Reference Information Specifications for Europe (RISE)
Specific Support Activity SSA3-2005-012141
Project co-funded by the European Commission within the 6th Framework Programme (2002-2006)
Integrating and Strengthening the European Research Area – Priority 4: Aeronautics & Space

Methodology & Guidelines on Use Case & Schema Development
Version 1.2

Deliverable number: 15 [4.1.1]

Date: 28 September 2007

Project started: 1st September 2005
Duration: 27 months

Project Coordinator: Nick Land, EuroGeographics

Bundesamt für
Kartographie und Geodäsie

Customer:	European Commission
Customer Reference:	Contract FP6-012141
Project Title:	RISE: Reference Information Specifications for Europe
Document Title	Methodology & Guidelines on Use Case & Schema Development
Principle Author	Clemens Portele
Principle Contributors	Andreas Illert, Dominique Laurent, Morten Borrebaek
Responsible Partner	OGCE
Reviewed By	RISE PCMG
Approved by	Nick Land, EGHO
Filename	RISE15_Methodology&Guidelines_V1.2
Distribution	PU
Contact	http://www.eu-rise.org/ Clemens Portele cportele@opengeospatial.org

Revision History			
Version	Author	Date	Comments
1a	Clemens Portele	2005-12-20	1st draft for review by RISE consortium.
1b	Clemens Portele	2006-01-04	Minor update based on document review.
1c	Clemens Portele (ed.), Andreas Illert, Morten Borrebaek, Dominique Laurent	2006-01-21	Update based on discussion in Paris meeting and subsequent comments.
1d	Clemens Portele (ed.), Andreas Illert, Morten Borrebaek, Dominique Laurent	2006-01-29	Updated based on results of 2006-01-26 telecon. Changes: Draft structure for checklist contents. Compilation of available input in temporary annex for further processing. Details added to the process steps based on Morten's input. Minor additions to Chapter 8.
1e	Clemens Portele (ed.), Andreas Illert, Morten Borrebaek	2006-02-19	Updated based on result of 2006-02-02 telecon. Changes: Adapted to RISE document template. Section 4.3 (steps in the development of a data specification) revised and extended. Chapter 7 (Checklist) updated and first draft of checklist spreadsheet provided. Chapter 9 added. Minor revision to Chapter 10.
1f	Clemens Portele (ed.)	2006-02-23	Final draft for comments. Updated based on result of 2006-02-22 telecon. Changes: Revision of executive summary. Chapter 1 (Terms) updated. Section 4.3.1 (steps in the development of a data specification) updated to clarify that typically multiple data product specifications are involved in a single use case / application.
1	Clemens Portele (ed.)	2006-03-15	Editorial update for publishing as version 1.
1.1a	Clemens Portele	2006-05-28	Draft update to include results of feedback meeting from 2006-05-18. Changes: Clarified scope in the executive summary and also introduced the scope content to the main body of the document (in Chapter 4). Chapters 4 and 7 updated. Placeholders for further editing before release of version 1.1 are highlighted.
1.1b	Morten Borrebaek	2006-06-09	Some edits in "Executive summary", "4.4.6 Data Product specification", and "7. Checklist for data harmonisation".
1.1c	Andreas Illert, Dominique Laurent, Clemens Portele	2006-07-06	Update of Annex 1 (Checklist).
1.1	Andreas Illert, Dominique Laurent, Clemens Portele	2006-09-16	Including comments from reviews, open issues are highlighted; version to be distributed to INSPIRE DT.
1.2a	Helen Eriksson, Jörgen Hartnor	2007-05-31	Document restructured, additions from INSPIRE DT document and RISE experiences.
1.2b	Andreas Illert	2007-08-07	Helen's additions aligned with current version of INSPIRE D2.6 (v1.2). Scope amended to reflect comments from external review.
1.2c	Wyn Cudlip	2007-09-03	Format and other minor changes.
1.2d	Wyn Cudlip	2007-09-11	Inserted INSPIRE Data Specification Drafting Team Checklist into Annex 2.
1.2e	JB & WC	2007-09-14	Reformatted to remove document size problem.
1.2f	Andreas Illert	2007-09-19	Comments on 1.2c and 1.2e considered.
1.2	Wyn Cudlip	2007-09-28	Final formatting.

Executive Summary

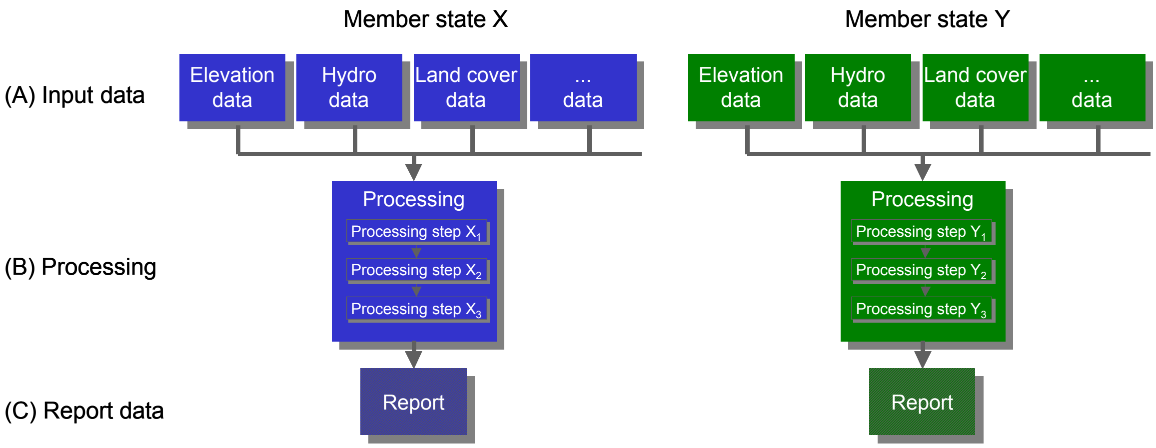
The overall aim of RISE is to facilitate the production of data product specifications on the conceptual and implementation level consistent with the relevant international and industrial standards. A key outcome from RISE is, therefore, the definition of a process – a repeatable methodology – for developing, adopting and maintaining data product specifications. The definition of the repeatable methodology, in particular, addresses issues concerning the harmonisation of heterogeneous data sources.

The objective of this document is to describe this repeatable methodology for the development of data product specifications supporting the principles of GMES and INSPIRE in terms of data harmonisation.

The process of developing a harmonised data specification based on the methodology described in this document will typically involve the following four roles: domain expert, GI architect, software engineer and facilitator. This document is mainly addressing the facilitator who is required to have an excellent understanding of GMES and INSPIRE, in particular the underlying user requirements and IT architectures. However, the methodology developed is applicable to the development of any kind of data product specification.

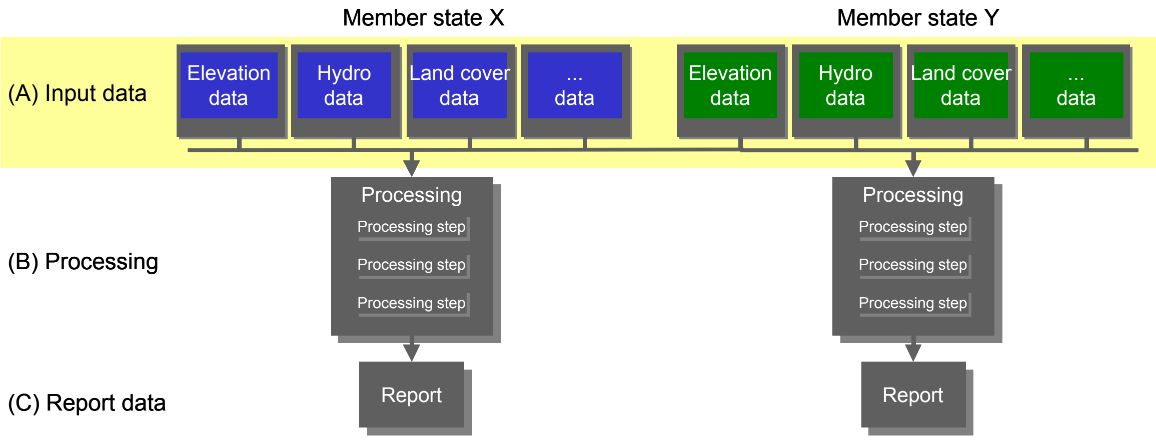
The scope of the methodology focuses on the data specifications of datasets that are used by other applications, i.e. “input data”, as shown in the following figures using reporting by EU member states as an example.

A simplified view to the processing of data for reporting today is shown in the following figure. In most cases, each member state uses input data according to different, often undocumented data specifications and uses different methods to process the input data to produce a more or less standardised report.



The methodology described in this document aims at a better understanding of the (common) requirements for Community-wide reporting. Based on these requirements, it focuses on the development of harmonised data specifications for the input data so that all input data from the different member states follows the same data specifications – and in principle could use the same processing steps to derive the reporting (although this harmonisation step is out of scope for RISE).

The updated figure based on data specifications developed using the methodology is as follows, where the highlighted area indicates the focus of the RISE Methodology:



The input data in the member states will typically be the same as before, but it will be seen through the harmonised view and provided in the member states by data services following the harmonised data specifications. This service-oriented approach is, and has proven itself as, extremely valuable in building spatial information networks comprising heterogeneous data sources.

A consequence of this is that it is important to look into the mapping rules, translating from the internal model(s) to the common application schema during the harmonisation process, to ensure that the resulting schema not only captures the concepts appropriately but also that the required information can be provided by the candidate data holdings. Hence the Methodology has an emphasis on the analysis of the as-is situation and existing “gaps”.

In looking at possible approaches to data harmonisation, it is also important to note that the focus of this Methodology is in particular on looking at generating harmonised data from existing data, not at the collection or creation of new harmonised datasets. However, to harmonise data across different datasets, changes to existing data may be required to some extent.

Finally, it is also worth noting that the Methodology includes provisions to develop theme specific data specifications, (e.g. data specifications for elevation data) incrementally, so that new requirements for a new application can be used to amend existing data specifications.

Table of Contents

Executive Summary 4
Table of Contents 6
1. Scope 8
2. References 9
3. Terms and abbreviations 10
3.1. Terms 10
3.2. Abbreviations 14
4. Overview 15
4.1. Structure of this document 15
4.2. The overall process model of RISE 15
4.3. Data harmonisation as part of the RISE process 17
4.4. Scope of the data harmonisation Methodology 18
4.5. Modelling geographic information 19
4.6. A framework for data harmonisation 20
4.6.1. Principles 20
4.6.2. Data harmonisation components 21
4.7. Steps in the development of a harmonised data specification 22
4.7.1. Overview 22
4.7.2. Use case description 23
4.7.3. Identification of requirements (including data harmonisation) 24
4.7.4. “As is” analysis 24
4.7.5. “Gap” analysis 25
4.7.6. Data Product Specification 25
4.7.7. Implementation, test and validation 26
4.8. Roles 26
5. Data specification and application schema development 27
5.1. Data and interoperability 27
5.2. The application schema and the feature catalogue – the conceptual level 27
5.3. Bridging the conceptual and the system level 28
5.4. Identification of requirements 28
5.5. Identification of feature types 29
5.6. Identification of properties and constraints of feature types 29
5.7. Compiling and evaluating the data product specification 29
6. Implementation and testing 31
6.1. Overview 31
6.2. On-the-fly conversion 31
6.3. Conversion to a derived INSPIRE Dataset 32
6.4. Testing 32
7. Data specifications 33
7.1. Overview 33
7.2. ISO 19131 33
7.2.1. General remarks 33
7.2.2. Identification Information 35
7.2.3. Content and structure 35
7.2.4. Reference system 36

7.2.5. Delivery........36
7.2.6. Data quality........36
7.2.7. Maintenance information (optional)........37
7.2.8. Data capture information (optional)........37
7.2.9. Portrayal information (optional)........37
7.2.10. Additional information (optional)........37
7.3. Service specifications........37

8. Checklist for data harmonisation components........38

9. Use case template........39

9.1. Introduction........39
9.2. The Template........40

10. Experiences & Recommendations from the RISE Project........42

10.1. Introduction........42
10.2. Methodology........42
10.3. Choice of use cases........43
10.4. Use case development........45
10.5. Specification development........46
10.6. RISE Test Environment........49
10.7. Cost Benefit Analysis........50

11. Conclusions........52

ANNEX 1: Data Harmonisation Components........53

A.1 Information Model........53
A.2 Operational Components........55
A.3 Guidelines and Best Practice........57

ANNEX 2: Data harmonisation checklist........59

1. Scope

This document describes a process – a repeatable methodology – for developing, adopting and maintaining data products specifications supporting the principles of GMES and INSPIRE in terms of the harmonisation of heterogeneous data sources.

The scope of the methodology focuses on the development of data specifications that can be used for the creation of stand-alone data products, or the provision of data services in which data is generated "on-the-fly" by request.

The methodology aims at a better understanding of the (common) requirements for Community-wide applications (e.g. reporting). Based on these requirements, it focuses on the development of harmonised data specifications for the input data so that all input data from the different member states follows the same data specifications. The input datasets in the member states will typically be the same as before, but they will be seen through the harmonised view and provided by the member states through data services following the harmonised data specifications.

The methodology includes provisions to develop theme specific data specifications, (e.g. data specifications for elevation data), incrementally so that new requirements for a new application can be used to amend existing data specifications.

The methodology is based on international standards from ISO and OGC. This document explains how to make use of these standards when creating harmonised data specifications. Conceptual schema and application schema are core elements of the RISE approach. Formal modelling is considered a prerequisite for harmonisation, including the mapping of harmonised data models to the heterogeneous data sources for the benefit of automated transformation tools.

The methodology is not executable without a good knowledge of the specific application and of the international standards that serve as a basis for the methodology. This document addresses the expert in standards in the first place, who needs to team up with domain experts and develop the harmonised data specification in a joint effort.

This document presents a general (theoretical) approach. The approach has been tested in RISE with the example of three simplified use cases. At the same time, INSPIRE has adopted the RISE Methodology for the development of theme specifications. Further experience will feed into the Methodology to improve it further, although this needs to happen now beyond the end of the RISE project.

2. References

The following referenced documents are indispensable for the application of this document.

[1] ISO/TS 19103:2005, Geographic Information – Conceptual Schema Language.

[2] ISO 19109:2005, Geographic Information – Rules for Application Schemas.

[3] ISO/DIS 19131, Geographic Information – Data Product Specification.

[4] ISO/DIS 19136, Geographic Information – Geography Markup Language (GML 3.2.0).

[5] ISO/IEC 19501:2005, Information technology — Open Distributed Processing — Unified Modelling Language (UML) Version 1.4.2.

[6] INSPIRE Glossary, Version 0.2, Paul Smits (editor), unpublished.

[7] INSPIRE D2.5 v1.0 Generic Conceptual Model.

[8] INSPIRE D2.6 v2.0 Methodology for the development of data specifications (first draft): http://www.ec-gis.org/inspire/reports/ImplementingRules/ inspireDataspecD2_6v2.0.pdf

[9] OGC Reference Model (ORM): http://orm.opengeospatial.org/

[10] RISE Conceptual Schema in UML. V1.1. September 2007. http://www.eu-rise.org/ or http://www.eurogeographics.org/eng/03_RISE.asp

[11] RISE Use Case Document V1.5. September 2007. http://www.eu-rise.org/ or http://www.eurogeographics.org/eng/03_RISE.asp

3. Terms and abbreviations

3.1. Terms

The terms in this sub-section are taken from the “Glossary of Generic Geographic Information Terms in Europe” that specifies the terminology used in the INSPIRE Implementing Rule documents, and from the INSPIRE document D2.6 v1.2 “Methodology for the development of data specifications (first draft)”.

application schema

Conceptual schema for data required by one or more applications [ISO 19101]

conceptual model

model that defines concepts of a universe of discourse [ISO 19101]

conceptual schema

formal description of a conceptual model [ISO 19101]

conceptual schema language

formal language based on a conceptual formalism for the purpose of representing conceptual schemas [ISO 19101]

EXAMPLE UML, EXPRESS and INTERLIS are examples of conceptual schema languages

coordinate reference system (CRS)

systems for uniquely referencing spatial information in space as a set of coordinates (x,y,z)

and/or latitude and longitude and height, based on a geodetic horizontal and vertical datum [INSPIRE Directive]

NOTE 1 ISO 19111 defines coordinate reference system as a coordinate system that is related to the real world by a datum

EXAMPLE 1 A national coordinate system with the datum ETRS89.

NOTE 2 There is an ISO work item to provide an addendum 19111-2 to define parametric CRS

EXAMPLE 2 the ICAO standard atmosphere; or ISO 2533:1975 which uses a pressure as a coordinate

NOTE 3 Although the definition in the INSPIRE Directive is strictly seen as restricted to spatial reference systems, temporal reference systems are understood as covered by the term coordinate reference systems as well, because temporal information has to be associated with a reference system, just like spatial geometries. ISO 19111 also recognises temporal reference systems explicitly

EXAMPLE 3 The Gregorian calendar is a temporal reference system

coverage

feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain [ISO 19123 - modified]

EXAMPLE Orthoimage, digital elevation model (as grid or TIN), point grids etc

data harmonisation

process of developing a common set of data product specifications in a way that allows 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 co-ordinate reference systems, classification systems, application schemas, etc.

data harmonisation components

structured collection of components that will be documented to support the interoperability and harmonisation of spatial data across Europe

NOTE Rules for application schemas, identifier management, terminology, etc. are examples of the components

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 19131:2007]

dataset

identifiable collection of data [ISO 19115]

dataset series

collection of datasets sharing the same product specification [ISO 19115]

ESDI

European spatial data infrastructure as built based on the INSPIRE framework directive

NOTE The ESDI is expected to include, for example, additional content beyond the data provided by those that are legally mandated to do so according to the directive.

feature

abstraction of real world phenomena [ISO 19101]

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. The term “spatial” is used in the Directive because the limitation to geography was too limited.

feature catalogue

catalogue(s) containing definitions and descriptions of the spatial 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 – modified]

feature concept

abstract specification of the semantics of a feature type [adapted from ISO/CD 19126]

EXAMPLE A ‘road’ feature concept describes the real world phenomena known as “motorway”, “boulevard” and “street”

feature concept dictionary

dictionary containing definitions and descriptions of feature concepts and feature-related concepts [ISO/CD 19126]

feature type

classification of features

EXAMPLE Cadastral parcel, road segment or river basin are all examples of potential feature types

NOTE In the conceptual schema language UML, a feature type will be described by a class with stereotype <<FeatureType>>.

GMES

concerted effort to bring data and information providers together with users, so they can better understand each other and make environmental and security-related information available to the people who need it through enhanced or new services [www.gmes.info]

NOTE GMES is the abbreviation of "Global Monitoring for Environment and Security".

INSPIRE

Framework directive for building an infrastructure for spatial information in the Community [inspire.jrc.it]

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 [INSPIRE Directive]

NOTE Interoperability in the ESDI context means that each member state maintains their own infrastructure, but adopts a framework that enables existing spatial data sets to be linked up from one member state to another (e.g. via transformation).

metadata

information describing spatial data sets and spatial data services, making it possible to discover inventory and use them [INSPIRE Directive]

NOTE A more general definition provided by ISO 19115 is "data about data"

multilingual

in or using several languages [Oxford Dictionary]

profile

set of one or more base standards or subsets of 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 [ISO 19106]

NOTE 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.

reference model

architectural framework for a specific context, e.g. an application or an information infrastructure

EXAMPLE ISO 19101 and the OGC Reference Model are reference models

register

set of files containing identifiers assigned to items with descriptions of the associated items [ISO 19135]

registry

information system on which a register is maintained [ISO 19135]

resolution

resolution expresses the size of the smallest object in a spatial dataset, that can be described. It refers to the amount of detail that can be discerned. It is also known as granularity. Resolution is also limited because geo-spatial databases are intentionally generalised. Resolution affects the degree to which a database is suitable for a specific application. [INSPIRE Position paper Reference Data and Metadata Position paper v4.2]

NOTE In a gridded coverage (e.g. raster data in 2D grids), resolution expresses the size of the raster cell in the real world.

scale

the relation between the dimensions of features on a map and the geographic objects they represent on the earth, commonly expressed as a fraction or a ratio. A map scale of 1/100,000 or 1:100,000 means that one unit of measure on the map equals 100,000 of the same unit on the Earth.

spatial data

data with a direct or indirect reference to a specific location or geographic area [INSPIRE Directive]

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 just 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 data set

identifiable collection of spatial data [INSPIRE Directive]

temporal reference system

reference system against which time is measured [ISO 19108]

theme

grouping of spatial data according to Annex I, II and III of the INSPIRE Directive

units of measurement

defined quantity in which dimensioned parameters are expressed [ISO/TC 211/N1791]

3.2. Abbreviations

BRGM	Bureau de recherches géologiques et minières
CORINE	Co-ordination of Information on the Environment
CRS	Co-ordinate Reference System
CSL	Conceptual Schema Language
DEM	Digital Elevation Model
DFDD	DGIWG (Digital Geospatial Information Working Group) Feature Data Dictionary
DNL	Diffuse nutrient leakage
DPS	Data Product Specification
DT	Drafting Team
DT DS	INSPIRE Drafting Team Data Specifications
EA	Enterprise Architect
EC	European Community
ESDI	European Spatial Data Infrastructure
ETRS89	European Terrestrial Reference System 89
EU	European Union
EVRF2000	European Vertical Reference Frame 2000
GFM	General Feature Model
GI	Geographic Information
GIMODIG	Geospatial info-mobility service by real-time data-integration and generalisation
GMES	Global Monitoring for Environment and Security
GML	Geography Markup Language
hDPS	Harmonised Data Product Specification
HTML	Hypertext Markup Language
ICAO	International Civil Aviation Organization
ICT	Information and Communication Technology
IDPR	Indice Développement Persistence des Réseaux
IGN-F	Institut Géographique National France
INSPIRE	Infrastructure for Spatial Information in the European Community
IR	Implementing Rules
ISO	International Organization for Standardization
IT	Information Technology
ITS	Intelligent Transport Systems
OCL	Object Constraint Language
OGC	Open Geospatial Consortium
ORM	Object-Relational Mapping
OWL	Web Ontology Language
PDF	Adobe Portable Document Format
RISE	Reference Information Specifications for Europe
RTE	RISE Test Environment
SDE	Spatial Data Extension
SDI	Spatial Data Infrastructure
SDIGER	Spatial Data Infrastructure (SDI) for geographic environmental resources
SMHI	Swedish Meteorological & Hydrological Institute
TIN	Triangular Irregular Network
TS	Technical Specification
UML	Unified Modelling Language
WCS	Web Coverage Server
WMS	Web Map Server
WFD	Water Framework Directive
WFS	Web Feature Server
WFS-X	Translating Web Feature Server
XML	eXtensible Markup Language

4. Overview

4.1. Structure of this document

This document consists of several Chapters. This Chapter provides a description of the Methodology and consists of the following parts:

- the process model of RISE (see 4.2),
- overview of data harmonisation and the scope of the Methodology (see 4.3 and 4.4),
- an introduction to modelling geographic information (see 4.5),
- the aspects that have to be considered as part of the harmonisation process (see 4.6), where the harmonisation components are explained in more detail in Annex A,
- the main steps in the development of a harmonised data specification (see 4.7),
- the roles required to develop data specifications (see 4.8).

Chapter 5 provides a description of the process of developing data specifications with a focus on the development of the application schema. Also included is a discussion of the gap between the harmonised application schema and the existing dataset(s).

Chapter 6 discusses the options for implementing the data specification and the testing of the data specification.

Chapter 7 provides the contents of a data specification on the basis of ISO 19131.

Based on the data harmonisation components (Annex 1) a checklist is used to facilitate the process of identifying and documenting requirements, the current situation, identified issues and harmonisation approaches (Chapter 8 and Annex 2).

Requirements are identified using use case descriptions. Chapter 9 offers a template for use case descriptions that may be used.

Chpter 10 summarises the experiences and recommendations from the RISE Project; and Chapter 11 gives some brief conclusions.

4.2. The overall process model of RISE

In order to produce sustainable results and provide best support for GMES, INSPIRE and related operational scenarios, RISE is following a standardised and well-documented approach which is described in the OpenGIS Reference Model (ORM) [9]. The RISE Methodology aims at a predictable and repeatable spiral development process model. It is based on a cyclic approach for incrementally growing RISE's degree of definition and implementation based on requirements in combination with a set of anchor point milestones for ensuring stakeholder commitment to feasible and mutually satisfactory system solutions.

A Process Model answers two main questions: What should the project do next and how long should the project continue doing it? The spiral model holds that the answers to these questions vary from project to project, and that the variation is driven by risk considerations. It emphasises the importance of having all of the project's success-critical stakeholders participate concurrently in defining and executing the project's processes (it uses risk considerations to ensure that progress is not overly slowed down by stakeholder over participation). It can be used to integrate software, hardware, and systems considerations, but is most important to use for software-intensive systems.

In RISE, each spiral in the Spiral Development process is defined by the following steps:

- Requirements expressed as capabilities are described in use case scenarios. These use cases will be described according to a template for use case descriptions and a checklist for data harmonisation aspects to be addressed in the description.
- These use cases describe the relevant universe of discourse which is then documented as an application schema which is a model based on feature types and their properties resulting in a common terminology and described in a conceptual schema language. The Unified Modeling Language (UML) as profiled by the ISO 19100 series of International

Standards will be used as the conceptual schema language. This application schema constitutes a core component of a Data Specification.

- The Data Specification will be documented according to ISO 19131, the (draft) International Standard specifiying the contents of Data Product Specifications in the field of geographic information. A Data Product Specification includes at least the following sections: specification scopes, data product identification, data content and structure, reference systems, data quality, data product delivery, and metadata.
- Based upon the application schema, a GML application schema (GML = Geography Markup Language = ISO 19136) will be generated following the normative rules for such conversions.
- These results will be tested within a prototype under real world conditions and brought forward to INSPIRE, OGC and industry for appropriate consideration and feedback.
- Incremental costs and benefits of the harmonisation efforts will be tracked and documented as repeatable sustainable capability within the Exploitation Guidelines document.

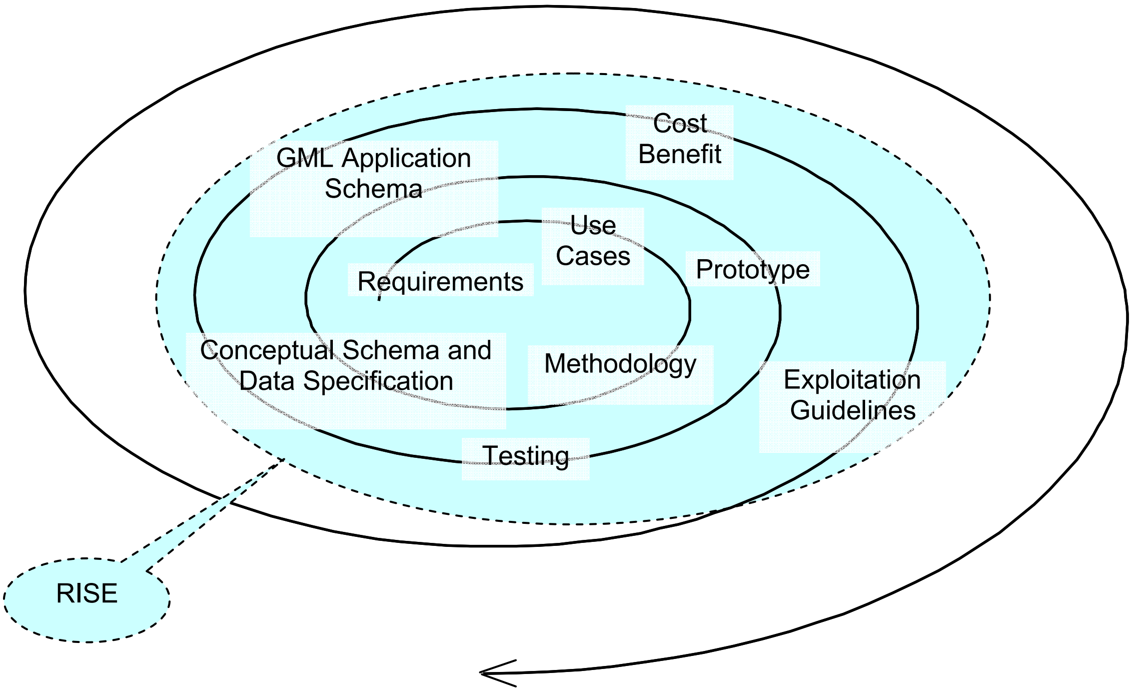


Figure 4.1. Representation of spiral development approach of the RISE Methodology.

This process also exemplifies a modelling approach where the requirements are first modelled on the conceptual/logical level and then converted to specifications on the implementation level. In RISE, the application schema in UML is converted to a GML/XML schema that can be used in conjunction with Web Feature Service interfaces.

To highlight the distinction between the two levels, OGC makes a distinction between an implementation specification (i.e. Data Product Specifications) and an abstract specification.

The general architecture is described in abstract specifications which provide the conceptual foundation across different implementation platforms. Implementation specifications are built and referenced against abstract specifications, thus enabling interoperability between different software components implementing these specifications. An interface specification is considered to be at the implementation level of detail if, when implemented by two different software engineers in ignorance of each other, the resulting components plug and play with each other at that interface.

Most standards of the ISO 19100 series are abstract specifications in the sense described above. Core standards from the ISO 19100 series, including ISO 19107, ISO 19115 and ISO 19119, are also abstract specifications of OGC.

Both types of specifications are complementary and both play an essential role in the RISE process. Therefore, it is key that the implementation specifications used in INSPIRE and GMES are conformant to the abstract specifications used in INSPIRE and GMES.

4.3. Data harmonisation as part of the RISE process

Data harmonisation requires a series of activities. These activities are to be undertaken by subject-matter experts possessing the knowledge about the reasons why data from a theme is collected, and what information is required and can be provided for the relevant applications / use cases. They are supported by interoperability experts, familiar with the underlying GI and ICT standards and specifications.

All Data Specifications shall in general be based on well-defined application requirements only, i.e. all information in the application schema shall in general be traceable to a specific requirement. However, it is also allowed to include, as optional elements, information available from existing datasets even if there is no specific application requirement on it if this element is considered relevant in other application contexts.

The Methodology for creating Data Specifications for harmonised data products consists of two main threads:

1. Specification of the general aspects that are relevant for the development of Data Specifications that support data harmonisation
   - a Identification of the principles and the components of data harmonisation.
   - b Development of a checklist based on these components that shall support the development of use cases in a way that allows the identification of the key harmonisation requirements that need to be supported by the Data Specification.

   → This is done once and becomes part of the Methodology.

2. Development of a Data Specification supporting one or more specific use cases involving data harmonisation.
   - a Development of use cases based on a use case template and the checklist provided by 1b above.
   - b Identification of the key harmonisation requirements that need to be supported by the Data Specification.
   - c Development of the Data Specification following ISO 19131 (Data Product Specification).
   - d Implementation of the Data Specification, testing based on the use case(s) and validation of the solution.
   - e Maintaining a consolidated feature concept dictionary and a common glossary of terms based on all GMES/INSPIRE data specifications to ensure consistent use of terminology.

   → This is done for every Data Specification to be developed.

At every step, potential issues are pushed back to the previous steps to enhance the process, if required (iterative development).

RISE tested the Methodology with use case(s) involving the hydrography, land cover and elevation themes.

The sequence of steps in the development of a Data Specification is a result of merging the approach for modelling geographic features as developed by ISO/TC 211 with the use-case-driven data harmonisation processes required by INSPIRE and GMES.

The checklist mentioned at 1b is considered a useful tool to monitor and document the harmonisation process, although it has not been declared a mandatory document in the development of INSPIRE data specifications.

4.4. Scope of the data harmonisation Methodology

The scope of the Methodology focuses on data specifications of datasets that are used by applications, i.e. where “input data” as shown in the following figures is processed to derive value-added information for the use case at hand. An example is reporting by member states.

A simplified view of the processing of data today is shown in the following figure. In most cases, each member state uses input data according to different, often undocumented or ill-documented data specifications and uses different methods to process the input data to produce more or less similar information relevant for policies within the Community.

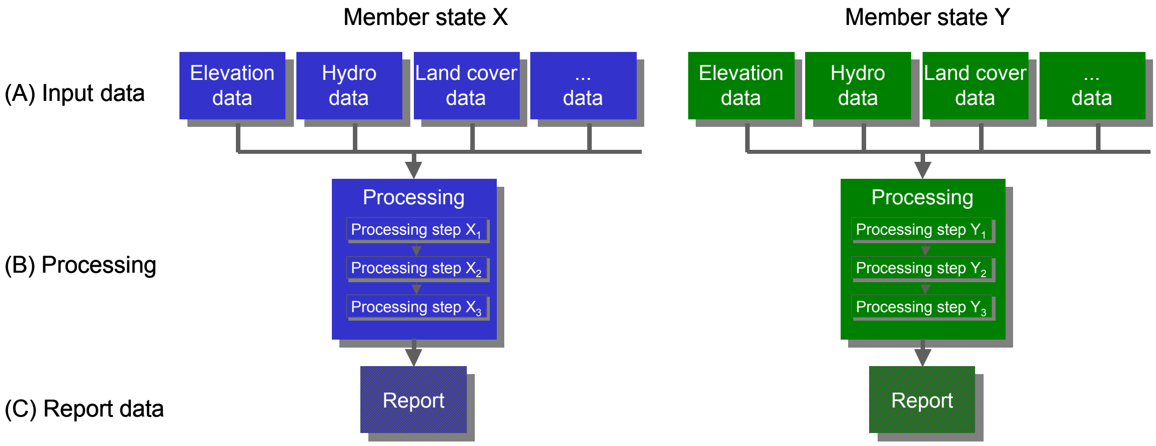


Figure 4.2. Current heterogeneous systems for European reporting.

The methodology described in this document aims at a better understanding of the (common) requirements for Community-wide reporting. Based on these requirements, it focuses on the development of harmonised data specifications for the input data so that all input data from the different member states follows the same data specifications – and in principle could use the same processing steps to derive the reporting (although this harmonisation step is out of scope for RISE, which is concerned with Data Specification harmonisation and not processing harmonisation).

The input data in the member states will typically be the same as before, but it will be seen through the harmonised view and provided in the member states by data services following the harmonised data specifications.

It is also worth noting that the methodology includes provisions to develop theme specific data specifications, e.g. data specifications for elevation data, incrementally so that new requirements from a new application are used to amend the existing data specification.

The updated figure based on data specifications developed using the methodology is as follows, where the highlighted area indicates the focus of the RISE methodology:

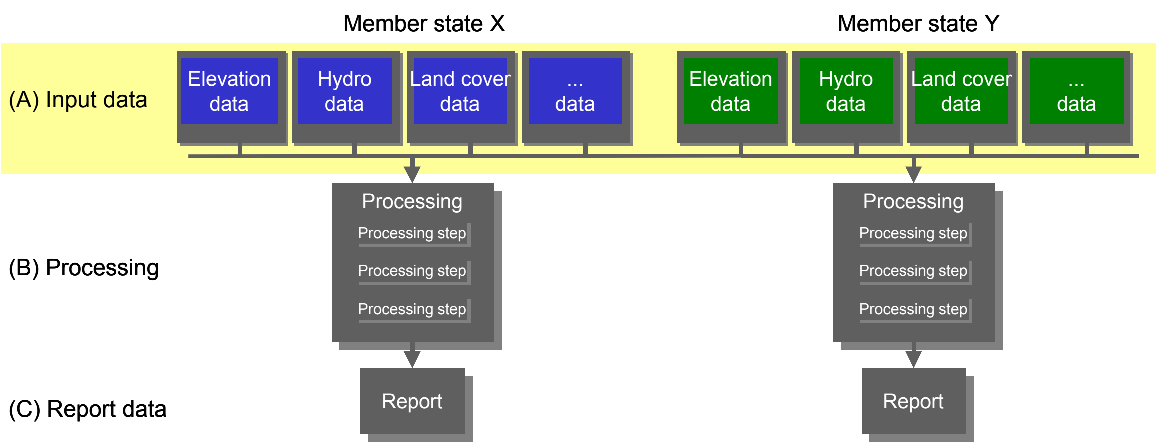


Figure 4.3. Potential to harmonise input data and allow standardised processing.

4.5. Modelling geographic information

Based on ISO 19109 (Rules for application schemas), the general process is the following:

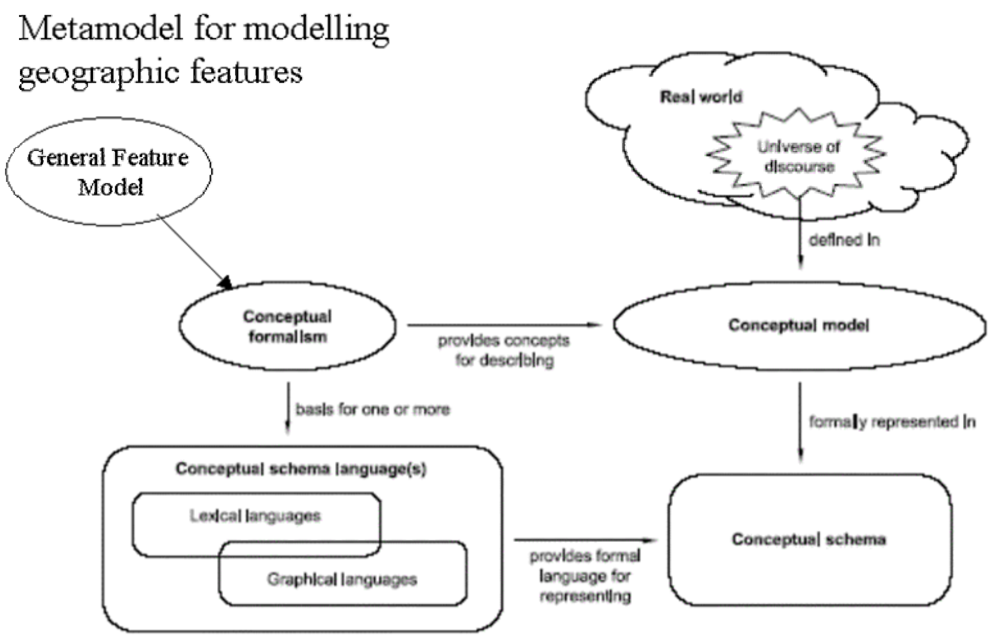


Figure 4.4. Schematic representation of steps involved in modelling geographic data.

The first step is to define the universe of discourse from the real world, e.g. the Water Framework Directive, into a Conceptual model.

To describe this conceptual model, a conceptual formalism is required. A part of this conceptual formalism is the so-called general feature model (GFM). The GFM is a model of the concepts required to classify a view of the real world. It is expressed in a formal language, i.e. a conceptual schema language (CSL), which in this methodology is a profile of UML restricted to model elements used in class diagrams as described by ISO/TS 19103 and/or ISO/DIS 19136 Annex E (NOTE: details are to be defined based on the recommendations of the CSL workshop from October 2005), but in principle it could be in any CSL that can be mapped to the GFM[1]. Note that UML has its own model of concepts (metamodel). As both the GFM and the UML metamodel deal with classification, the concepts are very similar. However, there is a significant difference: The concepts in the GFM establish a basis for the classification of features, whereas the UML metamodel provides a basis for the classification of objects of any kind.

The ISO geographic information standards focus on abstract, implementation-neutral UML models that can serve as specifications for implementations using various implementation platforms.

The next step will thus be to generate the GML Application Schema based upon the application schema, following the process documented in GML (ISO/DIS 19136) Annex E. This transforms the abstract specification of the application schema into a XML-based implementation specification that can immediately be used, for example, in a web feature

[1] The Conceptual formalism is the basis for the conceptual schema language, consisting of a lexical and/or a graphical notation. In the report drafted by CEN/TC 287 WG5 it is stated that if an information community applies a Conceptual Schema Language other than UML, it is the responsibility of that information community to map the ISO general feature model to the meta-model of the Conceptual Schema Language of choice, and to maintain the mapping rules, following the ISO/TS 19103:2005 conformance statement: Non-UML schemas shall be considered conformant if there is a well-defined mapping from a model in the source language into an equivalent model in UML and that this model in UML is conformant.

service. This will allow querying and selecting data according to the application schema as well as inserting, deleting and updating data in transactions.

The process described above is well documented in the context of the ISO 19100 series of standards. However, there is a need for additional guidelines on how to apply use cases as the methodology to capture and identify requirements, and the process to go from the universe of discourse to a conceptual model.

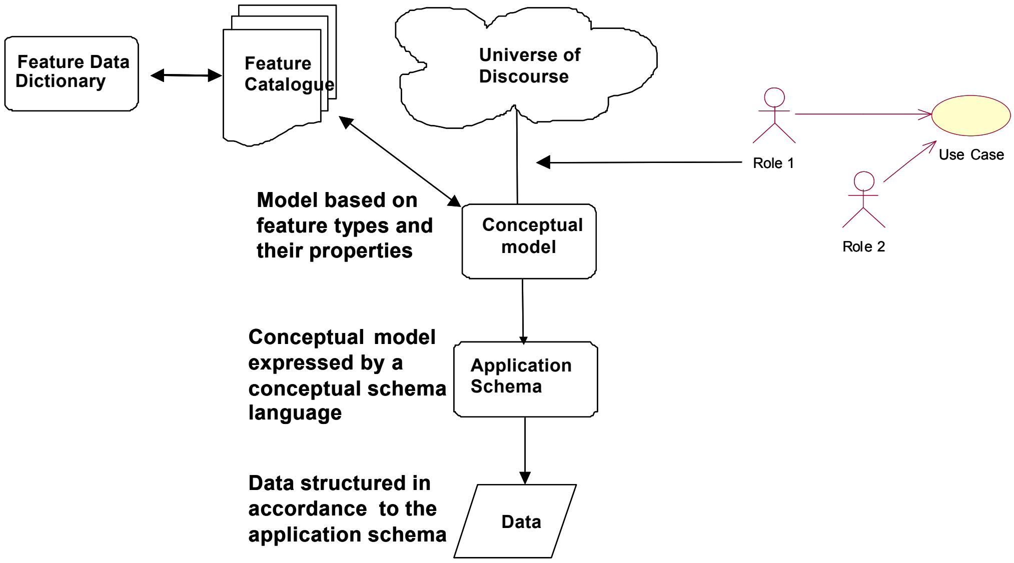


Figure 4.5. Issues for consideration in going from Universe of Discource to a Data Product.

Use cases will be used to model the business requirements, and are considered as a valuable methodology to describe the universe of discourse and how the business is supported by data or services related to other data.

4.6. A framework for data harmonisation

4.6.1. Principles

The following description identifies the principles for data harmonisation that are guiding the process to a Data Specification. They will clarify the meaning of data harmonisation in GMES and INSPIRE.

The framework is based on the following assumptions:

- All nations & organisations (within a nation) start from different positions in terms of data models etc. Due to different political, economic, cultural and organisational drivers, Europe will not achieve total harmonisation across every nation as part of the INSPIRE process.
- A mechanism that provides a common language to support needs at EU and other large-scale cross-border and cross-sector levels is required.
- The trend towards the integration of GI into the ICT mainstream will accelerate.
- The main goal at least for the foreseeable future will be “harmonisation” through interoperability in a service-based architecture rather than (full) harmonisation of the underlying data models.
- In this context, data harmonisation requirements have to be considered on different levels:
    - the schema level (use of common application schemas independent of the data model of the base data)
    - the data level (e.g. edge matching in border areas)

4.6.2. Data harmonisation components

The following figure provides an overview of the components relevant for data harmonisation and the grouping of these components.

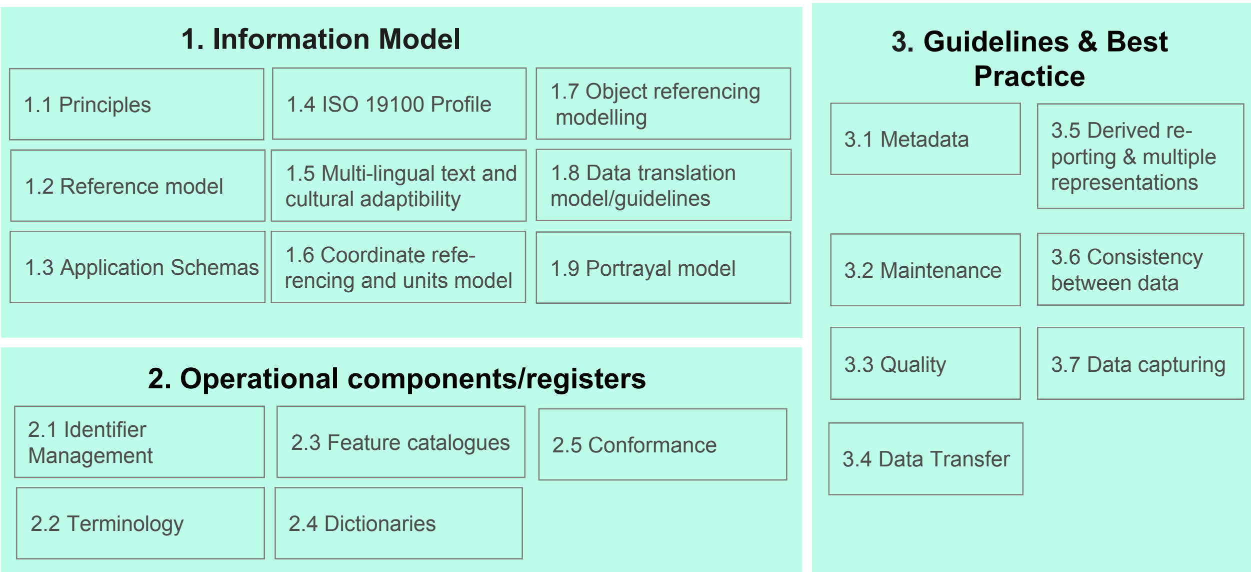


Figure 4.6. Main components of Data Product Specification.

These components apply to all types of spatial data including vector and coverage data (coverages are features too). However, for the different spatial representation types, the components will in general be different.

NOTE The components are work in progress. The INSPIRE Drafting Team on "Data Specifications" has already modified the list of components at the time this RISE document was finalised, and will develop the harmonisation components further.

A more detailed description of the data harmonisation components can be found in Annex 1. Annex 2 contains the corresponding version of the INSPIRE Drafting Team Checklist as of August 2007. Examples of completed Checklists can be found in the RISE Use Case Document [11].

4.7. Steps in the development of a harmonised data specification

4.7.1. Overview

Step	Result	Relevant standards, standing documents and chapters of this document
Use case description	Use case description, updated glossary[2]	Section 4.7.2 ISO/TS 19103 Glossary Use case template (Chapter 9)
Identification of data harmonisation requirements and feature types	List of feature types, list of requirements (structured according to the data harmonisation components described in the following Chapter)	Section 4.7.3 Data harmonisation checklist (Chapter 8) ISO 19131
As-is analysis	Description of the current situation with respect to the data harmonisation components (per source dataset)	Section 4.7.4 Data harmonisation checklist (Chapter 8)
Gap analysis	Description of data harmonisation issues derived from the identified requirements and taking the as-is analysis into account (per source dataset) Updated as-is analyses (additional data sources that have been identified) or updated/reduced requirements to ameliorate the identified gaps	Section 4.7.5 Data harmonisation checklist (Chapter 8)
Data Product Specification development, in particular feature concept dictionary update, specification of application schema and feature catalogue	Data product specification (per data theme) with a sections specified in ISO 19131 (including application schema in UML as well as the corresponding feature catalogue and GML application schema) Updated feature concept dictionary, updated glossary	Section 4.7.6 & Chapter 5 ISO/TS 19103 ISO 19109 ISO 19110 ISO 19126 ISO 19131 ISO 19136 (GML) Feature concept dictionary Glossary
Implementation, test and validation	Implementation of the application using services and the data specification; test report	Section 4.7.7 & Chapter 6 ISO 19136 (GML) ISO 19142 (WFS) ISO 19143 (FES)

No specific formal software development process model for how to create and analyse use cases and turn them into an application schema is mandatory for the data harmonisation process. This recognises the fact that in practice different approaches are being used successfully and it does not seem to be appropriate to prescribe a particular approach and disallow others. The approach taken by RISE is rather to offer "tools" that are intended to be adapted as appropriate in different contexts. In this step this includes:

- a use case template that contains information typically provided in use case descriptions (see Chapter 9).
- a checklist of aspects that need to be understood in the context of the use case to address the data harmonisation issues in the use case (see Chapter 8 and Annex 2). This checklist is based on the data harmonisation components discussed in the Section 4.6.

[2] The terms in Section 3.1 may be used as a starting point for the glossary. However, beside general terms included right now, the glossary shall eventually include, in particular, all theme specific terms and concepts that are not defined in the feature concept dictionary or in code lists.

- a feature concept dictionary and a glossary common to all data specifications developed in the GMES/INSPIRE context.

The results of every step may require an update in the results of the previous steps.

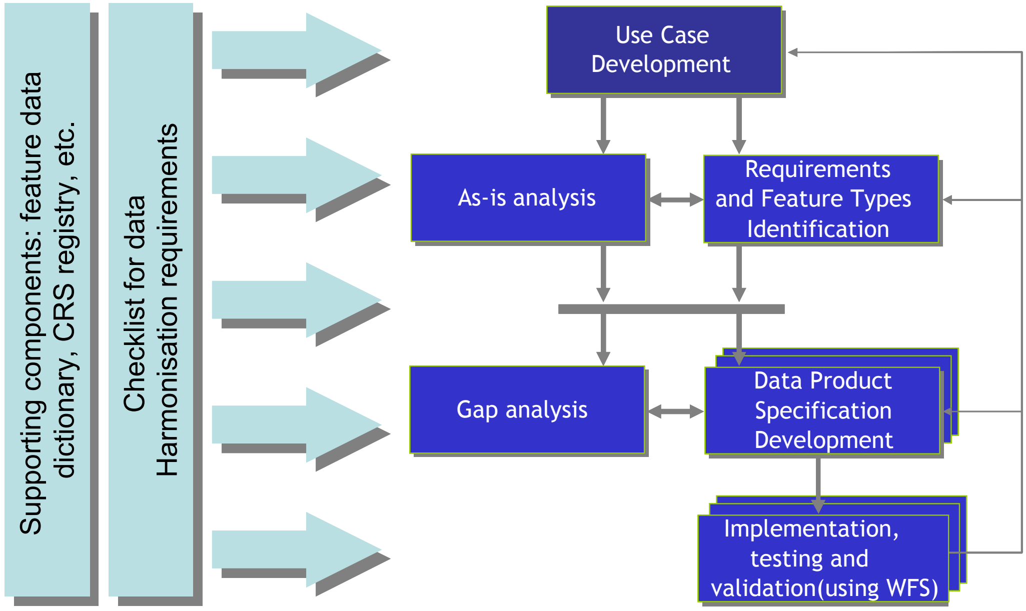


Figure 4.7. Diagrammatic representation of RISE Methodology.

The RISE approach allows the identification of issues relevant for data harmonisation as early as possible in the process. In particular, it is the goal to identify as many issues as possible before implementation. As data harmonisation is a complex issue, it is not realistic to expect that every issue is caught early in the process, so implementing and testing is also an integral part of the process.

The steps are not carried out sequentially, but with a considerable overlap to allow for rapid feedback. The following diagram illustrates roughly the flow of the development process. Note that feedback is intended to be propagated back to previous steps in every stage.

It is important to note that typically multiple data product specifications will be involved in implementing a use case. Identifying the appropriate segmentation into data product specifications is part of the analysis phase. Typically, for example, different data product specifications will be used to distinguish between source data from different themes and/or different organisational responsibilities.

Likewise, an analysis of the as-is situation and the gaps to the requirements typically has to be carried out separately for every input dataset involved. One particular difficulty in this step will often be that the existing data will not be properly documented and thus some re-engineering of the data may be required.

In addition, the steps reflect that the development of a data product specification in general does not start from scratch, but that in a growing ESDI increasingly existing data product specifications (e.g. for the different themes listed in the INSPIRE Annexes I, II and III) will instead be amended to address new requirements. Maintaining an ESDI-wide feature concept dictionary is intended to help in keeping an “holistic view” to the information available in the ESDI.

4.7.2. Use case description

A use case defines a goal-oriented set of interactions between external actors and the system under consideration. Actors are parties outside the system that interact with the system (UML 1999, pp. 2.113- 2.123). An actor may be a class of users, roles users can play, or other systems.

A use case is initiated by a user with a particular goal in mind, and completes successfully when that goal is satisfied. It describes the sequence of interactions between actors and the system necessary to deliver the service that satisfies the goal.

Generally, use case steps are written in an easy-to-understand structured narrative using the vocabulary of the domain. This is engaging for users who can easily follow and validate the use cases, and the accessibility encourages users to be actively involved in defining the requirements.

The use case describes the processes and this supports the analysis required to identify the requirements; in RISE the focus is on the requirements relevant to producing a harmonised Data Product Specification.

During a use case both ‘as is’ analyses and ‘Gap’ analysis typically need to be considered.

A Use Case Methodology should be applied. The RISE focus is on the proposed use case template and the checklist, but their application is optional as long as, as a result of the step, the requirements are identified and the data harmonisation components have been considered. The goal of the checklist is to assist during a discussion between a domain expert and a GI expert. The result of this discussion will be documented and this helps define the requirements. Most of the requirements may be documented according to the structure of ISO 19131 Data Product Specification. Any remaining requirements should be inserted as comments into the Data Product Specification.

4.7.3. Identification of requirements (including data harmonisation)

The results from the use case development constitute requirements. The requirements that relate to the input data are transformed by a GI expert to a first-cut of data product specifications (typically one per data theme), according to ISO 19131 Data Product Specification, but not fully documented at this early stage.

If a data specification for the relevant input data already exists, an initial analysis will be carried out to determine if the existing specification should be used as a basis and be amended, or if the requirements warrant the creation of a new data specification. Whenever possible, the reuse of an existing data specification should be considered.

Based upon the requirements and the checklist, the GI expert will have a fairly good understanding of the feature types and associated attributes, constraints and associations, as well as other relevant information like co-ordinate reference systems, metadata, etc. involved, and this information will be the basis for the further analysis.

4.7.4. “As is” analysis

After modelling the “first-cut” application schema(s), the next step is the “as is” analysis, identifying the available information. In addition to different kind of reference and thematic spatial data, there will be data documents in various kinds of formats, from PDF documents to general databases. Some of these data are referenced by geographic identifiers (place names, river identifiers, etc.).

In most situations, the existing source material is an important part of the use case and, in general, a user defining his/her requirements is well aware of existing data. In other situations, the existing source material is different from country to country and requires a search for potential input data based on the descriptions in the use case scenarios.

In general, it will not be required to also apply the “tools” and standards that are used to describe the harmonised data specifications to the existing datasets. If the available documentation of the existing data includes information about, for example, the feature types and their properties, there is no need to recapture this information in a data product specification according to ISO 19131.

However, sometimes the input data will not be documented sufficiently. This Methodology does not provide specific rules on how to deal with this fact, but if the documentation has to be created for the dataset, it is recommended to use the same technologies for the input data product specification as used for the harmonised data product specification. It is important to note that such documentation is not the result of the harmonisation process as such, but would only be a temporary result to support the gap analysis.

Also, complete English versions of the national/regional feature catalogues are in general not available (note that sometimes feature and attribute names are translated, but not the descriptive text). Therefore, in principle local experts from all member states are required to participate in the harmonisation process. The GiMoDig project, for example, recommends that all parts of the reference feature catalogues / product specifications that are relevant for the harmonised data specification should be translated into English. This would require the involvement of large number of high-profile experts (an expert must know both the language and the domain).

4.7.5. "Gap" analysis

This analysis compares the results of each "as-is" analysis with the first-cut application schema, and evaluates if the identified source material is sufficient to fulfil the requirements in the application schema. It also identifies how to extract information from these data sources into the application schema.

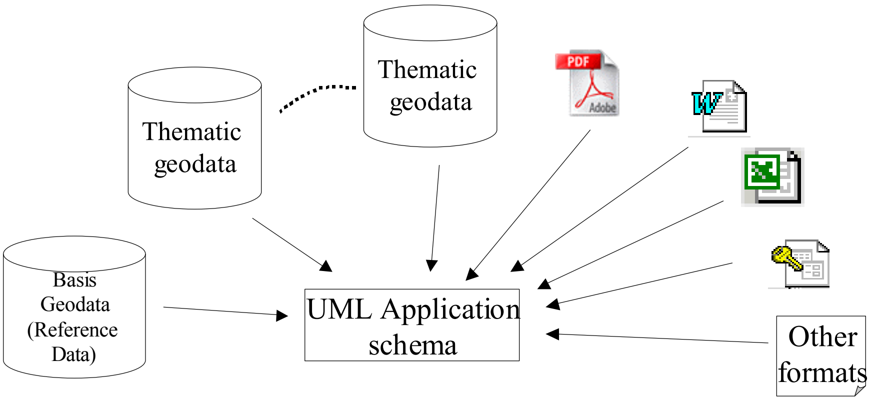


Figure 4.8. Potential sources of information required to develop UML Schema.

If the identified source material is not sufficient, a new "as-is" analysis, or change of requirements in the use case, may be required, or the provision of new data must be considered. Provision of new data may be capturing new data sets, but will more typically mean amending existing data (e.g. with new codes or classifications) or processing existing data (e.g. to derive the central axis of rivers or other generalisation algorithms).

The gap analysis is useful at two steps:

- between the existing data sources and the user requirements.
- between the existing data after harmonisation and the user requirements :
  - for the use case itself, to choose the best harmonisation approach if there are several options, to look for new solutions in the future, etc.
  - for the Methodology development itself, to assess different harmonisation approaches.

The use of "matching tables" can help in the processes of mapping the available input data to the required output data. In RISE, this was done using Excel spreadsheets, as described in the RISE Conceptual Schema Document [10].

4.7.6. Data Product Specification

The results from the use case and the analysis according to the checklist are formalised in a data product specification. The result of the checklist is an important starting point for creating a data product specification. But it is also clear that the result of the checklist itself may not be detailed enough, and further work must then be conducted in close cooperation between a domain expert and a "modelling" expert. However, the result of the checklist will guide the discussions, and ease the creation of the data product specification.

The following figure (ISO 19131, Figure B.1 — Relationship of data product specification to metadata) shows the sequence from a data product specification that specifies a data product, implemented as a dataset described by metadata.

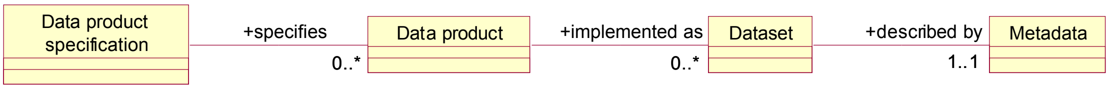


Figure 4.9. Relationship of data product specification to metadata.

A vital part of a data product specification is the ISO 19109 application schema specified in UML. A UML application schema is a conceptual schema for data required by one or more applications, modelled in UML. This application schema describes the conceptual model for the data that is supposed to fulfil the identified requirements.

A GML application schema is typically derived automatically from the UML application schema, applying the rules defines in ISO 19136 GML Annex E, and provides a syntax for a concrete representation of data according to the data product specification.

Since the Generic Conceptual Model allows for a more complex UML profile than the profile specified in ISO 19136 Annex E, a platform specific UML model has to be derived from the application schema, which is platform independent. In addition, other delivery formats besides GML, for example for coverage data (e.g., for imagery or DEMs), will be defined based on specific user requirements and state-of-the-art in the relevant communities. The data from the various sources are transformed/exported/imported into GML document(s) according to the GML Application Schema derived from the UML application schema.

4.7.7. Implementation, test and validation

These results are then tested within a prototype under real world conditions. The data product specifications are updated based on the results of the tests. In RISE, a Test Environment was created consisting of an Oracle database to hold the input data, and a translating Web Feature Server (WFS). This allowed the creation of test data products conformant to the RISE harmonised Data Product Specifications. This arrangement, in particular, was able to confirm the feasibility of mapping, "on-the-fly" from existing data sources to the required harmonised products.

4.8. Roles

The process of developing a harmonised data specification will typically involve the following roles:

- **Domain expert:** provides expertise about the thematic domain and the data to be used in the application.
- **GI architect:** provides expertise about geographic information specifications (ISO 19100 series, OpenGIS Implementation Specifications, other GMES/INSPIRE specifications) and information modelling; it is important to note that this includes the service interfaces used to provide access to the data products as these are instrumental to harmonised data products.
- **Software engineer:** provides expertise about implementation and deployment aspects of the relevant specifications.
- **Facilitator:** manages the process and oversees that all data harmonisation requirements (see the checklist) are identified and adequately addressed.

Depending on the complexity of the use case, a person may take several roles or several persons may be required for a single role.

The diversity in Europe implies that domain experts from a range of communities will in general be required to cover a theme.

5. Data specification and application schema development

5.1. Data and interoperability

Every information system works with data stored according to an internal data model based on its own set of requirements. As shown in the previous Chapters, a key step in the data harmonisation process is to achieve interoperability on the conceptual level (semantic interoperability) so that users and implementers of different information systems from the INSPIRE, GMES or any other information community can understand the semantics of the relevant information provided by the other system.

It is worth repeating that interoperability is the ability of two or more autonomous entities to communicate and co-operate among themselves in a meaningful way despite differences in language, context or content. This interaction should not require special efforts by the data producer or consumer - be it human or machine. There is no "interoperability" between data sets per se. The only things that can interoperate are services and systems. In the face of heterogeneous data sources, interoperability requires "wrapping" data sources into services that conform to standards. The output of these services is what needs to be harmonised, not their inputs (database schemas). Thus the legacy is maintained and can evolve to support the specified service interfaces.

5.2. The application schema and the feature catalogue – the conceptual level

The semantics are captured in an application schema, a schema for data required by one or more applications. An application schema defines content and structure of data (the data model), but may also specify services for accessing and manipulating data by an application. The data model will specify the domain specific feature types describing the specific view of the real world based on the information requirements of this domain. The feature types will name the core concepts of the domain in a meaningful way (e.g. “lake”, “parcel”, “road”) along with their definition, their properties, possible constraints, etc. To facilitate support for multilingualism the names and definitions in the application schema must be captured in the relevant languages if at all possible.

In the definition of these vertical, i.e. domain specific, feature types, commonly used horizontal types, like “point”, “surface”, “accuracy”, “time period”, etc. that have been previously specified by OGC and ISO in consensus-based standardisation processes, will be reused. This saves time (reuse of existing definitions) and increases interoperability (using the same standardized components across different application domains).

It is worth mentioning that the same entity in the real-world may be represented in different datasets by different features (because they have a different view of the world).

An application schema will be described in a formal language, the schema language, which captures the facts agreed upon by the different partners in the harmonisation process. The Consortium will use a profile of UML as specified in ISO 19103 and ISO 19109 to describe the application schema on the conceptual level – in line with the recommendations of the ISO 19100 series and the current practice in the GI community. It should be noted that ontology languages are currently a hot research topic and that in the future it may prove to be beneficial to express the ontology also in one of these emerging languages (e.g. OWL), but for the time being it is recommended to use a schema language that is well rooted and accepted in the GI and standards community.

A human readable version of the application schema is the feature catalogue. The catalogue can in general be automatically derived from the information contained in the application schema and styled for presentation to a human reader, e.g. as HTML, PDF or Word documents.

5.3. Bridging the conceptual and the system level

The second level of interoperability is interoperability on the system level so that the two systems can access and process data provided by the other system.

Since interoperability must be achieved stepwise on both levels (conceptual and system), schemas on both levels are required, too. To distinguish them, the schema on the conceptual level will be referred to as “application schema” or “ISO 19109 application schema” specified using UML, while the schema on the encoding level will be referred to as “GML application schema” specified using XML Schema.

Establishing and documenting a common, multilingual vocabulary by means of a conceptual schema is a major cornerstone in the harmonisation process. Nevertheless it must be taken into account that still the gap between the internal data model of today’s applications and the common schema needs to be bridged.

There are two general options for this: either a mapping from the internal to the common model and vice versa is defined or the internal data model is altered to fit the common model (data re-engineering). Naturally, a solution mid-way between these two extremes is also possible. The appropriate solution has to be determined on a case-by-case basis, but in general it can be expected that a cost-benefit-analysis will yield the result that a major change in the internal data model is not a reasonable option in many cases – at least in the short term.

A consequence of this is that it is important to look into the mapping rules translating from the internal model(s) to the vocabulary / features of the common application schema. This should be done early during the harmonisation process to ensure that the resulting schema not only captures the concepts appropriately but also that the required information can be provided by the candidate data holdings. Hence the emphasis is on the as-is and gap analysis in the previous steps. This analysis also takes into account that often the internal data model will contain much more detailed data than it will be required on the harmonised schema level, so the harmonised schema will in general be a generalisation of the internal models.

Finally, it must be emphasised that it is considered unlikely that a single European data model or specification for all domains is achievable. At the same time a traditional bulk transfer of data(sets) is often too inflexible and does not meet the requirements of the users. Therefore an approach focussing on providing services to the information that can be generated from the different data holdings is required. This service-oriented approach is and has proven itself to be extremely valuable in building spatial information networks comprising heterogeneous data sources.

5.4. Identification of requirements

The requirements considered by RISE came from an application that operates in a European context; first priority was given to applications required to support European Directives (e.g. Water Framework Directive) and policy making. However, the RISE Methodology is in most parts not restricted to applications on the European level and can also be used in different contexts.

The use case analysis based on the checklist is intended to capture the requirements of the harmonised data product specification. The result shall take all relevant aspects into account:

- the target information required.
- the current situation.
- known issues that are common in data harmonisation efforts.
- the reference model and architecture of the future ESDI as far as it is known today.
- the standardized structure of a data (product) specification.

The result of the requirements analysis is captured in statements comprising the requirements model.

5.5. Identification of feature types

Based on the modelling process described in the ISO 19109 (Rules for application schemas), the process of defining and structuring spatial data is shown in the following figure.

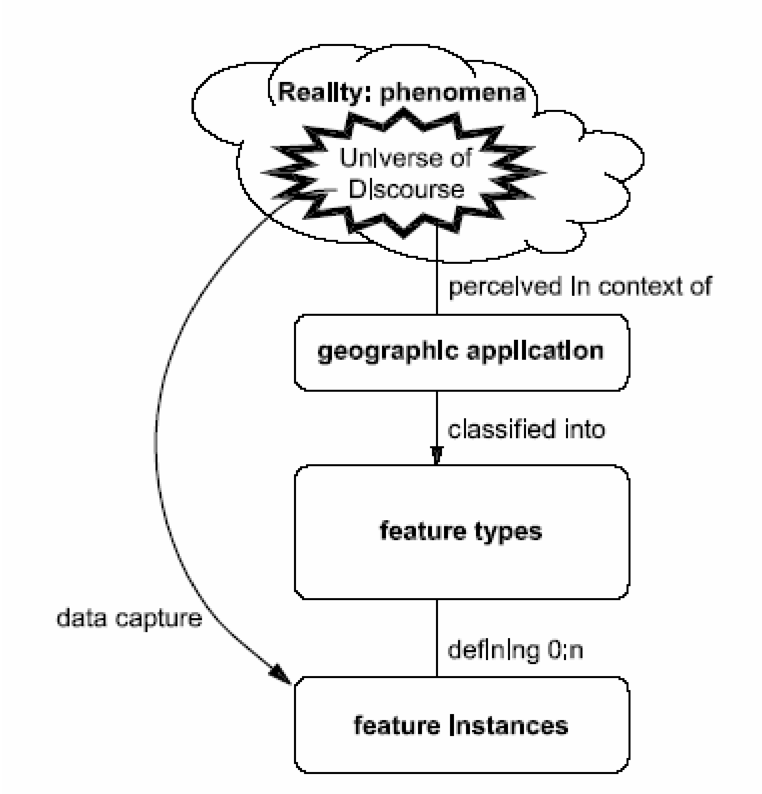


Figure 5.1. Process for defining and structuring spatial data.

As has been stressed before, recognising the fact that different approaches are being used successfully in practice, no specific formal software development process model for how to turn the use cases into feature types (i.e. identifying classes in terms of object-oriented analysis and design) or an application schema is mandatory for the data harmonisation process.

An example of identifying the relevant classes/feature types is to use the “noun” method based on the requirements model and the use case description(s). In them, the relevant nouns have been identified and – after elimination of duplicates and synonyms as well as an examination if the noun describes a feature type of another concept (e.g. a data type without any identity) – can be used as an initial list of feature types.

5.6. Identification of properties and constraints of feature types

By analysing the requirements and the use case descriptions, the properties (i.e. attributes and relationships with other types) as well as the constraints will be specified for the individual feature types using UML according to ISO/TS 19103. The result is the application schema.

5.7. Compiling and evaluating the data product specification

Another aspect that has to be considered in this process as elaborated above is the fact that the application schema is only part of the required harmonisation effort. To enable the seamless use of data provided by different sources, additional specifications are required beside a formal description of the data model. That additional information is captured in the data product specification, as explained in Section 7.2, through the harmonisation components discussed in sub-section 4.6.2.

The resulting data specification should be checked to verify that it meets the following expectations:

- For all feature types, attribute types and association roles in an application schema, it must be agreed whether the model element is required (i.e., must be provided by a data provider), or is optional (i.e., may be provided by a data provider). All these model elements have to be documented so that their meaning is clearly stated.
- All data specifications will include a normative encoding of the feature types specified in the application schema based on GML (ISO 19136) to be used in conjunction with services including the Web Feature Service (currently also in the standardisation process to become ISO 19142).
- The same entity in the real-world may be represented in different datasets by different features - because the datasets may have a different view of the world - but within a dataset, the same entity shall be represented only by a single feature.
- Verify the feasibility by testing the mapping from existing data sources to the harmonised European application schema.

In looking at possible approaches to data harmonisation, it is also important to note that the focus as discussed above is in particular on looking at harmonising existing data, not at the collection or creation of harmonised datasets. However, to harmonise data across different datasets, changes to existing data may be required to some extent.

Existing data will in almost all cases be structured in a way that differs from the application schema of a "harmonised" data specification.

Since the data in a dataset will often be collected for a local, regional or national task, a conversion of data to a European application schema will often involve conversions for several reasons:

- Conversion due to different encodings.
- Conversion due to different levels of detail as the European view may be substantially different from the view of the existing applications.
- Conversion due to different terminologies and concepts.

In addition, data integration between different representations of the same entity in different datasets (for example along or across borders, themes, sectors or at different resolutions) will be required. This will likely involve, for example, data matching algorithms.

It is therefore a necessary step of the data harmonisation process to look not only at the data requirements, but also at the conversion rules required to map from the existing application schemas to the harmonised application schemas to ensure that the resulting schema not only captures the concepts appropriately, but also that the required information can be provided by the existing data repositories in an acceptable way. This also has to take into account that often the existing data will contain much more detailed contents than it will be required for potential users, so data available in the ESDI is in general expected to be a generalisation of existing data.

Another aspect highlighted above and to be considered in the harmonisation process is the fact that the application schema is only part of the required effort. To enable the seamless use of data provided by different sources, additional specifications are typically required beside the description of the data model (this will include mostly those aspects covered currently in the 2nd and 3rd data harmonisation component groups including identifiers, dictionaries, rules for the consistency of spatial data across datasets, etc.). It has to be verified that this additional information is captured in the data specification, too.

6. Implementation and testing

6.1. Overview

It is expected that, in general, there will be two methods (or a combination of them) to publish harmonised data in the European Spatial Data Infrastructure (ESDI):

i. On-the-fly conversion.

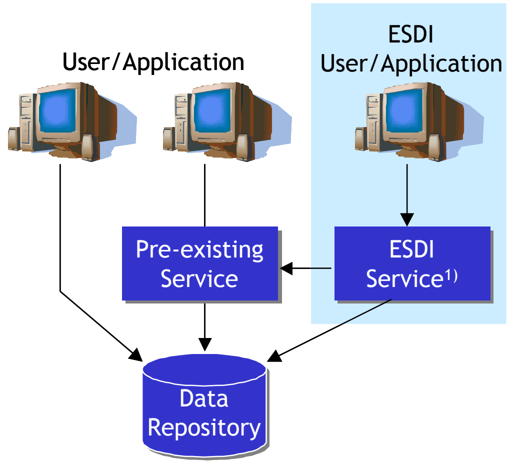
ii. Conversion to a derived dataset.

NOTE The following description is focussed on translating data from one schema to another. However, requirements that fall into the component "consistency of data" will often require changing data to create consistency with other data (e.g. edge matching). This is currently not yet reflected in the descriptions, because a better understanding of the requirements is necessary first. The present text refers mostly on schema translation aspects. For example, schema matching, i.e. conversion to a data specification, may be done on-the-fly whereas data matching is typically not so simple, because of geometrical and topological constraints, and will often have an impact on the data itself.

The need for supporting services helping in such translation processes, e.g. Coordinate Transformation Services or Gazetteer Services to convert between different reference systems or Registry Services to provide and disseminate information about the developed schemas or multilingual thesauri, should in principle be identified as part of the harmonisation process.

6.2. On-the-fly conversion

This method requires on-the-fly conversion of data (to map between the existing and the harmonised application schema) and allows support for queries (to map between the query models of the existing and the harmonised application schema).

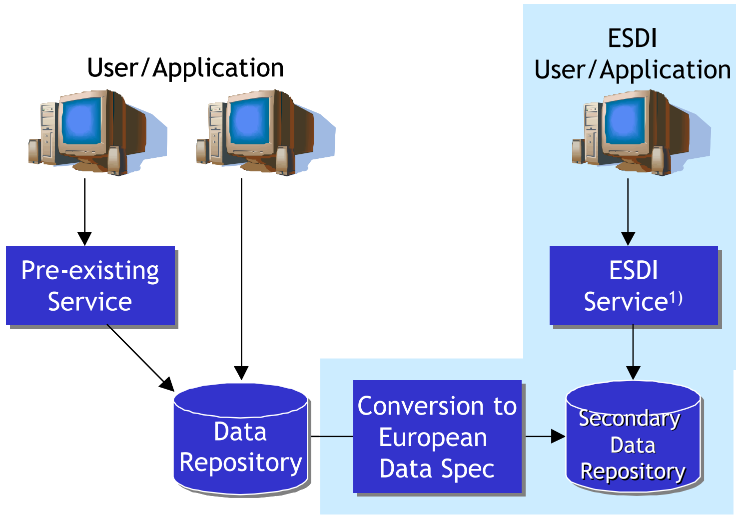


1) Service supporting one or more European Data Specifications, in particular services implementing the Web Feature Service and Web Coverage Service specifcations

Note that the service may access either the data repository directly or an existing service

Figure 6.1. On-the-fly method for creating harmonised data.

6.3. Conversion to a derived INSPIRE Dataset



[1)] Service supporting one or more European Data Specifications, in particular services implementing the Web Feature Service and Web Coverage Service specifcations

Figure 6.2. Pre-generation method for creating harmonised data.

The necessary conversion frequency depends on the data product specification requirements. It is also necessary to maintain a secondary data repository; and updates to the main data repository have to be replicated in the secondary data repository (or it has to be re-generated).

6.4. Testing

To test the harmonised data specification, the use case(s) – or at least a core part of them – should be implemented in conformance with the data specification. The results of testing can then be used to improve the data specification if required.

As mentioned earlier, in RISE, testing was carried out using a translating Web Feature Server (WFS-X).

7. Data specifications

7.1. Overview

Within information technology there are several examples on how to model a business model / use case. But what is essential to the RISE project is the spatial domain. Requirements for the spatial domain lead us to standards and specifications from ISO/TC 211 and OGC. ISO/TC 211 have a work item called ISO 19131 Data Product Specification[3].

From the introduction of ISO 19131:

A data product specification is a 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. It is a precise technical description of the data product in terms of the requirements that it will or may fulfil. However, the data product specification only defines how the dataset should be. For various reasons, compromises may need to be made in the implementation. The metadata, associated with the product dataset, should reflect how the product dataset actually is.

This means that the result of the business model and requirement model to a large extent will be reflected in a data product specification – and the applications schema is only a part of the product specification.

7.2. ISO 19131

7.2.1. General remarks

According to ISO 19131, Data Product Specification, the scope of the standard is to describe requirements for the specification of geographic data products, based upon the concepts of other ISO 19100 standards. It also provides help in the creation of data product specifications, so that they are easily understood and fit for their intended purpose.

It is stated in ISO 19131 that a data product specification may be created and used on different occasions, by different parties and for different reasons. It may, for example, be used for the original process of collecting data, as well as for products derived from already existing data. It may be created by producers to specify their product, or by users to state their requirements.

Some of the definitions in ISO 19131 give the understanding that some elements are only applicable for product specifications describing existing products. One example of this is identification information, where the definition of geographic description is stated to be “description of the geographic area within which data is available”. We consider this to be interpreted like “description of the geographic area within which data is or is intended to be available”. This is a logical conclusion since there are no notes or text in the standard stating that a mandatory element is only applicable for data product specifications describing existing data.

ISO 19131 specifies nine UML packages, some are mandatory, some optional.

[3] approved In November 2005 during the formal DIS vote

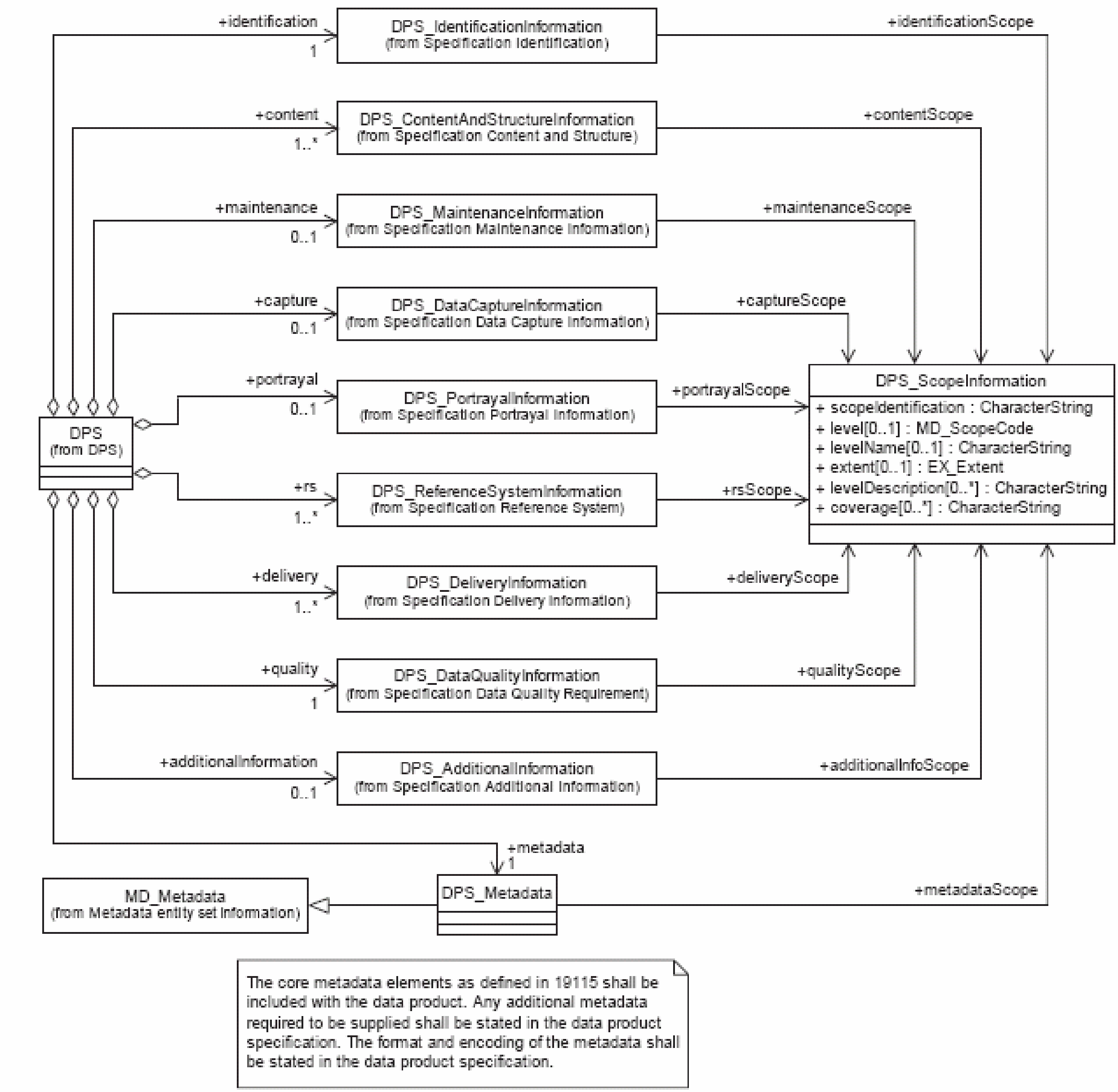


Figure 7.1. UML model for scope information.

The above figure is identical to Figure D.1 — UML model for scope information. In general, a product specification may have different requirements on different parts of a product. DPS_ScopeIdentification defines which part of a product a requirement is related to. It could be a geographic extent or part of a product identified by the scope code, like attribute, attributeType, feature, featureType, etc.

ISO 19131 states that all product specifications shall state the requirements for:

- Content and structure.
- Reference System.
- Delivery.
- Quality.
- Identification Information.
- Metadata.

7.2.2. Identification Information

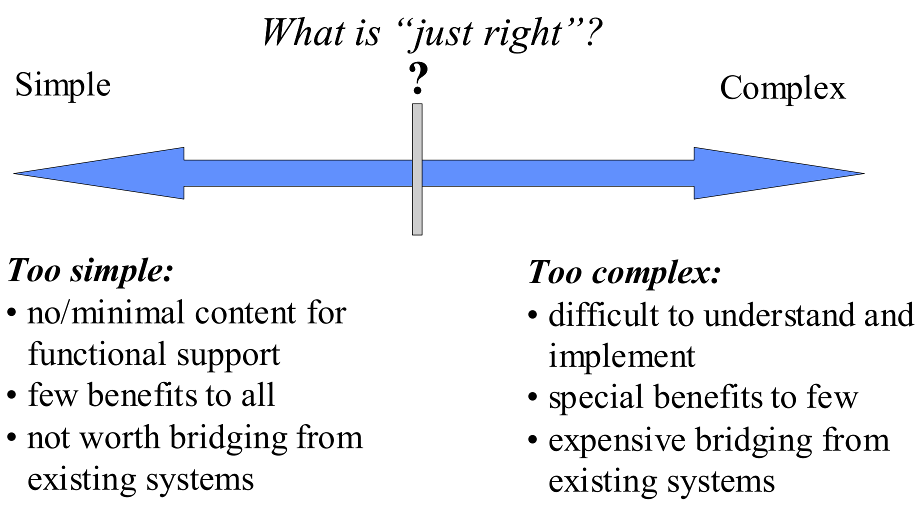
This package contains general information such as the title of the specification, a list of the main theme(s) and the geographic extent of the data.

The identification information allows for the description of spatial resolution. Resolution is an important criterion for the usefulness of data.

7.2.3. Content and structure

This package contains the application schema requirements according to ISO 19109 (Rules for application schemas) and ISO 19110 (Methodology for feature cataloguing) for feature data in general as well as ISO 19123 for coverage and imagery data.

One of the main questions regarding the framework design constraints on the content is if a data product should fulfil the requirements of a large and general community or a small and specialized community. See the figure below.



(from: Douglas Nebert, FGDC, 2005)

Figure 7.2. Issues relating to balancing the complexity of a Specification.

Some of the INSPIRE principles (see data harmonisation component 1.1) give overall guidance on the design constraints by stating:

- Data should be collected once and maintained at the level where this can be done most effectively.
- It must be possible to combine seamlessly spatial information from different sources across Europe and share it between many users and applications.
- It must be possible for information collected at one level to be shared between all the different levels, e.g. detailed for detailed investigations, general for strategic purposes.

Based upon these principles one should believe that the models should neither be too simple, nor too complex, and the requirement models should reflect this viewpoint. The checklist shown in Chapter 8 is intended to help in identifying the "right level" of complexity.

Relevant data harmonisation components (see 4.6.2): 1.1, 1.3-1.5, 1.7, 2.3.

7.2.4. Reference system

It is quite obvious that a user may have requirements related to which spatial referencing system that should be applied, often related to national or local systems that should be applied together with the data.

As mentioned before, most of these requirements may be fulfilled by a coordinate transformation service (a service offering coordinate conversions between two coordinate reference systems according to ISO 19111). The next question is if the quality of the transformation service will have any significant impact on the requirement on data quality as such.

Another aspect of reference systems is temporal reference systems. The widespread application of computers and geographic information systems has led to the increased analysis of geospatial data within multiple disciplines. Geographic information is not confined to a three-dimensional spatial domain. Many geographic information systems require data with temporal characteristics. It shall be questioned if there are any requirements for the temporal characteristics of geographic information including feature attributes, feature operations, feature associations, and metadata elements that take a value in the temporal domain.

Relevant data harmonisation components (see 4.6.2): 1.6, 2.4.

7.2.5. Delivery

Delivery information has two components, Delivery medium and delivery format. Both of these may be of interest for the requirement model.

In particular, this item is also related to the topic of service interfaces used to deliver the data products.

Relevant data harmonisation component (see 4.6.2): 3.4.

7.2.6. Data quality

Data quality is a difficult issue. Most of the data available has only a few quality statements, and hardly related to any standard. In the long term, this may change, also due to user requirements.

ISO/TC 211 has standardized a quality model, modelled as part of the metadata. The model consists of data quality information according to 5 quality elements, and lineage information. The 5 quality elements are:

- Completeness.
- Thematic accuracy.
- Logical consistency.
- Temporal accuracy.
- Positional accuracy.

It is possible to require this as a quantitative result or as a conformance result.

However, there is a challenge to formulate the questions for the requirement modelling in such a way that the results align with the standardized quality model. With the exception of positional quality, the awareness of the quality information is weak within the GI communities.

Requirement questions could be something like:

- Are there any requirements related to correctness of thematic classification?
- Are there any requirements related to the accuracy of non-quantitative attributes?
- Are there any requirements related to the accuracy of quantitative attributes?

However, each of these requires explanation and examples.

Lineage is information about the events or source data used in constructing the data specified by the scope or lack of knowledge.

Relevant data harmonisation component (see 4.6.2): 3.3.

7.2.7. Maintenance information (optional)

Where this section is included, it describes the principles and criteria applied in the maintenance of the data once it has been captured. This includes the maintenance and update frequency (frequency with which changes and additions are made to the data product).

This section, if provided, is simply a textual description and does not provide for / require any substructure in the information to be provided in the data specification.

Relevant data harmonisation component (see 4.6.2): 3.2.

7.2.8. Data capture information (optional)

Where this section is included, it contains a data capture statement, i.e. a general description of the sources and the processes to be used. It may allow freedom of choice for the data capture process, or may specify one particular data capture process.

This section, if provided, is simply a textual description and does not provide for / require any substructure in the information to be provided in the data specification.

Relevant data harmonisation component (see 4.6.2): 3.7.

7.2.9. Portrayal information (optional)

If included, this section provides information on how the data held within the dataset is to be presented as graphic output as a plot or as an image expressed as a reference to a set of portrayal rules in conformance with ISO 19117. Note that this section is not a focus area of the RISE project; however, it is acknowledged that this information is typically important for the use of geographic information in practice.

Relevant data harmonisation components (see 4.6.2): 1.9, 2.4.

7.2.10. Additional information (optional)

This may be one of the challenges and covers all kind of requirements that is not predefined in the ISO 19131 DPS or underlying standards. It is expected that a lot can be learnt from use cases / business models outside the spatial domain as well.

7.3. Service specifications

From a business / requirement model point of view, the model of ISO 19131 could in principle be extended to cover both data and services, which often are required to fulfil all user requirements.

Following the UML model for data product specifications, a similar model may be developed – taking ISO 19119 into account – to describe more precisely the output from the requirement / business model in form of services, too.

One example where a service is required is the following:

- User requirement: Data in a European datum.
- Data Product: Data is in a national datum.

The data product as such does not fulfil the user requirement, but a transformation service would definitely do so, transforming data from the national datum to the European datum.

There will be several examples where the data product as such does not fulfil the user requirements, but that different kind of services will transform data to the requested model, like semantic interoperability and ontology.

RISE as a project focuses on the data specification aspects. In the GMES/INSPIRE architecture a number of additional services are instrumental for a working infrastructure including catalogue/registry and processing/transformation services mentioned above.

8. Checklist for data harmonisation components

The harmonisation process consists of several steps, each addressing the data harmonisation components identified above. A checklist has been developed (see Annex 2) to help to identify the questions to be addressed in each step for each component.

The important steps supported by the checklist are:

- Requirements: the expected results as derived from the use case description
- As-is analysis: situation with respect to the existing data sources
- Gap analysis: identification of issues (gap between requirements and current situation)

  NOTE The gap analysis is useful at 2 steps :

  - between the existing data sources and the requirements
  - between the existing data after harmonisation and the requirements :
    - for the use case itself, to choose the best harmonisation approach if there are several options, to look for new solutions in the future, etc.
    - for the Methodology development itself to assess different harmonisation approaches
- Harmonisation approach: proposed approach for the harmonised data specification

The requirements column will typically be provided only once per use case. The as-is and gap analysis has to be executed once per input dataset. A harmonisation approach column will be used once per data product specification, i.e. typically once per data theme.

Not all components are relevant in every step.

It is proposed to document at least summary information of the analysis in a matrix representation so that it is easy to get an overview per step (e.g. an overview of all requirements) or per data harmonisation component (e.g. spatial profile required and used in all existing data sources).

To help this process the checklist is documented in Annex 2 as a spreadsheet where the individual cells shall be filled in and the checklist information shall be addressed and replaced by the use case specific information.

Note that it may be required/helpful to add additional columns; for example, if the analysis involves multiple data sources.

The checklist will assist in the process of creating a Data Product Specification, but will not be sufficient detailed to replace a discussion between the domain expert and the 'modelling' expert.

There is no specific sequence of questions in the checklist; the sequence may be changed from use case to use case, depending on the application domains.

9. Use case template

9.1. Introduction

A use case defines a goal-oriented set of interactions between external actors and the system under consideration. Actors are parties outside the system that interact with the system. An actor may be a class of users, roles users can play, or other systems.

A use case is initiated by a user with a particular goal in mind, and completes successfully when that goal is satisfied. It describes the sequence of interactions between actors and the system necessary to deliver the service that satisfies the goal.

Generally, use case steps are written in an easy-to-understand structured narrative using the vocabulary of the domain. This is engaging for users who can easily follow and validate the use cases, and the accessibility encourages users to be actively involved in defining the requirements.

According to ISO/TS 19103, a general structure of work products and models for a software development process using UML can be described for example as follows (adapted from the unified software development process):

- Business/Domain models
  - Domain model, business model, business use cases (reverse/as-is, forward/to-be)
- Requirements models
  - Use case model, actors, architecture view, glossary, user interface prototype, etc.
- Analysis/Architecture models
  - Architecture model (identify service/control objects, information/entity/feature objects and boundary/interface objects; analysis of packages/classes; use case realisation)
  - Design models (may be implementation neutral)
- Implementation models
  - Implementation model, components, etc.
- Test models
  - Test model, test cases, test plan/procedure, test component, test evaluation

In this step of the RISE process, it is the business model and the requirements model approach that are the most relevant aspects, captured in use case diagrams and textual use case descriptions.

The following is a use case template that may be used to describe use cases.

9.2. The Template

Describe the use case(s) with as much detail as possible. This information will be the input in order to (i) detect possible new requirements, and (ii) assess the services that need to be implemented according to the needs of the end users and stakeholders.

The description consists of three parts:

Part 1: UML use-case diagram to provide an overview of the use case(s) and the involved actors (example below).

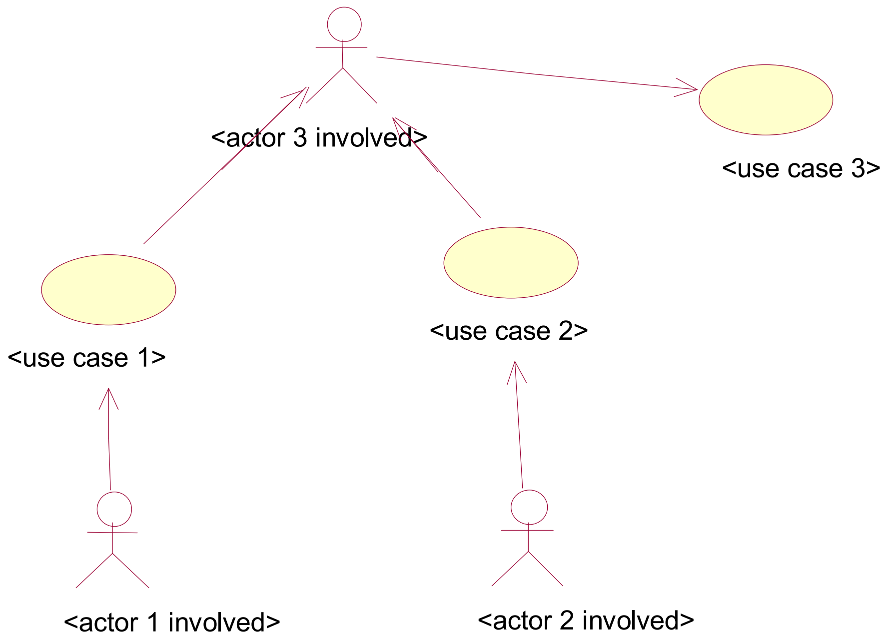


Figure 9.1. Representation of actors involved in use case.

Part 2: Narrative explanation of the use case(s)

(See RISE Use Case Document [11] for examples.)

Part 3: Detailed, structured description of the use case (template below)

Use Case Description	
Name	<name of use case>
Priority	<high/medium/low>
Description	<short description>
Pre-condition	<What are the pre-requisites from other use cases? What input is required?>
Flow of Events – Basic Path	
Step 1.	
…	
Step m.	
Step m+1.	
…	
Step n.	
Flow of Events – Alternative Paths	
Step m.	
Step m+1.	
Post-condition	<What is the output for other use cases? What are the anticipated next steps?>
Data source: <Name> *[repeat per data source]*	
Description	
Data provider	
Geographic scope	
Thematic scope	
Scale, resolution	
Delivery	
Documentation	

Repeat for additional use cases.

10. Experiences & Recommendations from the RISE Project

10.1. Introduction

In order to test the RISE Methodology the RISE project has conducted a use case analysis and used this as a basis for creating conceptual models in UML and finally ending up with harmonised data specifications. As a final test, the implementation and use of the harmonised data specifications has been tested in the RISE Test Environment.

10.2. Methodology

One of the first discussions in the RISE project concerned being precise about the objective(s) of the project: the RISE Methodology is about developing data product specifications in a specific context, which is mainly concerned with INSPIRE and other related activities (e.g. GMES). In other words, the objective is to provide **harmonised** data specifications: harmonisation being seen as the process to provide access to data through spatial data services in a representation that allows for combining it with other harmonised data in a coherent way.

So, the Methodology has had to take into account several viewpoints:

- Development of data specifications: this part deals mainly with the user requirements study, the identification of features and attributes and with the writing of a “data product specification”. This part has been done for years by data producers and is far from new. However, even about this traditional part, the RISE Methodology brings interesting input, by proposing and testing tools for use case development (to help to find the user requirements) and by recommending and testing the use of ISO 19131 for data product specification (which enable data specifications from different themes to be presented in a standardised way).
- Existing data: the aim is to have harmonised data product specifications, not to define a new product from scratch; so, existing data has to be taken into account when defining the product specification. The RISE Methodology has considered this point by including the as-is analysis, the gap analysis (gaps between requirements and existing data) and the harmonisation approach (ways to match existing data to harmonised specifications) and by recommending ISO standards which offer a common “language” to describe geographic data.
- Providing access to data through services: the aim is that data producers keep their own data specifications for their national/local data but supply Web services which make the translations or transformations from the local model to the harmonised model. Ideally, this transformation should be on-the-fly. The RISE project has dealt with this point by considering the system level (deriving GML schema from UML one) and in the testing phase (transforming national test data into the common data model).

The RISE Methodology has been closely developed in link with INSPIRE:

- the main objective of the RISE Methodology was to serve as starting point to the document D2.6 : “Methodology for the development of data specifications” and it has been effectively adopted and adapted by the INSPIRE Data Specification Drafting Team.
- There has been collaboration between the RISE project and the INSPIRE DT by joint meetings, by involvement of common experts, and by review of documents.

The steps of this collaboration have been the following:

- definition by the INSPIRE Data Specification Drafting Team of the “data harmonisation components”,
- adoption of these components in the RISE Methodology (especially in the check-list),

- adaptation by the INSPIRE Data Specification Drafting Team of the RISE Methodology.

This collaboration with the INSPIRE Data Specification Drafting Team has shown both that the RISE Methodology is a general methodology to develop harmonised data specifications and so, may be useful for other projects dealing with this topic. However, iIt is recognised that the RISE Methodology may not be adopted as a whole but may have to be adapted for each specific context. Specially, the Methodology has to be adapted to the "size" of the harmonisation initiative.

EXAMPLE 1: the RISE Methodology has had to be "extended" for the INSPIRE context. The adaptation to INSPIRE has mainly consisted in adding INSPIRE specific context (legal aspects, work programme) and in adding helpful annexes, such as a template and examples of data product specifications or examples of harmonisation approaches.

EXAMPLE 2: the RISE Methodology has had to be "reduced" in RISE itself: during the "exercise" of developing specifications, we decided that some processes might or even should be simplified to fit with the RISE context. For instance, some data harmonisation components are necessary for INSPIRE but are not applicable in the RISE context, as RISE has quite a more narrow scope than INSPIRE (see remarks on the check-list in the chapter on Use-Case development). Another example is given by the registries (feature conept dictionary, glossary) indicated in the Methodology but which have not been used in the RISE prototype (as it was considered it more relevant to just include terms and definitions in the data product specification).

The RISE Methodology has already been provided for external review; the general feeling being that it is that it is quite a theoretical document and that the check-list may supply a more accessible entrance point to this Methodology. Other points to note are:

- the core document is full of ISO concepts and most people are still not familiar with these concepts,
- the check-list is easier to understand by people just having general GI knowledge.

However, the check-list is just a tool and should not be considered as a summary of the Methodology.

It was discussed if the Methodology should or not propose some methods for doing the as-is analysis and the gap analysis. The main conclusions have been:

- it is necessary to understand existing data but not to have them described in the same formalism as the harmonised data (e.g. it is not necessary to have UML schema),
- it is not possible to completely formalise these two steps due to the large variety of potential cases. The experts involved in the harmonisation initiative will have to choose the way which best suits the context.

More generally, this Methodology is not a press-button tool; it will help data harmonisers but a large part of the work will remain dependant on human expertise.

10.3. Choice of use cases

One of the major challenges in RISE was the choice of the use cases. In its project proposal, the RISE consortium addressed the GMES data harmonisation issue with a focus on hydrography, and elevation data themes. The additional theme of Land Cover was added early in the project as this allowed additional aspects of the harmonisation process to be illustrated. During the first phase of the project, relevant GMES projects were analysed. As a result the RISE general use case topic was narrowed to 'Diffuse Nutrient Leakage' reporting in the context of the Water Framework Directive (WFD).

The following criteria were applied to the selection of use cases in RISE. Some are specific to the RISE project but some others are more generic and may be useful for other harmonisation initiatives:

- be a real use case, not an artificial one (generic criterion);
- deal with the themes identified in the project, i.e. for RISE: hydrography, land cover and elevation (generic criterion);

- be challenging enough, i.e. the use case must raise some harmonisation issues, as the core objective of RISE was to demonstrate how to harmonise data (specific criterion);

  EXAMPLE1: the choice of the Elevation use case. It was decided to work with a use case implying DEM and not just with an attribute about altitude as attached to hydro data in Sweden and Norway.

  EXAMPLE 2: as the aim was development of **harmonised** data product specification, the use case should involve trans-border issues, so It was decided chose to use test data for the prototype on an area between Sweden and Norway (for hydrography and land cover) and not only on Swedish data.

  EXAMPLE 3: for land cover, as the requirements for WFD in 2005 were rather weak concerning spatial data, it was decided to consider requirements for 2010 reporting (even if they were not yet perfectly defined) because these requirements involve more spatial data.

- not be too complex, due to the limited resources of the project (generic criterion);

  EXAMPLE: for hydrography and land cover, it was decided to deal only with a subset of WFD Nutrient Leakage Reporting procedure (the harmonisation of source data to the input data needed for the nutrient leakage modelling), not with general reporting for WFD.

- stakeholders are willing to participate, to spend time attending meetings, filling questionnaires etc. (generic criterion);

  NOTE: this turned out a major bonus with the Swedish-Norwegian use case.

- be convenient for people involved in the harmonisation initiative (generic criterion). For instance, use cases involving countries that are involved in the consortium, which makes communication with local experts easier (language!), enables access to local networks of expertise, and reduces travel costs.

  EXAMPLE: In RISE, the use cases did involve:

  - Sweden and Norway, as Mapping Agencies from both countries are involved directly or indirectly in the RISE consortium;
  - France, the elevation use case has been supplied by BRGM who was involved in the RISE project (BRGM being partner of OGC-E);

In RISE, some information has been added about the SDIGER project (even though it is not a real use case) because it was also easy to add this information through the involvement of IGN France in both RISE and SDIGER).

Unfortunately, the RISE Project was not funded sufficiently to undertake a complete end-to-end use case. RISE was limited to selecting focussed use cases which demonstrated the key aspects of the specification harmonisation process. A full end-to-end Use Case Project would be valuable for the WFD and existing national models, and this would be a substantial activity in its own right. It could perhaps be regarded as a possible follow up project.

After investigating quite a lot of information resources, it was obvious that the nutrient leakage report according to the Water Framework Directive will be a significant undertaking. One of the major inputs to the use case analysis was the EUROHARP project. EUROHARP was a research project supported by the European Commission under the Fifth Framework Programme and contributing to the implementation of the Key Action “Sustainable Management and Quality of Water” within the Energy, Environment and Sustainable Development Programme. One of the primary objectives was to evaluate nine contemporary quantification tools and their ability to estimate diffuse nutrient (N, P) losses to surface freshwater systems and coastal waters.

This review of EUROHARP methodologies revealed that approaches differ markedly in their complexity, resolution in time and space, and resource (data, time) requirements. Such differences reflect the history of the models’ development and original purpose. Most tools have only been applied to a limited range of catchment types. The acquisition of specific input data may also limit the applicability of some models[4].

Due to these complications, it was decided in RISE to take a more practical approach, looking into the requirements for the calculation according to the national models. The use cases for RISE describe the specific situation in Sweden and Norway (for hydrography and land cover) and France (for elevation). This was sufficient to test the RISE methodology. As a consequence from the limitations to the use cases, the RISE methodology results in a harmonised data product specification that cannot serve as “The European Data Specification for WFD reports”. However, the procedure could work if the necessary data were made available according to the RISE Methodology.

10.4. Use case development

It should be clear that the harmonisation requirements do not come out of the Methodology. The actual requirements come out of the use case analysis, and are formalised in the Conceptual Schema and Data Product Specification (DPS).

The checklist is a very important part of the Methodology, as it helps the facilitator in the use case development process to address aspects that are relevant for the harmonisation. However, RISE did not have the intention that the result of the use case and checklist analysis should alone be sufficient to create a harmonised data product specification. The checklist focuses on issues at a higher conceptual level, and the knowledge about this is beneficial for the further process. The process of developing a harmonised data specification based on the Methodology described in this document will typically involve at least a domain expert, GI architect, software engineer and facilitator. The checklist is mainly addressing the facilitator who is required to have an excellent understanding of GMES and INSPIRE, in particular the underlying user requirements and ICT and SDI architectures. The use case and checklist are provided as important background material to enable these resources to create a DPS, but it is not intended to be detailed enough to let the facilitator do this alone or just based on the checklist.

Some issues that came up while using the checklist:

- The initial check-list was full of questions about the use of ISO standards. The experience has proven that this way of phrasing the checklist was not helpful.
  - The ISO standards are appropriate to express the result (the Data Product Specification)
  - but they were not helpful/relevant for asking questions to the stakeholders involved in the use case; generally, these stakeholders are domain experts with some GIS knowledge but they are not fully aware of ISO standards. So, it is better to ask questions with more common vocabulary they can understand (e.g. even within the project team, we have had misunderstandings about “coverages” because of different meanings in ISO and in GIS-vendor terminologies).
- The check-list has been based on the data harmonisation components identified jointly by RISE and the INSPIRE Drafting Team on ‘Data Specifications’. As INSPIRE is a significantly wider initiative than RISE, some components are necessary for INSPIRE, but were not relevant for RISE:
  - For RISE, the main components involved in the use case have been terminology, application schemas, multilingual aspects, spatial aspects, coordinate reference systems, identifier management, data transfer, consistency between data, and data capturing rules. For these components, RISE has had substantial response from the interviews. It is expected that these components are of interest for most of the use cases in any harmonisation initiative.
  - Some components may be of interest in other use cases (e.g. object referencing modelling, portrayal model, data quality, metadata, maintenance, multiple representations etc.).

[4] see EUROHARP second-newsletter.pdf

  - Some other components (e.g. INSPIRE principles, Reference Model, conformance) are only relevant for wide harmonisation initiatives and even in this case, it is likely that many use cases won't put requirements on these components.

Other issues that came up during the implementation of the use cases are:

- The use cases have also shown that it is important to agree upon terminology and address multilingual issues at an early stage. Terminology was an issue even at high level. The consortium discussed not only the names for the land cover classes but also found it difficult to agree on appropriate terms for the general concepts.
- The use cases have proven the importance of properly identifying the requirements and thus to have an engaged user.
- Good relationship with users is a key point during this phase. Of course, this point has to be taken into account in the choice of use cases but also during all the use case development:
  - Need of physical meetings: exchanges by mail and telephone are useful but not enough, meetings and interviews are also required (e.g. to explain the aim and context of the harmonisation initiative, to collect information from users).
  - It is necessary to encourage users to participate:
    - Listen to them, first try to understand what their work is, and only then ask technical questions.
    - Use a language they can understand (e.g. use rather GIS concepts than ISO ones).
    - Make them comfortable (e.g. they may be reluctant to show their work if they do not consider it to be good).
    - Don't give them lessons.
    - Explain that the harmonisation initiative may facilitate their work later (even if it requires some extra effort now).

The Methodology, when defining the roles, describes only the technical skills which have to be present in the harmonisation initiative. However, relational skills should not be forgotten: people involved in harmonisation initiatives have to be able to work together and with users; to listen to each other and to speak a common language, generally English if in an international context.

10.5. Specification development

It needs to be emphasised that the data product specification does not describe an existing data product. It describes a harmonised specification for a (virtual) data product, which will be created by transforming the existing data on the fly, potentially by invoking other services in a service chain. The complexities of these mechanisms are dependent on the gap analysis, which represents the gap between the existing data (as identified in the as-is analysis) and the harmonised data product specification.

One main discussion topic during the project has been the link between use case and data specification. In real-life, a use case generally involves several data themes (e.g. hydrography, transport, elevation, etc.) and this requires several data product specifications. In addition, a data product specification is generally designed to fulfil the requirements of several use cases.

The other discussion topic has been about how to extract data requirements from the use case itself. The experience has shown that:

- the template for use case development is mainly useful to understand the use case, and to identify the processes and the main issues,

- the identification of requirements comes rather from the check-list (either directly or by the documents identified in the check-list),
- In the RISE project, it was decided at the beginning to choose only one overall use case: the “Diffuse Nutrient Leakage reporting to the WFD” for the 3 themes to be considered (hydrography, land cover and elevation). Nevertheless, some adaptations have had to be done during the project:
  - For elevation, it was decided to select the use case about computation of IDPR (“Indice Développement Persistence des Réseaux”) which is indirectly used for WFD reporting (but not for Nutrient Leakage). Moreover, the data specifications for elevation were developed keeping in mind the requirements on DEMs for orthophoto production in the GMES programme.
  - For land cover, the reporting for 2010 has been selected for data specification development because it was considered as more relevant for the project than the reporting for 2005, which is the one described in the RISE use case document.

The first point just illustrates that the specification development is an iterative process (as explained in the RISE Methodology) as one may have to change one’s mind if the initial implementation does not prove to be fit for purpose.

The second point is more problematic and opens the question of how to proceed when user requirements are not yet defined or are only very weakly defined. In this case, the template for use case development proposed by RISE may require modification as this template aims at describing existing processes.

The processes to extract specifications from the three themes involved in RISE have been quite diverse:

- The specification process for the theme “hydrography” has been really “requirements driven”; there is a document “Implementing the GIS elements of WFD” which gives more or less specifications of data required for WFD. The work within RISE has mainly consisted in extracting the sub-set of requirements for “Nutrient Leakage Reporting”. We can say that the RISE specifications on Hydrography are more or less a profile of the requirements identified by the WFD document, while they have been reduced according to the criteria mentioned above under ‘choice of use case’.
- The specification process for the theme "elevation" has been based more on existing data; BRGM uses the French national DEM (BD ALTI from IGN F) to compute its IDPR index and this DEM is convenient for this application. As Sweden and Norway had DEMs with similar specifications (same grid size, similar vertical accuracy), the RISE specifications have just adopted these common characteristics and the harmonisation issues are mostly concerned with coordinate reference systems and data formats. In practice, the RISE specifications on elevation are more or less an adoption of the French DEM characteristics with adaptation to the European context by the choice of a common coordinate reference system (both vertical and horizontal).
- The specification process of the theme “land cover” is probably the most complex and as a result, the most interesting:
  - As it is based on reporting for WFD in 2010, the requirements were not yet perfectly known; there have been three different sources for requirements (the EUROHARP model, the Swedish TRK model and the Norwegian Teotil model).
  - The Methodology has really involved both the user requirements study (by considering the three requirement models), the as-is analysis (by considering the potential data sources: CORINE and RedMap for Sweden, Markslag from Norway) and the gap analysis (by comparing requirements to existing data in matching tables).
  - As the case was more complex, the experience and knowledge of the domain experts involved in the use case has been crucial in taking decisions about the specifications to be adopted.
  - Taking into account existing de-facto standards in the domain, the resulting specifications are more or less an aggregation of the CORINE specifications.

- One other conclusion from this use case may be that it is easier to use already harmonised data, when they exist, rather than the heterogeneous national data sources.

The Methodology recommends the use of ISO 19131 (Data Product Specification), and the RISE project has used this standard as a template for its specifications. This puts some challenge on the Methodology, as ISO 19131 applies to real existing products, while the harmonised data specification in RISE describes a virtual product which is derived from heterogeneous national sources, each with its own specification. There has been intensive discussion on this issue in the consortium. Finally the partners agreed that even if ISO 19131 does not fit perfectly to the scope of RISE, this standard should be used, as the known difficulties in implementation could be overcome and would not warrant a deviation from established templates.

EXAMPLE 1:

> The geographic description is a mandatory element in ISO 19131. In the context of RISE this can not be described by exact area, because the RISE Methodology is considered applicable for the whole of Europe, while the RISE Test Environment only uses data from a small area at the Swedish-Norwegian border.
>
> The text introduced to the RISE data product specification is: “Geographic description: Europe. Note: this is the geographic area as intended in the RISE Use Case of Diffuse Nutrient Leakage Reporting to the Water Framework Directive. The actual tests in the RISE prototype will be carried out with data of the cross-border water basin Enningdalsälven / Enningsdalselva in Sweden and Norway.”

EXAMPLE 2:

> A data product specification includes a section on data capture. Usually this section specifies requirements such as the minimum size of feature to be portrayed in the dataset. These requirements need to be applied to the original data sources from which the RISE virtual product is derived,
>
> The RISE Data Product Specification will, in general, inherits the rules on data capture and does not define additional requirements.
>
> This keeps the RISE procedure flexible. It means that the capture criteria are applied on the source data, which is channelled into the harmonisation process as it is. Another option might be to specify capture criteria which would then override the data specification of the source data. In that case the harmonisation cannot be performed on the fly, it would require the specification of the source dataset to be adapted and the dataset to be re-engineered.

Some other topics:

- The conceptual schema has been highlighted in the RISE process as the core part of the harmonisation approach. However, the Methodology addresses all other elements of the data product specification also, and they may be described according to ISO 19131.
- The data product specification does not go into detail on a number of topics. For example, it does not specify requirements on the geometry such as the density of vertices or interpolation methods. Instead, the Methodology allows several interpolations, both for curves and surfaces. However, complex interpolations or other topological or geometric primitives are not specified because the requirements from the underlying models, in for example the Norwegian/Swedish use case, do not require it. But to understand the requirements for ‘complex’ geometry/topology, the underlying models need to be fully understood.
- Proper identification of requirements turned out to be crucial to the development of application schema.
- In the Methodology document, the single entry point to the development of application schema is the use case description, with the requirements to be derived from that document. In reality many different documents had to be used to get the required information.

- The matching table turned out the most important document in the preparation of the conceptual model. In this role it complements the checklist:
    - The identification of features and attribute generally involves existing documents (describing user requirements and existing data); the comparison is made in matching table and may be considered as the first phase of the gap analysis.
    - The check-list is mainly useful for the other components of data specification (coordinate reference system, geometry/topology, quality, etc.)
- In comparing the national data models, the language posed a significant obstacle. Even when national and English terms appear to coincide, it required a native language speaker with domain expertise to confirm (or reject) the match.
- The RISE description of work included a multilingual feature catalogue. This was incorporated in the data specification by adding 'aliases' to names of feature types, attribute types and attribute values, in the languages Swedish, French and German.
- Another issue for debate was whether there should be one data specification for the RISE prototype, or three separate specifications for Hydrography, Land cover and Elevation? ISO 19131 would have allowed the first option by introducing so called 'specification scopes' that divide the product into partitions. This structure was considered too complicated, as the themes do not have very much in common. Therefore, the RISE data specification document is split into three data product specifications for Hydrography, Land cover and Elevation, each of them repeating the clauses from ISO 19131.

10.6. RISE Test Environment

The RISE Test Environment (RTE) demonstrates the creation of harmonised data products "on the fly" and allows users to test data structured according to the associated Application Schema generated by the RISE Project. The user feedback allows further development of the Application Schema, if required.

An interactive website is used where users can view or download products (datasets) whose production has been generated based on the RISE harmonised Data Product Specifications (hDPS). The design and operation of the RTE is described in the RISE Report on the Operation of the RISE Testing System (http://www.eu-rise.org/.

Test data were obtained covering the hydrography, land cover and elevation themes for two river-basin test areas, namely, Vindån, in the East of Sweden, and Enningdalsälven, a cross-border area in the West of Sweden / East of Norway.

Hydrography

There were two components to the hydrography use case. The first was the creation of River Basin District Overview Information, and the second was the creation of Surface Water Body Categorisation information, both as required to satisfy Water Framework Directive reporting requirements. The input data came mainly from the General Map in Sweden (provided by Lantmäteriet, the National Land Survey of Sweden), a refined hydrographical layer (provided by SMHI, the Swedish Met Office), the REGINE dataset and National Water Theme Layers in Norway (provided by Statens Kartverk, the Norwegian Mapping Agency).

Transformation to the harmonised Data Product Specifications was developed for both input data sources; and an output Data Product Service was implemented through the use of a translating Web Feature Server (WFS-X). The transformations were implemented without changing or adding to the data in the underlying Oracle database through a combination of WFS and ORACLE SDE functionality.

It should also be recognised that RISE addressed the issue of harmonising the data product specification and did not address the other harmonisation issues such as edge matching at national boundaries. However, one of the Test Environment external tests does demonstrate how the harmonised data products can be used to help address these other harmonisation issues.

Land Cover

The Land Cover tests demonstrate the creation of a dataset suitable for diffuse nutrient leakage modelling derived from two completely different land cover datasets. The Swedish input came from the Swedish CORINE, which contains land cover data classified in 15 main classes, each with a number of sub-classes. The Norwegian data came from the N50_Areal25 dataset (again provided by Statens Kartverk).

In the same way as for hydrology, the transformation to the harmonised Data Product Specifications were developed for both input data sources; and an output Data Product Service was implemented through the use of a translating Web Feature Server (WFS-X). The transformations were implemented without changing or adding to the data in underlying Oracle database through a combination of WFS and ORACLE SDE functionality.

Elevation

Elevation data, in the form of a Digital Elevation Model (DEM), has a raster based structure and so has to be handled in a different way to the hydrography and land cover data.

The elevation requirements for harmonising the data projection and sampling interval can be addressed with a standard OGC compliant Web Coverage Server (WCS). A WCS was deployed within the RTE and test data for the Enningdalsälven river basin was made available.

The external testing involved the generation of a theoretical river network to compare to the real river network as an aid to river basin flow modelling.

RTE Issues

Some issues that came up during the implementation and testing or the RTE were:

- The automatic schema translation cannot handle all types of missing information, for example, names of features (such as river sub-basin name), if it is not in the original input dataset.
- Care must be taken when defining re-projection parameters (e.g. from the Swedish RT-90 projection to straight long/lat). Changes to the reference datum (e.g. WGS 84) in addition to the definition of projection, must also be taken into account.
- RISE chose to use the latest GML Standard (GML 3.1.1) for the definition of its data products. Unfortunately, some of the software tools used in the test Environment are not able to handle this version of GML. However, updates to the software can be expected in the short term as the use of this GML Standard becomes more widespread.
- There were problems in handling the large GML files in the software tools used in tests of the RTE.

10.7. Cost Benefit Analysis

The costs and benefits of the harmonisation efforts are tracked based on the actual efforts which have been spent in the project to create initial samples of harmonised data products and specifications following the RISE Methodology. The RISE Cost Benefit Analysis thus provides a template for organisations who wish to deploy the RISE Methodology. By outlining the necessary steps and related efforts, each third party will be able to do a cost estimate based on its very own internal structure, processes and procedures as well as financial framework. In the same way the benefits are outlined on a generic level, thus allowing an organisation to calculate e.g. potential savings on the production side or gains due to newly addressable market segments based on their very own cost structures and marketing strategies.

Some issues that came up during the development of the cost benefit analysis were:

- Many data providers are governmental organisations and have legal obligations to implement, for instance, the INSPIRE directive, even though the cost benefit analysis shows a long pay-off time for their investments.

- Data providers can make an investment calculation to get more accurate figures for their costs and benefits. For other stakeholder categories it can be more difficult to get quantitative measures. In these cases, the cost and benefits have to be described in qualitative terms instead.
- It is recommended that cost benefit analysis work begins at the start of the project so that costs and benefits can be traced during the project.

11. Conclusions

The RISE project has documented a Methodology for developing harmonised data product specifications. This Methodology is applicable for both European harmonisation initiatives (e.g. in response to European Directives) and more local harmonisation issues in, for example, cross-border river-basin scenarios.

The RISE Methodology exploits international standards wherever possible, in particular those coming from ISO and the OGC. Furthermore, the Methodology embraces the concept of producing harmonised data products on-the-fly from existing data without the need to re-engineer the underlying datasets.

The Methodology has been tested in the RISE project using a real-world use case. The experiences have been documented and should be taken into account in any future revision of the Methodology.

The current proposal for a methodology for the development of INSPIRE Implementing Rules for spatial data themes [8] is based on the RISE Methodology. While the INSPIRE methodology is tailored towards the specific context of the INSPIRE Directive, most changes or additions are general in nature and not specific to INSPIRE. It is therefore recommended for parties interested in applying the RISE Methodology to consider the additional work done in the INSPIRE process, too. Note that the current INSPIRE document referenced above will be revised as a result of a public review and further validation through testing.

ANNEX 1: Data Harmonisation Components

from: INSPIRE D2.5 v1.0 Generic Conceptual Model [7].

A.1 Information Model

This group describes all aspects that need to be modelled, agreed and published. This ranges from the basic principles, to co-ordinate reference systems to application schemas. As a result, this group will define the ESDI in terms of achieving a level of harmonisation through interoperability. The components in this group can then be used to test whether or not (or how much) a dataset is consistent/compliant with the GMES and INSPIRE framework.

Component	Description
Principles	The INSPIRE principles are considered to be the general basis for the data harmonisation efforts – also in GMES. The first three of the five INSPIRE principles are relevant to data harmonisation: • Data should be collected once and maintained at the level where this can be done most effectively • It should be possible to combine seamlessly spatial data from different sources and share it between many users and applications • Spatial data should be collected at one level of government and shared between all levels
Reference model	This component will define the framework of the technical parts including topics like data models (i.e. conceptual modelling framework with rules for application schemas) and data administration (i.e. reference systems). The reference model of ISO 19101 shall be the reference model for INSPIRE. So far, no explicit reference model for GMES has been specified. Therefore, a working assumption is that the implicit Reference Model would be a yet unidentified subset of (ISO 19101, OGC Reference Model). Additional detail is provided by the other data harmonisation components.
Application schemas	The purpose of this component is: Provide a computer-readable data description defining the data structure - enabling automated mechanisms for data management. Achieving a common and correct understanding of the data, by documenting the data content of the particular application field, thereby making it possible to unambiguously retrieve information from the data. Data structures used by multiple themes are expected to be modelled in a common base application schema used throughout the ESDI. The expected approach is: Start with a minimal base application schema and move types to this schema as they turn out to be of general use. Abstract types will be used to model properties shared by objects across the different themes. The conceptual model will include generic mechanisms to support simple integration or interlinking of objects based on a common methodology for references to other features (explicit relationships between feature types of different themes shall be used based on clear requirements only and need specific maintenance procedures whenever they are defined). Application schemas will conform to ISO 19109. A CSL will be used in-line with the ESDI CSL Workshop recommendations: • Use UML as the CSL and restrict the UML profile. • Restrict General Feature Model to a profile. • Restrict allowed types from 191xx harmonized model to a limited set . Start with strict and limited profiles and extend the profiles if needed → a methodology is needed for this. Application schemas will include elements from the next components (types from the ISO 19100 profile, multi-lingual and cultural adaptability support, coordinate reference systems, units of measurement, object referencing).

Component	Description
Profile of the ISO 19100 model	Conceptual schema for describing the spatial and temporal characteristics of geographic features. Spatial vector geometry and topology. Temporal vector geometry and topology. Coverage features (examples of coverages include rasters, triangulated irregular networks, point coverages, and polygon coverages). Specified as a profile of ISO 19107, ISO 19108 and ISO 19123. Start with a simple profile and extend it based on requirements by applications. In selecting the profile, one needs to take the capabilities of software systems into account, too. Initial set of questions for the spatial model: • Which interpolation types are required for curves and surfaces? Is linear/planar interpolation sufficient? • Dimensionality of the geometries (0D, 1D, 2D, 3D)? • Sharing of geometry objects required? • Topology required? If yes: • How is topology related to geometry? • Is topology required "inside the system" (to support edge matching, consistency across themes, generalisation, etc.) or also for the user application? • Initial set of questions for the temporal model: • Support for features that move or change geometry with time? • Support for multiple versions of a feature (historic data)? Or even versioning of properties? • Initial set of questions for the coverage model: • Which coverage types are required? Additional parts of the ISO 19100 series may be added, e.g. for quality information (ISO 19113-19115).
Multi-lingual text and cultural adaptability	Conceptual schema for multi-lingual character strings (likely to be based on the model developed in ISO/TS 19139). To be used in all application schemas and as a result in data instances: all string valued properties that may be provided in a language shall use this type To be used in the dictionary model so that dictionaries may be multi-lingual, e.g. the feature catalogue or code lists. Cultural differences have to be taken into account, e.g. not all terms may be translatable from one language to another. In general, it still needs to be clarified, which parts of a European Data Specification and which parts of the data need to be multi-lingual.
Coordinate referencing and units of measurement model	This component will describe methods for spatial and temporal reference systems as well as units of measurements – including transformations and conversions. The focus is on reference systems that are valid across Europe (in case of projected systems split into zones this will be a collection of such systems covering the different zones). Specify units of measurements used by the reference systems or to be used in the harmonised data.
Object referencing modelling	This component will describe how information is referenced to existing objects rather than directly via coordinates. The model will define how “user defined objects” (e.g. property parcels, the River Rhine, sensors, etc) and application information (e.g. property owner, river quality samples) are based on the reference information and how to maintain such geographies and links. The model will support the generation and maintenance of “user geographies” based on common reference data. The aim is to promote the easy and reliable exchange of users application level data (e.g. river quality sample records) across several users who use a common base (this avoiding massive data transfers to support regular

Component	Description
	reporting). Example applications areas include: • Streets/highways • Rivers and Water • Property/addresses, Land • Utility Assets • Administrative areas
Data translation model / guidelines	This component is about translating from a national/local data model to a European application schema and vice versa. Translations are required for data and for queries. No well-defined set of translation capabilities has been standardised in the GI community yet. It is not yet clear, if there will be a need to specify translations also between different European application schemas, e.g. for different representations or for creating specifc information products, e.g. reports, from base data. Open question: Is there a need to identify a set of translation capabilities that should be sufficient to map between a national/local data model and European application schemas (as a help for the member states as well as a more explicit measure for the complexity of translations that are considered acceptable)? Since the ESDI uses a service-oriented architecture and the application schemas will only be exposed at the service interface, this is not strictly necessary. However, it is important to understand how and if data can be translated at all.
Portrayal model	This component will define a model for portrayal rules for data according to a Data Specification. It will clarify how standardised portrayal catalogues can be used to harmonise the portrayal of data. A portrayal model in conjunction with a generic view service would allow the use of user controlled portrayal rules for data accessible through network services (map service vs. feature portrayal service). Note that in the RISE project, this component will not be addressed in detail.

A.2 Operational Components

This group describes components that will be required to support an operational ESDI. This ranges from identifier management, object classification, terminology dictionaries, multi-language definitions to a test for conformance. This effectively could result in a registry of active operational components in the ESDI, including compliant datasets, objects and who holds them, schema translations, etc.

Component	Description
Identifier management	Every feature will have a unique identifier. This component will define the role and nature of unique identifiers (or other mechanisms) to support unambiguous object identification in: • Linking application information to reference objects • Defining user-defined geographic features • Sharing of Application information • Maintenance of reference objects Mechanisms required to ensure uniqueness (see note below) To ensure uniqueness some form of management system will be required. This does not mean that all organisations need to adopt a common form of identifier or other mechanism but the identifier management mechanisms (e.g. registers) in use at national level will need to be synchronised/mapped to ensure pan European integration. This component will also clarify: • how to treat copies of features; • how to link/x-reference objects;

Component	Description
	• how to find features in the European SDI based on their ID; • how to deal with feature versioning (if part of 1.4). Note that the same real-world phenomenon may be represented by different feature (with their own identifiers).
Terminology	This component will support the use of a consistent language when referring to terms via a glossary and dictionary. This needs to be registered and managed through change control with multi-lingual support. The ESDI needs to select a common terminology from all of the existing terminologies and/or their translations. Assumption: Use ISO/CEN as starting point, add as required. Start with terminology for generic concepts; theme specific terms will be added later. Terminology should be maintained at least in the official European languages (to be verified).
Feature catalogues	Feature catalogues define the types of features and their properties and are indispensable to turning the data into usable information. Feature catalogues promote the dissemination, sharing, and use of geographic data through providing a better understanding of the content and meaning of the data. The full description of the contents and structure of a geographic dataset is given by the application schema which is expressed in a formal conceptual schema language. The feature catalogue defines the meaning of the feature types and their properties specified in the application schema. This component will be a register of the harmonised catalogues - fully synchronised with the application schemas. In addition, the register should provide feature catalogues of the schemas at national level. The catalogues will define types developed by different user domains and these will not necessarily correspond from domain to domain e.g. a taxable hereditament classification will not necessarily correspond with an ownership classification or physical object classification. Reuse of existing types and classifications cross domains where appropriate. This forms part of the metadata within the ESDI. From the registry it should be possible to determine who or where land ownership objects are defined and maintained, where river objects are held (topographic or network), etc. Again, text elements in the Feature Catalogues should be maintained at least in the official European languages.
Dictionaries	Dictionaries maintained by registers will at least be required for: • all supported reference systems; • all supported units of measurement; • all code lists / thesauri used in the application schemas (multi-lingual, at least in all official European languages) • the data dictionary for elements used by application schemas (multi-lingual, at least in all official European languages) Metadata on dataset level will be available through services. In addition, portrayal catalogues and symbols will be available through services.
Conformance	To determine whether any data(set) is “harmonised” with the data specification, it will be necessary to apply certain tests for conformance. Ideally these should be automated but should include tests as appropriate against all components in the Information Model as well as against other information of a Data Specification. Since a Data Specification is specified by a ISO 19131 Data Product Specification, conformance would require also passing the ISO conformance Chapters.

A.3 Guidelines and Best Practice

This group of components supports the previous two groups by defining guidelines and best practice. This shall support the consistent implementation of ISO/CEN/OGC standards and other components of the infrastructure to promote data harmonisation and will include aspects like metadata (exploration and exploitation level), maintenance, deriving lower resolution abstractions, transfer and other items as these emerge.

Component	Description
Metadata	This component will cover metadata on the following levels: • Discovery level • Exploration level • Exploitation level Object level metadata will in general be described as part of the application schemas. The general aim should be to identify and document best practice. Metadata will be multi-lingual, at least in all official European languages. The metadata will also express the relationship of a dataset / product with the different data harmonisation components; partially this is covered by ISO 19115, but it needs to be checked if this is complete. Metadata will be used to specify which options are selected and will describe any potential deviations.
Maintenance	This component will define best practice in ensuring that application data can be managed against updates of reference information without interruption of services. This will require the definition of mechanisms by different stakeholder areas to manage: • Change only updates • Versioning of objects (and their properties) • Object lifecycles Maintenance requirements are part of the data specification. Propagation of changes across scale to maintain consistency of the data (automatic or manual processes); same for cross-border.
Data & Information quality	This component will advise the need to publish quality levels of each dataset using the criteria defined in the ISO standards, including completeness, consistency, currency and accuracy. This will include methods of best practice in publishing: • Acceptable Quality Levels of each dataset • Attainment against those levels for each dataset Quality requirements for dataset or information products are specified as part of a Data Specification. This is an important issue – as a result of further discussions this component may be moved to another component group, e.g. if it eventually would go beyond best practice and a quality model would be specified (on the basis of the ISO 19100 series). It is expected that this component will require further input and testing.
Data transfer	This component will describe methods for encoding application and reference data as well as information products. Support for Access to and update of data; requirements for change-only updates will be addressed. The encoding will in general be model-driven, i.e. fully determined by the application schema in UML.
Derived reporting / multiple representations	This component will describe best practices how data can be aggregated • across time and space • across different resolutions (“generalisation” of data) Such aggregation processes are used in particular to create the following results:

Component	Description
	• Multiple representations. • Derived reporting (example: typically water samples at 1 Km intervals are reported to the European level). This will describe how to create specific aggregated information directly from base data while retaining the integrity of the information of both the record and the geography.
Consistency between Data	This component will describe guidelines how the consistency between the representations of the same entity in different datasets (for example along or across borders, themes, sectors or at different resolutions) shall be maintained. The custodians of such datasets will decide by mutual consent on the depiction and position of such common features.
Data capturing rules	This component will describe data specification and feature type specific rules for capturing features.

ANNEX 2: Data harmonisation checklist

See RISE19_M&G_Annex2_V1.2.xls