GIS4EU

Provision of interoperable datasets to open GI to EU communities

Deliverable D-3.6
Rules for Merging Cross-border Datasets

Gobe Hobona, Mike Jackson
as editors

Magistrato alle Acque di Venezia

e**Content***plus* This project is funded under the *e*Content*plus* programme a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

Resume

Title	**Rules for Merging Cross-border Datasets**
WP to refers	WP 3
Document Code	D 3.6
Creator	Gobe Hobona, Mike Jackson
Responsible Partner	UNOTT
Creation date	29/10/2008
Subject	GIS4EU Merging Rules
Status	Realised
Publisher	Repository
Type	Deliverable
Description	Rules for merging cross-border datasets
Contributor	Gobe Hobona, Robert Laurini , Carmelo Attardo, Mike Jackson, Maria Pla, Mauro Vasone, Silvano De Zorzi, Pál Lévai, Marco D'Orazi, Lubica Buchelova, Stefania De Zorzi, Rui Reis, Sergio Farruggia, Piotr Krock, Catherine Roussey, Roberta Luca, Mauro Vasone, Tamas Schmauder, Otto Tancsik, Alina Kmiecik, Juraj Valis, Giovanni Rocca
Format	Doc
Source	Gis4Eu Project
Version	1.0
Dissemination level	Restricted
File Namer	d3_6_merging_rules_final.doc
Keywords	Gis4eu, harmonisation rules, guidelines, cross-border, aggregation
Language	English
Relation	D2.2 Data Descriptions, D3.1 Requirements for a Common Data Model, D2.3 Data Provider Requirements, D2.4 User Requirements

Table 1 - Document classification resume

INDEX

1 Summary

The GIS4EU project aims at providing base cartography datasets (administration units, hydrography, transportation networks and elevation themes) for Europe, and to ensure its cross-scale, cross-language and cross-border interoperability and accessibility according to International Standards and INSPIRE requirements.

The focus of this document is to identify different situations that could be encountered during the **merging** of datasets and to recommend rules for addressing those situations.

2 Document Scope

The GIS4EU Description of Work states the following specification regarding merging and degradation:

T3.3 Data harmonisation/aggregation processes specifications by the definition of the rules to harmonise and aggregate data in order to:

1. Merge cross-border datasets (TWGs-A)

2. Degradation rules in order to pass to less detailed scale (TWGs-A)

T3.4 Correlation at the multilingual scale by definition of rules to describe elements and to aggregate them at different scale and geographic regions. Defining the related ontology.

This document describes only those rules that relate to aggregation of datasets at the same scale from different geographic regions (i.e. cross-border). A related document, D3.7 discusses aggregation of datasets of different scales.

The document presumes that the two datasets are of similar qualities, either geometric (coordinates, same geometric representation and topology) or temporal (approximately same updating dates). The document presumes that the datasets have similar feature types and attributes though potentially named differently (e.g. in different languages).

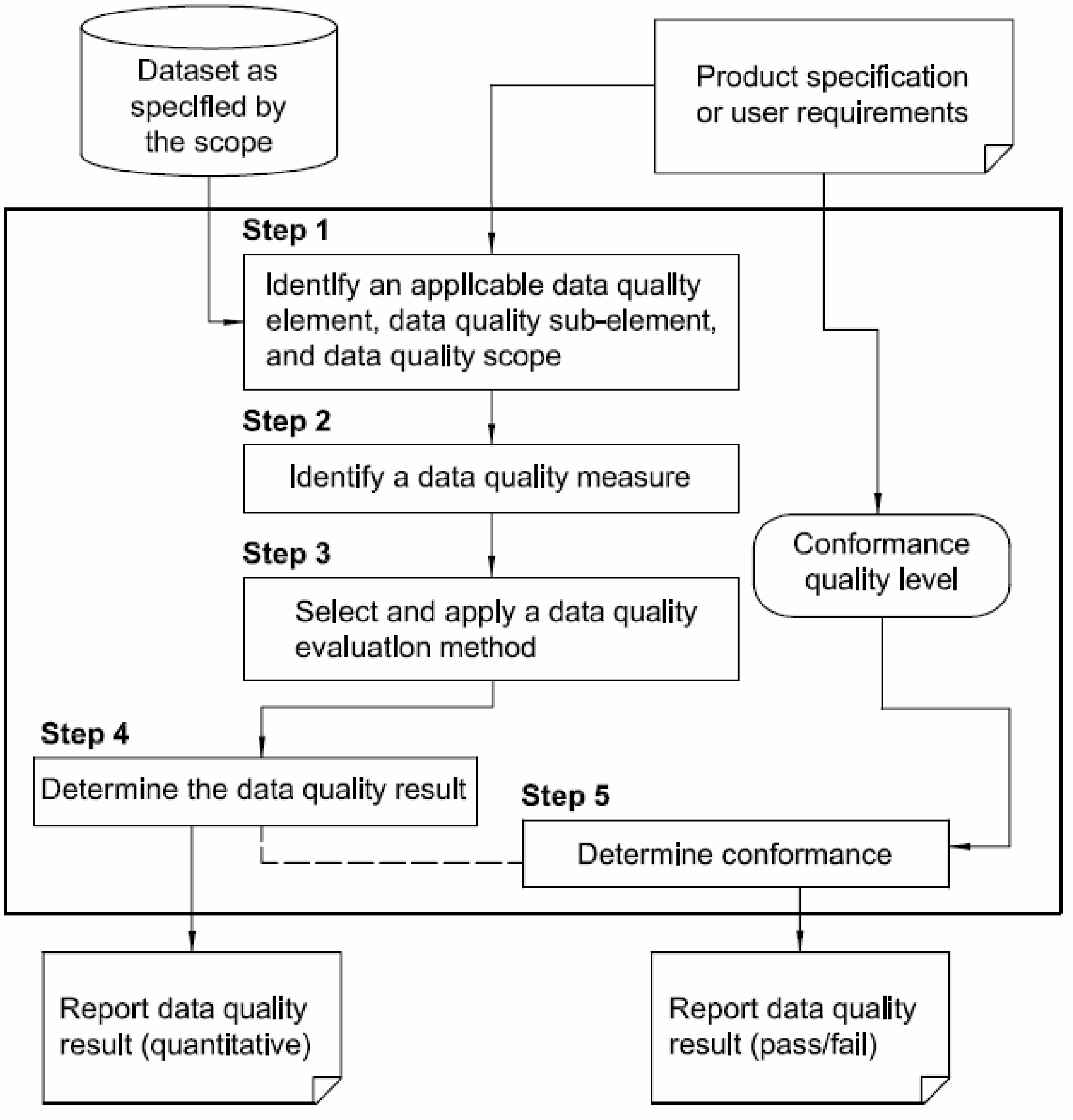
3 Concepts of Geographic Data Integration

Geospatial datasets potentially differ in several ways. First, the coordinate reference systems could differ thereby requiring that the coordinates be transformed into the same coordinate reference system. Without transforming the coordinates, the datasets being integrated are likely to be misaligned as they fall within different reference systems. Most GIS software will automatically convert data between different coordinate reference systems. Second, the geometries of datasets being integrated are not likely to topologically touch but instead will present overshoots and undershoots, if they have been surveyed by different organisations. This can present problems for themes such as Administrative Units particularly as administrative boundaries determine jurisdictions. Unfortunately, at present addressing undershoots and overshoots currently requires a user to validate and edit the geometries. Third, if the datasets have been created by different organisations may not use the same taxonomies, classifications on ontologies. Therefore even if the feature properties are harmonised into the same application schema, it is necessary to reclassify codelists and possibly apply conversion factors to some numeric values in order to harmonise them. Further, the language and character encodings used in the input datasets may differ thereby requiring translation into a common language and a common character encoding.

This document presents a use case and explanations of rules for addressing these variables.

3.1 Evaluation of Data Quality

Before carrying out any merging of datasets, it is necessary to evaluate the quality of each dataset that is to be merged with another. Assessing data quality will help determine which feature properties should be retained and which should be discarded if there are any geometric and attributive properties duplicated during aggregation. The five elements of data quality identified by ISO 19114 are completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy. The following figure presents a workflow for evaluating and reporting data quality. In data aggregation, we consider Step 5 to be a determination of preference between the input datasets (where the highest quality dataset is the most preferred).

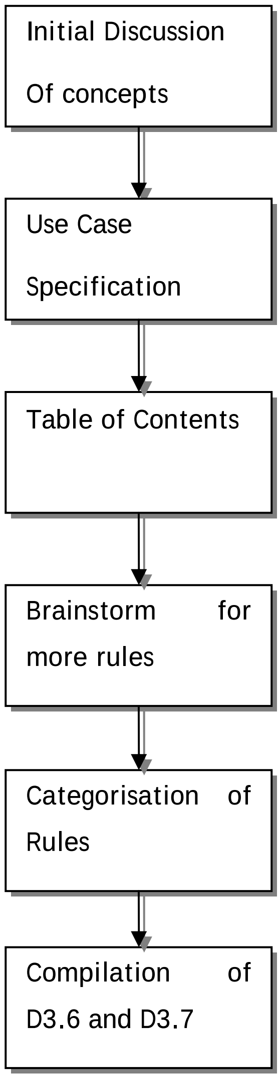


Process for Evaluating and Reporting Data Quality (Source: ISO 19114)

RULE 1: Each dataset produced through aggregation should be accompanied by metadata that describes the lineage of the aggregated dataset.

RULE 2: The quality of the input datasets shall be evaluated before any aggregation is undertaken by examining the metadata of the input datasets according to the five elements of data quality listed in ISO 19114.

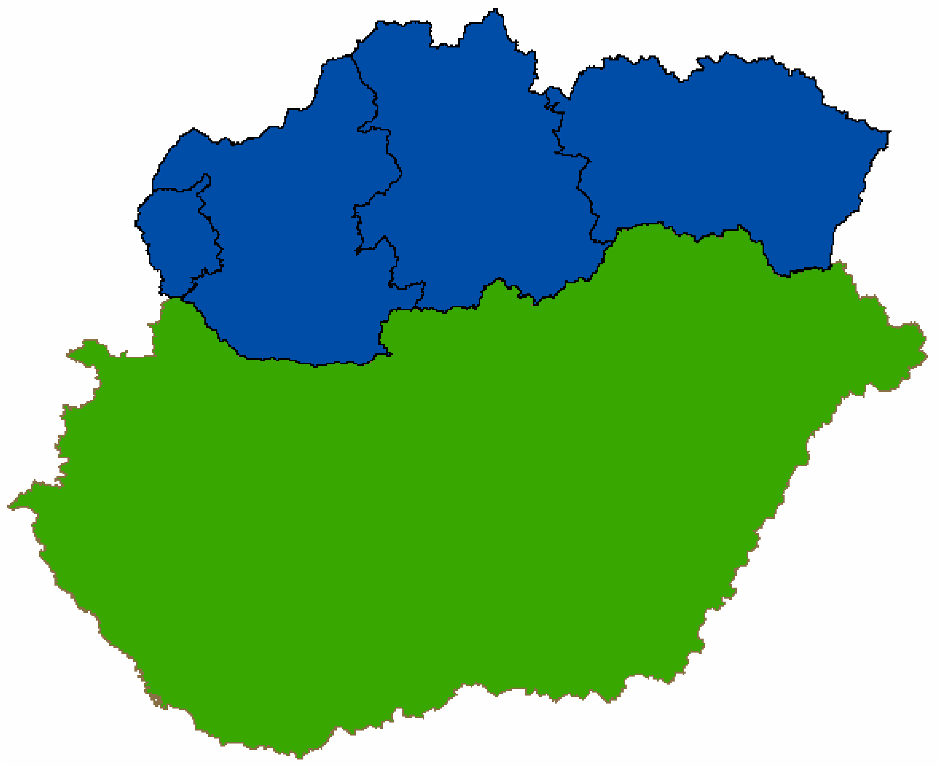
4 Methodology



The rules described in this document were developed through teleconference discussions within WP3 on January 16th, February 11th and March 2nd 2009. It was decided to use the same methodology for D3.6 and D3.7 because they both aimed to produce rules for aggregating datasets. After a teleconference to kick off tasks T3.3 and T3.4, draft documents were prepared listing possible situations as the table of contents for the deliverable. Use cases were then designed to provide a focus for the merging and degradation rules. Participants thereafter brainstormed and suggested rules for both D3.6 and D3.7. Afterwards, a teleconference was arranged during which rules were categorised into merging and degradation. An initial draft deliverable was prepared and revised within WP3. A further revision was distributed at the IV Technical Meeting, on March 16th and 17^{th,} to all participants for further discussion and distributed electronically to all project partners for review and commenting thereafter. Therefore, the final deliverable has been prepared through consultation internally and externally of WP3.

5 Use Case Specification

This chapter presents a use case specification that covers some of the aspects of the integration of heterogeneous geographic datasets. It should be noted that several aspects of data integration are applicable to all Aggregation TWGs, therefore this document presents a single use case that addresses various aspects of integration of geospatial data.



The sample data used in this use case is from separate datasets showing Hungary and Slovakia. In the previous figure, Slovakia is presented in blue and Hungary in green. The following use case presents an illustration of rules for merging cross border datasets. It should not limit the use cases adopted by Aggregation TWGs. Instead it aims to provide guidance on how and where to apply the rules described in this document.

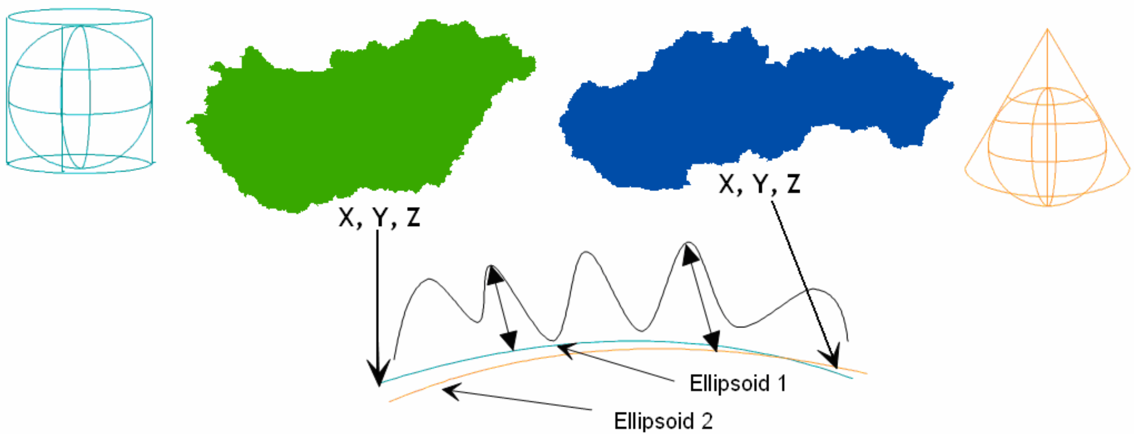
Name	Merging of Cross Border Datasets
Description	In this use case a sample dataset showing counties of Slovakia is to be merged with a dataset showing the country outline of Hungary. Additional data showing roads at a border between Slovakia and Hungary will also be merged. The final result of the use case should be a dataset with the country outlines of Hungary and Slovakia; including merged roads.
Pre-condition	Each data provider uploads their shapefiles into a Web Feature Service (WFS) and the data is then published in Geography Markup Language (GML).
Flow of Events	

Step 1	The Hungary data is retrieved from the WFS and the GIS4EU portal converts the data into ETRS89. If necessary an affine transformation is applied to translate or rescale the features.
Step 2	The Slovakia data is retrieved from the WFS and the GIS4EU portal converts the data into ETRS89. If necessary an affine transformation is applied to translate or rescale the features.
Step 3	One of the WFS is set to export data in ISO 8859-16 character encoding. In order to ensure that the data is interoperable with the other dataset, the GIS4EU portal converts the data into UTF-8.
Step 4	The user retrieves the transformed data from the GIS4EU portal and if any gaps exist between features, the gaps are 'filled' through the moving of features to a Hausdorff midline.
Step 5	The Slovakia dataset shows counties and the topology of the counties is known (i.e. we know that the county boundaries touch). Therefore, through the GIS4EU portal, the counties are merged into a single feature showing the county outline.
Step 6	The same roads in both datasets do not touch. Therefore, an artificial line is added to 'connect' the roads.
Step 7	The user wishes to eventually interpolate the altimetry of the roads from elevation data from the two data providers. One elevation dataset is encoded as a regular grid of rectangular cells and the other as a triangular irregular network (TIN). The two elevation datasets are merged through conversion of the rectangular grid to a TIN, interpolating new points between the two TINs.
Step 8	The user requires the data to be output to the INSPIRE data models. The attributes are mapped to their INSPIRE equivalents and codelists are reclassified.
Step 9	Any numeric attributes showing the same measures (e.g. distance) but in different units (e.g. kilometres, metres) are harmonised by applying a conversion factor.
Post-condition	The input datasets have been harmonised and merged into the INSPIRE data specification.
Data Source	

Description	Slovakia counties
Data Provider	VUGK
Geographic Scope	N: -892652 S: -154790 E: -1339821 W: -601958 S-JTSK Krovak projection (EPSG: 2065)
Delivery	Shapefile
Data Source	
Description	Hungary country boundary
Data Provider	FOMI
Geographic Scope	N: 1302868 S: -896154 E: 1781410 W: -417612 Contents of the Projection (PROJ) file are: PROJCS["unnamed",GEOGCS["unnamed",DATUM["D_European_Datum_1950", SPHEROID["International-1924",6378388.0,297.00000000006]],PRIMEM["Greenwich",0.0], UNIT["degree",0.0174532925199433]],PROJECTION["Transverse_Mercator"], PARAMETER["false_easting",650000.0],PARAMETER["false_northing",200000.0], PARAMETER["central_meridian",19.048569],PARAMETER["scale_factor",0.99993], PARAMETER["latitude_of_origin",47.144394],UNIT["METER",1.0]]
Delivery	Shapefile

6 Rules for Geometric Properties

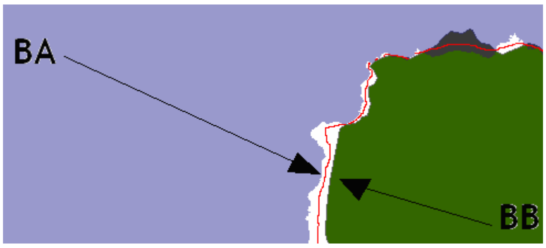
6.1 Coordinate Reference Systems



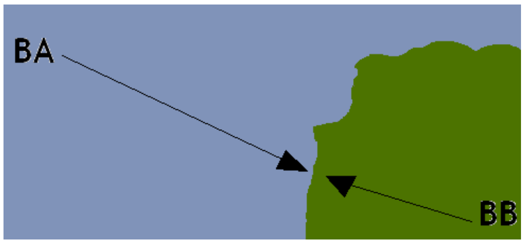
Geographic datasets have historically been collected and maintained in a variety of coordinate reference systems. For example, the United Kingdom uses OSGB 1936(EPSG:4277) whereas France uses the New Triangulation of France Lambert zones I, II and III (EPSG:27561,27562 and 27563). When datasets in different coordinate reference systems are being harmonised they do not align in an accurate representation of the earth. It is necessary to ensure that the datasets are in the same coordinate reference system to ensure that the spatial objects align accurately. Mathematical algorithms are applied to coordinates of geometries to effect a coordinate transformation; this capability is built into most modern Geographic Information Systems. The INSPIRE directive lists coordinate reference systems as an Annex I theme and consequently the INSPIRE TWGs have developed a data specification for harmonising coordinate reference systems.

RULE 3: If an input dataset is not in ETRS89, then the requirements listed in the INSPIRE specification for Coordinate Reference Systems should be applied.

6.2 Boundary Force fitting



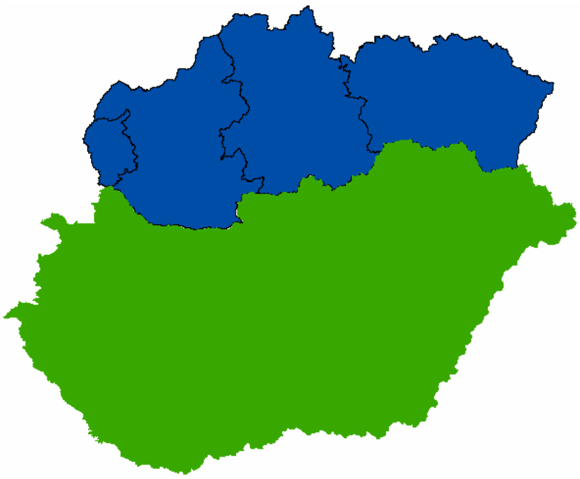
BEFORE



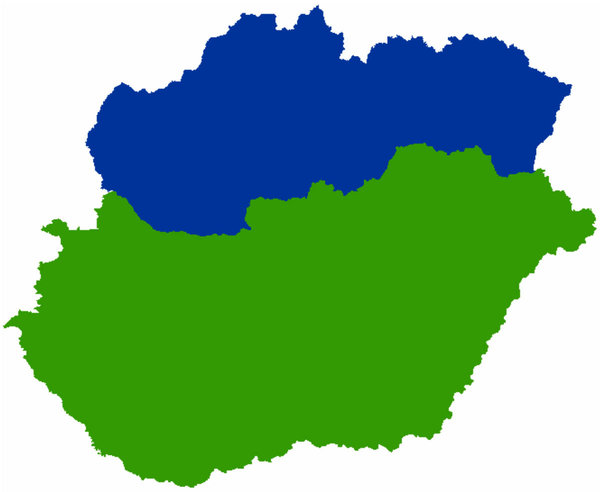
AFTER

When integrating cross-border datasets it is possible to find that some of the polygon or linear features overlap. Where the datasets have been produced by different surveys it is likely that parts of a polygon will cover part of an adjacent polygon. Further, if the input datasets are in different coordinate reference systems, when they are reverse projected into the same coordinate system gaps may be observed between borders due to different projection scale factors. To address this scenario Boundary force-fitting could be applied using the Hausdorff distance between the two overlapping boundaries [Aspert]. In the previous figure the midline is coloured red between the boundaries. However, any boundary force fitting will make the data inadmissible in any legal context. Therefore, any modified dataset should not be used in any legal context.

RULE 4: If Hausdorff Distance between BA and BB < epsilon, then BA and BB are both replaced by Midline(BA,BB); if and only if, the aggregated datasets are **not** to be used in a legal context.



BEFORE

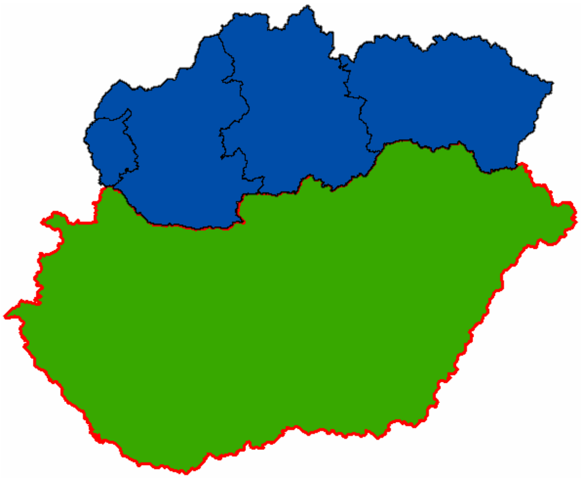


AFTER

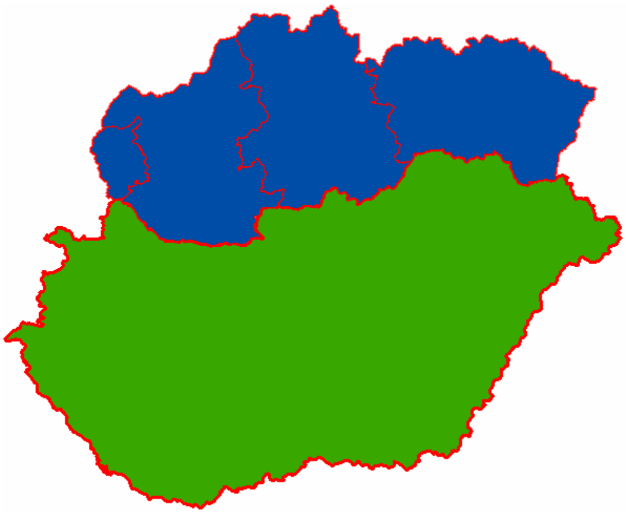
It is also possible to find that two artificially cut datasets may have features along the cutting line that have exactly the same attribute types, names and values. In such a scenario, the artificially cut features may be replaced by their geometric union. Care should be made to ensure that features with different unique identifiers are not merged, even if all the attributes are exactly the same. This is because the identifiers may relate to different records in other databases.

RULE 5: If two adjacent polygonal features have been force fit and they are of EXACTLY the same type, then they can be replaced by their union.

6.3 Topological Integration



BEFORE

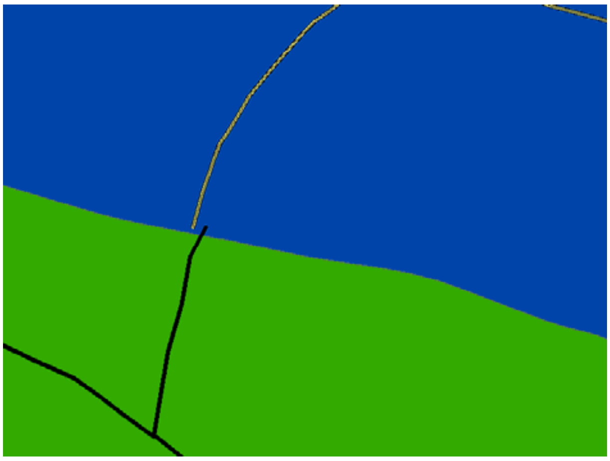


AFTER

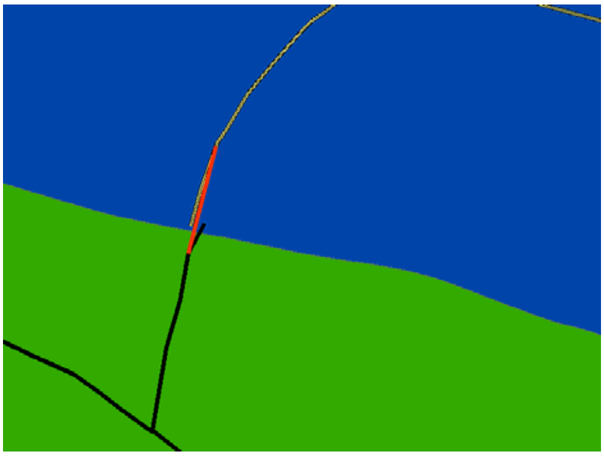
It is possible that the datasets being integrated may contain geographic features that share topological edges. Where the topological connectivity is known (for example, the boundaries of adjacent countries or municipalities) the boundary of one of the adjacent polygons may be dissolved to create an edge shared by the adjacent polygons. This rule can only be applied if the edges are known to be topologically equal, for example if the data geometries were surveys by the same National Mapping Agency but the attributes maintained by different municipalities.

RULE 6: If topological connectivity between two zones is explicitly known and the quality of the input datasets is different, then the topologically equal edges may be replaced with a shared edge taking the boundary data with the higher level of quality.

6.4 Linear Extremity Integration



BEFORE



AFTER

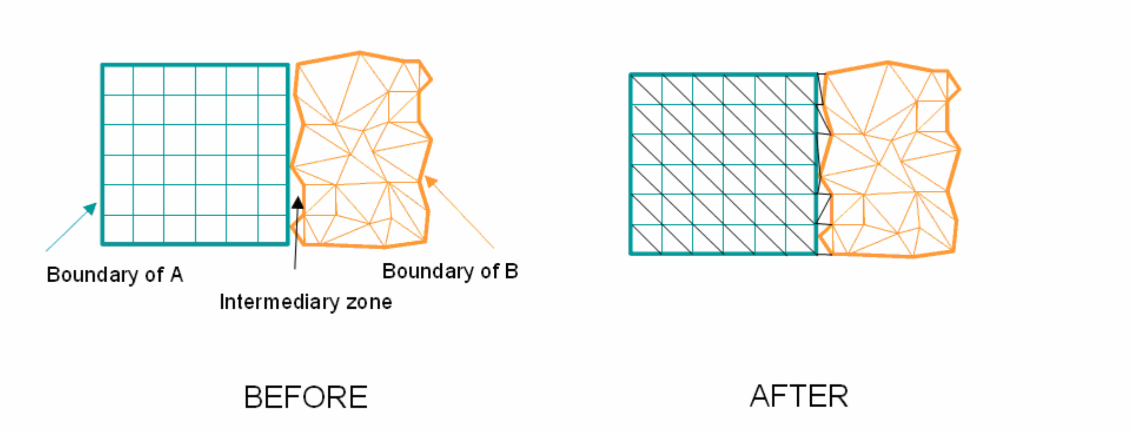
Where two adjacent datasets contain a geographic feature representing the same linear feature (e.g. river or roads), it may occur that the nodes of the linear features do not coincide. To address this situation, coordinates of the nodes may be translated to the midpoint of the nodes. Moving the nodes may consequently require other vertices to be moved in order to ensure a smooth line is generated.

RULE 7: If a linear feature is represented in two adjacent datasets, then coincidence of the end nodes should be checked.

RULE 8: If the end nodes are not coincident, then the nodes should be translated to their midpoint.

RULE 9: If the generated line is too jagged(i.e. zig zag formation), then other vertices should be moved to make the line smooth.

6.5 Geometric Integration of Raster Datasets



Raster datasets such as elevation or surface models from two adjacent geographic extents may need to be integrated. Geometrically, the datasets may be encoded as regular grids or triangular meshes. If the geometries of cells are both regular grids then using raster algebra, the two datasets can be combined. If the geometries of cells are NOT the same then they should both be converted to triangular meshes and any intermediary spaces filled with triangles. Regardless of which approach is used to geometrically integrate the datasets, it will be necessary to reclassify the datasets so as to harmonise the values of the pixels.

RULE 10: If adjacent raster datasets are of different geometries, then transform both into triangles and fill the intermediary zones by additional triangles. A regular grid may then be interpolated from the triangular irregular network (TIN); if and only if, the aggregated datasets are **not** to be used in a legal context.

7 Rules for Alphanumeric Properties

Some of the alphanumeric attributes/properties contained in datasets supplied to GIS4EU include placenames, quantities, unique identifiers and classification labels. Although there may be other types of attributes, the use cases specified in this document consider only the aforementioned.

7.1 Different Languages

When integrating datasets with attribute values expressed in different languages for example "országhatár" is Hungarian for "national boundary", it is necessary to convert the words to the same language. Therefore, the use of a look-up table or ontology is required. A look-up table is the simplest solution as it presents a spreadsheet-like mapping between words. A multilingual glossary expressed as an ontology offers more capability as words can be transformed based on inheritance trees e.g. a skyscraper is a building and a building is a structure.

RULE 11: If the meaning of a feature type or attribute is known to be similar to one in the INSPIRE data model, then the feature type or attribute name will be modified to adopt the name of its INSPIRE equivalent. Similarly the **data types** (floating point, character strings, integers) will adopt those of the INSPIRE data model.

RULE 12: If the complete codelist of an input attribute is known to be a subset of the multilingual INSPIRE/GIS4EU glossary AND if the attribute does **NOT** hold placenames, then the values shall be translated to English.

RULE 13: if a placename has several names, then use the local name.

7.2 Different Units

When integrating datasets with expressions of quantities or magnitudes, it is necessary to ensure that the quantities refer exactly to the same phenomenon and also in the same units. For example, distance could be expressed in either metres or miles. It is therefore necessary to apply a conversion factor to numeric attributes that represent quantities or magnitudes if they are not based on the same units as the output schema. Please note that numbers are sometimes used in codelists, therefore it is important to make sure that the numeric values are quantities or magnitude and NOT codelists.

RULE 14: If the attributes of features in input datasets show quantities or magnitudes AND if the quantities or magnitudes are not in the same units as the output data, then a conversion factor shall be applied to the attributes on the input datasets to match the unit used by the output dataset.

7.3 Duplicated Features

When integrating datasets there may be some features or attributes duplicated, particularly if the datasets were originally created by the same source. Features are considered duplicated if there is another feature in the same dataset that has exactly the same attribute values. In such a scenario, it is necessary to identify unique attributes and then delete the duplicated attributes. Care should be made to identify duplicated attributes accurately, i.e. both the data type, name and codelists of the extra attribute must match those of the other attribute. Care should also be made to ensure that the semantics of the feature are not changed by the deletion of the duplicated attributes.

RULE 15: If any features contain exactly the same attributes AND the attributes have the same names and type, then the duplicated features should be deleted If and Only If the semantics of the dataset are not affected.

7.4 Different Raster Classifications

When integrating raster coverages it is possible to find that pixels in one coverage adopt a different classification(or codelist) to pixels in another coverage. Once the coverages have been geometrically merged and resampled, it is then necessary to reclassify the pixel values to ensure that classifications(or codelists) match. Reclassification requires that the codelist and the meanings of the labels in the codelist be known. It is necessary to ensure that the attribute values of the higher quality dataset are preserved.

RULE 16: If two raster datasets being merged are known to have different classifications (or codelists), then the pixel values of one of the datasets has to be reclassified into the classification of the output raster; if and only if the semantics of the classification labels are known.

7.5 Character Encodings

When integrating datasets it is necessary to determine whether the datasets are in the same character encoding. Although currently the Unicode (UTF-8) is the most popular, some datasets may adopt other character encodings such as those defined in the ISO 8859 standard. The default encoding for XML documents is UTF-8 and as GIS4EU is going to use WFS as data sources, it is expected that the data will already be in UTF-8 encoding.

RULE 17: If character encoding of an input dataset is not in Unicode UTF-8, then the dataset will be converted to UTF-8.

7.6 Handling Identifiers

Originally all objects have a local identifier in both datasets; for instance there can be two different objects having the same identifier in both datasets. In order to avoid ID collision, fresh global identifiers must be conferred to each merged object.

RULE 18: if the objects in input datasets do not have identifiers, then they will be given unique identifiers in the merged dataset

RULE 19: if the objects have been given identifiers prior to merging, then new unique identifiers will be generated for each spatial object.

7.7 Feature Reconstruction

Several times, a geographic feature can be artificially cut in several database objects, for instance roads, rivers or lakes crossing a border. More there exist buildings artificially-cut in two pieces, for example along the French and Swiss borders. In this case, in the merging procedure those objects must be merged into a single object. So, several issues arise:

- the newly-reconstructed object must be topologically consistent (unique geometric representation, for instance a river must be a single polyline),
- the newly-reconstructed object must receive a unique placename and all its attributes must be merged,
- the newly-reconstructed object must replace the different parts existing in both datasets which must be discarded; as a consequence it will receive a single ID, and old ID's must be discarded.

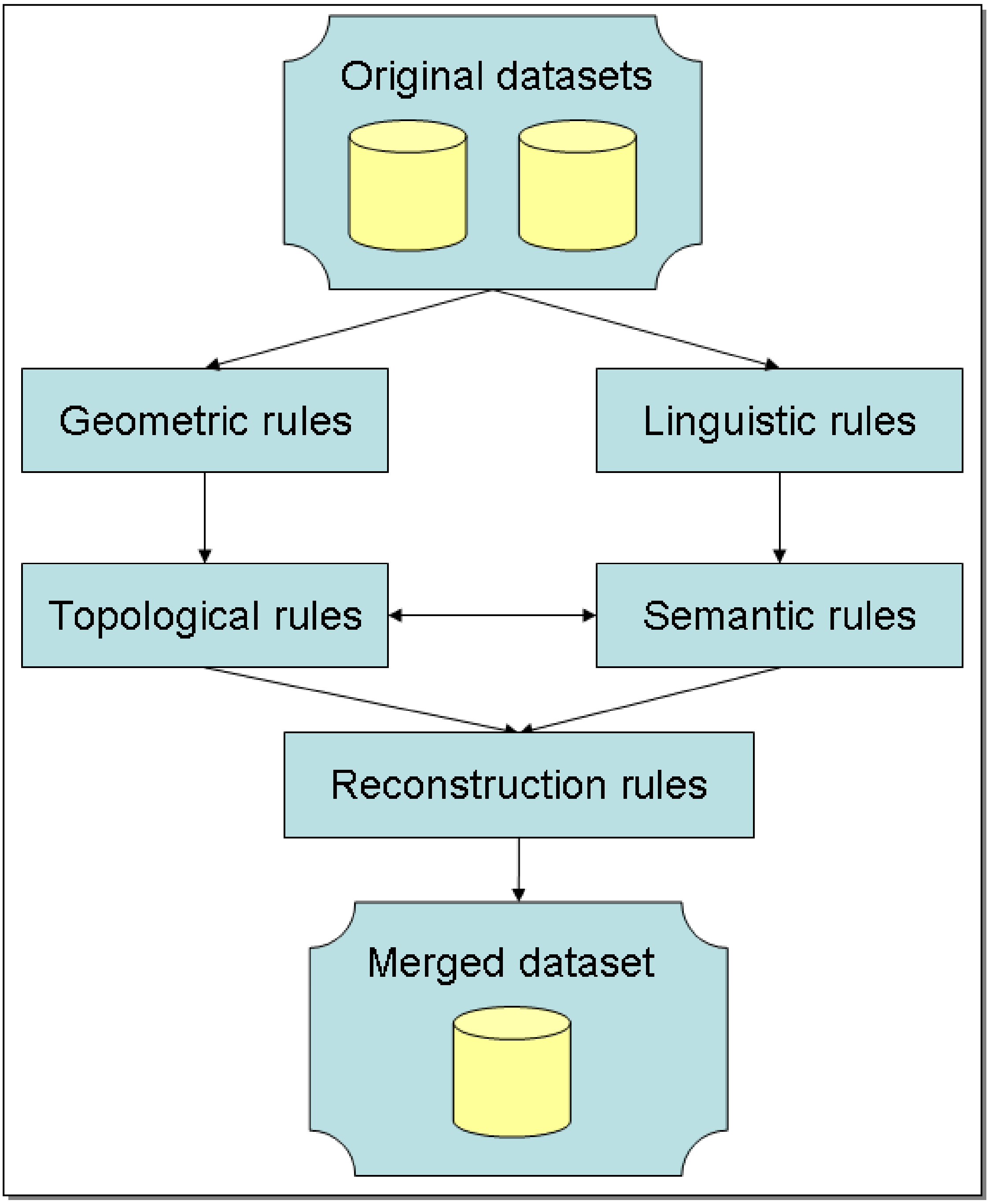
As a test for linear features, a minimum path algorithm can be run successfully between a city located in one original dataset, and a second city located in the second original dataset. The test succeeds if the path returned equals (or approximately equals) the known distance between the two cities. If the test does not succeed then some edge connections are missing and Rules 7, 8 and 9 on **Linear Extremity Integration** should be applied.

A test for polygonal features would be to determine whether the area of the artificially cut features is equal or (or approximately equal) to the known value. In all cases it is necessary to ensure that the attribute values of the higher quality dataset are preserved.

RULE 20: if a single feature is artificially cut into two object parts, each belonging to a different dataset, then the two objects should be reconstructed into a single new object through a geometric union with attributes copied from one dataset to the other. The newly reconstructed object should have a unique ID.

8 Conclusion

Rules applicable in the merging of cross border datasets can be classified according to whether they are geometric, semantic, linguistic, topological and reconstruction-oriented. However, some rules will fall into more than one category for example identification of artificially cut features may involve examination of semantic properties but will have an effect on the geometries. As a guide, the following figure illustrates the order in which the rules should be applied. In conclusion, great care should be taken not to lose much information when datasets are being integrated by preserving the semantic and geometric characteristics of the datasets where possible.


Original datasets
Geometric rules
Linguistic rules
Topological rules
Semantic rules
Reconstruction rules
Merged dataset

9 Appendix

9.1 Abbreviations

CEN	Comité Européen de Normalisation
CEN/TC287	CEN Technical Committee 287 Geographic Information
CRS	Coordinate Reference System
CSL	Conceptual Schema Language
DCP	Distributed Computing Platform
DIS	Draft International Standard
DNS	Domain Name System
DT	Drafting Team
DT-DS	Drafting Team “Data Specifications”
DT-DS TWG TN	Drafting Team “Data Specifications” Thematic Working Group on Transport Networks
DTI	Draft Implementing Rules
EC	European Commission
EPSG	European Petroleum Survey Group (now OGP Surveying & Positioning Committee)
ESDI	European Spatial Data Infrastructure
EU	European Union
EUREF	Reference Frame sub commission for Europe (IAG Commission I)
FGDC	Federal Geographic Data Committee
GCM	Generic Conceptual Model
GEODRM	Geospatial Digital Rights Management
GI	Geographic Information
GML	Geography Markup Language
GNM	Generic Network Model
ICT	Information and Communication Technology
IRS	Implementing Rules

INSPIRE	INfrastructure for SPatial InfoRmation in Europe
ISO	International Standardization Organisation
ISO/TC211	ISO Technical Committee 211 Geographic information/Geomatics
JRC	European Commission Directorate General Joint Research Centre
LBS	Location Based Services
LMO	Legally Mandate Organisation
LoD	Level of Detail
OCL	Object Constraint Language
OGC	Open Geospatial Consortium
OGP	international association of the Oil & Gas Producers
OSF	Open Software Foundation
OWL	Web Ontology Language
PSI	Public Sector Information
SDI	Spatial Data Infrastructure
SDIC	Spatial Data Interest Communities
SI	International System of Units
SOAP	Simple Object Access Protocol
TIN	Triangular Irregular Network
UML	Unified Modelling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Universal Resource Names
UUID	Universally Unique IDentifier
WCS	Web Coverage Service
WFS	Web Feature Service
WMO	World Meteorological Organization
WMS	Web Map Service
XML	eXtensible Markup Language

Table 2 - Abbreviation list of document content

9.2 Glossary

Term	Definition
Application data	Data in support of user requirements
Application schema	Conceptual schema for data required by one or more applications [ISO 19101:2002(E)]
Class	Description of a set of objects that share the same attributes, operations, methods, relationships, and semantics [ISO 19107:2003(E)]
Code list	Value domain including a code for each permissible value [N1784]
Conceptual model	Model that defines concepts of a universe of discourse [ISO 19101:2002(E)]
Conceptual schema	Formal description of a conceptual model [ISO 19101:2002(E)] Note: ISO 19107 contains a formal description of geometrical and topological concepts using the conceptual schema language UML.
Conceptual schema language	Formal language based on a conceptual formalism for the purpose of representing conceptual schemas [ISO 19101:2002(E)] Notes: UML, EXPRESS, ORM and INTERLIS are examples of conceptual schema language
Coordinate reference system	Coordinate system that is related to the real world by a datum [ISO 19111:2003(E) – Modified] Note: ISO19111 defines coordinate reference system as coordinate system that is related to the real world by a datum 2: Following ISO19111, temporal reference systems are understood as covered by the term coordinate reference systems as well. Examples are: ETRS89 and any formally defined national coordinate system such as the ITM (Irish Transverse Mercator).
Coverage	Spatial objects that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain. [ISO 19123:2005(E) – Modified] Examples are Orthoimage, digital elevation model (as grid or TIN), point grids etc
Data	Reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing [ISO/IEC 2382-1]. Note 1: Data can be any form of information whether on paper or in electronic form. Data may refer to any electronic file no matter what the format: database data, text, images, audio and video. Everything read and written by the computer can be considered data except for instructions in a program that are executed (software). Note 2: Services can provide things like WMS (a picture of a map), WFS (GML) and WCS (an image). Then there are services where a user supplies a coordinate and the service transforms it to another coordinate, or a user supplies an image and the service transforms or performs image processing. These are all something that can be read and written by the computer and are in accord with note 1 data.
Data harmonization	Providing access to data through network services in a representation that allows for combining it with other harmonized data in a coherent way by using a common set of data product specifications this includes agreements about coordinate reference systems, classification systems , application schemas etc.
Data interchange	Delivery, receipt and interpretation of data [ISO 19118].

DATA MODEL	A model that defines in an abstract way how data is represented in an information system or a database management system
DATA PRODUCT SPECIFICATION	Detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party [ISO/DOS 19131].
DATA SPECIFICATION	Data product specification that describes datasets of a specific theme in a harmonized way [N1786].
DATA TRANSFER	Movement of data from one point to another over a medium [ISO 19118].
DATASET	Identifiable collection of data [ISO 19115:2003(E)].
DATASET SERIES	Collection of datasets sharing the same product specification [ISO 19115].
DISCOVERY METADATA	The minimum amount of information that needs to be provided to convey to the inquirer the nature and content of the data resource Note: The above definition falls into broad categories which answer the "what, why, when, who, where and how" questions about spatial data.
E-GOVERNMENT	Application of information and communication technology to enhance the effectiveness of a legislature, judiciary or administration, either to improve efficiency or to change the relationship between citizen and government, or both
ENCODING	Conversion of data into a series of codes [ISO 19118].
ENTITY	Real-world phenomenon
ESDI	European Spatial Data Infrastructure as built and based on the INSPIRE framework directive]
EVALUATION	Providing sufficient information to enable an inquirer to ascertain that data fit for a given purpose exists, to evaluate its properties, and to reference some point of contact for more information (adapted from GSDI Cookbook). Note: metadata include those properties required to allow the prospective end user to know whether the data will meet the general requirements of a given problem.
EXCHANGE FORMAT	Structured representation of data in a document for exchange between systems In most cases, a machine readable schema will document the structure of the data in the exchange document. Example: GML encodes the application schema in XML schema
EXONYM	Name used in a specific language for a spatial object situated outside the area where that language is spoken, and differing in its form from the name used in an official or well-established language of that area where the geographical feature is located UNGEGN Glossary of Terminology: http://unstats.un.org/unsd/geoinfo/glossary.pdf - Modified
EXTERNAL [OBJECT] IDENTIFIER	A unique [object] identifier which is published by the responsible body, which may be used by third parties to reference the spatial object
FEATURE	Abstraction of a real-world phenomena. Note: The term "(geographic) feature" as used in the ISO 19100 series of International Standards and in this document is synonymous with spatial object as used in this document. Unfortunately "spatial object" is also used in the ISO 19100 series of International Standards, however with a different meaning: a spatial object in the ISO 19100 series is a spatial

	geometry or topology. [ISO 19101].
FEATURE CATALOGUE	Catalogue(s) containing definitions and descriptions of the feature/object types, their attributes and associated components occurring in one or more spatial data sets, together with any operations that may be applied [ISO 19110:2005(E) – modified].
FEATURE DATA DICTIONARY	Dictionary containing definitions and descriptions of feature concepts and feature-related concepts [ISO/CD 19126].
GAZETTEER	Directory of instances of a class or classes of features containing some information regarding position A gazetteer can be considered as a geographical index or dictionary of spatial objects [ISO 19112].
GENERAL FEATURE MODEL	Metamodel for spatial object types and their property types [ISO 19109]
GEOGRAPHIC FEATURE	Synonymous with spatial object
GEOGRAPHIC IDENTIFIER	Spatial reference in the form of a label or code that identifies a location [ISO 19112:2003(E)]. Example 1: Paris, [river] Rhine, Mont Blanc Example 2: Postal codes: 53115, 01009, SW1, IV19 1PZ
GEOGRAPHICAL GRID SYSTEMS	Harmonized multi-resolution grid with a common point of origin and standardized location and size of grid cells. Note: Geographical grid systems are not limited to rectified grids or grids using cell axes parallel to the meridians
GEOMETRIC PRIMITIVE	Geometric object representing a single connected, homogeneous element of space [ISO 19107].
GLOSSARY	An alphabetical list of words often defined or translated: dictionary, lexicon, vocabulary, wordbook
HOMOLOGOUS SPATIAL OBJECTS	Set of spatial objects that correspond to the same real world entity, but are represented differently according to different levels of details or point of views
INSPIRE APPLICATION SCHEMA	Application schema specified in the INSPIRE implementing rules
INSPIRE DATA SPECIFICATION	Data product specification for a spatial data theme from Annex I, II or III of the INSPIRE Directive
INSPIRE INFORMATION MODEL	A structured collection of components that will be documented to support the interoperability and harmonization of geographic information across Europe. Note: rules for application schema, identifier management, terminology etc are examples of the components.
INTEROPERABILITY	Possibility for spatial data sets to be combined, and for services to interact, without repetitive manual intervention, in such a way that the result is coherent and the added value of the data sets and services is enhanced.
LINEAR REFERENCE SYSTEM	Reference system that identifies a location by reference to a segment of a linear spatial object and distance along that segment from a given point [ISO 19116:2004(E) – modified]. Example: kilometer markers along a motorway or railway, references along the center line of a river object from the intersection with a bridge object. Note: synonymous with linear referencing system
METADATA	Information describing spatial data sets and spatial data services and making it

	possible to discover, inventory and use them [ISO 19115:2003(E)] The more general term as defined by ISO19115 is "data about data"
METADATA ELEMENT	Discrete unit of metadata [ISO 19115]
MULTICULTURAL	Multiplicity in systems of values held by different groups: ethnic, regional, or professional [Hofstede G. 1980. Culture's Consequences, Sage: London – modified].
MULTILINGUAL	In or using several languages
MULTIPLE REPRESENTATION	Representation of the relationship between homologous spatial objects
OBJECT	In this document is synonymous with spatial object
OBJECT IDENTIFIER	A unique identifier associated with a spatial object
OBJECT REFERENCING	A method of referencing thematic or other spatial objects to existing spatial objects describing their location to ensure spatial consistency across the spatial objects associated in this way in this way
PORTRAYAL	Presentation of information to humans [ISO 19117]
PRODUCT DESCRIPTION	Detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party [ISO 19113].
PROFILE	Set of one or more base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function. A profile is derived from base standards so that by definition, conformance to a profile is conformance to the base standards from which it is derived [ISO 19106].
REFERENCE DATA	Spatial objects that are used to provide location information in object referencing
REFERENCE MODEL	Architectural framework for a specific context, e.g. an application or an information infrastructure
REGISTER	Set of files containing identifiers assigned to items with descriptions of the associated items [ISO 19135].
RESOURCE	Asset or means that fulfills a requirement Example: dataset, service, document, person or organisation.
SERVICE	Distinct part of the functionality that is provided by an entity through interfaces [ISO 19119].
SPATIAL DATA	Any data with a direct or indirect reference to a specific location or geographic area NOTE The use of the word "spatial" in INSPIRE is unfortunate as in the everyday language its meaning goes beyond the meaning of "geographic" – which is considered by the Drafting Team as the intended scope – and includes subjects such as medical images, molecules, or other planets to name a few. However, since the term is used as a synonym for geographic in the draft Directive, this document uses the term "spatial data" as a synonym for the term "geographic information" used by the ISO 19100 series of International Standards.

Spatial object	An abstract representation of a real-world phenomenon related to a specific location or geographical area. NOTE It should be noted that the term has a different meaning in the ISO 19100 series. It is also synonymous with "(geographic) feature" as used in the ISO 19100 series.
Spatial object type	Classification of spatial objects NOTE In the conceptual schema language UML a spatial object type will be described by a class with stereotype <<FeatureType>>.
Spatial reference systems	System for identifying position in the real world, which does not necessarily use coordinates [ISO 19112:2003(E) -Modified]. EXAMPLE Geographic coordinates describing positions on the Earth surface (coordinate reference system), linear measurements along a river centreline from the intersection of a bridge (linear reference system), postal codes identifying the extent of postal zones (gazetteer)
Spatial schema	Conceptual schema of spatial geometries and topologies to be used in an application schema
Temporal reference systems	Reference system against which time is measured [ISO 19108;2002(E)].
Thematic application schema	INSPIRE application schema for an INSPIRE theme
Thematic data	Synonymous to application data
Thematic identifier	A descriptive identifier applied to spatial objects in a defined information theme EXAMPLE an administrative code for administrative area objects in the administrative units theme, a parcel code for parcel objects in the cadastre theme
Theme	Grouping of spatial data according to Annex I, II and III of the INSPIRE Directive
Topology	Topology is the science and mathematics of relationships used to validate the geometry of vector entities, and for operations such as network tracing and tests of polygon adjacency [Longley et al, 2001: page 190]
Transfer protocol	Common set of rules for defining interactions between distributed systems [ISO 19118]
Unique object identifier	A piece of data, usually in the form of printable characters, that unequivocally identifies a spatial object
Units of measurement	Defined quantity in which dimensioned parameters are expressed [ISO/TC211/N1791].
Use	Information required to access, transfer, load, interpret, and apply the data in the end application where it is exploited (adapted from GSDI Cookbook). Note: This class of metadata often includes the details of a data dictionary, the data organization or schema, projection and geometric characteristics, and other parameters that are useful to human and machine in the proper use of the spatial data.
Version	A particular form of something differing in certain respects from other forms of the same type of thing

VERSIONING	Applying a process to ensure that one version of something can be distinguished from another
XML SCHEMA	Means for defining the structure, content and semantics of XML documents

Table 3 - Table of abbreviation

9.3 References

9.3.1 Paper references

[Longley] P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, Geographic Information Systems and Science. West Sussex England: Wiley, 2005.

[Aspert] Nicolas Aspert, Diego Santa-Cruz, Touradj Ebrahimi(2002) "MESH : MEASURING ERRORS BETWEEN SURFACES USING THE HAUSDORFF DISTANCE" In Proceedings of the IEEE International Conference in Multimedia and Expo (ICME), vol. 1, pp. 705-708, Lausanne, Switzerland, August 26-29, 2002

[ISO 19114] International Standard for Geographic information — Quality evaluation procedures

[ISO 19138] Draft International Standard for Geographic information — Data quality measures

9.3.2 Web reference

- INSPIRE, 2007: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, 25.4.2007, L 108/1. April 25th, 2007.

  http://inspire.jrc.ec.europa.eu/directive/l_10820070425en00010014.pdf

- INSPIRE D2.3, 2008: Drafting Team "Data Specifications" – deliverable D2.3 - Definition of Annex Themes and Scope - Version 3.0. Drafting Team “Data Specifications” (DT-DS) - INSPIRE. March 18th, 2008.

  http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pdf

- European Petroleum Survey Group (EPSG) Coordinate Reference System code http://www.epsg.org

- INSPIRE Model, 2008: INSPIRE Consolidated UML Model - 1 October 2008 -1st draft, Revision 258 (corresponding to INSPIRE Data Specifications v1.00). Drafting Team "Data Specifications" (DT-DS) - INSPIRE. October 1st, 2008.

  https://inspire-twg.jrc.it/inspire-model/

List of Tables

9.4 Partner list

PARTICIPANT NAME	SHORT NAME	COUNTRY
Consorzio per il coordinamento delle Ricerche sul Sistema Lagunare di Venezia	CORILA	Italy
Intergraph (Deutschland) GmbH	GERIN	Germany
Vyskumny ustav geodezie a kartografie v Bratislave	VUGK	Slovakia
Universitat de Girona	UDG	Spain
Institut Cartogràfic de Catalunya	ICC	Spain
Geographical Information Systems International Group	GISIG	Italy
Földmérési és Távérzékelési Intézet	FÖMI.	Hungary

Regione Liguria	RLIG	Italy
Regione Piemonte	RPIE	Italy
University of Nottingham	UNOTT	United Kingdom
Comune di Genova	CGE	Italy
University Of Rome "La Sapienza"	UNISAP	Italy
Intergraph Polska sp. z o. o.	INGR	Poland
Instituto Geográfico Português	IGP	Portugal
Institut National des Sciences Appliquées de Lyon	INSA	France
INSIEL Informatica per il Sistema degli Enti Locali Spa	INSIEL	Italy
CSI-Piemonte - Consorzio per il Sistema Informativo	CSI	Italy
Institute for Geoinformatics of the University of Muenster	UNIMUN	Germany
Intergraph Italia LLC	INTIT	Italy
Regione Veneto	RVEN	Italy
Magistrato alle Acque di Venezia	MAV	Italy
Università IUAV di Venezia	IUAV	Italy

Information Society and Media
eContent*plus* programme

CONTACTS

Dott. Stefania De Zorzi
Project Coordinator

CORILA. Consorzio per il Coordinamento
delle Ricerche sul Sistema Lagunare
di Venezia

S. Marco 2847. 30124 Venezia - Italy

Tel: + 39.041.2402523
Centr: +39.041.2402511
Fax: +39.041.2402512

e.mail: dezorzi@corila.it
www.gis4eu.eu