Cost-Benefit Considerations in Establishing Interoperability of the Data Component of Spatial Data Infrastructures

Katalin Tóth, Paul Smits
European Commission DG Joint Reseach Centre
21027 Ispra, via E. Fermi, TP 262
Italy
katalin.toth@jrc.it

Abstract
The cultural, historical, and economical aspects are fundamental when making spatial data sets available as a component of Spatial Data Infrastructures (SDIs). Besides of technical feasibility the arrangements on interoperability of spatial data sets and services should be based on careful analysis of the related costs and the benefits gained. The task must be carried out with utmost attention, being data the most sensible and costly part of spatial information systems and infrastructures. It is prudent to take appropriate measures throughout the specification process of harmonised data to avoid excessive costs; on the other hand it is necessary to make the benefits visible to improve the acceptance of the necessary measures.
This paper proposes the multi-criteria-based Cost-Benefit Considerations instead of the classical Cost-Benefit Analysis for evaluating the impact of the data harmonisation process. Using the example of INSPIRE (Infrastructure for Spatial Information for Environment) it shows how cost-benefit aspects are considered at the different steps of data specification development. We conclude that such methodology yields balanced results in terms of objectives to be reached and the related costs.

Introduction
The challenges regarding the lack of availability, quality, organisation, accessibility, and sharing of spatial information are common to a large number of entities and activities and are experienced across the various levels of local, regional, national and global decision making. In order to solve these problems it is necessary to take measures of coordination between the users and providers of spatial information. Such coordination is manifested in Spatial Data Infrastructures (SDIs). The Directive 2007/2/EC of the European Parliament and of the Council adopted on 14 March 2007 aims at establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) for environmental policies, and for policies and activities that have an impact on the environment.

INSPIRE defines a SDI as a system where *metadata, spatial data sets and services, network services and technologies, agreements on sharing, access and use, coordination and monitoring mechanisms are established, operated or made available in an*

interoperable manner. Amongst these components spatial data has crucial role. In Geographic Information Systems data is considered to be the most resource-consuming part. Data collection can account for 60-80% of the total costs of setting up a system. Consequently, reusing data to be best possible extent is of prime economic interest in every SDI. INSPIRE does not require collection of new data; rather, it focuses on making available the existing ones through interoperability arrangements. Interoperability in INSPIRE means the possibility to combine spatial data and services from different sources across the European Community in a consistent way without involving specific efforts of humans or machines. It is important to note that "interoperability" is understood as providing access to spatial data sets through services, typically via Internet. Interoperability may be achieved either by changing (harmonising) the structure of existing data sets or by transforming them before they are made available as part of the INSPIRE infrastructure.

This paper focuses at a less discussed, but rather important aspect of SDIs: how the potential costs and benefits related to data interoperability can be captured and considered during the period of establishing the infrastructure. It outlines a methodology suitable for INSPIRE, which equally considers issues of technology and the methodology offered by economic science.

Methodologies for assessing costs and benefits in SDIs

For public enterprises, Cost Benefit Analysis (CBA) is an instrument for evaluating the feasibility or desirability of a given intervention in markets, which attempts to measure the efficiency of effectiveness of the intervention in achieving a stated public goal relative to the existing situation. The costs and benefits of the intervention are evaluated in terms of the public's willingness to pay for the actions required (benefits) or willingness to pay to avoid them (costs). In order to compare intervention alternatives, including remaining with the status quo situation, the expected costs and benefits are converted in comparable units, usually monetary values. Listing all the costs and benefits of an intervention or big project is quite difficult, which is the case of large infrastructure projects, where it becomes more complex to quantify them. It is usually possible to identify and quantitatively express the costs, but quantifying benefits is much more complex for the following reasons:

- Benefits are often gained in places other than where the costs occur, e.g. opening up new business or markets or boosting the value of existing ones.
- In many cases, benefits are not immediate and may accrue in different sectors of the economy not initially targeted by the intervention.
- Benefits are enjoyed by groups that do not create economic output directly connected to the intervention, e.g. increased security of citizens, improved quality of life, etc.

- The damage caused by failing to put in place the project or intervention may not be directly visible.
- It may be simply impossible to assign a monetary value to an expected benefit, for example, what is the value of a human life saved by an improvement in infrastructure or, similarly, the value of general improvement in quality of life for a community due to the intervention.

These difficulties strongly manifest themselves in cases of infrastructure investments and interventions for environmental protection. The INSPIRE Directive lays down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Community for the purpose of improving information availability, dissemination and use to aid in planning, implementing and monitoring environmental policies of the European Community and policies or activities which may have an environmental impact.

In preparation of the INSPIRE Directive, an extended impact assessment has been carried out. Concerning the Implementing rules laying down technical arrangements for the interoperability and, where practicable, harmonisation of spatial data sets and services, the Directive requires that "feasibility and cost-benefit considerations shall be taken into account in the development of the implementing rules". This is the reason why it has been decided to focus on Cost Benefit Considerations (CBC). CBC is akin to accepted Multi Criteria Analysis (MCA) methodologies, which are better suited to include qualitative as well as quantitative aspects in the decision making process regarding major market or organisational interventions.

For proper consideration of cost and benefits related to interoperability of spatial services a methodology is needed that instead of trying to convert each cost-benefit aspect in comparable (monetary) units **contains statements of where and how cost and benefits are likely to occur and gives recommendations on how to avoid or decrease costs** with appropriate decisions and technical measures. Likewise, CBC analysis **highlights the possible benefits** and how to make them more visible to stakeholders and decision makers at all levels of government. Consequently CBC relating to INSPIRE data specifications development and implementation **is not a single analysis task, but rather a chain of prudent and intuitive decisions** by experts that are present throughout the specification process.

Interoperability of Spatial Data Sets and Services in INSPIRE

The technical measures and means necessary to reach the interoperability of spatial data sets and services within INSPIRE will be regulated through an implementing rule that sets legally binding .obligations for the Member States when they provide data belonging to the data themes listed in table 1.

Annex I		Annex II	
1.	Coordinate reference systems	10.	Elevation
2.	Geographical grid systems	11.	Land cover
3.	Geographical names	12.	Ortho-imagery
4.	Administrative units	13.	Geology
5.	Addresses		
6.	Cadastral parcels		
7.	Transport networks		
8.	Hydrography		
9.	Protected sites		
Annex III			
14.	Statistical units	24.	Area management/restriction / regulation zones & reporting units
15.	Buildings		
16.	Soil	25.	Natural risk zones
17.	Land use	26.	Atmospheric conditions
18.	Human health and safety	27.	Meteorological geographical features
19.	Utility and governmental services	28.	Oceanographic geographical features
20.	Environmental monitoring facilities	29.	Sea regions
21.	Production and industrial facilities	30.	Bio-geographical regions
22.	Agricultural and aquaculture facilities	31.	Habitats and biotopes
23.	Population distribution – demography	32.	Species distribution
		33.	Energy Resources
		34.	Mineral resources

Table1: Data themes of the INSPIRE Directive

The implementing measures to be gradually developed within the period of 2006-2012 will be based on data specifications describing the targeted objectives for each spatial data theme. Thus data interoperability will be achieved through conformance to the specifications. Since INPSIRE will be based on SDIs established and operated by the Members States, it is expected that data sets need to be transformed in a certain extent that involves costs. In order to assess the return on this investment both the technical feasibility and the cost and benefit issues should be carefully considered.

Cost-benefit information sources for INSPIRE

Information for assessing cost-benefit of INSPIRE data interoperability comes from numerous sources. In early 2008 a user requirement survey was carried out that yielded 93 full descriptions of different application scenarios where spatial data within the scope of INSPIRE is used. This overview of data usage complemented with the information stemming from different application development projects gave a rather good understanding of data requirements.

The INSPIRE stakeholder communities have proposed numerous references comprising national specifications as well as documentations of data harmonisation initiatives. While national specifications help to understand the status quo situation the resources used for data harmonisation projects/initiatives give a first approximation of the possible costs. In order to collect further information the Member States were asked on the following points:

COSTS	BENEFITS	
▪ Costs related to the development of the specifications ▪ Costs in reengineering the databases ▪ Hardware and software costs if new systems were required ▪ As an alternative, costs in developing schema mapping from old to new specifications ▪ Costs in running checking/validating the transformation	**a. Direct User Value/Benefit** ▪ Increased data availability ▪ Increased ease of use ▪ Better data sharing ability ▪ Reduced cost of integrating data **b. Social Value** ▪ Enables better decision making ▪ Reduces barriers between organizations ▪ Increases institutional effectiveness ▪ Promotes more efficient use of (taxpayer) funds	**c. Institutions operational benefits** ▪ Promotes intra-institutional collaboration ▪ Promotes inter-institutional collaboration ▪ Reduces data integration cost across institutions ▪ Promotes re-use of existing datasets ▪ Decreases cost of IT / information management **d. Institutional Financial Value** ▪ Overall cost savings for info management ▪ Achieves cost avoidance (as opposed to savings) **e. Strategic and Political Value** ▪ Fosters closer working relationships ▪ Support improved decision making ▪ Supports other information infrastructure

Table 2: Questions of the survey on cost-benefit aspects of data interoperability

For assessing the feasibility of specifications a specific testing phase has been organised. Although it is just one of the steps of specification development where CBC takes place, the output of testing is the main source of quantifiable information both on costs and benefits.

The reference materials, the outcome of the surveys and testing is being analysed by the group of experts responsible for the development of the data specifications. These experts all have long experience in their thematic field. In this respect, they assume the role of "knowledgeable facilitators" employed in Multi Criteria Analysis initiatives, especially in relation to judgements regarding unquantifiable benefits, as well as representatives of key stakeholder groups for specific data themes.

Generic cost-benefit issues to consider in the data specification process

For every instance of the cost-benefit consideration, e.g. use case development, specification testing, Member State situation comparisons, etc., the following issues should be addressed:

- Direct and indirect generic costs and benefits of harmonising vs. cost of not harmonising
- Type of costs. (one-off costs vs. recurring costs, full vs. marginal cost, set-up costs and/or operational costs, direct vs. indirect costs)
- Types of benefits (immediate gains, future gains, benefits at data provider level, benefits at local, cross-border, national, and pan-European user level, benefits within and outside the scope of INSPIRE). The benefits can be both quantifiable and qualitative.
- Impact issues (social implications, e.g. better management especially in disaster prevention, public health, security of citizens, shortening response time of time-critical decisions, e.g. mitigation of consequences of disasters)
- Economic implications, (i.e. quantifiable costs/benefits from the absence/presence of harmonized information, e.g. damage containment, boosting new business and information products)
- legal implications, (e.g. obligatory/optional use of specifications, impact of and on national standardisation regulations, including prescribed standards for e-Government initiatives)
- Long-term environmental impacts, (e.g. contribution to tackle global problems such as carbon sinks/sources, deforestation, air/water/soil quality or pollution)
- Technical issues (level of detail required for the specifications, possibility for on-the-fly harmonisation via transformation services, possibility for data set harvesting, need for data set duplication, possibility for update propagation including for data and metadata)
- Implementation aspects (time necessary for implementation, interoperability based on the existing situation vs. infrastructure for the future?)
- Assessing risk (i.e. risk that input information is inaccurate or incomplete, that important cost or benefit elements have not been identified, or that assumptions about the future (impact on costs or benefits) are inaccurate).

CBC in INSPIRE data specification steps

The INSPIRE Methodology for the Development of Data Specifications defines a set of repeatable steps how to arrive from user requirements to a data specification through use-case development, initial specification development, analysis of analogies and gaps for further specification refinement. Cost-benefit considerations can be directly linked to each specific step proposed in this methodology; they are inherently present in the decision process that drives the data specification effort.

Use-case development

The user driven specification development according to the methodology comprises the development (or selection of existing) use cases. The number of use-cases, for practical

reasons should be small, never the less they have to be representative (i.e. reflect the multiplicity of possible uses of data within the infrastructure). For proper weighting of requirements they have to be ranked according to priority. Use cases that make part of many user scenarios should be assigned to the high priority group. Satisfying the requirements stemming from them brings definite benefits to the users: using harmonised data reduces ad-hoc work necessary for data integration within the application. Based on qualitative criteria, the time-critical use cases that require cross-border / European intervention (disaster management, flooding, etc.) should be also ranked as high priority.
Use-cases that target solving cross-border or pan-European scenarios are likely to have addressed already some interoperability issues. If these issues have been solved by a wide group of stakeholders and the corresponding interoperability arrangements have been applied, taking these as a starting point for INSPIRE data specifications that targets at a wider thematic scope could reduce the related costs.

Identification of user requirements and spatial object types
Based on the requirements stemming from the use-cases theme specific data requirements are identified at this step. First, the optimal Level of Details (LoD) should be identified. It is highly desirable to use as few LoD as possible. More LoD with separate application schema may increase the burden of responsible organisations in terms of data transformation, requires additional efforts to keep different LoDs consistent, and may leave residual data integration problems for users.
When deriving data requirements (feature types and attributes) from different use cases special attention should be paid to the re-use of concepts. In order to benefit from the endeavours of international standardisation bodies, organisations established under international law, and international initiatives their standards and technical means have been referenced, whenever possible. Likewise, the existing, already harmonised elements within the infrastructure have to be considered. For this purpose, INSPIRE will maintain a feature concept dictionary, a glossary, and a consolidated data model repository containing the application schema of each theme.
"As-is" analysis
The "as-is" analysis yields information on the existing situation, i.e. how the requested data can be supplied by the potential data providers. For example if there is no one-to-one relationship between the proposed harmonised schema and the theme related datasets of the data providers, additional cost of data integration may occur.
Cost also may depend on the stability of the data sets, i.e. frequency with which information changes and the relative size of updates (in terms of data size and number of features). The intensity of use of the data sets, i.e. how many public authorities/persons make use of the data set, how often, in which way (e.g. for spatial analysis, for visualisation only), gives further input on how beneficial data interoperability might be.

Gap analysis

The Gap analysis identifies user requirements that cannot be met by the current data offer. The task is done for each theme considering all the relevant data sets in the Member States. The baseline for comparison is the state of the art studies regularly prepared on the initiative of European Commission and the reference materials made available by the Member States.

As filling the gaps can be one of the most substantial sources of costs, specific attention should be paid to proposing technically sound, but cost effective interoperability approaches. The availability of automatic methods and tools for data transformation should be considered. It should be noted that filling gaps will be one of the main added values for users that definitely creates benefits by eliminating/simplifying ad-hoc measures to be taken each time users use data for the purposes identified in the requirements.

The gap analysis should yield information per theme on the gap situation (i.e. the average situation, worst case, best case, etc.), whether an approach for "bridging the gap" represent a one-time cost or a recurring cost, which is the estimate of the costs to cover the gaps, and the time (savings) and economic benefits of systematic data harmonisation versus *ad hoc* harmonisation efforts.

Data specification development

When formalising application schema, the following aspects need to be considered:

- Modelling vs. object referencing. Before modelling a specific schema element, the option for object referencing should be investigated. This may reduce redundancy and cost, while increasing consistency.
- Levels of Detail (LoD). When a recommended INSPIRE schema is targeted at a lower LoD than is used in some Member States, the generalisation-specialisation hierarchies between the source and target as well as the availability of easy-to-apply generalisation algorithms (like line simplification) should be considered. The same principle should be applied in developing multi-resolution models for the same theme.
- Mandatory vs. optional elements within the application schema. Application schemas should be delivered that fulfil the full set of identified user requirements. However, when based on the results of the "as-is" analysis, testing, and/or the stakeholders' review, it becomes clear that the implementation of certain provisions would create a heavy burden for data providers, it is appropriate to classify that element as optional Using optional elements don't impose immediate actions on the data provider never the less may contribute to future coherent development of SDIs.
- Data quality and conformance. Data quality requirements shall be justified by user requirements. A prudent approach is recommended, i.e. non-conformant data or data of lower quality is better than no data at all. This approach is justified by the fact

that ‘fitness for purpose’ depends upon the data user’s application(s), not on the quality of data available from the provider in the first place. The specifications may include a recommendation of using the highest quality data available from the data owner for the purposes of the infrastructure.

Implementation, test and validation

Testing implementation aspects of the specifications is the main arena for collecting qualitative and quantitative data on costs and benefits. Setting up the test environment, as stated in the RISE project cost-benefit analysis report [RISE 2007], creates implementation benefits both for the data specification process and the participants. It verifies if the data specification contains enough information to support the implementation and if it is actually fit for purpose. It also allows an iterative approach to data product specifications providing feedback from one iteration to another to aid in the development of an upgraded version. The cost of implementing the test environment can be used to help estimate more accurately the cost of developing a full-scale implementation. Test environment development provides a training environment allowing staff to gain experience before setting out on a major system implementation.

In INSPIRE, two types of testing are foreseen: transformation testing and application testing. Transformation testing serves in the first place to asses the effort required by data providers related to transforming their data according to the INSPIRE data specifications for a given theme. As such, the testing phase can yield not only qualitative but also quantitative information on:

- efforts needed to implement the transformation, e.g. as person-hours per dataset or similar measures,
- initial investments necessary to implement the data transformation, e.g. hardware, software, training of specialists, etc.,
- how and to what extent existing tools and “know-how” have been exploited,
- resources needed for maintaining an operational transformation service,
- time efficiency of the on-the-fly transformation services,
- comparison of alternative harmonisation approaches, e.g. on-the-fly transformation versus storing data according to INSPIRE specifications and associated update propagation issues,
- demand for the data being tested, input for estimating the benefits both on the provider and the user side,
- to what extent INSPIRE specifications can be used in the future within organisations for the purpose of ongoing developments, such as establishing object oriented databases, specification development, standardisation, etc.,

- how participation in testing helps stakeholders to identify corresponding or missing data and the processes necessary for implementation of INSPIRE within the Member State.

Application testing focuses on assessing how much the work of users has been made easier due to harmonisation of data. Using data delivered according to INSPIRE data specifications, participants in this kind of test are invited to perform their usual tasks and to keep records of their observations regarding ease of use, increased efficiency, decreased time costs, etc.. This test is expected to demonstrate:

- economies in regard to preparatory work for employing data for the given application, e.g. reduction of human effort spent on data transformation and integration, reduced manual work because of semantic harmonisation, data matching, etc.,
- economies in terms of the cost of tools not now needed for performing the task because of the use of interoperable/harmonised data,
- interaction of cross-theme harmonisation, e.g. making use of coordinate transformations or consistent geographical names,
- feasibility and fitness of proposed use cases, in practice,
- how and to what extent the existing tools and "know-how" have been exploited.

Information obtained in the testing phase also gives general verification of the cost-benefit considerations of the previous steps. The results of testing will be carefully analysed and if required, the proposed specifications will be modified accordingly.

Conclusion

The main effect of introducing harmonised specifications for data within an SDI is that users are likely to spend less time and efforts on understanding and integrating data when they build their applications based on data delivered within INSPIRE. However, arranging interoperability of spatial data sets and services may create costs related to data transformation.

From point of view of acceptance by the stakeholders it is crucial to carefully balance the cost and benefits of the exercise finding the proper boundary of the level of ambitions. Careful analysis of representative use-cases sets acceptable objectives, the "as-is" and the gap analysis provides information about feasibility and the likely technical and economical burdens. The application schema that comprises the results of these steps can be further improved by cost-effective technical and modelling means. This prudent approach need to be complemented by testing that supplies not only a qualitative feedback, but also quantitative information that can be later on used as a starting point for detailed cost benefit analysis. Meanwhile, the role of multi-criteria based cost benefit analysis remains

important, as it gives input of such aspects of SDI that are not convertible in monetary units.
The methodology elaborated for INSPIRE cost-benefit considerations supports the entire cycle of the data specification process and also opens up further possibility for more rigorous analysis when the results of testing will be available.

Acknowledgement
The authors gratefully acknowledge the input of Dr. Max Craglia, who has compiled the questionnaire referred in table 1 as well as the expert of the Member States who have provided their input.

References

Assessing the impacts of Spatial Data Infrastructures (2006). Report of International Workshop on Spatial Data Infrastructures' Cost-Benefit / Return on Investment. European Communities., http://www.ec-gis.org/sdi//ws/costbenefit2006/

Commission Staff Working Document (2004). Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), Extended Impact Assessment. European Communities, http://www.europarl.europa.eu/registre/docs_autres_institutions/commission_europeenne/sec/2004/0980/COM_SEC(2004)0980_EN.pdf

Craglia, Max and Nowak, Joanna (editors) (2006) Report of International Workshop on Spatial Data Infrastructures “Cost-Benefit / Return on Investment”. European Communities, http://www.ec-gis.org/sdi/ws/costbenefit2006/reports/report_sdi_crossbenefit%20.pdf

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007establishing an Infrastructure for Spatial Information in the European Communities (INSPIRE) http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

Guide to cost-benefit analysis of investment projects (2008). European Communities, http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf

Methodology for the development of data specifications (2007). INSPIRE Drafting Team "Data Specifications" http://www.ec-gis.org/inspire/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf

PANEL-GI Compendium (2000), Office for Official Publication of the European Communities, EUR 19360en

RISE (2007) Reference Information Specifications for Europe: Cost-Benefit analysis http://www.eurogeographics.org/eng/documents/RISE30CostBenefitReportV1.1.pdf