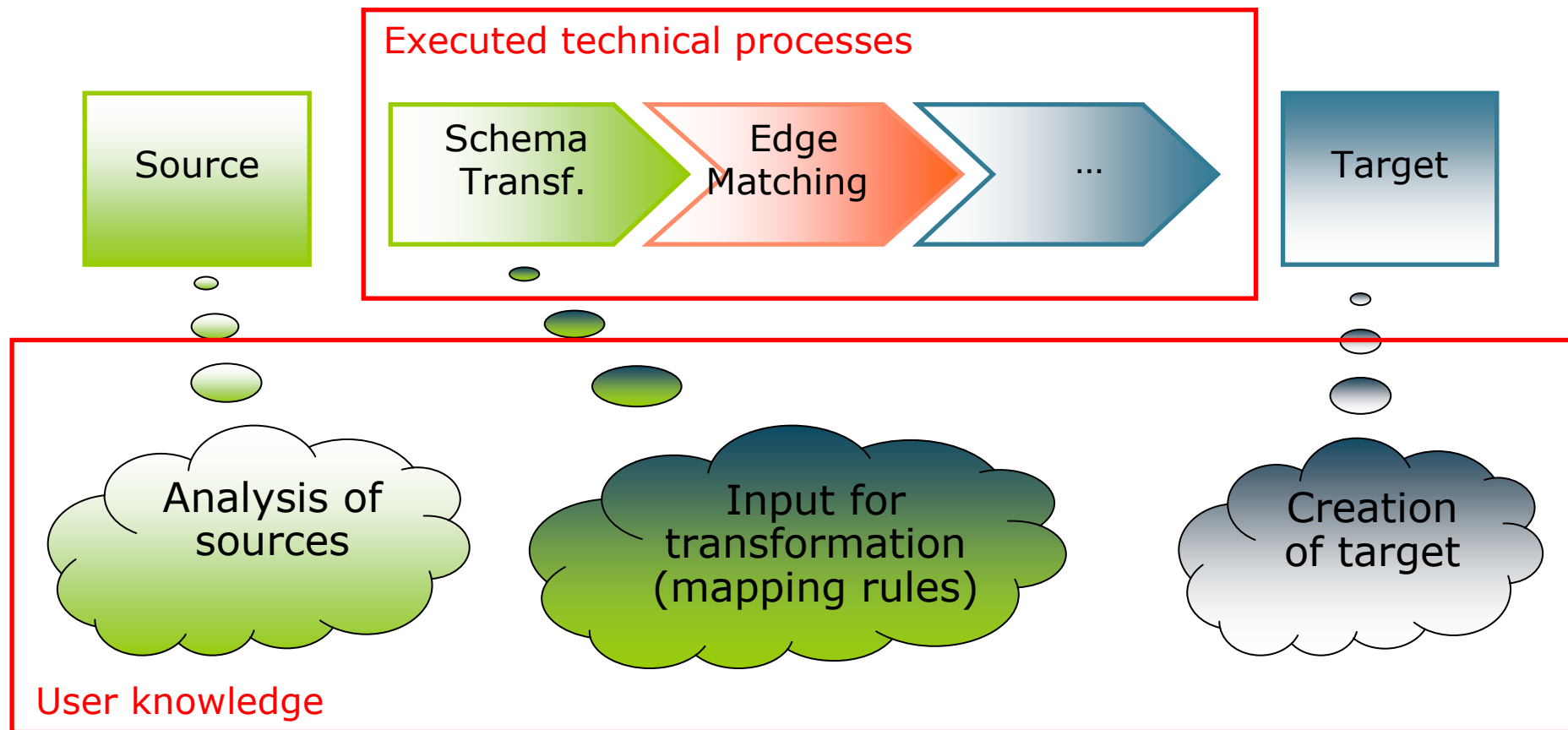
- Filtering: conditional statements applied to source data to filter features (extract sub-sets)
- Reclassification of attribute values
- Renaming of feature classes or attributes
- Merge / split of features or attributes
- Change of attribute order
- Type conversions
- Value conversions
- Reclassification
- Augmentation







- Functionalities for covering the data harmonisation process as a whole



Questions answered in HUMBOLDT

“What” questions:

- ▣ What is an ‘application-specific’ harmonised data model?
- ▣ What are the requirements for a “good” harmonised model (e.g. usage of standards: ISO/OGC, INSPIRE, ..)?
- ▣ What is the role of these harmonised models in data harmonisation in the HUMBOLDT Scenarios and the HUMBOLDT project as a whole?

“How” questions:

- ▣ How to specify a data model for the Scenario?
- ▣ How to carry out the transformation of existing datasets ?



Thank you for your attention!

→ Technical Results