

Basic Principles and Aspects of Spatial Data Harmonization

Madrid (Spain), 24. 10. 2011

Outline

Who?

How?

When?

Why?

What?

Spatial Data Harmonization

Definition

- **Harmonised data product specifications** – set of data product specifications that support the provision of access to interoperable spatial data through spatial data services in a representation that allows for combining it with other interoperable spatial data in a coherent way
- Note 1: The harmonised data product specifications will be based on the data interoperability components.
- Note 2: The harmonised data product specification is not intended to replace or deprecate existing data specifications that are currently in use.



Definition

- Data harmonisation – providing access to spatial data through network services in a representation that allows for combining it with other harmonised data in a coherent way by using a common set of data product specifications



Definition

(Creating) the possibility to combine data from heterogeneous sources into integrated, consistent and unambiguous information products.

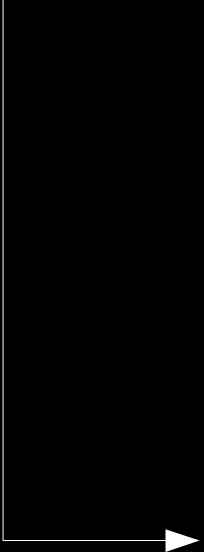


Definition

- ISO 860:1996 – Terminology work – **Harmonization of concepts and terms**: Concept harmonization is an activity for reducing or eliminating minor differences between two or more concepts which are already closely related to each other.



Spatial Data Harmonization



Activity for elimination
or reduction of
heterogeneities of
various properties of
spatial data to support
interoperability.

Spatial Data Harmonization

Active
process

Not
complete
removing

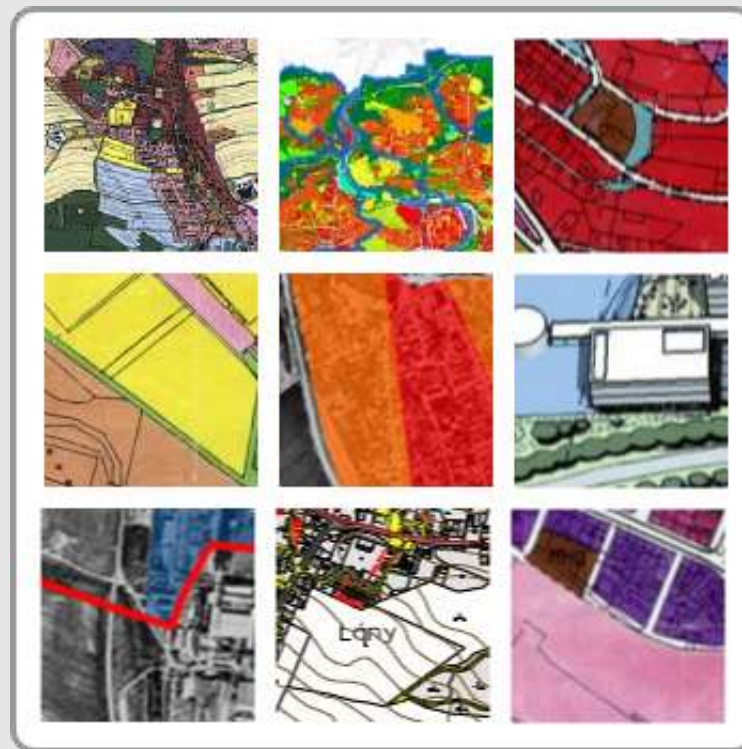
Activity for elimination
or reduction of
heterogeneities of
various properties of
spatial data to support
some aspects
interoperability.

Differences

Not
complete
data

Not whole
interoperability
-
Particular
purposes

How can heterogeneous spatial data look?



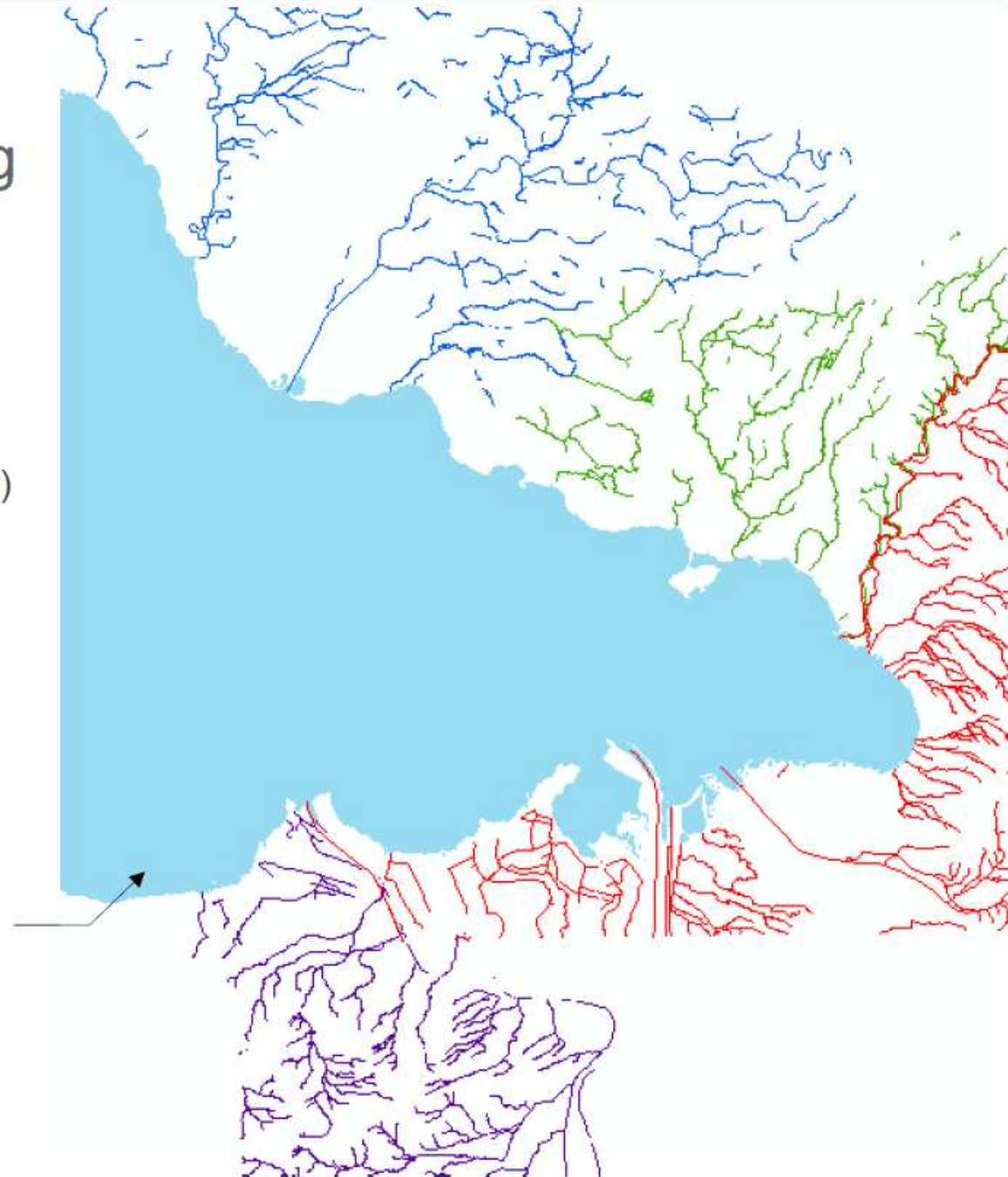
Data used in the testing

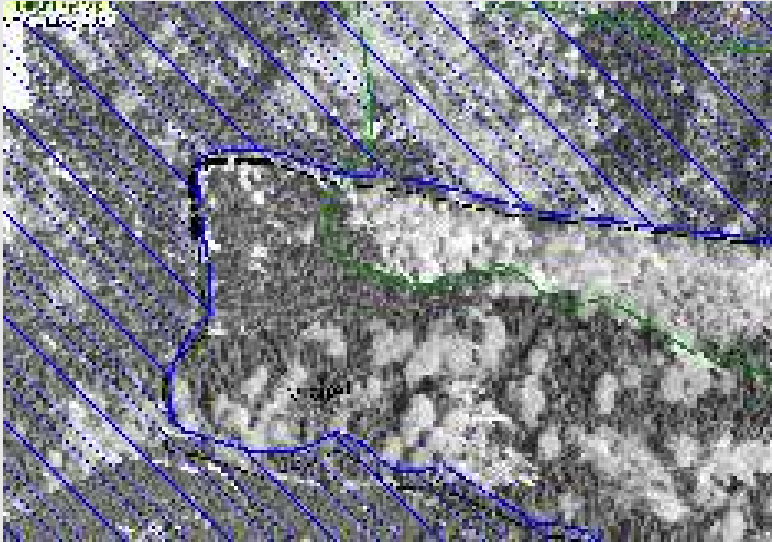
Topographic vector data
(streets and rivers) from

- ▣ Germany: **Baden-Wuerttemberg** (BW) and **Bavaria** (BY)
- ▣ Austria: **Vorarlberg** (VA)
- ▣ **Switzerland** (CH)

Data sources:

- © Bayerische Vermessungsverwaltung
- © Landesamt für Geoinformation und Landentwicklung Baden-Württemberg
- © Land Vorarlberg
- © swisstopo





CNN, 25.09.2001



CBC news, 14.9.2006

- An incorrect map and communications failure led to an Israeli air strike on a UN observer post that killed four peacekeepers, including a Canadian...
- ..."The UN position – we had it as a Hezbollah target. This was an error."...said Israeli Foreign Ministry spokesman Mark Regev

Back to Harmonization...

The elimination of the aspects of spatial data heterogeneity cannot be based on a creation of some **uniform rules and data models**, because, there are too many subjects with individual requirements – formats, precision, reference systems, terminology...

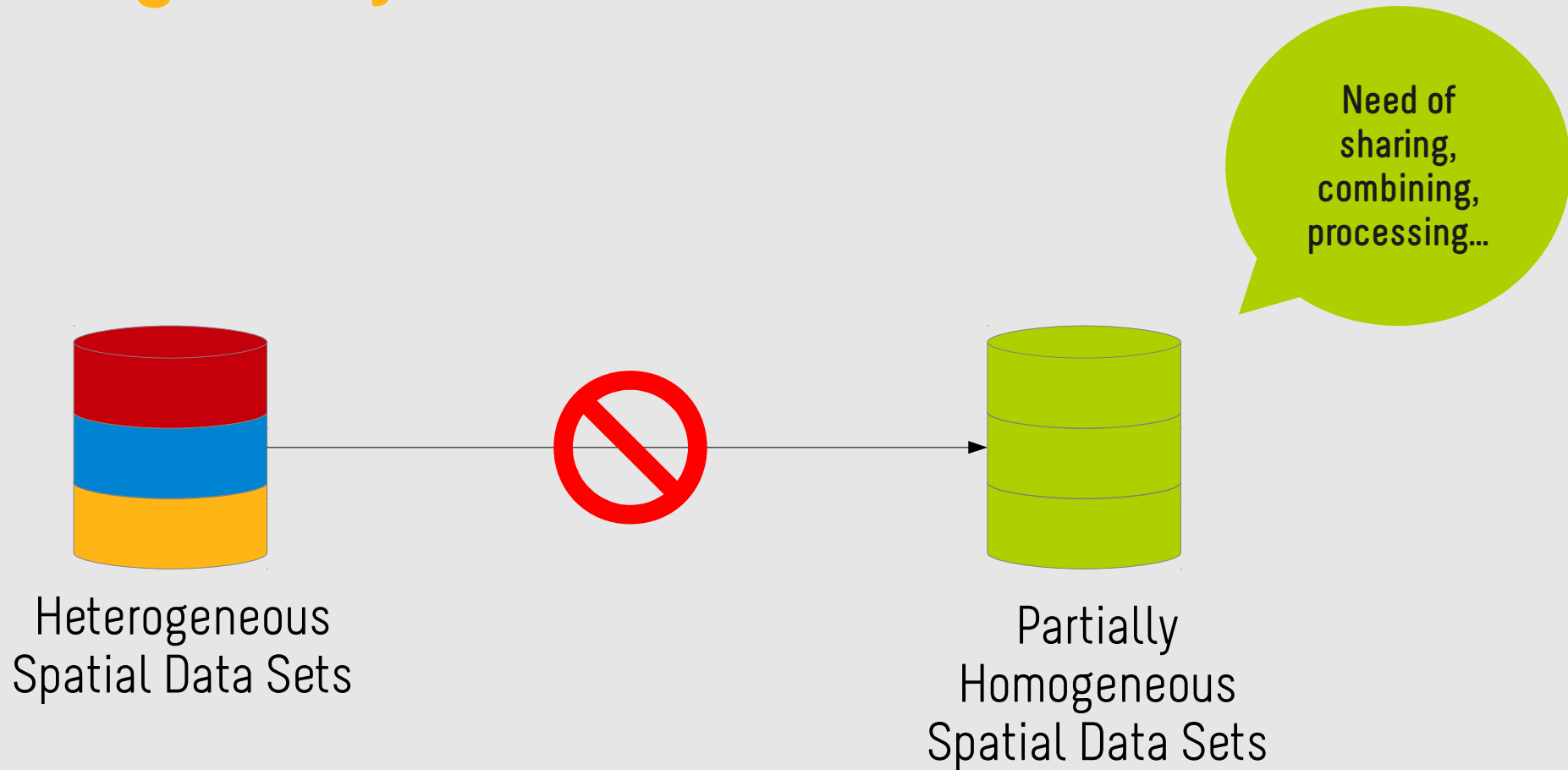
Why to harmonize?

- To enable a sharing, combining and publishing of data
- To re-use existing sources
- To improve data quality
- To use web services and other automatic tools (SaS)
- To keep data interoperability (it's cool!)
- To increase the number of stakeholders
- To meet legislation requirements

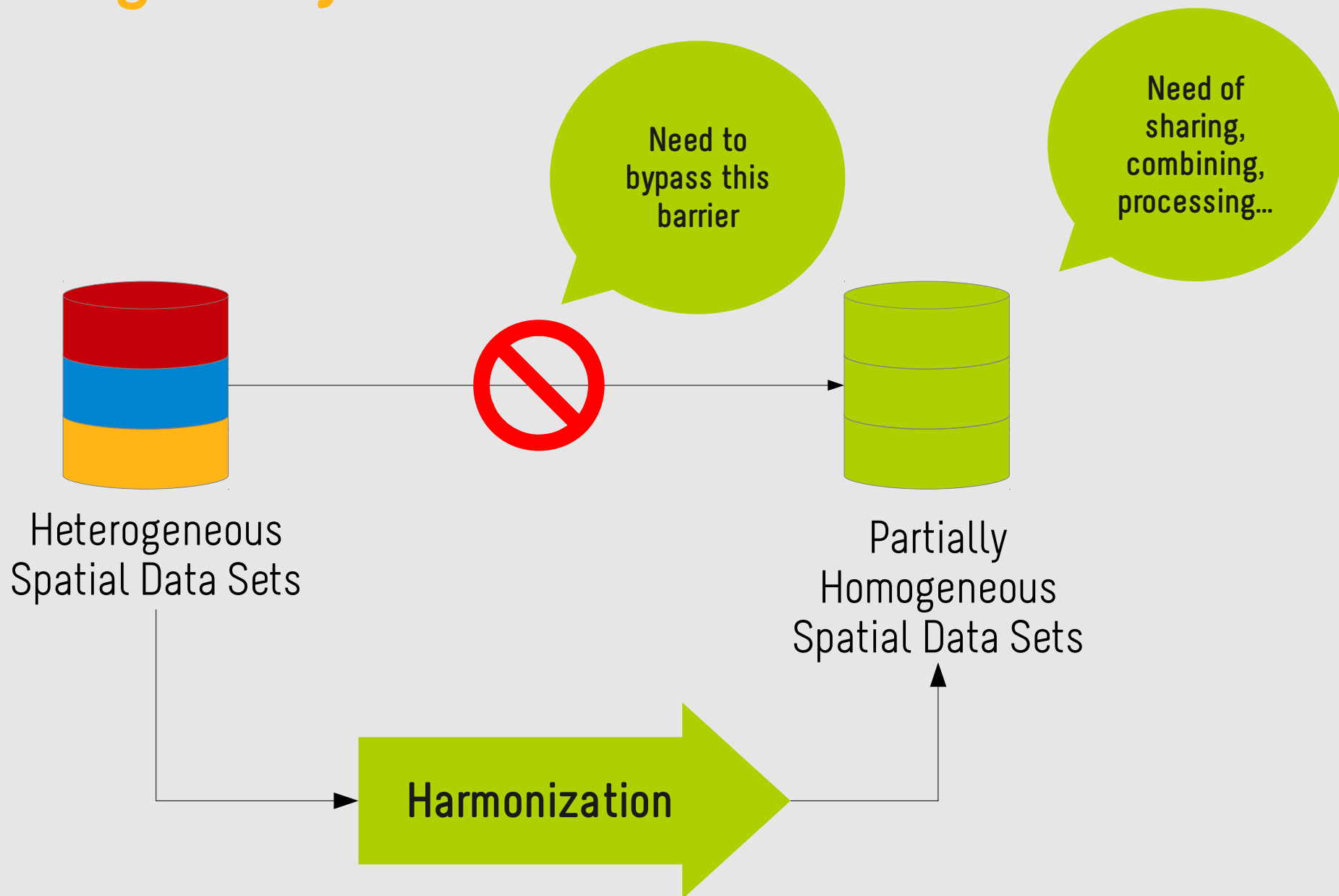


**All reasons are
strongly
interconnected**

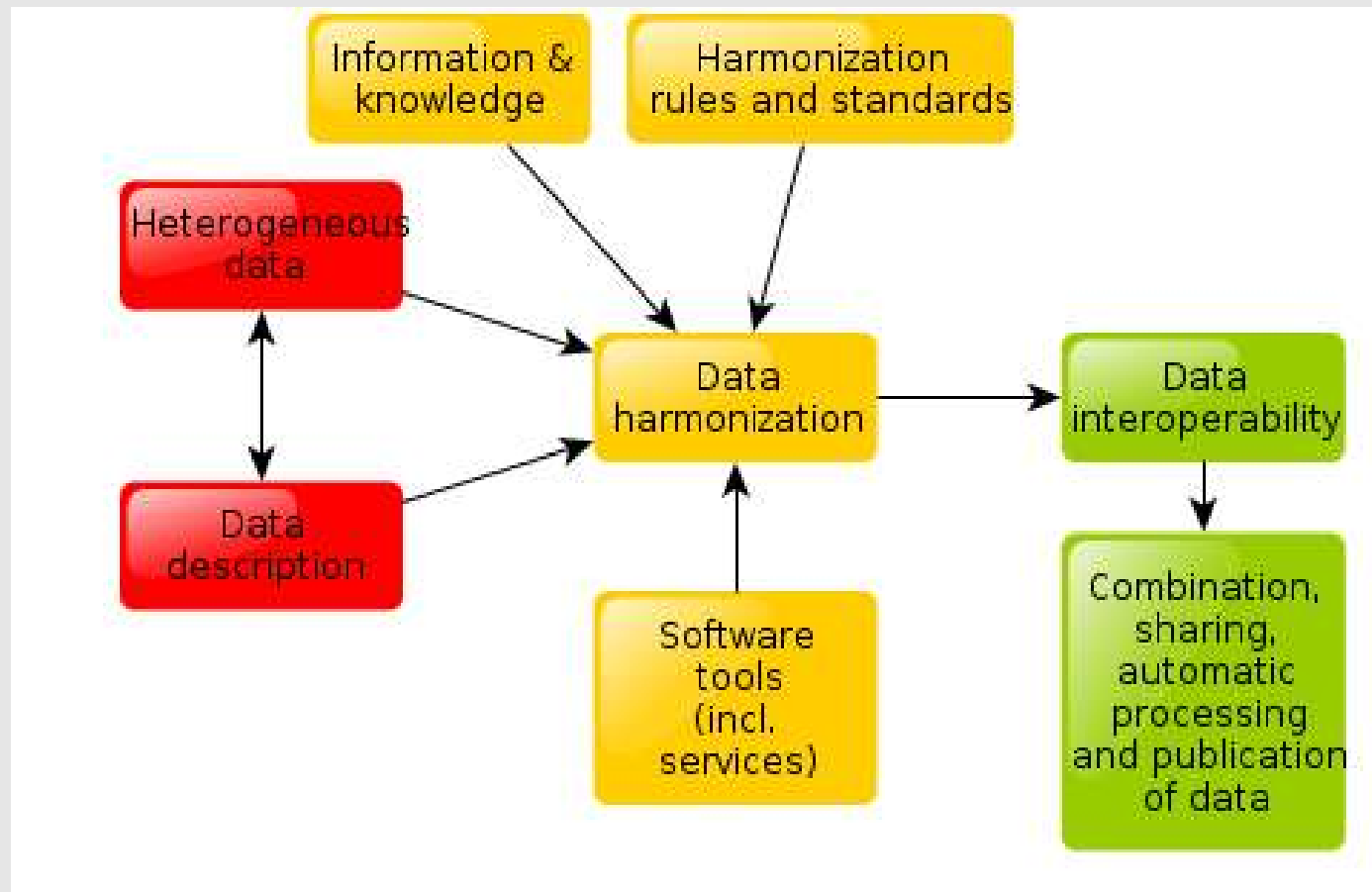
Heterogeneity vs. Harmonization



Heterogeneity vs. Harmonization



How? - Architecture of Harmonization



Harmonization Process



Harmonization

- To declare the target of harmonization process
- To describe input and output data (including data models, metadata)
- To define necessary harmonization sub-processes
- To find appropriate technology (technologies)
- To establish the sequence of harmonization sub-processes
- Testing, feedback, iterative development...
- To realize harmonization process

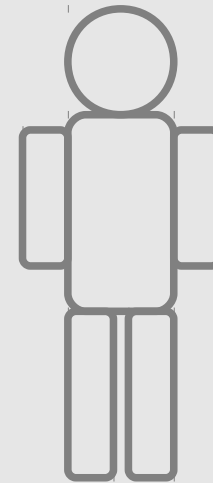
Harmonization Process - Requirements



Harmonization

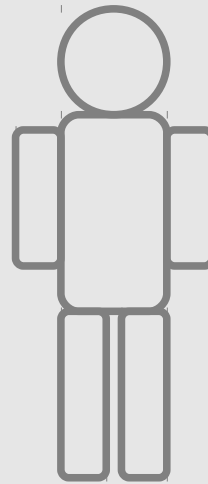
- Education & experience (spatial data, domain, GIT... harmonization)
- Enough information (user requirements, data description, standards...)
- Understanding & knowledge → Willingness to think about
- Communication
- Willingness to do compromises

Who?

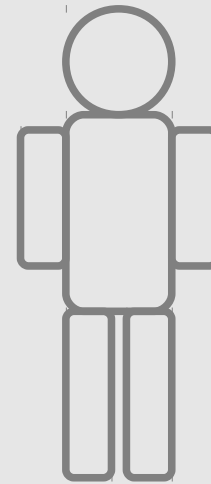


GIT
Experts

Who?

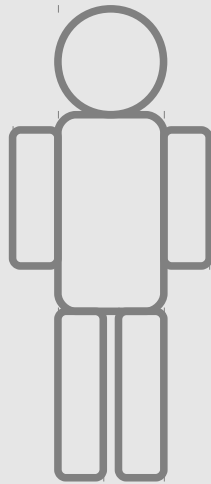


Data
Experts

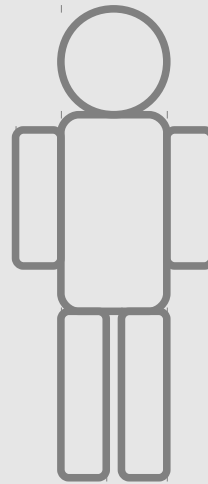


GIT
Experts

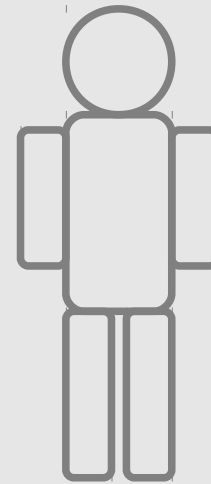
Who?



Domain
Experts

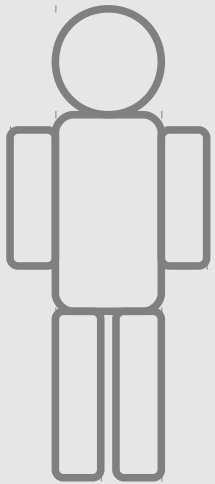


Data
Experts

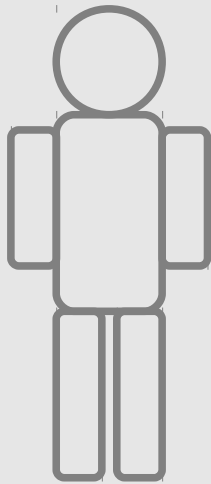


GIT
Experts

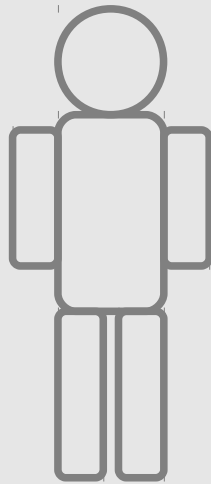
Who?



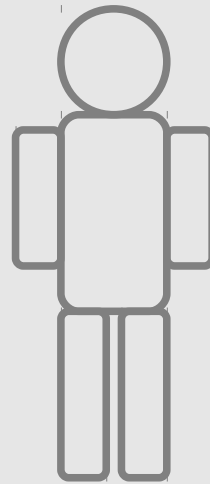
Data
Providers



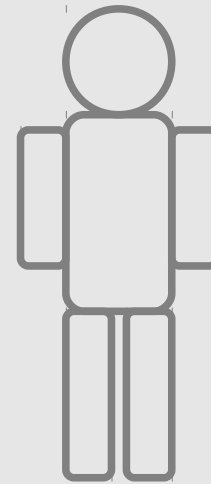
Data
Users



Domain
Experts

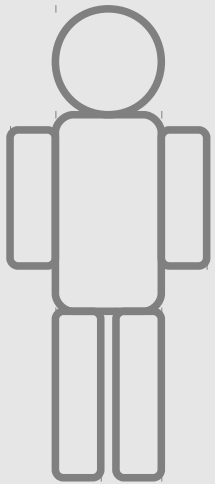


Data
Experts

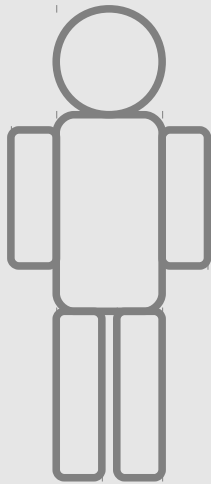


GIT
Experts

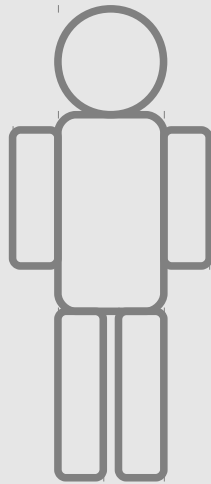
Who?



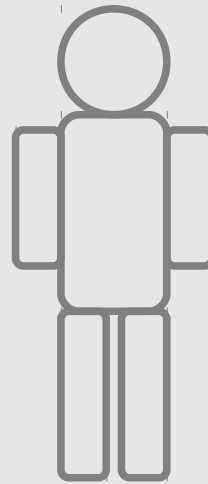
Data
Providers



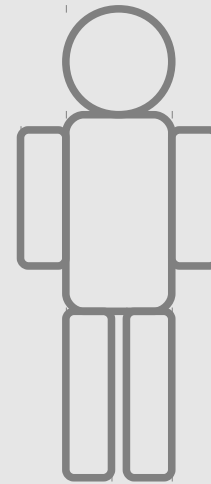
Data
Users



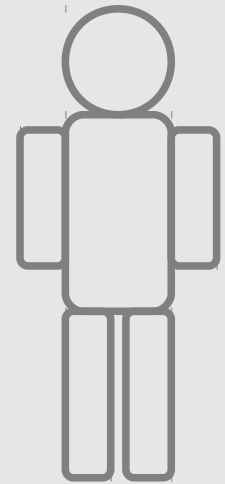
Domain
Experts



Data
Experts



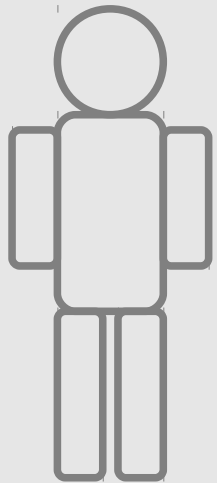
GIT
Experts



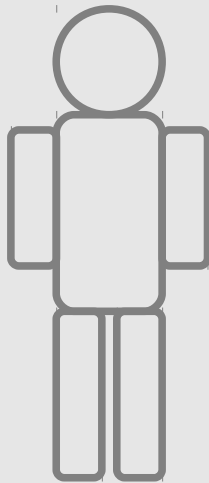
Others
(Cartographers,
Programmers,
Managers,
Integrators...)

Who?

Requests



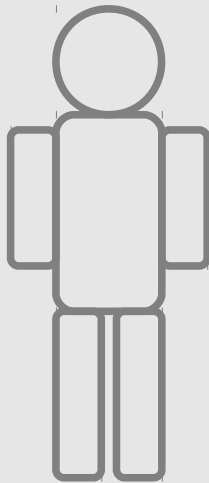
Data
Providers



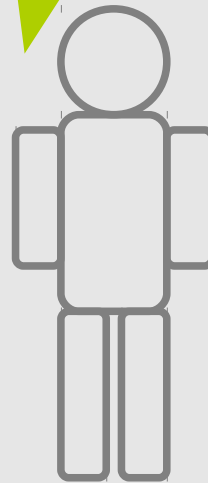
Data
Users



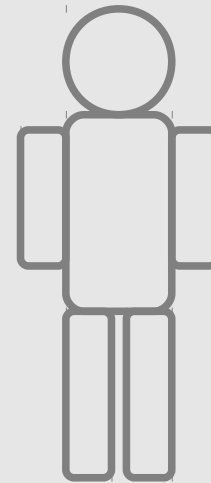
Proposals



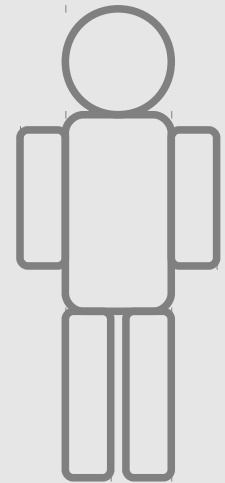
Domain
Experts



Data
Experts

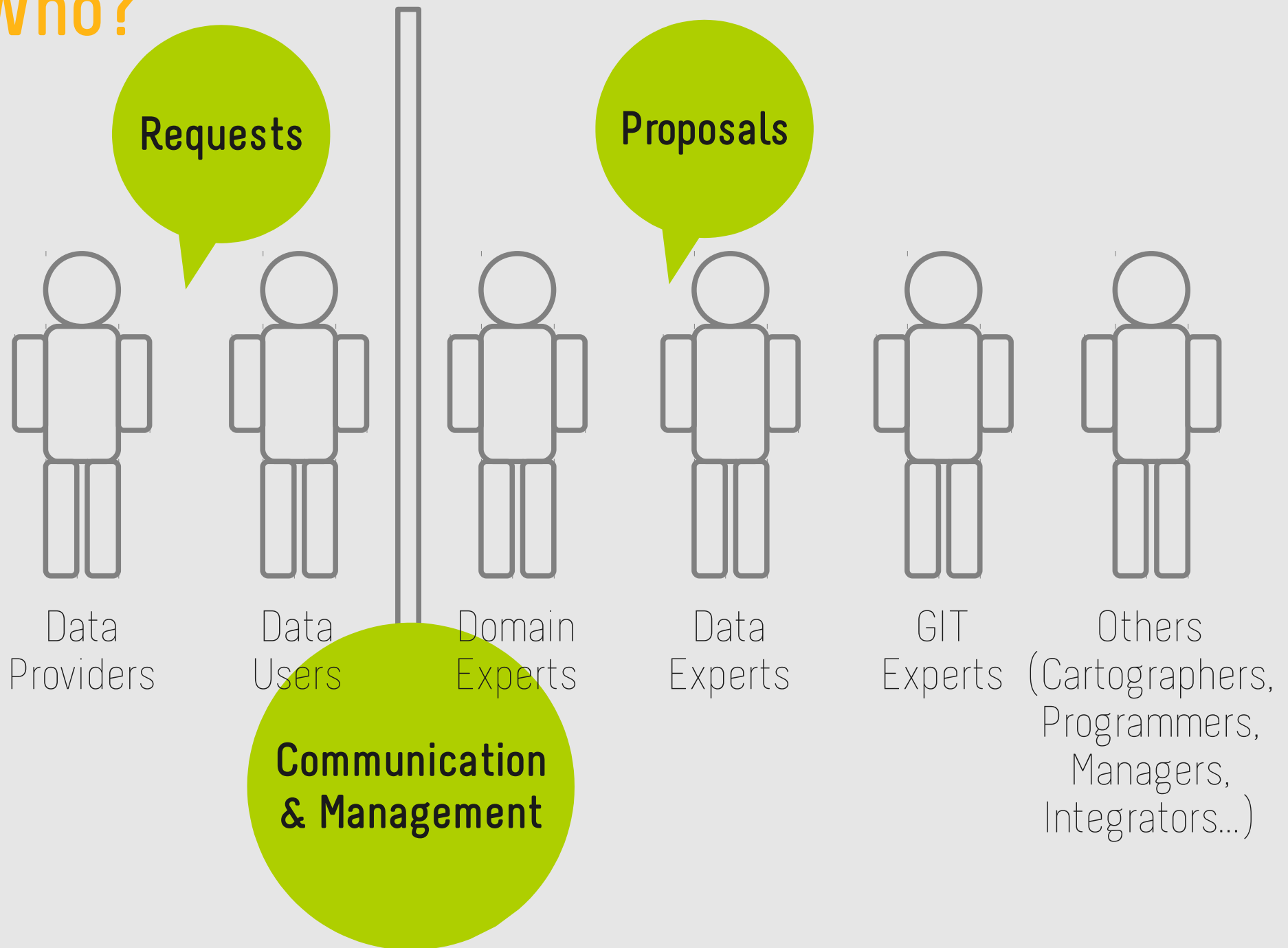


GIT
Experts

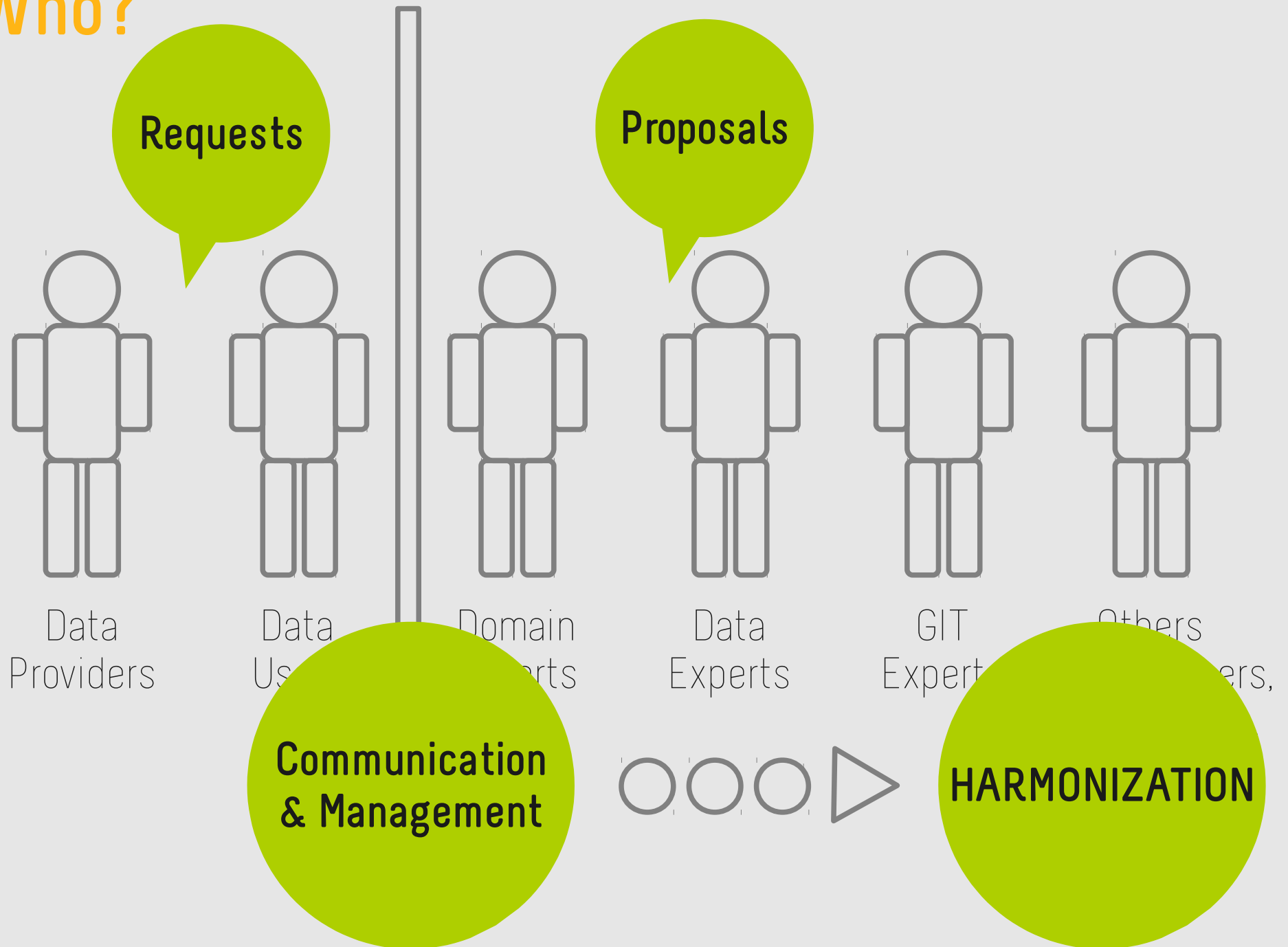


Others
(Cartographers,
Programmers,
Managers,
Integrators...)

Who?



Who?



Magical Blue Table

Harmonization Types (Sub-processes)

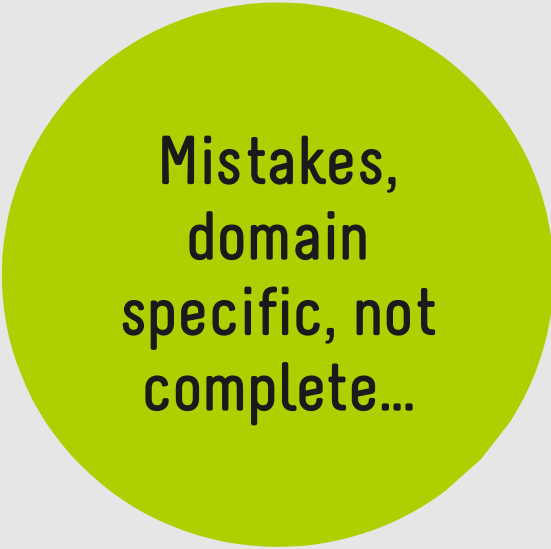
Data interoperability components
Potential heterogeneities

(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) Identifier Management
(J) Data transformation	(K) Portrayal model	(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	



Clear Harmonization Sub-processes

- Coordinate references and unit model → EPSG Codes, International System of Units
- **Terminology** → thesauruses, codelists, taxonomies
- Rules for application schemas and Feature catalogues → data model and their description
- Metadata → Standards and profiles
- Portrayal Model → SLD
- INSPIRE principles → INSPIRE documents



Mistakes,
domain
specific, not
complete...

Complicated Harmonization Sub-processes

- Spatial and Temporal Aspects → geometry vs. time
- Multilingual Text and Cultural Adaptability
- Maintenance
- Conformance
- Quality



Open & Not Clear

Results of state-of-the-art analysis

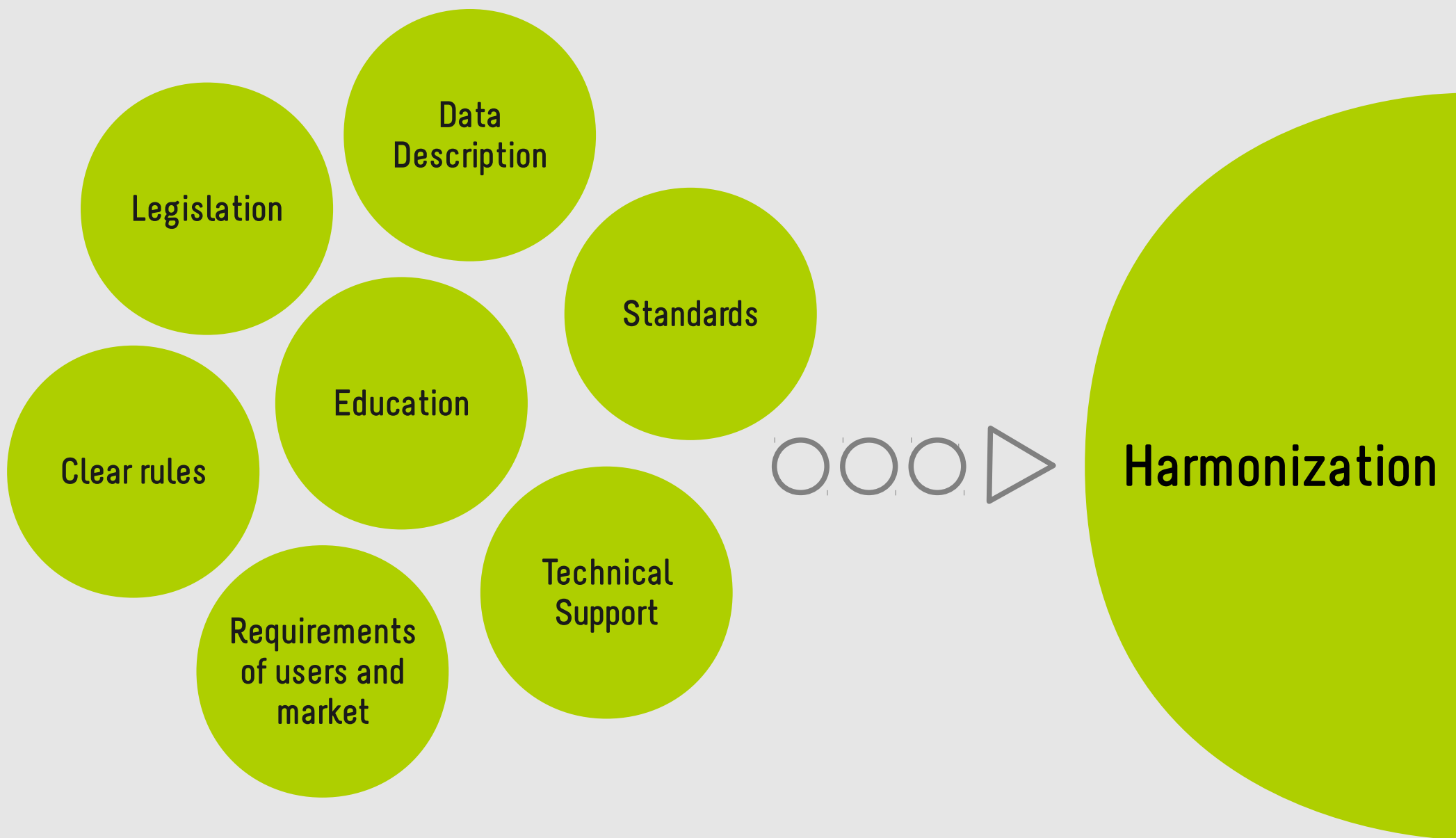
The top five requirements are:

- ▣ 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)

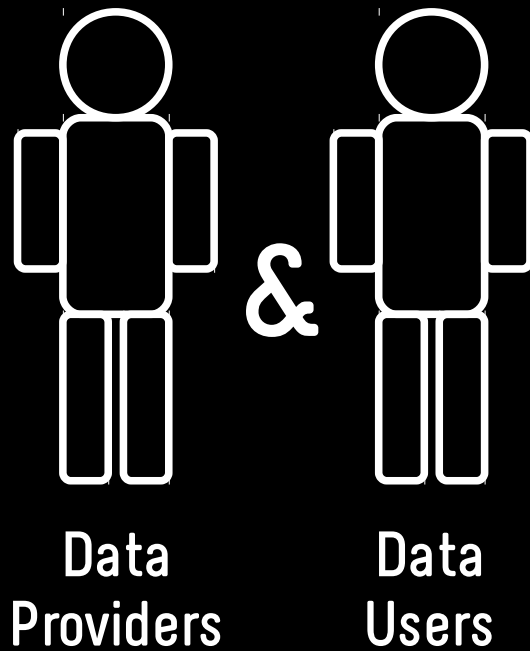
Benefits of Harmonization

- Limitation of duplicities in data
- Clear origin and assurance of quality of the data
- Data structure standardisation
- Data purity, security and structure uniformity
- Better data manipulation
- Reciprocal data accessing per web services – preservation data up-dating (possibility of on-line actualisation)
- Fall of cost for data updating and maintenance
- Better software development
- Better source exploitation
- Improvement of chances in communication with authorities
- Better utilization and commercialization of spatial data
- Increasing activities, e.g. education
- Accordance with reality

How to Push Harmonization Activities



Main Problems of Harmonization



We need to harmonize, we don't know targets and reasons

We need to harmonize all data sets by implementation of one system

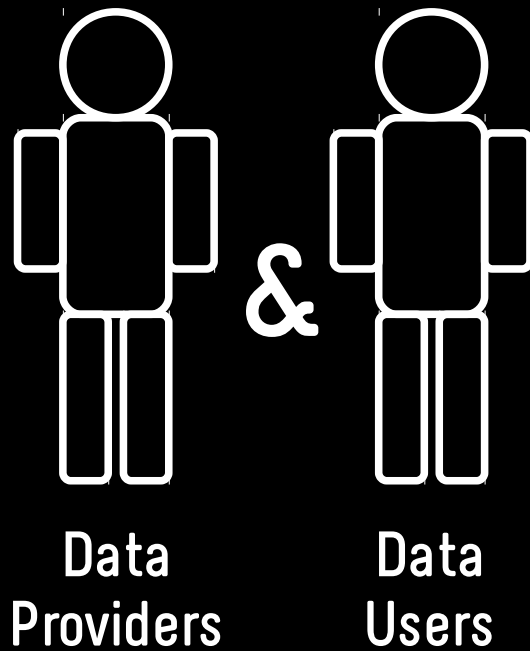
Harmonization is...

...complicated system of many interconnected steps (dependent on a level of heterogeneity of source data).

These steps have a huge number of input parameters.

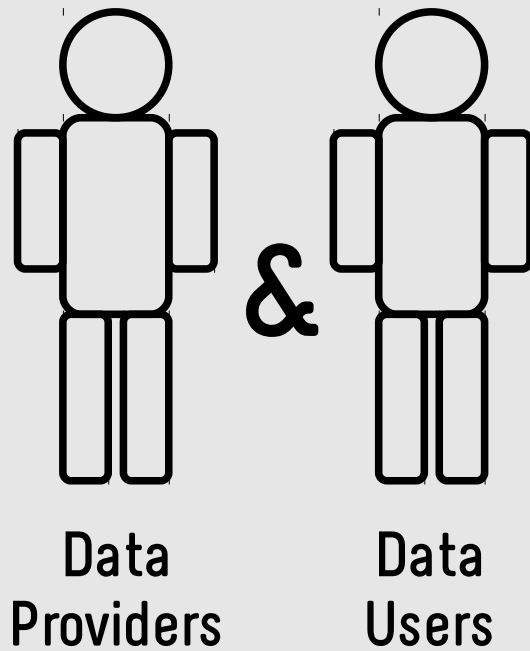
There is still very important the role of humans, because an automation is possible just for selected particular aspects.

Don't ask



We need to
give an advice
on spatial data
harmonization
of our data

Ask



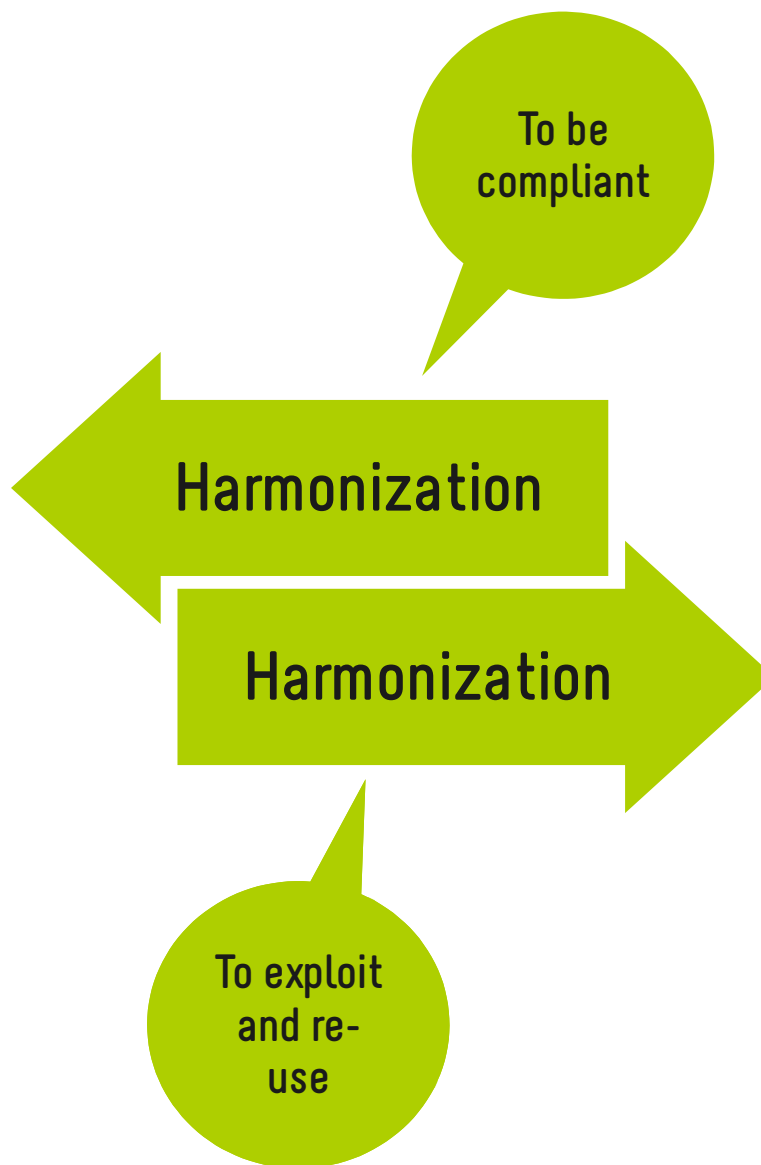
We know what,
where, how...
We need to
give an advice
how to do.

Harmonization & INSPIRE

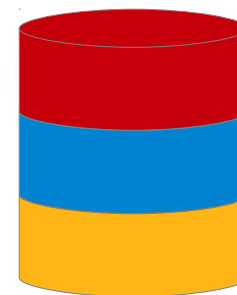
INSPIRE Data Specifications



Minimal common
set of classes and
attributes



Your Data



Conclusion

- You need to know reasons of harmonization.
- The cooperation of participated subjects is necessary.
- The role of harmonization expert (manager) is irreplaceable in case of large harmonization projects.
- Harmonization is strongly connected with experience, knowledge and information on data, domain and technologies.
- The harmonization is one coherent process, it is composed of various sub-processes.
- It is not possible any complex automation of harmonization.
- Use well-describe components – your data will be harmonizable.
- Sometimes is cheaper and more efficient to create new data than harmonize