



Report

**Joining tabular and geographic data – Merits and possibilities of the Table Joining Service**

Michel Grothe

Thijs Brentjens

**date**

9-9-2013

**Version**

1.1

## Version management

### Revision history document:

Versie	Datum	Status	Auteur	Omschrijving
V0.05	10-01-2013	Draft	Michel Grothe	First draft of contents
V0.1	10-01-2013	Draft	Thijs Brentjens	Added text for Chapter 2
V0.15	23-01-2013	Draft	Michel Grothe	Added several paragraphs chapter 1, 2 and 3
V0.2	11-02-2013	Draft	Thijs Brentjens	Added and edited several sections in chapter 2, cleanup / editorial work in chapter 2
V0.25	14-2-2013	Draft	Michel Grothe	Added and edited several sections in chapter 1 and 2
V0.3	2-4-2013	Draft	Michel Grothe	Edited several sections in chapter 3.
V0.31	4-4-2013	Draft	Thijs Brentjens	Edited several sections, answered questions. Chapter 2 and 3
V0.4	10-4-2013	Draft	Michel Grothe	Added chapter 4.
V0.5	9-6-2013	Draft	Michel Grothe	Edited several sections, Added chapter 5 Concluding remarks.
V0.6	12-6-2013	Draft	Michel Grothe	Edited several sections after meeting with P. Bresters (Statistics Netherlands). Edited TCO-Model and update TCO-model results figures. Added Acknowledgement and, management summary.
V0.7	18-6-2013	Draft	Michel Grothe	Edited after last comments P. Bresters (Statistics Netherlands) and comments Peter Schut (paragraph 2.4).
V0.8	19-6-2013	Draft	Michel Grothe	Edited some editorial comments of P. Bresters (Statistics Netherlands)
V0.9	22-6-2013	Draft	Nancy Christiaans	Language check and edits.
V1.0	23-6-2013	Final	Michel Grothe	Final minor edits in all chapters.
V1.1	9-9-2013	Final	Michel Grothe	Edits in several chapters, after comments of Ekkehard Petri (Eurostat) .

## Contents

<b>ACKNOWLEDGEMENT .....</b>	<b>5</b>
<b>MANAGEMENT SUMMARY.....</b>	<b>7</b>
<b>CHAPTER 1 INTRODUCTION AND CONTEXT .....</b>	<b>9</b>
1.1 INTRODUCTION .....	9
1.2 GOAL OF THIS STUDY .....	9
1.3 OUTLINE OF THE STUDY .....	10
1.4 PARTIES INVOLVED .....	10
1.5 TERMS AND DEFINITIONS USED.....	12
1.6 REFERENCES.....	14
<b>CHAPTER 2 INTRODUCTION TO THE TABLE JOINING SERVICE.....</b>	<b>15</b>
2.1 INTRODUCTION .....	15
2.2 THE INSPIRE PERSPECTIVE .....	15
2.3 THE TABLE JOINING SERVICE .....	20
2.3.1 <i>TJS Data Access and Joining Service.....</i>	<i>20</i>
2.3.2 <i>Overview of the operations .....</i>	<i>21</i>
2.3.3 <i>How a Table Join Service works .....</i>	<i>22</i>
2.4 ALTERNATIVE APPROACHES TO TABLE JOINING.....	25
2.4.1 <i>Spatial database, ETL-based and GIS-based table joining.....</i>	<i>25</i>
2.4.2 <i>WFS based joining.....</i>	<i>25</i>
2.4.3 <i>Linked data.....</i>	<i>26</i>
2.5 ADVANTAGES AND DISADVANTAGES OF TJS.....	27
<b>CHAPTER 3 THE IMPLEMENTATION(S) OF TJS.....</b>	<b>29</b>
3.1 INTRODUCTION .....	29
3.2 CONCEPTUAL ARCHITECTURE(S) FOR TJS .....	29
3.2.1 <i>Conceptual architecture.....</i>	<i>29</i>
3.2.2 <i>TJS and the INSPIRE network services .....</i>	<i>31</i>
3.3 THE ORGANIZATIONAL ASPECTS OF TJS IMPLEMENTATION.....	31
3.3.1 <i>Actors involved in TJS implementation.....</i>	<i>31</i>
3.3.2 <i>TJS and the user perspective.....</i>	<i>32</i>
3.3.3 <i>The need for a spatial (identifier) framework.....</i>	<i>34</i>
3.4 TJS SOFTWARE IMPLEMENTATIONS .....	34
3.5 CLIENT APPLICATIONS OF JTS.....	37
3.6 CONCLUDING REMARKS.....	39
<b>CHAPTER 4 THE COST-BENEFITS OF TJS, EXAMPLE OF THE NETHERLANDS.....</b>	<b>41</b>
4.1 INTRODUCTION .....	41
4.2 EXAMPLE OF TABLE JOINING PRACTICE AT STATISTICS NETHERLANDS .....	41
4.3 COST FACTORS FOR TCO-ANALYSIS .....	44
4.4 TCO-ANALYSIS SCENARIOS FOR THE NETHERLANDS .....	46
4.5 RESULTS OF TCO-ANALYSIS FOR THE NETHERLANDS .....	49
4.6 BENEFITS OF TJS .....	50
<b>CHAPTER 5 CONCLUDING REMARKS.....</b>	<b>51</b>
<b>ANNEX A – USE CASE SCENARIO (UML) INTERACTION DIAGRAM .....</b>	<b>52</b>
<b>ANNEX B – LIST OF EXPERTS CONSULTED .....</b>	<b>53</b>





## Acknowledgement

The authors would like to acknowledge the work of several experts and stakeholders that contributed to this study. We would like to thank the following stakeholders of Statistics Netherlands: Johan van der Schuit (PBL), Maarten van Ham (Technical University of Delft), Jan Jaap Harts (University of Utrecht), Gerard Hazeu (Alterra), Peter Louwerse, (Ministry of Infrastructure and Environment - I&M), Tom Merkx (VSO), Matthijs van Niel (Province of Zuid-Holland), Ioulia Ossokina (CPB), Martin Ruiterman (ABF Research), Lotte Vermeij (SCP), Erik van der Zee (PDOK).

We would particularly like to thank Tom Merkx and Ineke de Haas (municipality of Nijmegen), Gert-Jan van der Weijden and Rob Gieb (municipality of The Hague), and Hans van Amsterdam and Johan van der Schuit from the Netherlands Environmental Assessment Agency (PBL). They all provided valuable information about the daily practice of handling thematic mapping and data joining.

We equally thank Peter Schut, the editor of the OGC Table Joining Standard, for his comments and suggestions on worldwide TJS implementations. We also thank Geocuba for providing information about the TJS Geoserver implementation.

We also thank experts of the Statistics Netherlands for their valuable workshop input: Niek van Leeuwen, Dennis Ramondt, Michiel Eijkhout, Dick Windmeijer, Henrico Witvliet, Erik Fokke, Caroline van Houwelingen and Chantal Melser.

We especially appreciate the work of Pieter Bresters of Statistics Netherlands, who has systematically reviewed and provided the authors with his comments, questions and queries that contributed to improving the work of this study.

Finally we would like to thank Eurostat for providing a grant to the Statistics Netherlands to examine the TJS concept.

Michel Grothe and Thijs Brentjens  
Amersfoort, june 2013



## **Management summary**

There is a strong need for merging and linking socio-economic, environmental, health and other types of data on the social and physical environment to a location. Simply put, this implies linking people, business, physical objects and environmental processes to a particular place and improve and empower the understanding issues and problems of complex social, economic and spatial systems by data analysis, spatial analysis and thematic mapping.

From a technical perspective, the linking of social-economic data, so called tabular (attribute) data and often statistical data, to location or better geographic data, can be done in many ways. Several techniques and approaches exist and different types of tools are available and in use in various user communities. In the spatial or geo-information community, geographical information systems (GIS) are widely used for joining tabular (attribute) and geographic data. GIS are known for their ability to use location as the common factor to integrate data; to link and merge various data sources by their geography and provide all kinds of maps.

There is an increasing demand for linking tabular data to geographic data, driven by different user demands. One of these demands is the European INSPIRE directive, which aims at establishing a spatial data infrastructure in Europe for solving environmental problems. This spatial data infrastructure is characterized by the integration of spatial data from different, multiple sources among several socio-economic and statistical themes, like statistical and administrative units, population distribution, health statistics and energy resources. For the integration of socio-economic data from the themes population distribution, health statistics and energy resources (seen as tabular data) with administrative boundaries and/or statistical units (geographic data), INSPIRE recommends that the Table Joining Services concept be used.

The general aim of the study is to explore the merits and possibilities of the Table Joining Service, a geospatial standard of the Open Geospatial Consortium (OGC). TJS offers a standardized web service interface for automatic joining of tabular data (e.g. statistical data) to geographic data (administrative boundaries, postal codes, statistical units) by distributed access. The study examined the concept and functional possibilities of TJS, the existing TJS implementations and client applications. The implications, roles and tasks, and cost-benefits of the adoption and implementation of TJS are also considered. Therefore, the main goal of this study is to gain insight into the merits and possibilities of the TJS for merging or joining tabular (attribute) data to geographic data.

This report is prepared by Geonovum in partnership with Statistics Netherlands through the support of a grant awarded by the Eurostat (Grant no. 50502.2012.001-2012-518).

The TJS concept is explained through its functional merits, the operations and some of the specific characteristics. A short overview of common table joining methods (SQL and databases, GIS-based, WFS-based) and more modern approaches (linked data) to joining data are discussed as well. TJS can be seen as an alternative method in the context of the adoption and use of data from spatial data infrastructures. TJS has several potential advantages, such as its use in service oriented architectures, being an open standard that increases interoperability, it is simple and powerful, uses the power of distributed computing and it is easy to find data through registries. At the same time, some potential disadvantages are illustrated as well, like another service specification and another specific format for encoding data, the lack of a mechanism for uploading and creating tabular (or geographic) data.

TJS can be considered a supporting concept in spatial data infrastructures for joining data from various, distributed sources or infrastructure nodes in order to view and download datasets. There are several ways to implement TJS in organizations, in case of INSPIRE in European Member States. This study also explores the organizational aspects of TJS implementation and examines the possibilities for

implementation of TJS software. The adoption of TJS will have an impact on the organizations and its services infrastructure of the different data providers involved. Three types of data providers are distinguished.

As mentioned in the previous chapter, table joining is based on a spatial framework and the common geographic identifier. Therefore, the concept of the table joining in a distributed service oriented architecture requires that organizations implement their services according to a common spatial (identifier) framework. Therefore, organizations that want to adopt and use TJS with long-term perspective need to agree and adopt standards for spatial data (concepts and models) and geographic identifiers or, better yet, a common spatial (identifier) framework. A common spatial and identifier framework consist of several aspects:

- Data models for spatial framework data based on a generic conceptual model
- Generic approach for handling and encoding unique, persistent geographic identifiers;
- Registry for maintenance of data models and identifiers.

So, for data providers offering data in TJS infrastructures it is vital to agree upon a common spatial (identifier) framework.

There are only a few TJS implementations up-and-running at the moment and publicly available. As TJS is a back-end technology, it is not always obvious that TJS is used. However at this moment, it is clear that this OGC standard is not implemented broadly.

Nowadays, IT investments in the public sector, like the adoption and implementation of a TJS infrastructure, require business case development. So, one of the questions in this study is, what are the costs-benefits of the adoption and implementation of TJS in the public sector in the Netherlands? In this study, the cost analysis approach is based on a simplified Total Cost of Ownership method (TCO) with 5 variables and 10 constants in the TCO-model. For the purpose of cost estimation, three scenarios have been developed and cost estimations have been done for a period of 10 years. The variable and constants values are estimated by experts.

The costs-benefits of the adoption and implementation of TJS bringing tabular and geographic data together are roughly estimated for the public sector in the Netherlands. The cost-benefit analysis based on a Total Cost of Ownership method (TCO), shows that –although roughly estimated – the costs of TJS adoption will have a positive business case in the long-term (after 4 years for the minimum and after 8 years for the maximum TJS scenario) from the costs perspective. Besides the costs perspective several potential, indirect benefits have been identified during expert workshops.

Major finding is that TJS has the potential to replace the 'manual' data joining operations in the daily practice of data management for thematic mapping and spatial statistics. However, in order to stimulate the large scale adoption and implementation, there is a need to explore the possibilities and cost benefits of TJS concept in more detail and to put TJS in practice by implementation of TJS infrastructure nodes. It is therefore recommended to:

1. Test the TJS concept in a real-world case study; to implement the minimum scenario for INSPIRE Health statistics theme in the Netherlands. This requires cooperation between several INSPIRE data providers in the Netherlands. Also, a minimum scenario case study offers the possibility to validate the TCO model (see point 3);
2. Establish a national spatial (identifier) framework in close cooperation and based on common agreement(data concepts and models) with public sector organizations. The spatial (identifier) framework is necessary to establish a commonly agreed situation for organizations to join their tabular data to geography, knowing that it can rely on a stable spatial (identifier) framework;
3. Further refine the presented TCO-model with real world model values based on the case study (point 1) and recalculate the TCO-model in order to have a more reliable business case for the Netherlands and as a general format for individual organizations to develop their business cases.



## Chapter 1 Introduction and context

### 1.1 Introduction

There is a strong need for merging and linking socio-economic, environmental, health and other types of data on the social and physical environment to location. Simply put, this implies linking people, business, physical objects and environmental processes to a particular place and improve and empower the understanding issues and problems of complex social, economic and spatial systems by data analysis, spatial analysis and thematic mapping [1].

From a technical perspective, the linking of social-economic data, so called tabular (attribute) data and often statistical data, to location or better geographic data, can be done in many ways. Several techniques and approaches exist and different types of tools are available and in use in various user communities. In the spatial or geo-information community, geographical information systems (GIS) are widely used for joining tabular (attribute) and geographic data. GIS are known for their ability to use location as the common factor to integrate data; to link and merge various data sources by their geography and provide all kinds of maps.

There is an increasing demand of linking tabular data to geographic data, driven by different user demands. One of these demands is the European INSPIRE directive, which aims at establishing a spatial data infrastructure in Europe for solving environmental problems. This spatial data infrastructure is characterized by the integration of spatial data from different, multiple sources among several socio-economic and statistical themes, like statistical and administrative units, population distribution, health statistics and energy resources. For the integration of socio-economic data from the themes population distribution, health statistics and energy resources (seen as tabular data) with administrative boundaries and/or statistical units (geographic data), INSPIRE recommends that the Table Joining Services [2] concept be used [3].

This study regarding the merits and possibilities of the Table Joining Services concept is introduced in this chapter. Paragraph 1.2 explains the goal of this study and summarizes the questions to be answered. The outline of this study in paragraph 1.3 introduces all the chapters. Next, in paragraph 1.4, the parties involved in this study are introduced and in paragraph 1.5, the terms and definitions used are summarized. Finally, in paragraph 1.6, the references used for this study are listed.

### 1.2 Goal of this study

The general aim of the study project is to explore the merits and possibilities of the Table Joining Service, a geospatial standard of the Open Geospatial Consortium (OGC). TJS offers a standardized web service interface for automatic joining of tabular (attribute) data (e.g. statistical data) to geographic data (administrative and statistical units) by distributed access.

The study will examine the concept and functional possibilities of TJS, the existing implementations and applications, and will look at implications, roles and tasks, and costs-benefits of the adoption and implementation of TJS. The main goal of this study is to gain insight into the merits and possibilities of the TJS for merging or joining tabular (attribute) data to geographic data.

In order to achieve the goal of this study, several questions were formulated:

1. What is a Table Joining Service? When and for what purposes would their use be beneficiary?
2. How and for which INSPIRE themes can TJS help to reach the INSPIRE objectives
3. How and for which general statistical themes can TJS help statistical offices

4. How could TJS work for Statistics Netherlands and the Dutch organization for hosting geoservices from the government (PDOK)
5. What impact would the TJS have on the organization infrastructure for data providers and data users? How would they interact? Do we need new applications or can they be implemented in existing geoviewers?
6. Are private GIS-software companies anticipating on the TJS? Should we develop things or can we rely on the private sector?
7. Which table formats can be used?
8. Can TJS be demonstrated as prototypes?
9. Which operational TJS-applications are available?
10. What is the business case for a TJS?

These questions are explored in this study in the next chapters; questions 1-3 are explored in chapter 2, questions 4-9 in chapter 3 and question 10 is explored throughout all chapters and especially chapter 4.

### 1.3 Outline of the study

First, chapter 2 examines the merits of the TJS concept, outlines the TJS concept and explains the operations of TJS as a web service standard. More common approaches to table joining are also considered. Finally, in the last paragraph of chapter 2 possible advantages and disadvantages of TJS are discussed.

After the introduction to the TJS concept in chapter 2, chapter 3 explores the ways to implement TJS in organizations and examines the possibilities for implementation of different TJS software. For that purpose, the conceptual architecture of TJS is outlined first. Next, the organizational aspects of TJS implementation are discussed from the perspective of three types of actor organizations involved. The worldwide available TJS software implementations are explored as well. Finally, aspects of TJS client applications are discussed.

In chapter 4 the costs-benefits of the adoption of TJS in the public sector of the Netherlands is examined. The existing 'manual' approach to joining tabular and geographic data is examined by an example of the Statistics Netherlands. Next, the cost factors are introduced for analysis of the Total Cost of Ownership for TJS adoption in the public sector in the Netherlands. This TCO-model is run using three scenarios; a zero scenario (no TJS), a minimum TJS adoption and a maximum TJS adoption scenario. The results of the estimation of these three TCO-analysis scenarios for the Netherlands are presented. Finally, some expected indirect benefits (by expert opinions) are summarized.

Finally, chapter 5 offers some concluding remarks, including recommendations for further study and implementation of TJS.

### 1.4 Parties involved

This report is prepared by Geonovum in partnership with Statistics Netherlands through the support of a grant awarded by the Eurostat (Grant no. 50502.2012.001-2012-518). Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the Statistics Netherlands.

#### **Statistics Netherlands - Centraal Bureau voor de Statistiek (CBS)**

Centraal Bureau voor de Statistiek (CBS) or Statistics Netherlands is responsible for collecting and processing data in order to publish statistics to be used in practice, by policymakers and for scientific research. In addition to its responsibility for (official) national statistics, Statistics Netherlands also has the task of producing European (community) statistics. The legal basis for Statistics Netherlands and its work



is the Act of 20 November 2003, last amended by the Act of 15 December 2004, governing the central bureau of statistics (Statistics Netherlands).

The information published by Statistics Netherlands incorporates a multitude of societal aspects, from macro-economic indicators such as economic growth and consumer prices, to the incomes of individual people and households. Statistics Netherlands' statistical programmes (the long-term statistical programme and the annual work programme) are set by the Central Commission for Statistics. This is an independent commission that watches over the independence, impartiality, relevance, quality and continuity of the statistical programme. The Director-General decides autonomously which methods to use to make these statistics, and whether or not to publish results.

Statistics Netherlands aims to reduce the administrative burden for companies and the public as much as possible. To this end, it uses existing administrative registrations of both government and government-funded organizations. The information from these files is supplied to Statistics Netherlands free of charge. Only if they do not contain sufficient information, Statistics Netherlands is allowed to conduct supplementary surveys among companies and private persons. Companies are usually obliged by law to supply information to Statistics Netherlands and can be forced to cooperate under certain circumstances; Statistics Netherlands may use sanctions such as administrative fines. For its part, Statistics Netherlands is obliged to keep all individual data confidential. As an exception, data sharing with Eurostat, NSIs in EU member states, Dutch Central Bank and academic researches is allowed under certain restructures.

Source: <http://www.cbs.nl/en-GB/menu/organisatie/default.htm>

### **Geonovum**

Geonovum is the National Spatial Data Infrastructure (NSDI) executive committee in the Netherlands. The organization devotes itself to providing better access to geo-information in the public sector. Geonovum develops and manages the geo-standards necessary to implement this goal. The foundation of Geonovum marks an important step to improve coordination of public sector geographic data in the Netherlands. In its role as enabler of the national geo-information infrastructure, Geonovum connects public sector managers with the work floor professionals. Geonovum's strength is that it discusses, coordinates and monitors arrangements between the parties concerned with regard to how activities are to be carried out and completed. Geonovum is recognizable and professional, result-driven and independent. Its main goals are:

1. To develop and standardize the geo-information infrastructure while also being innovative;
2. To build up and disseminate knowledge in the area of geo-information infrastructure;
3. To make the geo-information infrastructure more accessible to administrative bodies, institutions and departments in the Netherlands and the European Union.

Source: [www.geonovum.nl](http://www.geonovum.nl)

### **Eurostat**

Eurostat is the statistical office of the European Union situated in Luxembourg. Its task is to provide the European Union with statistics at the European level that enable comparisons between countries and regions. This is a key task. Democratic societies do not function properly without a solid basis of reliable and objective statistics. On the one hand, decision-makers at the EU level, in Member States, in local government and in business need statistics to make those decisions. On the other hand, the public and media need statistics for an accurate picture of contemporary society and to evaluate the performance of politicians and others. Of course, national statistics are still important for national purposes in Member States, whereas EU statistics are essential for decisions and evaluation at the European level.

Statistics can answer many questions. Is society heading in the direction promised by politicians? Is unemployment up or down? Are there more CO2 emissions compared to ten years ago? How many women work? How is your country's economy performing compared to other EU Member States? International statistics are a way of getting to know your neighbors in Member States and countries outside the EU. They are an important, objective and down-to-earth way of measuring how we all live.

Eurostat was established in 1953 to meet the requirements of the Coal and Steel Community. Over the years, its task has broadened and when the European Community was founded in 1958, it became a



Directorate-General (DG) of the European Commission. Eurostat's key role is to supply statistics to other DGs and to supply the Commission and other European Institutions with data so they can define, implement and analyze Community policies.

The result: Eurostat offers a whole range of important and interesting data that governments, businesses, the education sector, journalists and the public can use for their work and daily life.

With the development of Community policies, Eurostat's role has changed. Today, collecting data for EMU and developing statistical systems in candidate countries for EU membership are more important than ten years ago. Source: [http://epp.eurostat.ec.europa.eu/portal/page/portal/about\\_eurostat/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/about_eurostat/introduction).

## 1.5 Terms and definitions used

For the purposes of this study, the following terms and definitions are used (see figure 1.4).

Terms	Definitions	Source of definition
application or client	software component that can invoke an operation from a server	[1]
attribute	a set of values that describe some aspect of a spatial framework. Population, temperature, and income are all examples of attributes that could apply to a spatial framework.	[1]
capabilities	service-level metadata describing the operations and supporting content available from a server	[1]
client or application	software component that can invoke an operation from a server	[1]
geographic data or spatial framework data	a GIS representation, either point, line, or polygon, of any collection of physical or conceptual geographic objects. Municipalities, postal code areas, telephone area codes, ecoregions, watersheds, road segments, fire stations, and lighthouses are all examples of geographic data or spatial frameworks	
geographic identifier	unique reference for a spatial object found in a spatial framework dataset	[1]
tabular data	a table of data consisting of a collection of attributes each with a geographic identifier that enables those attributes to be joined to a spatial framework.	[1]
operation	a transformation or query that a server may be requested to execute	[1]
request	invocation of an operation by a client	[1]
response	result of an operation returned from a server to a client	[1]
server	physical instantiation of a service	[1]
service	a suite of functionality that is provided by an entity through a standardized set of operations defined by interfaces	[1]
spatial analysis	Spatial analysis or spatial statistics includes any of the formal techniques which study entities using their topological, geometric, or geographic properties. The phrase properly refers to a variety of techniques, many still in their early development, using different analytic approaches and applied in fields as diverse as astronomy, with its studies of the placement of galaxies in the cosmos, to chip fabrication engineering, with its use of 'place and route' algorithms to build complex wiring structures.	[8]
spatial data infrastructure	A spatial data infrastructure (SDI) is a data infrastructure implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way. Another definition is the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data	[8]
spatial framework data	a GIS representation, either point, line, or polygon, of any collection of physical or conceptual geographic objects. Municipalities, postal code areas, telephone area codes, ecoregions, watersheds, road segments, fire stations, and lighthouses are all examples of spatial frameworks	[1]
thematic map	a type of map or chart especially designed to show a particular theme connected with a specific geographic area. These maps "can portray physical, social, political, cultural, economic, sociological, agricultural, or any other aspects of a city, state, region, nation, or continent.	[8]

Figure 1.4 Terms and definitions

## 1.6 References

The following references are used in this study:

- [1] United Nations - Economic and Social Council, 2012, Report by the Australian Bureau of Statistics on Developing a Statistical-Geospatial Framework, E/CN.3/2013/2.
- [2] Peter Schut (Ed.), 2010, OpenGIS® Georeferenced Table Joining Service (TJS) Implementation Standard v1.0.0, OGC 10-070r2.
- [3] D2.8.III.1 INSPIRE Data Specification on Statistical units – Draft Technical Guidelines. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_SU\\_v3.0rc3.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SU_v3.0rc3.pdf)
- [4] D2.8.III.10 INSPIRE Data Specification on Population distribution – demography – Draft Technical Guidelines. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_PD\\_v3.0rc3.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PD_v3.0rc3.pdf)
- [5] D2.8.I.4 INSPIRE Data Specification on Administrative units – Guidelines. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_AU\\_v3.0.1.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AU_v3.0.1.pdf)
- [6] D2.8.III.5 INSPIRE Data Specification on Human health and safety – Draft Technical Guidelines. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_HH\\_v3.0rc3.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HH_v3.0rc3.pdf)
- [7] D2.8.III.20 INSPIRE Data Specification on Energy Resources – Technical Guidelines. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_ER\\_v3.0rc3.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_ER_v3.0rc3.pdf)
- [8] INSPIRE Generic Conceptual Model D2.5: Generic Conceptual Model, Version 3.4rc3. Download: [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/D2.5\\_v3.4rc3.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf)
- [9] Wikipedia.
- [10] Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2009). "Linked Data—The Story So Far" (<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>). International Journal on Semantic Web and Information Systems 5 (3): 1–22. doi:10.4018/jswis.2009081901. ISSN 15526283. Retrieved 2010-12-18.
- [11] Technical Guidance for the implementation of INSPIRE View Services v3.11. Download: [http://inspire.jrc.ec.europa.eu/documents/Network\\_Services/TechnicalGuidance\\_ViewServices\\_v3.11.pdf](http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.11.pdf)
- [12] Technical Guidance for the implementation of INSPIRE Download Services v3.0. Download: [http://inspire.jrc.ec.europa.eu/documents/Network\\_Services/Technical\\_Guidance\\_Download\\_Services\\_3.0.pdf](http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_3.0.pdf)

## Chapter 2 Introduction to the Table Joining Service

### 2.1 Introduction

Generally, working with data often requires working with data tables and therefore joining tables in order to integrate, analyze and display data from databases. A join clause combines records from two or more tables in a database. It creates a set that can be saved as a table or used as it is. A join is a means for combining fields from two tables by using values common to each.

Also when data is obtained from other organizations, different departments in your organization, purchased commercially available data, or downloaded as open data from the Internet, joining data is often evident for further data processing. Joins can be performed in many ways, often by means of a computer application, e.g. a spreadsheet, database, GIS, statistical application, etcetera. Database design principles often recommend organizing databases into multiple tables, each focused on a specific topic, instead of having one large table containing all the necessary data. Having multiple tables prevents the user from duplicating data in the database; storing the data only once in one table. When the data that is needed isn't in the current table, joining tables is then a common operation.

In order to join a tabular dataset to a geographic dataset, both datasets must contain the same identifier (i.e. key field). The identifier, a so called "geographic identifier", refers to a geographic or spatial feature in a geographic dataset. An example of tabular data is a collection of population counts by city. The table includes the city name, but does not include any other geographic identifier. The city names can be used to join the population counts data to the geographic dataset for cities, that contain the boundary (represented as xy-coordinates or geometries) for each city and the names as well.

In spatial applications, table joining is often necessary to join tabular (attribute), often statistical data to a location in order to map the data or perform some kind of spatial analysis or produce thematic maps.

Thematic maps serve three primary purposes:

1. to provide specific information about particular locations;
2. to provide general information about spatial patterns;
3. to compare patterns on two or more maps.

So, thematic maps consist of geographical features and their thematic, often statistical, characteristics or attributes.

Once the data is stored in a table, it can be associated with the geographic features for display the data on the thematic map. The Table Joining Service (TJS) is a mechanism that supports this.

This chapter examines the merits of TJS. The incentive for this study, the European INSPIRE directive, is introduced in paragraph 2.2. Next, in paragraph 2.3, the TJS concept is introduced and the TJS operations outlined. Some existing approaches to table joining are considered briefly in paragraph 2.4. Finally, in paragraph 2.5, the possible advantages and disadvantages of TJS are discussed.

### 2.2 The INSPIRE perspective

The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organizations and better facilitate public access to spatial information across Europe. The INSPIRE directive came into force on the 15<sup>th</sup> of May 2007 and will be implemented in various stages, with full implementation required by 2020. A European Spatial Data Infrastructure will assist in policy-making across boundaries. INSPIRE is based on the infrastructures for spatial information as established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. This makes INSPIRE a unique



example of a legislative “regional” approach. The spatial information considered under the directive is therefore extensive and includes a great variety of topical and technical themes.

At least, for three INSPIRE themes the specification of geographic data is strongly related to statistical data. These INSPIRE themes require statistical datasets that are linked to a location or spatial reference. For this purpose, INSPIRE has defined a theme Statistical Units (SU).

In [3] statistical data is defined as “any numerical representation of a phenomenon”. A statistical unit informs on the statistical data location. In [3] only the specification of the Statistical Unit is addressed; the specification of statistical data is out of the scope of [3]. The principle is to provide stable and identified representations of the statistical units – and statistical data refers to these objects through their identifier. Some recommendations are provided on how to give stable identifiers to statistical units and how to use these identifiers to attach statistical information to them. This is particularly important if the responsibility for tabular and geographic data is shared between different institutions, like national statistical offices and national mapping agencies. Other INSPIRE data specifications, such as Population distribution - demography [4], use this reference mechanism.” Furthermore, Statistical units have links to Administrative units [5], Population distribution - demography [4], Human health and safety [6] and Energy Resources [7]. In Figure 2.1, the three statistical data themes are summarized and related to the INSPIRE theme of Statistical Units.

<b>INSPIRE themes for geographic data</b>	<b>Theme definition</b>	<b>Relation with theme Statistical units (SU)</b>
Statistical units (SU)	Units for dissemination or use of statistical information.	
Administrative units (AU)	Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries.	The application schema “Administrative units” shall provide an explicit connection to statistical units defined by the EU-administration to link statistical information to administrative units.
<b>INSPIRE statistics themes for tabular data</b>	<b>Theme definition</b>	<b>Relation with theme Statistical units (SU)</b>
Population distribution/demography (PD)	Geographical distribution of people, including population characteristics and activity levels, aggregated by grid, region, administrative unit or other analytical unit.	The theme Population distribution/demography contains attributes related to statistical units. This means that this theme has no direct spatial features such as many other INSPIRE Annex III themes like buildings, health care institutions etc., it needs to be linked to these features by the use of statistical units for example NUTS-codes or grid identifier.
Human health and safety (HH)	Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment.	This theme provides a generic data model applicable across statistical units (as presented in SU theme) available in the Member States. The human health theme contains mainly data attached to statistical units. Health data and biomarkers have no direct spatial features, and need to be linked to these features by the use of statistical units, for example NUTS-codes or grid coordinates.
Energy resources (ER)	Energy resources including	Detailed, complete, timely and



	hydrocarbons, hydropower, bio-energy, solar, wind, etc., where relevant including depth/height information on the extent of the resource.	reliable statistics are essential to monitor the energy resources and security at a country level as well as at an international level. This application schema supports the provision of aggregated data on Energy Resources and Energy statistics (i.e. all types of energy products and flows). The geometry representation for Energy Statistics support different geometrical representations and is described by the geometry descriptor in the Statistical Units application schema.
--	---	---

Figure 2.1 INSPIRE themes for statistical datasets

### Linking statistical data to Statistical Units (SU)

So, INSPIRE is asking for linking statistical data from three INSPIRE themes to spatial objects known as INSPIRE Statistical Units (US). For this purpose, statistical data should be linked to Statistical Units (see figure 2.2). The link from statistical data to Statistical Units should use the statistical unit identifier. In INSPIRE, this is called referencing by external object identifiers or 'object referencing'.

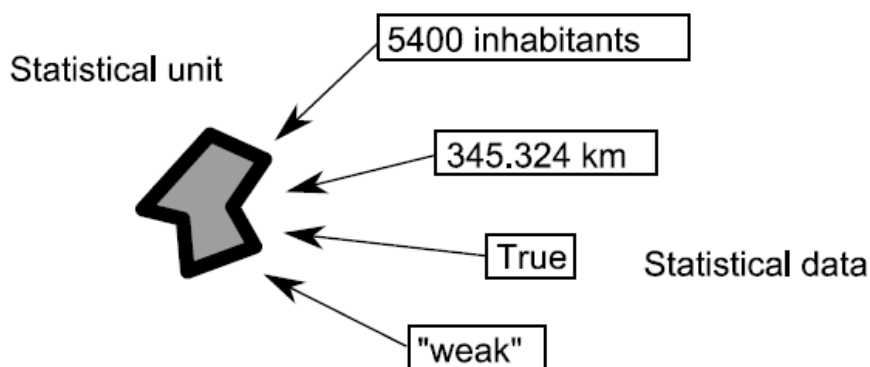


Figure 2.2 Statistical data linked to an INSPIRE Statistical Unit [4]

Several requirements and recommendations regarding object referencing and identifier management that are of importance to linking data are formulated in INSPIRE.

### Implementation considerations

Considerations for implementation of the linking mechanism between Statistical Units (geographic data) and statistical data are given in [3] as well. In Recommendation 8 [3] recommends: "If encoding formats and infrastructures exist for statistical data, INSPIRE should only focus on the spatial units, and actions to develop interfaces between these infrastructures and INSPIRE should be undertaken based on the previous recommendation. For example, existing formats for statistical data should be extended to allow the linkage to statistical units published under INSPIRE. The intention behind this recommendation is to build bridges between statistical and spatial data infrastructures."

As an example to illustrate the lack of a bridge between statistical and spatial data infrastructures, the example of the publication of statistical information is used [3]: "European statistical data are already published by National Statistical Offices through the Census Hub maintained by Eurostat. The SDMX format allows the publication of statistical information, but it does not cover spatial representations of the

statistical units. The link to the statistical unit is implicit through code lists. The explicit INSPIRE statistical units could be provided by extending these code lists to INSPIRE statistical unit identifiers.

In case no format and infrastructure exist for specific statistical units, it is recommended to use TJS (Table Joining Services) and GDAS (Geographic Data Attribute Set) defined in "Georeferenced Table Joining Service Implementation Standard, OpenGIS standard 10-070r2".

So INSPIRE recommends in [3, Recommendation 9]: "TJS service and GDAS format defined in: "Georeferenced Table Joining Service Implementation Standard, OpenGIS standard 10-070r2" should be used to publish statistical data to be linked to INSPIRE statistical units" and "If statistical data are reported on statistical units (such as Population Distribution data), then one should consider publishing through the same service" [3, Recommendation 10].

### INSPIRE approach to data linking/joining

Relations between INSPIRE data are defined in the data specifications. Relations between object types are explicitly modeled in three themes. INSPIRE allows for linking of objects by referring to them using INSPIRE identifiers [8]. How to construct these identifiers to enable the linking between datasets, even provided by different organizations or systems, is defined in INSPIRE. Namespaces play an important role in these identifiers.

Although not required, INSPIRE recommends the usage of URIs for referring to other objects. This means that joining data could be done by resolving the URIs.

Thus, links to other objects are defined on the data level. It is up to a client application or web service to actually joining the data, by following the INSPIRE identifiers. A Table Joining Service could be used for this.

### Organization of INSPIRE (statistical) data providing in the Netherlands

Several different organizations in the Netherlands do have responsibilities regarding the provision of statistical data for INSPIRE. Figure 2.3 provides an overview of Dutch organizations that are involved in the INSPIRE themes that are related to statistical data.

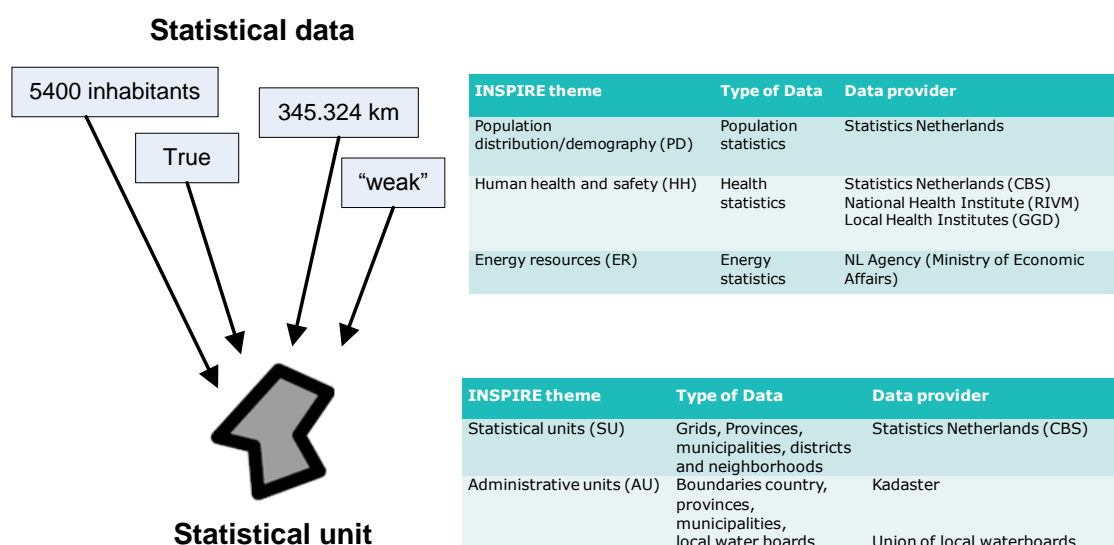


Figure 2.3 INSPIRE themes related to statistical data and the Dutch data providers

All the above datasets have a relation with Statistical Units. However, the data do not contain the data or objects from Statistical Units themselves, but refer to them. Objects in the dataset(s) for Human health

and safety, for example, have references to Statistical Units. In the Netherlands the INSPIRE data is provided by different organizations. The geographic data for the INSPIRE theme Administrative Units in the Netherlands are made available by centralized web services offered by Kadaster. The Statistics Netherlands is data provider for the INSPIRE theme Statistical Units, population and health.

So, the incentive for Statistics Netherlands to conduct a study about the Table Joining Service concept can be found in the European INSPIRE Directive. Statistics Netherlands is responsible for providing datasets for the INSPIRE themes Population Distribution and Human Health, but also Statistical Units. It would be easy if the statistical datasets, which fit the description of these themes, could just be presented as tables, without including the corresponding geographical data, as far as maintaining this geographical data is not the responsibility of the Statistics Netherlands. The geographical data could then be supplied by organizations that are responsible for this, like The Netherlands' Cadastre, Land Registry and Mapping Agency - in short Kadaster - through their spatial data infrastructure called "Publieke Dienstverlening op de Kaart" (PDOK). It is not as simple as that, because Statistics Netherlands is also responsible for the INSPIRE theme Statistical Units, which contains the geographical data for many statistics, but still, it would make responsibilities more clear and it would reduce duplication of geographical data services. Next to INSPIRE datasets, there are many more datasets that could benefit from this concept. Like the Statline database from Statistics Netherlands (<http://statline.cbs.nl/>). It contains many tables with a geographical (identifier) link, but it does not contain all the geographical data (see figure 2.4).

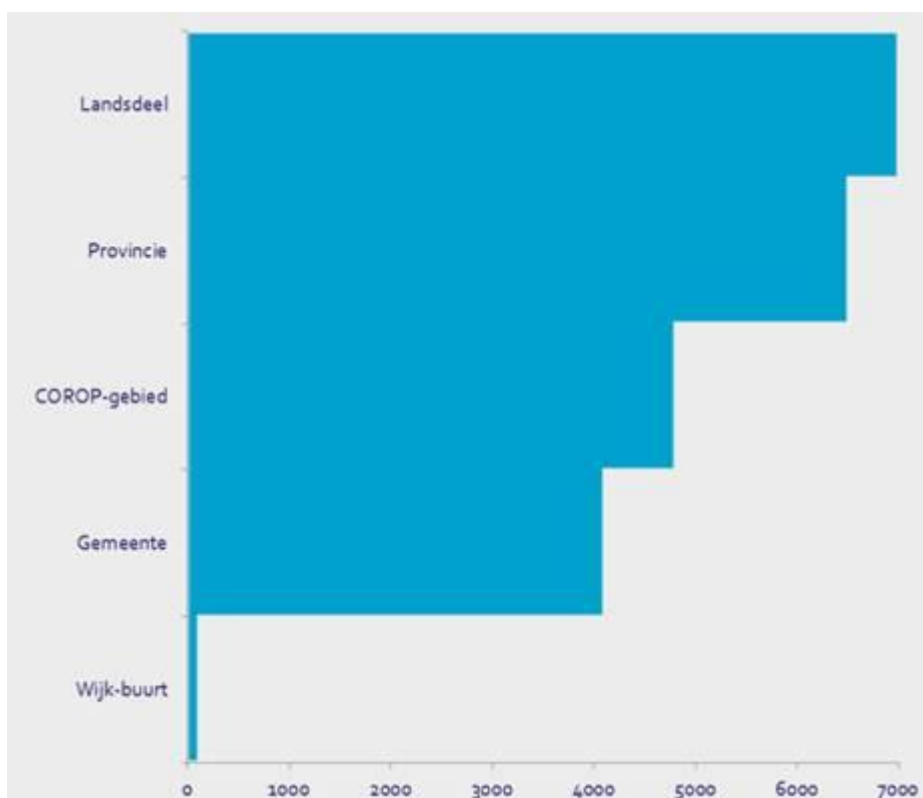


Figure 2.4 Dutch statistical units and their (number of) statistical variables

The concept might also make it easier within organizations. Within organizations like Statistics Netherlands, the responsibility for tables and geometry is also spread over different teams. The TJS concept could enable the teams responsible for tables to create and publish maps without the help of the team responsible for the geographical data.



### 2.3.2 Overview of the operations

The OGC TJS specification [2] describes some mandatory and some optional operations. The specification summarizes them as follows: "Like all OGC services, the Table Joining Service interface specifies a service discovery operation: *GetCapabilities* (required implementation by all servers) – This operation allows a client to request and receive back service metadata (or Capabilities) documents that describe the abilities of the specific server implementation. This operation also supports negotiation of the standard version being used for client-server interactions."

Describing the available datasets and the capabilities of the service in metadata and by the service itself is important for users to be able to find and evaluate the TJS and the data offered by the TJS for their use. The operations of TJS listed above and metadata on the datasets and service also facilitate the (automatic) joining of data.

In addition, TJS specifies two distinct sets of operations that can be requested by a client and performed by a server: data access operations and data joining operations. For geographic data the term 'spatial framework data' is used in the TJS specification.

#### Data access operations

Operations related to encoding and delivering tabular data in a format that can be incorporated by other services include:

1. *DescribeFrameworks* – This operation allows a client to obtain a list of the spatial framework data for which tabular data is available from the server;
2. *DescribeDatasets* – This operation allows a client to obtain general descriptions of the attribute data tables that are available from the server;
3. *DescribeData* – This operation allows a client to obtain a list describing the specific data contents of the attribute data tables (i.e. the attributes) that are available from the server;
4. *GetData* – This operation allows a client to obtain a specific set of tabular data.

If a server supports data access, then all of the data access operations are mandatory.

#### Data joining operations

Operations related to incorporating tabular data and joining it to its spatial framework (i.e data joining) include:

1. *DescribeJoinAbilities* – This operation allows a client to obtain the list of spatial frameworks to which the server can join tabular data, and the forms of output products supported that are supported by the server;
2. *DescribeKey* – This operation allows a client to obtain the list of geographic identifiers for a spatial framework supported by the server;
3. *JoinData* – This operation allows a client to request the joining of a specified tabular dataset to its spatial framework and receive references to the products of that join.

If a server supports data joining, then all of the data joining operations are mandatory.

*DescribeJoinAbilities* is to be used to support such service discovery and populate services registries [2]. This operation is important because it allows clients to explore the available frameworks in detail and explore the output mechanisms the service offers.

*DescribeFrameworks*, *DescribeDatasets* and *DescribeData* are important for clients to understand the data and for the preparation of the *JoinData* operation.

So basically from the implementation point-of-view, there are three types of Table Joining Services:

1. Data access services, that offer attribute data and/or geographic data;
2. Data Joining services, that offer functionality to join attribute data with spatial framework data, offered by that service;
3. A TJS can also function as both a Data Access service and a Data Joining service.

A TJS describes its data and joining abilities (if offered), in order to allow clients to determine if the data is suitable to make a join. For example, if Data access service A offers data about municipalities, the other Data access service B should have geographic data on municipalities that use the same geographic identifiers (keys) to make a join. A client application can use one or more Table Joining Services to join data and get the joined data back in a map or as a joined dataset. If supported by the TJS, a client application could send styling information to style the result map.

### 2.3.3 How a Table Join Service works

A typical workflow for a client application using a Table Joining Service is (figure 2.6):

1. The client searches for a service that offers tabular (attribute) data and a service that offers geographic data to join;
2. The client finds out how to join data, using information provided by the services. Which elements are mutual in the administrative data and geographic framework data? Which are the keys to join data on? Both services should describe these attributes, so the client can explore them and find an attribute to join on. For example, to make a join on the (official) name of provinces, a code for municipalities etc. The client defines how to join the data and requests the service with the geographic data to join the geography with the attribute data. The client tells the service where to get the attribute data from. The join is performed by the service that offers the framework data. If supported by the TJS, the join request could include information on how to style or display the result data;
3. After successful processing by the service, the client receives a reference to the joined data. This response returns the connection information to acquire the result of the join. So the client does not directly get the result back, but gets a location (URL) where the result can be found. For example, the response can be a URL to a map (to be retrieved by a web service like WMS for example) or a URL to get raw, vector data back (to retrieve using WFS for example).

A note concerning this last step. Geospatial web services can be used to offer maps and geospatial data to (end-) users. Commonly used standards are WMS to offer maps and WFS to offer the data. TJS can be used in conjunction with WMS and WFS services. The result of a TJS Join could be a WMS layer, that is dynamically added to a WMS by a TJS. Or a feature type offered by a WFS.

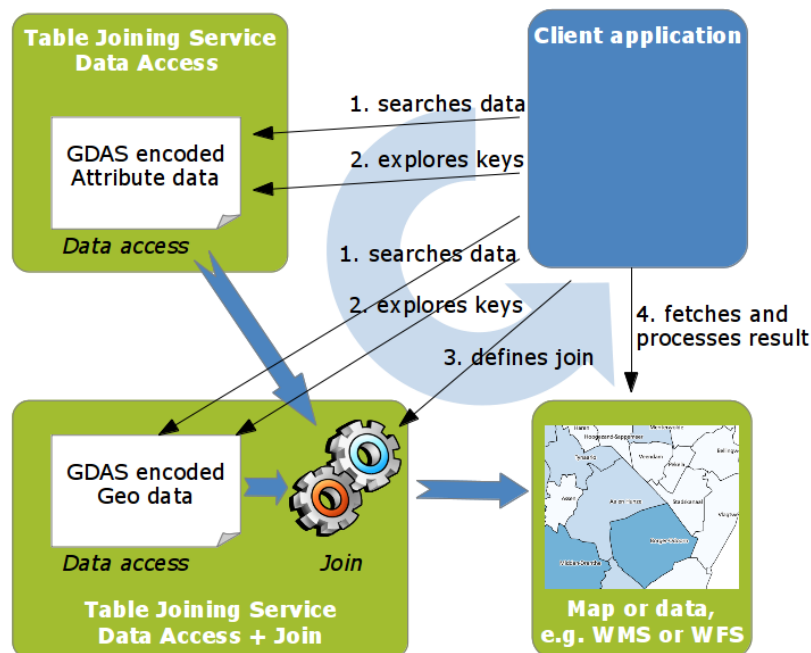


Figure 2.6 Typical workflow for joining data using Table Joining Services



Figure 2.6 illustrates the workflow above. However, different workflows exist when interacting with TJS services. For example, it might also be possible to have one service, offering both tabular (attribute) and geographic data, that is used to first explore geographic data and then find out which attribute data could be mapped onto that. See also the UML interaction diagram in annex A that shows the interaction between a TJS client and TJS servers for joining data for a typical use case.

### Matching objects

Matching objects to each other can become quite complicated. For example, how does one deal with one-to-many relations? To some extent, TJS and GDAS offer solutions. GDAS is able to deal with one-to-many (1:N) and many-to-many relations (M:N). In the encoding, it allows for a specification of how to deal with attribute values when mapping. For example, if a 1:N relation is found, one could specify that certain numerical values are to be summed. TJS also allows for describing composite keys, meaning keys consisting of multiple values (columns).

To find out which keys are matched and which not, TJS offers the possibility to request all keys of spatial framework data by means of the *DescribeKeys* operation. A client application has to compare these cases with the keys in the attribute data. Thus, in case an exception occurs on a *JoinData* request caused by missing keys (so when matching objects), a client can analyze which keys should be provided and fix the issues, where possible.

If matching objects does not succeed (entirely) with the *JoinData* operation or if a client would like to find out which identifiers (values) for the key to use in the attribute data, the *DescribeKey* operation can be used. This operation thus is useful for debugging *JoinData* or to build a new attribute dataset.

### Uploading data for data access

TJS operates on services on (pre)configured data. The client application can access these services and request data. The question now is, is a client capable of uploading data to a Data Access TJS, in order to join the tabular data with its geography (from another server for example)? This aspect is not covered by TJS. One can conclude that it is up to the TJS implementation to deal with this functionality. Uploading tabular (attribute) data by the end user is considered a very useful user requirement, but is not part of the TJS specification and it would require an extra effort on the part of a TJS service provider to provide such functionality.

### TJS data encodings (GDAS) and input formats for TJS

For processing data, Table Joining Services encode spatial framework data and attribute data in the GDAS format. GDAS stands for Geographic Data Attribute Set. This XML format is defined in the Table Joining Service specification and optimized for joining data using Table Joining Services. So when a TJS retrieves tabular (attribute) data to join with spatial framework data, it will expect the data to be in the GDAS encoding. The GDAS encoding is specifically designed for exchanging data for TJS operations. The main use of GDAS is for attribute data. Geographic data is typically dealt with internally by the TJS, so there is no need to encode that in GDAS.

The TJS specification does not prescribe any format to use for storing data or adding data. Attribute data is almost in all cases stored or exchanged in other formats, like CSV, spreadsheets or XML formats, like GML for geographic data. It is up to the data access service to transform the data from these formats to GDAS. The figure below illustrates this. In practice, TJS implementations therefore could support different formats as input formats.

Technically, a TJS deals with data as shown in figure 2.7.



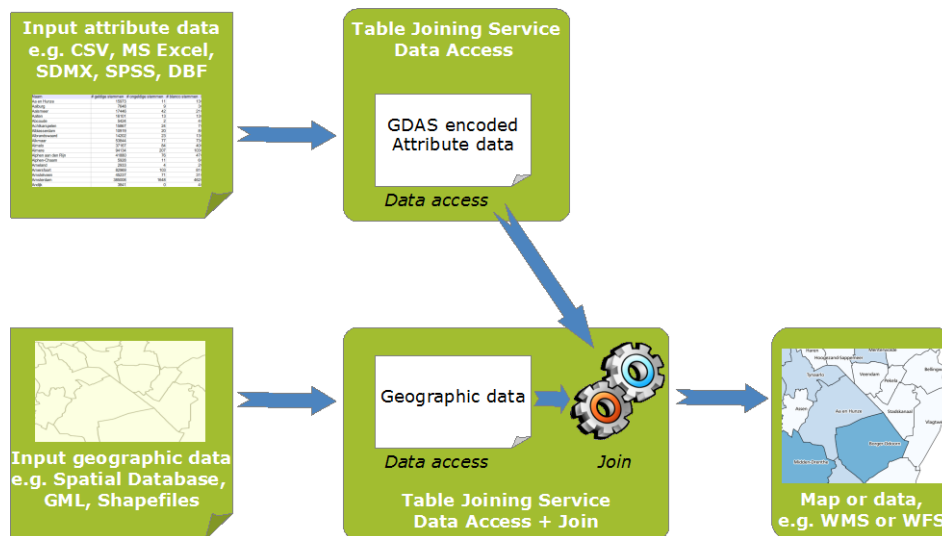


Figure 2.7 Input of data to create GDAS encoded data for Table Joining Services

Data formats that are of interest for attribute data, for statistical usage, are amongst others:

- CSV, Comma Separated Values. A simple table-like format, in plain text using commas or other characters to separate (column) values from each other;
- Fixed format (plain text);
- Microsoft Excel, spreadsheet format of Microsoft;
- SDMX (<http://sdmx.org/>), where SDMX-ML is an XML encoding of SDMX. Other SDMX encodings exist as well, for example a JSON encoding is worked on.
- SPSS SAV, a statistics format.

The Statistical Data and Metadata Exchange (SDMX) initiative (<http://www.sdmx.org>) sets standards that can facilitate the exchange of statistical data and metadata using modern information technology, with an emphasis on aggregated data. Expressing geographic data in SDMX is not standardized. The lack of this makes SDMX hard to use for geospatial applications, in a distributed, heterogeneous environment. Other formats and processing are needed to join SDMX encoded data to geographic data. SDMX itself is not sufficient to make joins, but could serve as an input format for joining.

Other relevant formats for the near future could be RDF (RDF/XML) from the Linked Data family and JSON, which is being used more and more in web applications and as exchange format for data in Web API's. Furthermore, the open data protocol ([www.odata.org](http://www.odata.org)), a standardized protocol for creating and consuming data APIs, is worth considering.

Using common XML technologies such as XSLT, data from XML formats like SDMX-ML and GML could be transformed to the GDAS format, GDAS is an XML language..

For geographic data, these formats could be relevant for TJS as input formats, since they are commonly used in spatial applications:

- Spatially enabled databases, like Postgis, Oracle spatial, ArcSDE;
- ESRI Shapefiles, a file based format;
- GML, Geography Markup Language, a standardized XML language for the encoding of geospatial objects.

Since Data access services do not offer filter or query functionality, it could be quite straightforward to offer Data access functionality. For example by transforming the data to GDAS encoding once and serve the resulting data as static files.



## 2.4 Alternative approaches to table joining

### 2.4.1 Spatial database, ETL-based and GIS-based table joining

The most common approaches to data joining are joining of tabular (attribute) data and geographic data in a spatial database environment, with ETL tooling or with a Geographical Information System (GIS). All three ways can be characterized through a set of operations, resp. spatial database joins, ETL joins or GIS-based table joining.

In some cases, GIS-functionality is integrated in data statistics applications, like ArcGIS mapping in Microsoft Excel in conjunction with ArcGIS online as a publishing platform or in SPSS Maps. Online solutions to join and process data are available as well, for example Google Fusion tables (<http://www.google.com/drive/start/apps.html#fusiontables>), including an option to create maps as output.

Also, business intelligence software is able to display tabular (attribute) data on a map and (dynamically) create custom maps with visualization tools. Tableau is an example of this (<http://www.tableausoftware.com>). Another data visualization tool is Gapminder (<http://www.gapminder.org/>), an application with which to visualize many types of data on a map. Gapminder itself does not allow for joining data provided by other, distributed sources.

All these common solutions use their own protocols and techniques for retrieving and joining data. Joining data across these systems is thus not standardized and therefore requires custom processes and possibly custom programming.

### 2.4.2 WFS based joining

At first sight, Web Feature Services only serve geographic data. WFS implementations could also be used for several joining approaches on geographic data:

- By using Join filters when retrieving data from a WFS (ImplementsStandardJoins). This could work on data that is all offered by the same WFS.
- By using links in GML. In GML, the properties of an object can be defined by linking these to another property or feature. This requires that the attribute data is encoded as GML (without a geometry included) and refers to geographic data for the geometries. The attribute data then contains explicit links to the geographic data, for the geometry. To model this, links need to be established / found and provided in the attribute data. The attribute data needs to be encoded as GML.
- By so-called cascading WFS implementations, see <http://www.galdosinc.com/cascading-and-federated-wfs-and-the-concept-of-geolinking/> for an explanation. In such a setup, one WFS acts as a client that retrieves attribute data from a WFS and geographic data from another WFS. The client-WFS joins the data and offers it to a user as a WFS. The join needs to be defined in the client WFS. Figure 2.8 shows this setup. This approach also requires conversion of tabular, attribute data to GML in one of the WFSes.

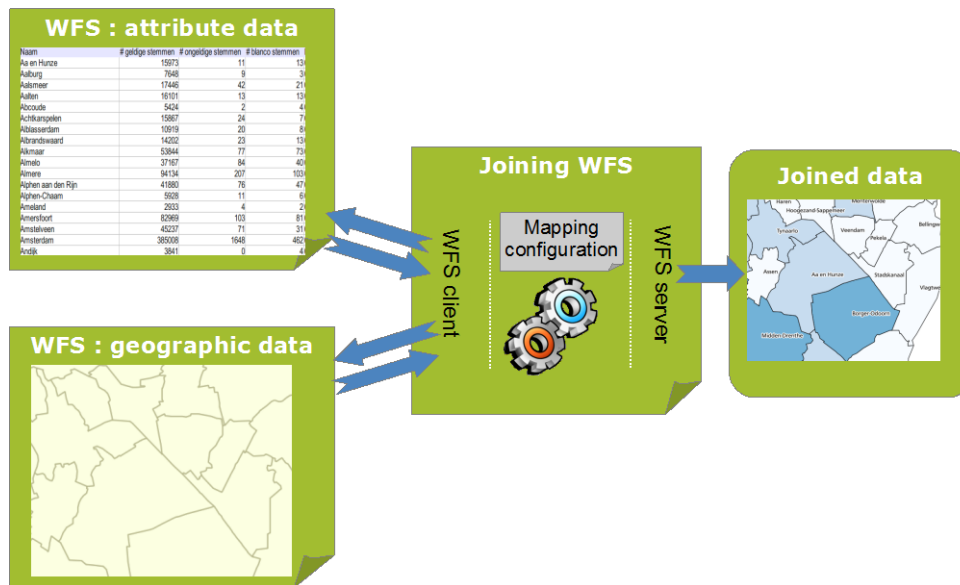


Figure 2.8 Cascaded WFS, where a Joining WFS accesses two WFSes

These approaches require that the attribute data and the geographic data are available as GML. So to allow for joining geographic data (in GML) with attribute data in a tabular format, like CSV or Microsoft Excel, processing is still required. This processing is conceptually similar to a TJS, where attribute data is also processed to another format (GDAS). The difference between a TJS approach and a WFS approach would be the format of the geographic data. WFS uses GML by default and this might already be available, while TJS uses a specific encoding, GDAS. Another difference is that TJS offers the ability to explicitly describe which attributes can be used as keys for joining and how to deal with attributes when joining, while WFS does not have this ability.

Note that TJS implementations could also use GML or other formats when retrieving data, but this would be implementation specific. The TJS specification does not prescribe implementations to support other formats.

### 2.4.3 Linked data

The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and these can in turn be linked to it [10]. Web technologies as RDF and HTTP are used for linked data.

Current interest in Linked Data is growing rapidly in the Netherlands and in other (European) countries and is also advocated in INSPIRE [8]. Adoption in practice, at a production-level, is at the moment still low. However, because of the nature of linked data, it makes sense to look at linked data in the context of joining attribute and geographic data.

Linked data aims to explicitly include relations between data in the data itself. Linked data uses URIs and vocabularies to express these relations. Take for example attribute data containing statistics on administrative units. If this data is modeled as linked data, then links in the statistical data point to the geographic data of the administrative units. These links are expressed on the object-level.

Compared to TJS, the links are thus more explicitly defined. Although the links can be followed, tooling is still required to combine the administrative and geographic data to one dataset or map (visualization). Conceptually TJS offers such functionality. As such, TJS could be used as a server-side solution to resolve the links in linked data. In other words: linked data could be input data for a TJS.

The encoding of linked data often is an RDF encoding, in XML or Turtle for example. To transform the RDF encoded data to GDAS would add an extra processing step, which does not seem to be necessary. In addition to linked data, the concept of a TJS is valuable, but the specification of TJS itself would require some modifications to allow for other data encodings.

The question arises whether TJS itself is needed or linked data tooling could be used. Some tooling is available to automatically create links. This could be used to resolve them and combine the data to create a map for example. SILK (<http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/>) is an example of software to assist the creation of links. Currently there is not much tooling for end-users to consume linked data directly. There are libraries available to consume linked data, but in most cases it requires some programming to use them in an application.

## 2.5 Advantages and disadvantages of TJS

### Advantages

Using TJS has several, potential advantages (source: <http://geoprocessing.info/tjsdoc/> and Peter Schut), such as:

1. The main advantage of TJS is interoperability. It allows organizations to host their corporate data on systems that are optimized for the management of that data, and yet allow themselves and others to take advantage of Geographic Information System (GIS) technology to examine and analyze that data. The TJS standard defines a set of data access operations that can be supported by any database management system, and a set of data joining operations that can be implemented by any GIS. Thus, TJS allows the latest data to be obtained when an analysis is being performed, regardless of whether or not the geospatial system is compatible with or directly connected to the corporate data management system;
2. Simple yet powerful. TJS uses standard HTTP and XML (eXtensible Markup Language) as a mechanism for data exchange; this makes it very lightweight, yet highly scalable to be used for mapping, analysis, calculations, or data replication. TJS owes its simplicity and power to the XML GDAS encoding it uses. All TJS operations create or use XML documents that are based on GDAS;
3. Exploits the power of distributed computing. TJS is designed to facilitate distributed data management, enabling distributed processing of geospatial data located anywhere on the Internet;
4. Fast, reliable access to "near real-time" information. Because it can be generated dynamically, the most up-to-date data can be accessed directly from the source organization responsible for its maintenance;
5. Flexibility with security. Exposing the data through TJS allows corporate data managers to change their underlying database design and security safeguards without compromising access to data that should be accessible by other systems;
6. Easy to find data. Exposing the data through TJS allows its metadata to be harvested into registries, thus making it easy to publish, find, and access attribute data and data joining services;
7. TJS enables fully automated calculations and mapping because of the metadata embedded in it - specifically the identification of the type of data (i.e. Nominal/Ordinal/Measure/Count). This is crucial in determining what kind of calculations and mapping can legitimately be applied. This avoids the human errors that are all too prevalent in human-mediated table joins and mapping; The GDAS format turns out to be quite useful for data exchange between data stores;
8. GDAS contains all the information needed to create charts;
9. TJS can be used as a data format for modeling and scenario development; to generate GDAS as the output from (environmental) models, and stores the resultant GDAS files in order to compare results from different inputs and (climate change) scenarios. The extensive metadata embedded in each GDAS file makes it easy to understand where each dataset came from.

### **Disadvantages of TJS**

TJS offers some (claimed) advantages over other solutions. There are some potential disadvantages as well. The following could be considered disadvantages and possible improvements for TJS:

1. Another service specification. There are already many service specifications operating on spatial data. The TJS specification itself lists, in chapter 6 of [2], some reasons why a separate specification has been established, instead of reusing or extending already existing ones as WFS. Still, one can question if another, new specification is necessary. Offering more service specifications might be hard to develop implementations, to maintain and to have clients capable of working on all these service types;
2. Another specific format for encoding data. TJS might require conversion to GDAS from known formats, which are used already. For example, if an organization already offers attribute data as CSV or JSON because these formats are used and readable for many applications and software libraries, one can argue if another encoding is necessary. Another, additional encoding introduces extra complexity in developing and maintaining software and data;
3. TJS assumes data is already available on a web server. TJS does not define a mechanism for uploading and creating attribute (or geographic) data by a user to a Data Access service. If this is wished for, the TJS implementation should provide functionality for this. Considering a common workflow of an end-user, wishing to combine for example a spreadsheet with geographic data, the fact that TJS does not define a mechanism to input or upload data, is an important limitation for the use of TJS in practice. Such a mechanism would be very valuable and thus requires some effort of the implementing party and might require (partially) dedicated clients. This might be straightforward to implement, but if the TJS specification would have described an (optional) mechanism for uploading data, this would have helped interoperability in some use cases;
4. When combining data, on keys, one can encounter many cases of non-matching data. For example because of misspellings, other naming conventions, abbreviations etcetera. While joins on other keys than names could be considered more exact, names could still be used. For TJS or any other joining technique, it is hard to solve these issues. However, TJS does not have a direct mechanism to show or report which objects or rows did not have a match. TJS does offer an operation to retrieve all keys. It is then up to the client application to find out which keys are matched and which are not. Matching issues are thus not reported in detail by TJS itself;
5. Mechanisms to use encodings other than GDAS to join data. TJS focuses on GDAS. To enhance and ease the usage of TJS, it would be helpful if TJS implementations are able to directly read encodings other than GDAS, without converting them to GDAS first. The TJS specification does not foresee this, but implementation could take other (input) formats into account.

## Chapter 3 The implementation(s) of TJS

### 3.1 Introduction

One of the incentives for this study is the European INSPIRE Directive. INSPIRE is a European law to establish a pan-European spatial data infrastructure to provide data for environmental applications (<http://inspire.jrc.ec.europa.eu>). To ensure that the spatial data infrastructures of the European Member States are compatible and usable in a transboundary context, INSPIRE requires that common principles and technical guidelines are adopted and implemented. These common principles and technical guidelines deal with the essentials of the spatial data infrastructure: metadata, data models, network services for discovery, view and download, agreements for sharing data and monitoring and reporting.

TJS can be considered as a supporting concept in the spatial data infrastructure that joins data from various, distributed sources or infrastructure nodes in order to view and download datasets. There are several ways to implement TJS in organizations – in the case of INSPIRE in European Member States – and this chapter explores some organizational aspects of TJS and examines the possibilities for implementation of different TJS software. In paragraph 3.2, the conceptual architecture of TJS is outlined. Next, in paragraph 3.3, the organizational aspects of TJS implementation are discussed from the perspective of three types of actor organizations involved. Paragraph 3.4 outlines the TJS software implementations that are available worldwide. Finally, in paragraph 3.5, aspects of TJS client applications are discussed.

### 3.2 Conceptual architecture(s) for TJS

#### 3.2.1 Conceptual architecture

The conceptual architecture of TJS is designed as service oriented architecture with service consuming and service providing components communicating via an (enterprise) service bus. The primary aim of the service infrastructure is to provide seamless and interoperable information exchange between data stores of different organizations. The conceptual architecture of the TJS implementation is shown in figure 3.1. It is important to note that it is assumed that all kinds of data and metadata access and processing to the data is performed using web services. All services are described by service descriptions (service metadata, as part of the metadata), allowing humans and software applications to discover specific service instances in the infrastructure and invoke them automatically through the TJS-Data join service.

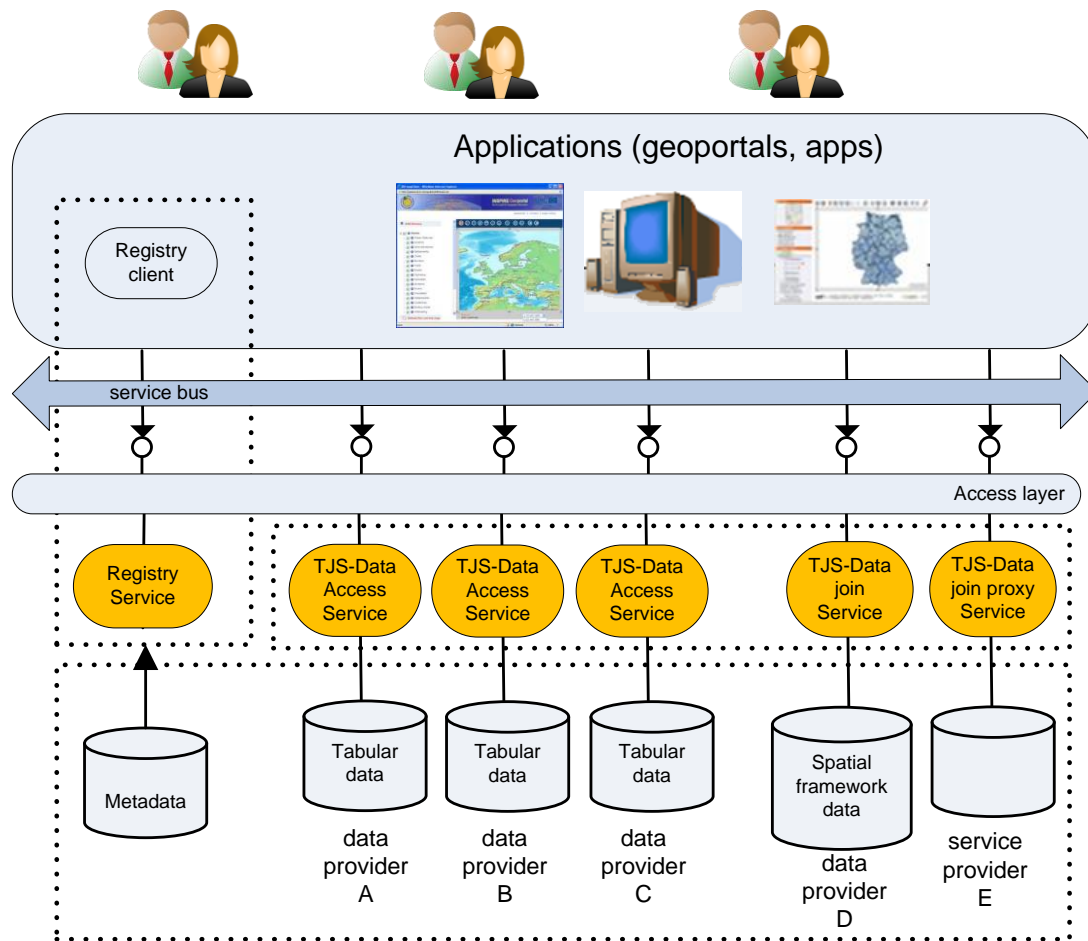


Figure 3.1 TJS conceptual architecture

Data providers can provide a TJS-based infrastructure with three kinds of TJS services:

1. TJS-access services, that provide attribute data in GDAS format. These services support the TJS Data access operations;
2. TJS-Data join services, that often provide spatial data and offer the TJS Data join operations;
3. TJS-Data join proxy services, that do not provide data, but only offer the TJS join operations needed to connect and join tabular (attribute) and geographic data (kind of proxy service). For this setup, the TJS needs to know the sources of the spatial data up front, because the TJS Join operation does not have a mechanism to get spatial data from an arbitrary remote resource and the other operations assume some knowledge of the data itself, e.g. for describing the keys.

TJS allows data to be maintained and updated closest to its source, and yet allows the latest data to be obtained when data analysis is being performed, regardless of whether or not the spatial framework system can make a direct connection to the data management system. Exposing the data through a TJS allows data management to change their underlying database design and security layer without compromising access to data that should be accessible by other systems. In effect, TJS supports both distributed data management, as well as the distributed processing of spatial data.

To facilitate such a SOA-styled architecture, well-defined, type-safe and machine-readable interfaces are essential. In this context, the communication protocol and interface binding technology for the TJS service have to be defined. The OGC TJS service bindings serve as a guideline and support a mix of protocols and technology bindings. These are Key-Value-Pairs sent via HTTP/GET, XML sent via HTTP/POST, SOAP via HTTP/POST and/or combinations.

In addition, it is envisioned that the service bus provides additional management services, like a service registry and access control functionalities. Besides the TJS service instances, the architecture holds their service metadata in a service registry; the user is able to find, evaluate and use TJS service instances through the metadata published in the service registry. Access control through an access control layer gives the data provider the possibility to restrict access to their data in case of access and procurement policies.

The architecture needs to support not only point-to-point communication between clients and services, but as workflows utilizing TJS services in a chain, for instance with WMS, WFS and WPS.

### 3.2.2 TJS and the INSPIRE network services

INSPIRE requires that data is made accessible using View Services [11] and Download Services [12]. View Services allow for the visualization of data as a map; Download Services for the retrieval of the raw (geo)data encoded as GML or coverages, for example. View Services could be implemented as Web Mapping Service (WMS) with some additional requirements; Download Services as Web Feature Service (WFS). Metadata of the data and the services shall be published in Discovery Services, allowing users to find the appropriate data and services and the use of both.

View Services of INSPIRE require that data have a spatial component, a geometry. Otherwise the data can't be shown on a map. This means that data that don't have geometries themselves, but include only references to objects with geometries of other objects, need to be joined before the data can be visualized on a map. Table Joining Services creating WMS as output of a join process, could therefore be a useful and critical component when View Services need to be created.

For Download Services it is allowed to just keep the references to other objects. Creating the join is not required by INSPIRE as such. But for clients to use the data in their systems and work processes, for example by using GIS software, it is generally very useful to have data that is already joined. Table Joining Services therefore could be a useful service type in the INSPIRE architecture.

INSPIRE recognizes that services other than View and Download could be important in the spatial data infrastructure. For example, services that allow for the processing of spatial data, for example for calculations using complex models or aggregation of data, and services for the transformation of data. In the INSPIRE framework, these are called Spatial Data Services and Transformation Services respectively.

Implementations of Spatial Data Services (SDS) could be available using custom protocols or using standardized protocols, such as Web Processing Services (WPS), an OGC specification to execute computational processes on geospatial data. TJS could be considered a SDS as well. WPS and TJS could work in conjunction. For example, to first aggregate data on a lower administrative level to a higher level (e.g. municipalities to provinces) and then combine that data with attribute data available on that level. Arguably, WPS could be used for joining data as well. However, WPS is very generic and would require clients to have more knowledge of the data and specific WPS implementation to successfully join the data. TJS is a dedicated specification and as such more suitable to execute joins.

## 3.3 The organizational aspects of TJS implementation

### 3.3.1 Actors involved in TJS implementation

The adoption of TJS will have an impact on the organization and its services infrastructure of the different data providers involved. This paragraph introduces the general activities for the different data providers involved in TJS adoption and TJS implementation. Three types of data providers are distinguished:

1. Tabular (attribute) data provider;
2. Spatial framework data provider;
3. TJS service provider.



### **Tabular (attribute) data provider perspective**

The tabular (attribute) data provider makes its data available on the web as follows:

1. Identify the data and determine which data and associated metadata the organization wants to publish. An important aspect of this step is defining the identifier(s) to be used for joining;
2. Install the TJS API's through the *GetCapabilities* and *GetData* operations and application program interfaces (API's) to interact with the database system and produce XML streams (GDAS) according to the TJS specification.
3. Register the TJS-Data access service. Create a service description document and publish it to a service registry.
4. Finally, the tabular (attribute) data provider maintains the TJS-Data access service and its infrastructure. This means regular maintenance (like event, incident and problem infra management) and providing updates of the data tables.

### **Geographic data provider perspective**

The geographic data provider makes its geographic data available on the web as follows:

1. Identify the framework data and determine which framework data and associated metadata the organization wants to publish. An important aspect of this step is defining the identifier(s) to be used for joining;
2. Install the TJS-Join data API's through the *GetCapabilities*, and *JoinData* operations and application program interfaces (API's) to interact with the spatial database system and TJS-Data access service and produce a joined data stream according to the TJS specification.
3. Register the TJS-Data join service. Create a service description document and publish it to a service registry.
4. Finally, the geographic data provider maintains the TJS-Join data service and its infrastructure. This means regular maintenance of the spatial (identifier) framework and maintenance of the services infrastructure (like event, incident and problem infra management) and providing updates of the spatial data.

### **TJS service provider perspective**

The TJS service provider makes a TJS-Data join service available on the web as follows:

1. Install the TJS-Join data services API's through the *GetCapabilities*, *GetData* and *JoinData* operations and application program interfaces (API's) to interact with the TJS-Data access and TJS-Data join services and produce a joined data stream according to the TJS specification.
2. Register the data service. Create a service description document and publish it to a service registry.
3. Finally, the TJS service provider maintains the TJS-Join data service and its infrastructure. This means regular maintenance of the services infrastructure (like event, incident and problem infra management).

#### **3.3.2 TJS and the user perspective**

Attribute data delivered from a TJS can be used in a variety of ways, including visualization as a web map or even for the use of models to perform calculations,. TJS services are more or less back-end services that load and join attribute and spatial data as infrastructure services and offer data to client applications.

The user can access the data client applications in two ways:

1. Access to attribute data only; or
2. Access to the attribute and geographical data (joined data) for display in a map view.

For access of only attribute data TJS service, often the data will be transformed (converted) at the server side into the client format, such as HTML, SDMX, CSV, Microsoft Excel or as XML. This user perspective of the TJS-Data access service instances is shown in figure 3.2.



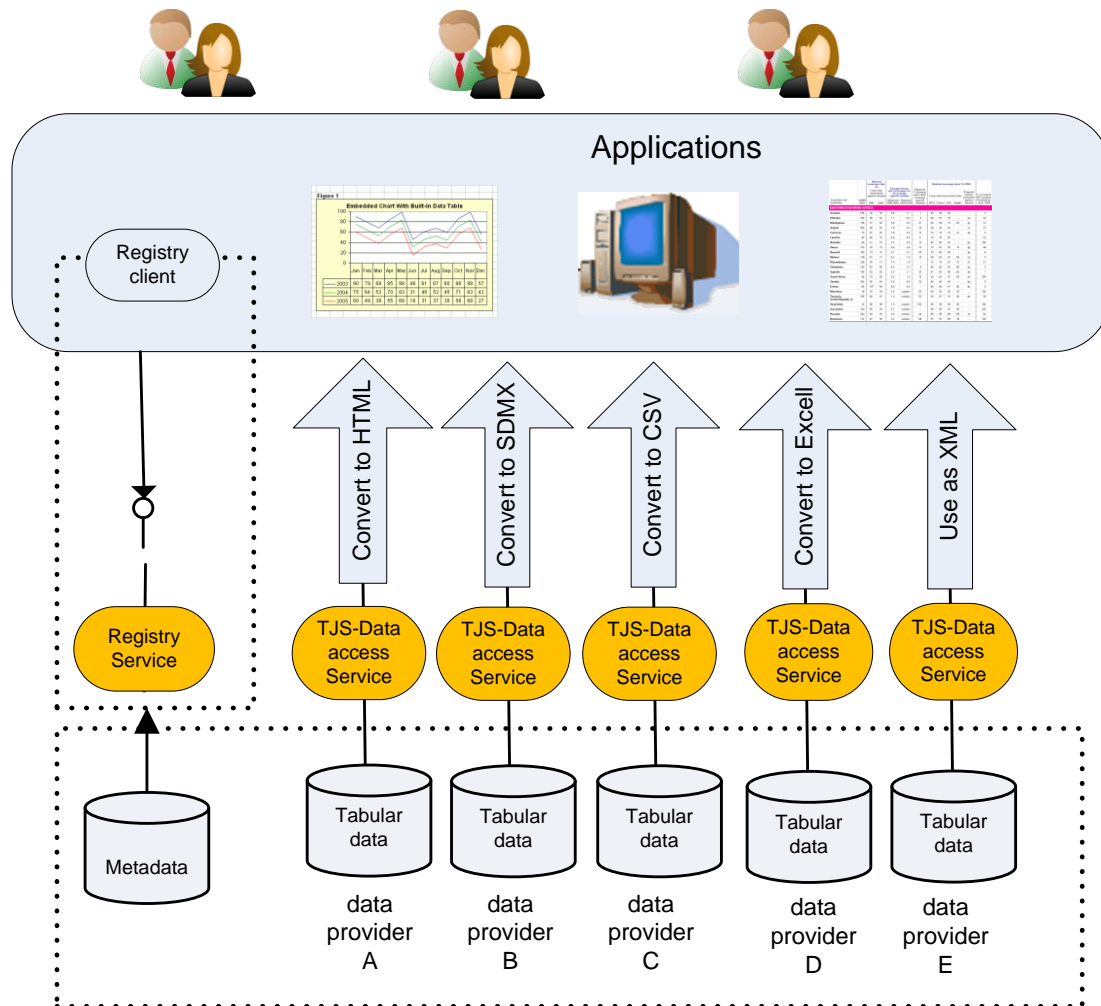


Figure 3.2 TJS-Data access user perspective

Which formats are offered by the TJS-Data access service is implementation-specific. This will depend on the software implementation the data provider uses in the service infrastructure. The TJS output format is GDAS, which can of course be converted to other supported data formats as well. To allow for conversion of GDAS data directly, the GetData operation of TJS allows for the (optional) inclusion of an XSLT-file, which can be used to transform XML data to other formats / encodings, for example to HTML, CSV or XML-encodings.

The perspective of the user regarding the use of the TJS-Data join service is a request map data. The TJS-Data join gives access to the attribute and geographical data or the joined data mostly for display in a map view, or for further calculations or database storage. For access of joined data through the TJS service, the data will often be stored in a database or the joined table is created on-the-fly and temporarily stored for further processing through web services standards, like WMS, WFS and WPS or as a data download service. As stated in chapter 2, TJS could work well with existing geospatial standards. TJS does allow for the publication of results in standardized geospatial web services like Web Map Services (WMS) and Web Feature Services (WFS). If this approach is used, existing web viewers and libraries can be used. Note that for executing join operations, clients possibly need to be modified. These modifications require development work, but could be straightforward.

This user perspective of the TJS-Data join service instances is shown in figure 3.3 .

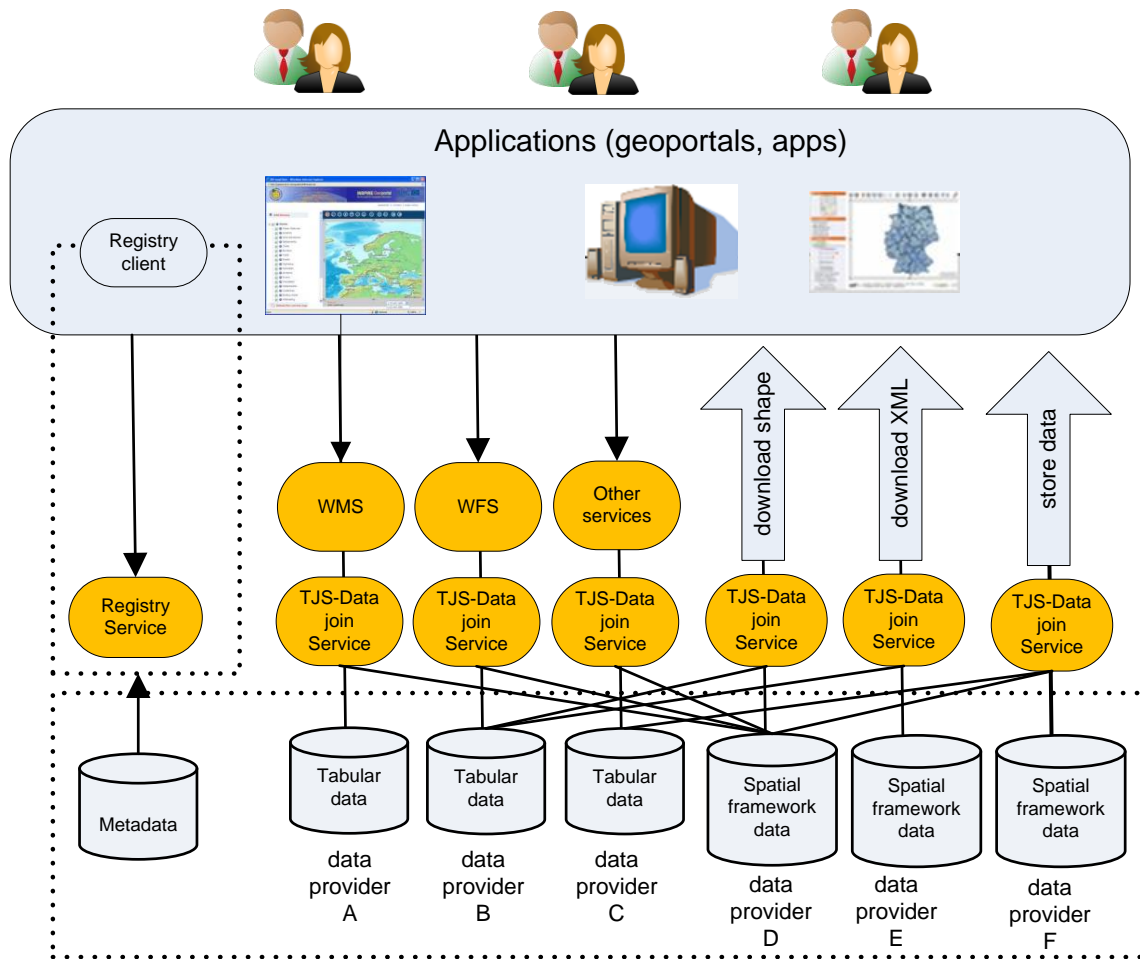


Figure 3.3 TJS-Data join user perspective

### 3.3.3 The need for a spatial (identifier) framework

As mentioned in the previous chapter, table joining is based on a spatial framework and the common geographic identifier. Therefore, the concept of the table joining in a distributed service oriented architecture requires that organizations implement their services according to a common spatial (identifier) framework. Therefore, organizations that want to adopt and use TJS with long-term perspective need to agree and adopt standards for spatial data (concepts and models) and geographic identifiers or, even better, a common spatial (identifier) framework. A common spatial and identifier framework consist of several aspects:

- Data models for spatial framework data based on a generic conceptual model;
- Generic approach for handling and encoding unique, persistent geographic identifiers;
- Registry for maintenance of data models and identifiers.

So, for data providers offering data in TJS infrastructures it is vital to agree upon a common spatial (identifier) framework. The study does not offer such a spatial (identifier) framework. It is not the objective of this study to provide an identifier framework. An interesting European spatial (identifier) framework can be found in INSPIRE Directive (see <http://inspire.jrc.ec.europa.eu/>) and [8].

## 3.4 TJS software implementations

For organizations, like statistical offices, to implement TJS, it is necessary to have a broad range of implementations in software. So that organizations have easy access and different possible



implementations to choose from. However, there are no OGC TJS compliant products registered at the OGC website. Furthermore, a Google search showed only a few software implementations; each different from the level of maturity. And only a few production implementations of these software products were found. In this paragraph, we give an overview of the TJS software products that were found so far, and some of their characteristics.

### OGC TJS compliant products

The OGC portal (<http://www.opengeospatial.org/resource>) provides information on the products that are implementing OGC standards, including products that have been certified as OGC Compliant. Today, there are no registered products on the OGC portal/website that are implementing and/or compliant with the OGC TJS specification (figure 3.4). The fact that there are no registered products for the OGC TJS standard, can probably be explained by a few reasons: 1. There is no automated test suite for TJS, 2. TJS is still relatively new as for OGC standards specifications (it seems to take at least five years for a new standard to see any significant adoption), and 3. It's a lot of work to register an implementation at OGC and many software developers don't bother to register their products.

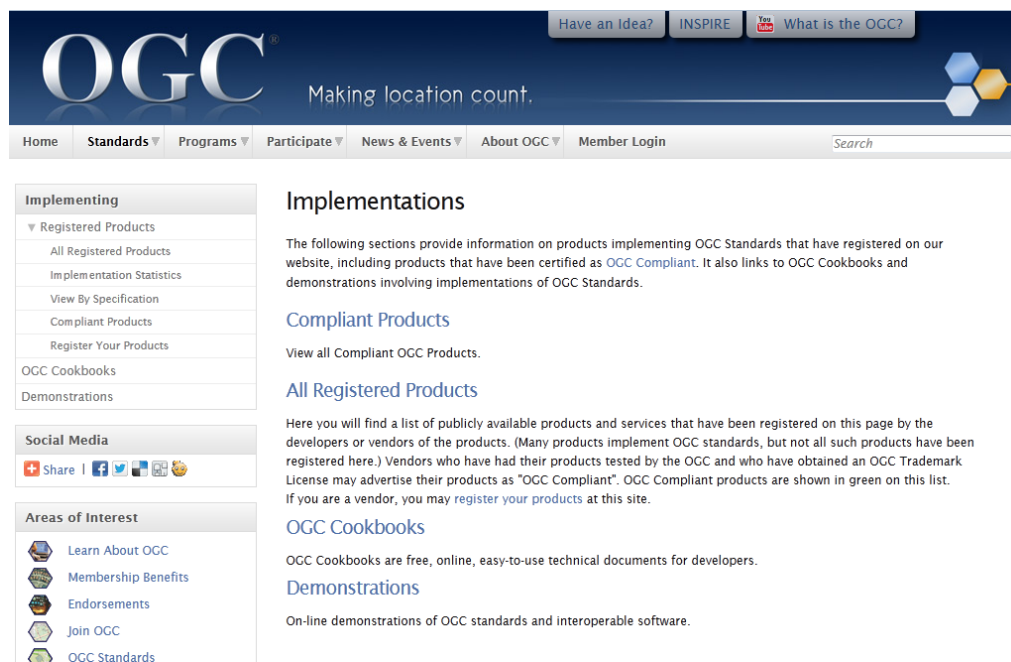


Figure 3.4 TJS Compliant products (<http://www.opengeospatial.org/resource>)

### Some early implementations (GLS and GDAS)

Early implementations of the GLS (Geo Linking Service, a predecessor of TJS, see section 2.2.1) and GDAS specifications can be found at:

- <http://code.google.com/p/mapattributes/>
- <http://ogsadai.sourceforge.net/documentation/ogsadai4.1/ogsadai4.1axis/OverviewSEEGEO.html>
- <http://geoapp.nottingham.ac.uk/GLSClientInstallation/index.aspx>

There is no recent implementation references for updates to TJS.

### Overview of TJS implementations

Furthermore, based on a Google search, five TJS software implementations were found (see figure 3.5). All five implementations support TJS server side, only two implementations implemented TJS client side software. The programming language technology used for the TJS software differ. Only one software product is proprietary, called Géoclip from the French company called emc3. The other implementations are open source software implementations.

Software product	Client/server	TJS access - TJS join	Technology	Type of software
Metamapper	server	TJS-access	Ruby on Rails (and MySQL)	Open source
Géoclip	client, server	TJS-access, TJS-Join	?	Proprietary
Canadian Geo-linked Data Access Service	client, server	TJS-access, TJS-Join	Java	Open source
TJSServer	server	TJS-Join	GO (and PostgreSQL)	Open source
Geoserver TJS extension	server	TJS-access, TJS-Join	Java	Open source

Figure 3.5 TJS Software implementations

#### Metamapper

Metamapper is an OGC TJS implementation of Peter Schut, the editor of the OGC TJS standard. Metamapper is written in Ruby on Rails and MySQL. MetaMapper is currently off-line, and on demand available as an open-source package.

#### Géoclip

Géoclip is a platform for geospatial visualization of statistical data on the web. Géoclip applications, particularly France découverte, began to implement this OGC TJS standard in January 2010. Emc3, the French company behind Géoclip, encourages its customers to embrace this standard, particularly statistical data providers such as INSEE or statistical services within governments. Géoclip offers flexibility and the capacity to handle large data sets, with interactive vectorial cartography that can represent millions of entities. A Géoclip Server application has its own database. Thanks to TJS, each one of those databases can be opened up to other applications. Géoclip's has implementation in different countries, such as France, Switzerland, Canada, Spain and the USA. emc3 promotes TJS actively, so that statistical organizations will launch their own TJS servers. One of the few TJS client demonstrators can be found on the Géoclip website: [http://www.geoclip.fr/an/p43\\_tjs.php](http://www.geoclip.fr/an/p43_tjs.php).

#### Canadian Geo-linked Data Access Service

As part of the Canadian Geospatial Data Infrastructure (CGDI), the specifications of the Geo-linked Data Access Service (GDAS) were developed with close cooperation among government institutions and private sectors. This comprehensive implementation consists of the Canadian Geo-linked Data Access Service, a GDAS Service Client and Extended Applications. For Agriculture Agri-Food Canada (AAFC), a GDAS Service Client has been developed to dynamically retrieve information from the GDAS registry and any GDAS data repository. Based on the dynamically generated index page of available GDAS Framework, Dataset and Attribute, this application provides users with several options to retrieve the attribute. The options are to



view the attribute of a selected framework as XML, Text, HTML Table, or Web Mapping Service (WMS) map. The AAFC GDAS API is also available for advanced users. More information on the Canadian Geo-linked Data Access Service can be found at: [http://wms1.agr.gc.ca/gdas/aafc\\_gdas\\_main.php](http://wms1.agr.gc.ca/gdas/aafc_gdas_main.php).

### **TJSServer**

TJSServer is a github project of Guilhem Vellut (<https://github.com/gvellut/TJSServer>). The goal of this project is to implement all TJS operations (in GO programming language) using the following data sources:

- CSV data files for the data access operations;
- PostgreSQL (with PostGIS) databases for the data join operations.

The output of the *JoinData* operation is pushed to the PostgreSQL database. First, by creating a temporary table from the data obtained *through* a *GetData* operation. Then, by creating a temporary view joining the temporary table and the spatial framework table. The join uses the common framework key. Finally, a layer is configured in the WMS using the temporary view. The duration of how long a WMS layer stays valid (along with its backing view and table) can be configured. A WMS is not provided. TJSServer integrates with Geoserver to create a temporary layer, using the REST configuration API. The WMS functionality is then provided by Geoserver.

### **Geoserver TJS plugin**

Geoserver is a popular open source software implementation to supply geospatial data to the web in several standards. For example, Geoserver offers OGC's WMS, WFS and WCS interfaces for building spatial data infrastructures. At the upcoming Foss4g conference in September 2013, a TJS implementation for a Geoserver (plugin) will be presented. This Geoserver plugin is currently being developed by Geocuba. The TJS plugin for Geoserver implements the Join operation and adds some user-friendly management tooling for publishing data. The output of the Join operation can be accessed by WMS.

## **3.5 Client applications of JTS**

TJS is most powerful in situations where data providers of tabular (attribute) and geographic data are distributed. However, distributed maintenance of tabular (attribute) and geographic data also occurs within organizations. So, TJS can be applied for data sharing and exchange within (larger) organizations as well. TJS can be used for:

- Data download. A TJS can offer GDAS data streams for the downloading of regular data content;
- Data replication. A TJS can use GDAS data streams to regularly update the contents of a data warehouse and its associated metadata tables, based on the latest data available from the primary data warehouse;
- Web mapping. One of the most applicable uses of TJS is to support web mapping, in particular, the dynamic creation of Web Map Service (WMS) layers. Dynamic data from an XML stream produced with a TJS-Data Access service can be seamlessly merged with the associated spatial framework data through the TJS-Data Join service, and delivered through a WMS;
- Desktop GIS. A TJS can add an attribute field or amend the contents of an existing field in a dataset in the spatial data so it can be used in a GIS. The TJS could provide the output as data in a file or through a web service interface as WFS.

There are only a few TJS implementations up-and-running at the moment and publicly available. As TJS is a back-end technology, it is not always obvious that TJS is used. However, at the moment, it is clear that this OGC standard is not implemented broadly. A TJS client demonstrator offers the possibility to connect to TJS servers. The most compelling demonstrator are some Géoclip applications

([http://www.geoclip.fr/an/p43\\_tjs.php](http://www.geoclip.fr/an/p43_tjs.php)), like an application built for health care and hospitals in New Jersey (see figure 3.6 and 3.7).

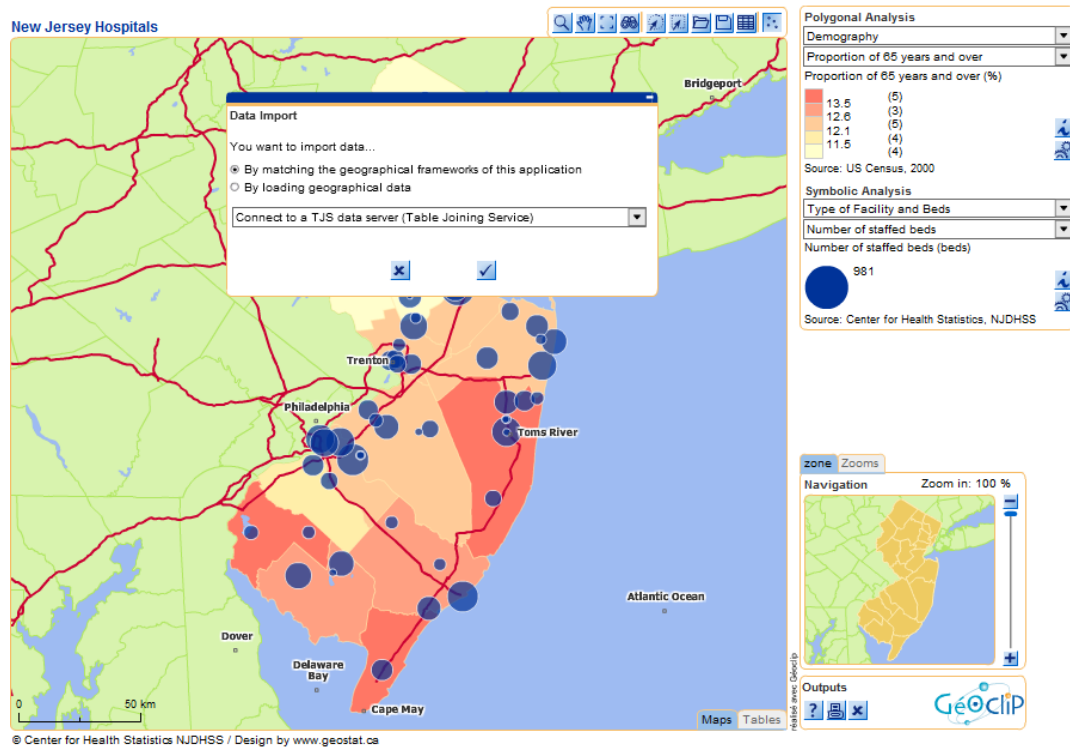


Figure 3.6 Data import in Géoclip application for health care in New Jersey

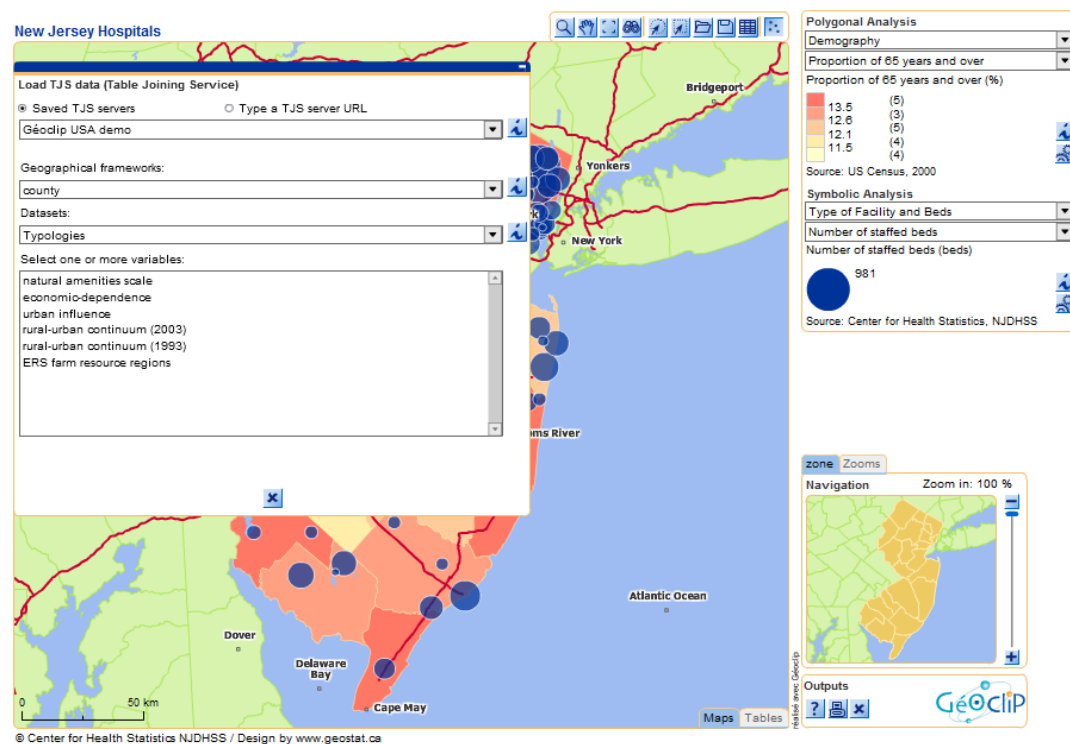


Figure 3.7 TJS connectivity in Géoclip application for health care in New Jersey



### 3.6 Concluding remarks

Several software implementations of the OGC TJS exists, each seems to have a different level of maturity. And only a few production implementations of these software products, proprietary as well as open source, were found. One the questions to be answered in this study was, if private GIS-software companies were anticipating on OGC TJS implementation. So, yes there are private companies that have implemented OGC TJS software, but there is not yet much to choose from for (national) statistical offices. Also the question if new TJS software needs to be implemented depends on the software infrastructure and strategy of the individual organization. For the Statistics Netherlands the Geoserver OGC TJS implementation offers a very promising choice, as the Statistics Netherlands offers their spatial data to the Dutch national spatial data infrastructure node (PDOK) that uses Geoserver as data serving platform. But, at the same time the small number of OGC TJS implementations found and the lack of large GIS vendors in the list of TJS software implementations, will require at least some investment in testing of existing and development of new TJS software and client applications that make smart use of TJS.





## Chapter 4 The cost-benefits of TJS, example of the Netherlands

### 4.1 Introduction

Nowadays, IT investments in the public sector, like the adoption and implementation of a TJS infrastructure, require business case development. So, one of the questions in this study is, what are the costs-benefits of the adoption and implementation of TJS in the public sector in the Netherlands?

There are many approaches to analyzing the costs-benefits of IT investments. In this study, the cost analysis approach is based on a simplified Total Cost of Ownership method (TCO-model) with 5 variables and 10 constants in the TCO-model. This TCO-model has been applied to the adoption and implementation of TJS infrastructure for the public sector in the Netherlands. For the purpose of cost estimation, three scenarios have been developed and cost estimations have been done for a period of 10 years. This study combines two approaches. A costs analysis has been done based on a Total Cost of Ownership method with 5 variables and 10 constants. The variable and constants values are estimated by experts. Furthermore, other indirect benefits were also examined during these expert meetings.

This chapter examines the costs-benefits of the adoption of TJS infrastructure in the public sector of the Netherlands. Paragraph 4.2 examines the existing 'manual' approach of joining tabular and geographic data by an example of the Statistics Netherlands. Next, in paragraph 4.3, the cost factors for TCO-analysis and the TCO-model are introduced. Paragraph 4.4. outlines the three scenarios including their characteristics as model variables and their estimated values. Paragraph 4.5 outlines the results of the estimation of three TCO-analysis scenarios for the Netherlands. Finally, in paragraph 4.6 the expected indirect benefits are summarized.

### 4.2 Example of table joining practice at Statistics Netherlands

In order to have an idea about the table joining activities with statistical data in public organization and the time spent on these activities, cases from different Dutch organizations have been consulted. These are the use cases behind the table joining activities needed in order to view statistical data in a thematic map. Here we describe the use case of the Statistics Netherlands.

The Statistics Netherlands uses the application PXmap for thematic mapping of statistical data. For mapping statistical data in PXmap two approaches exist:

1. Mapping data from existing Statline tables, the online statistical database from Statistics Netherlands (<http://statline.cbs.nl/statweb/>);
2. Mapping data from user-defined, often new tables (not available in Statline) e.g. for data quality assurance checks.

For the first approach the following steps are necessary to take:

1. Statline: In Statline choose theme "NL / regional"
2. Statline: Select tabular data based on attribute dimensions
3. Statline: Check geographic codes are added through subject "lokaliseren"
4. Statline: Download tabular data as csv or excel file
5. Text editor: Next, pre-process the table in text editor
6. Text editor: Delete unnecessary texts
7. Text editor: Rename joining table column to "STATCODE" and make this the first column
8. Text editor: Add column name ini-file that links to the appropriate geographic identifier column

9. Text editor: Save file as csv
10. PXMap: Load the csv file
11. PXMap: Choose map symbols
12. PXMap: Edit Legend
13. PXMap: Export map to WMF-file
14. Import WMF-file for further map editing (in office applications like Word, Excel).

The first approach is outlined in figure 4.1 that shows the necessary data processing steps and applications used for data joining and further processing into thematic maps.

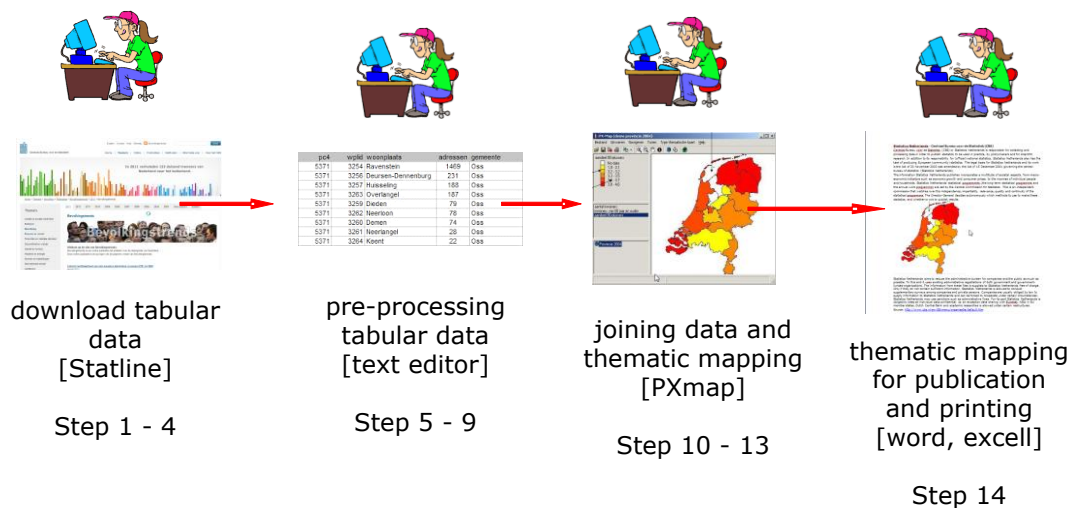


Figure 4.1 Joining data tabular and geographic data at Statistics Netherlands

In the second approach, steps 1-9 are different from the first approach. To prepare own table requires:

1. Intranet File server: Download geographic data from
2. Intranet File server: Choose type of geographic region per reference year
3. Intranet File server: Download tabular data in csv format

The above use case of Statistics Netherlands can be seen as model example, where basically manual processing is needed to join tables (figure 4.2); sometimes needed to prepare thematic maps preprocessed for mapping applications, sometime for further, more ad hoc analysis and thematic map printing. The other organizations that are studied for this purpose have more or less the same two approaches and procedures.

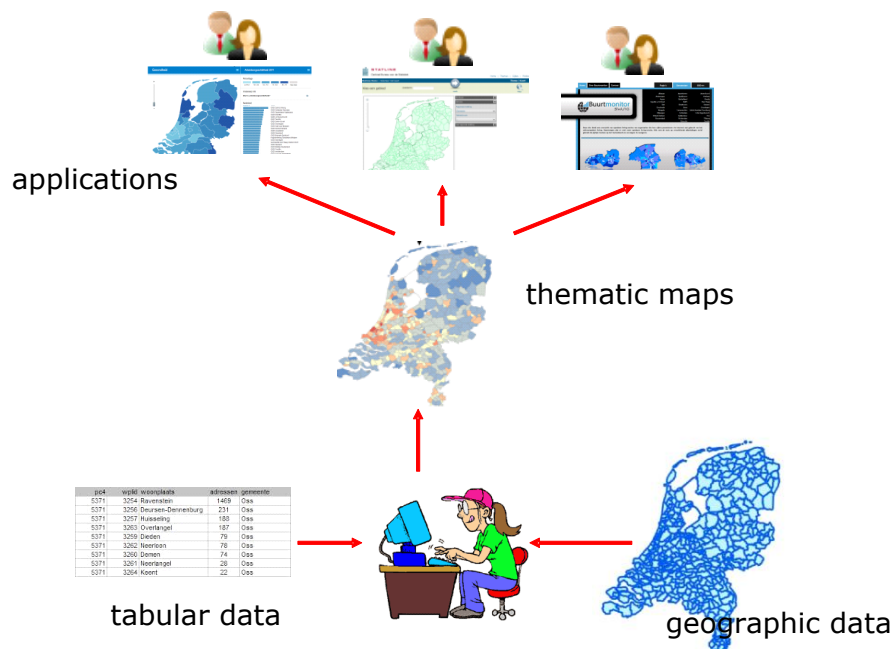


Figure 4.2 Joining data tabular and geographic data through manual processing

The processing of statistical data in thematic maps is cumbersome and TJS could offer simplicity through automatic joining data. This is shown in figure 4.3.

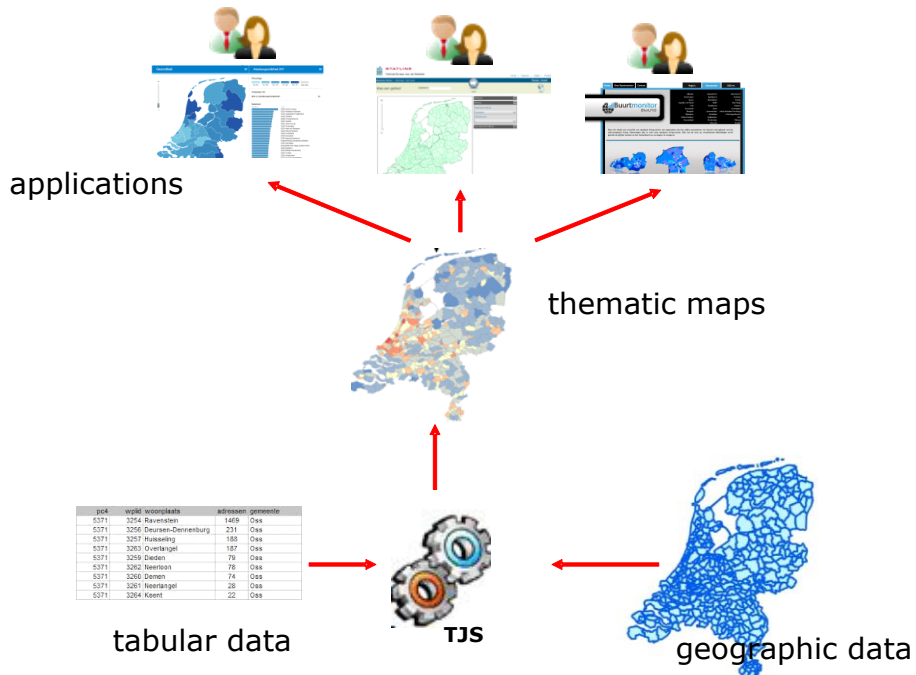


Figure 4.3 Joining data tabular and geographic data by automatic service oriented data joining

### **Expert workshop Statistics Netherlands**

During an expert workshop held at the Statistics Netherlands, employees from different parts of the organization mentioned the following areas where TJS might be useful to apply for Statistics Netherlands, both internally as externally (i.e. for cooperation with other organizations and the general public).

There are several applications and use cases for this, including applications following from the INSPIRE directive and the way INSPIRE requires data to be published. Statistics Netherlands employees have mentioned several applications of TJS on data offered by Statistics Netherlands for internal use (think of exploring data on maps for research and analysis) and external use (think of adding more maps and creation of maps for Statline data dynamically, easier publishing of maps using a regional division).

Internal applications of TJS for Statistics Netherlands:

- Dynamically joining data for exploring / analyzing data for research activities. Visualizing data on a map can be very valuable when analyzing data;
- Because of the internal security policy, using third-party data directly from the source is hard in some cases. It would be beneficial if these third-party datasets could be available at the CBS (without modifying them) for joining and visualizing the data with CBS's own data. TJS could play a role here.

External applications of TJS for Statistics Netherlands:

- Provide more and more easily publications based on the regional data. Visualizing and publish more data in its geographic context is considered valuable;
- Add (more) commonly-used maps and regional data for Statline, to enrich Statline for educational purposes and use of statistical data at schools.

## **4.3 Cost factors for TCO-analysis**

### **Total Cost of Ownership**

Total cost of ownership (TCO) is a financial estimate, the purpose of which is to help determine the direct and indirect costs of a product or system. TCO, when incorporated in any financial benefit analysis, provides a cost basis for determining the total economic value of an investment. A TCO analysis includes the total cost of acquisition and operating costs. A TCO analysis is used to gauge the viability of any capital investment. An enterprise may use it as a product/process comparison tool. It is also used by credit markets and financing agencies. TCO directly relates to an enterprise's asset and/or related systems total costs across all projects and processes, thus giving a picture of the profitability over time. TCO analysis was popularized by the Gartner Group in 1987. TCO tries to quantify the financial impact of deploying an information technology product over its life cycle. These technologies include software and hardware, and training [9].

The cost factors of the adoption and implementation of TJS are fourfold:

1. Development of server and client side TJS software
2. Data management costs related to TJS
3. Maintenance of (hardware and software) infrastructure.
4. Hidden costs (training).

### **TCO-model description**

The TCO-model introduced in this paragraph has been applied to the use of TJS in the public sector of the Netherlands. A time span of 10 years is taken into account in the TCO-model. As the TJS standard is nowadays not implemented in widely available spatial or statistical software (server-side and client-side), the development of TJS server and client software has to be considered as well and therefore also taken into account in the TCO-model. There is no competitive market as yet with out-of-the-box TJS software. It is assumed that these investments in TJS server and client software (development) shall be done in the first 3 years. The adoption of TJS will start after three years. So, the first three years, the manual data

joining still needs to be taken into account in the TCO-Model together with investments in software development. The costs will go before revenues will come.

An important aspect of the TCO-model are costs involved in data management or data joining activities. As shown in figures 4.2 and 4.3, TJS introduces a new way of data management; from manual data joining to automatic, service oriented data joining. So, the TCO-model needs to represent the data management activities for both the manual approach and TJS approach. For instance for TJS joins, attribute data should be made available on the web in a specifically defined XML encoding for TJS: GDAS. This requires some conversion and the provision of metadata. Data conversion could be part of a TJS implementation or could be done by converting data from CSV, Microsoft Excel or SDMX using separate tooling. TJS data management also includes metadata provision, which basically concerns manual editing of metadata necessary to find and bind the TJS services.

The TCO-model used for TJS implementation and adoption in the Netherlands is as follows:

**TCO\_TJS** = CostsDataManagement + CostsTJSServerDevelopment + CostsTJSClientDevelopment + CostsInitialTJSSetup + CostsTJSDatamanagement + CostsTJSMaintenance + CostsTJSServiceHosting + HiddenCosts

**CostsDataManagement** = (#Organizations \* #Users \* #HoursforDataJoins \* AvgHourRate)

**CostsTJSServerDevelopment** = (#TJSServer \* CostsTJSServerDevelopment)

**CostsTJSClientDevelopment** = (#TJSClient \* CostsTJSClientDevelopment)

**CostsInitialTJSSetup** = (#TJSProviders \* AvgHoursInitialTJSSetup \* AvgHourRate)

**CostsTJSDatamanagement** = (#TJSProviders \* AvgHoursTJSDatamanagement \* AvgHourRate)

**CostsTJSMaintenance** = (#TJSServer \* #AvgHoursTJSMaintenance \* HourRate)

**CostsTJSServiceHosting** = (#TJSServer \* CostsHosting)

**HiddenCosts** = CostsTraining

with,

#Organizations = Total number of organizations

#Users = Total number of users in an organization that are working with spatial statistics and need to join spatial and statistical data manually with their application

#HoursforDataJoins = Total number of hours spent by data professionals on non-TJS based data joins

AvgHourRate = Average Hour rate in Euros

#TJSServers = Total number of TJS server software

CostsTJSServerDevelopment = Average Estimated Costs for Development of TJS Server software in Euros. As there is no competitive market yet for out-of-the box TJS server software, it is assumed that custom implementation of TJS server software (software development) is necessary. Therefore, it is also assumed that the costs of an out-of-the-box TJS server is equal to the costs of the development of new TJS server software.

#TJSClient = Total number of TJS client software

CostsTJSClientDevelopment = Average Estimated Costs for Development of TJS client software in Euros

#TJSProviders = Total number of TJS data providers

AvgHoursInitialTJSSetup = Average Estimated Hours for initial setup of TJS Server software

AvgHoursTJSDataManagement = Average Estimated Hours for of TJS Datamanagement

AvgHoursforTJSMaintenance = Average estimated number of hours for TJS Server maintenance

CostsHosting = Costs of TJS Server network (datatraffic) and hosting

CostsTraining = TJS users training

## 4.4 TCO-analysis scenarios for the Netherlands

The TCO-model introduced in the previous paragraph has been applied to the adoption of TJS in the Netherlands. For this purpose, three scenarios were analyzed. These scenarios were developed based on expert opinions to illustrate the expected scenarios that TJS might be implemented in the Netherlands. These three scenario's and their main characteristics (variable values) are described below. Next, the outcomes of the TCO-model runs are described for a time span of 10 years. The outcomes of the three TCO-analysis scenarios can only be considered as first rough estimations that give an idea of the cost development of the adoption of TJS in the Netherlands.

### Three scenarios for TCO-analyses

The TCO-analysis for TJS adoption in the Netherlands is calculated for three different scenarios:

1. Scenario A; no TJS development and adoption;
2. Scenario B; a minimum adoption of TJS by a few organizations;
3. Scenario C; a maximum adoption of TJS by many organizations.

#### Scenario A The No-TJS scenario

This is the 'business as usual' scenario. There is no TJS infrastructure available. Data joining of spatial and attribute data is done manually or automated by non-TJS mechanisms. In figure 4.4, the TCO-model variables of Scenario A, The No-TJS scenario, are summarized.

TCO-Model variable	Estimated value	Explanation
#Organization	400	There are approx. 800 public organizations in the Netherlands of which 50% have data professionals who perform data joining activities with spatial data. In the Netherlands approx. 400 public organizations have data professionals who need to perform data joining activities.
#TJSServer	0	0 TJS servers exist in the No-TJS scenario.
#TJSClient	0	0 TJS clients exist in the No-TJS scenario.
#TJSProviders	0	0 TJS providers exist in the No-TJS scenario.
AvgHoursTJSData-Management	0	No time is spent on TJS data management in the No-TJS scenario.

Figure 4.4 Estimated values for Scenario A, the NO-TJS scenario

### Scenario B The Minimum-TJS scenario

This is the minimum scenario. There are 2 TJS infrastructure nodes available. These are main TJS nodes in the data infrastructure in the Netherlands:

1. The data offerings of the Statistics Netherlands
2. The geospatial framework node where all main spatial data is offered.

From the client perspective also 2 important clients exist. The first one is the Statline client application of the Statistics Netherlands, the second client is the proprietary Swing application that is used by many regional and local public organizations (provinces and municipalities).

However, in this scenario still data joining of spatial and attribute data is done manually or automated by non-TJS mechanisms by a number of organizations. In this scenario the Netherlands approx. 200 public organizations still have data professionals that need to perform data joining activities. In figure 4.5 the TCO-model variables of Scenario B, The Minimum-TJS scenario, are summarized.

TCO-Model variable	Estimated value	Explanation
#Organization	200	There are approx. 800 public organizations in the Netherlands of which 50% have data professionals who perform data joining activities with spatial data. Because TJS is deployed at the main infrastructure
#TJSServer	2	There are two TJS server infrastructure components. One for serving all spatial framework data, the second node is serving all attribute data of the Statistics Netherlands.
#TJSClient	2	There are two client applications that can process request and responses of TJS services.
#TJSProviders	2	Two TJS data providers (2 nodes) offer their data through TJS services.
AvgHoursTJSData-Management	30	The estimated average hours for TJS data management are 30 hours per year.

Table 4.5 Estimated values for Scenario B, the Minimum-TJS scenario

### Scenario C The Maximum-TJS scenario

This is the maximum scenario. There are 15 TJS infrastructure nodes available at different organizations. These are main TJS nodes in the data infrastructure in the Netherlands:

1. The data offerings of the Statistics Netherlands, National health Institute, large municipalities, ...
2. The geospatial framework node where all main spatial data is offered.

Ten clients also exist from the client perspective. Several other TJS client applications exist in addition to the Statline client application of the Statistics Netherlands and the proprietary Swing application that is used by many regional and local public organizations (provinces and municipalities).

In this scenario, some data joining of spatial and attribute data is still done manually or automated by non-TJS mechanisms by a number of organizations. In this scenario, only approx. 50 public organizations in the Netherlands still have data professionals that need to perform data joining activities.

The TCO-model variables of Scenario C, The Maximum-TJS scenario, are summarized in figure 4.6.

TCO-Model variable	Estimated value	Explanation
#Organization	50	There are approx. 800 public organizations in the Netherlands of which 50% have data professionals that perform data joining activities with spatial data. Because TJS is deployed at the main infrastructure
#TJSServer	15	There are 15 TJS server infrastructure nodes. One for supplying all spatial framework data, the other nodes supply attribute data of different statistics data providers.
#TJSClient	10	There are 10 client applications that can process request



		and responses of TJS services
#TJSProviders	15	Two TJS data providers (15 nodes) offer their data through TJS services.
AvgHoursTJSData-Management	60	The estimated average hours for TJS data management are 60 hours per year. In the Maximum-TJS scenario, more datasets are offered by the data providers through TJS services than in the Minimum-TJS scenario.

Figure 4.6 Estimated values for Scenario C, the Maximum-TJS scenario

#### Yearly reduction in developments costs

The development costs of TJS client and server software are considered short-term investment costs. It is assumed that the first three years these developments costs shall be taken into account. After three years, all development costs are considered zero. The TCO-model calculations are performed for a time span of 10 years and therefore take into account these short-term developments costs.

#### TCO-model constants

The TCO-model for the Netherlands contains 5 variables and 10 constants. The constants and their estimated values are summarized in figure 4.7, including a short explanation of the estimation of the constant values.

TCO-Model constant	Estimated value	Explanation of the estimated value
#Users	3	It is estimated that in each organization three professional users perform manual data joining activities.
AvgHourRate	80	The average (gross) hour rate for data management and infrastructure maintenance in the Netherlands is estimated at 80 Euros.
#HoursforDataJoins	10	Number of hours average per year spent on manual (non-TJS) data joining activities by one user. This is a rough estimation based on the estimation of experts in statistical data processing.
CostsTJSServerDevelopment	75,000	Cost of TJS server development in Euros. The costs estimated for TJS server software development are between 50,000 and 100,000 Euros as estimated by software developments experts.
CostsTJSClientDevelopment	25,000	Cost of TJS client development in Euros. The costs estimated for TJS client software development are between 20,000 and 30,000 Euros estimated by software developments experts.
AvgHoursInitialTJSSetup	160	This constant is the average number of estimated for initial setup of TJS server software. Infrastructure specialist need to get acquainted with TJS setup and tuning TJS deployments. This is estimated 20 days on yearly basis by software developments experts.
AvgHoursforTJSMaintenance	80	The average estimated number of hours for TJS Server maintenance per year is 80.
CostsHosting	12,000	The average costs of server network (datatraffic) and hosting per year.
CostsTraining	12,500	Average cost in Euros for user training for each TJS client application.

Table 4.7 TCO-model estimated values of constants



## 4.5 Results of TCO-analysis for the Netherlands

The results of the TCO-model runs for the Netherlands are shown in the three figures below. The model variables and assumptions are described in the previous paragraph. In figure 4.8, the estimated yearly costs in Euros of Scenario A, B and C in the Netherlands are shown.

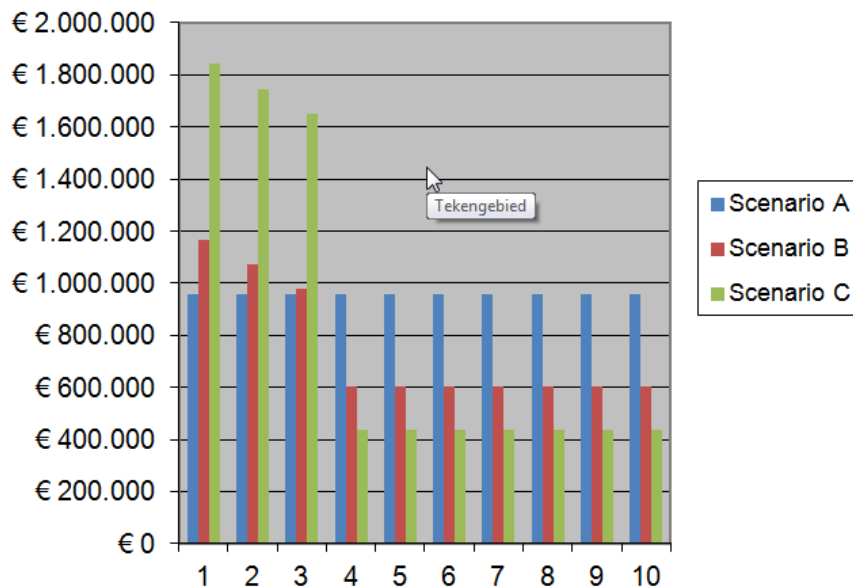


Figure 4.8 Estimated costs in Euros of TJS Scenario A, B and C in the Netherlands

In figure 4.9 the estimated cumulative costs in Euros of Scenario A, B and C in the Netherlands are shown over a period of 10 years.

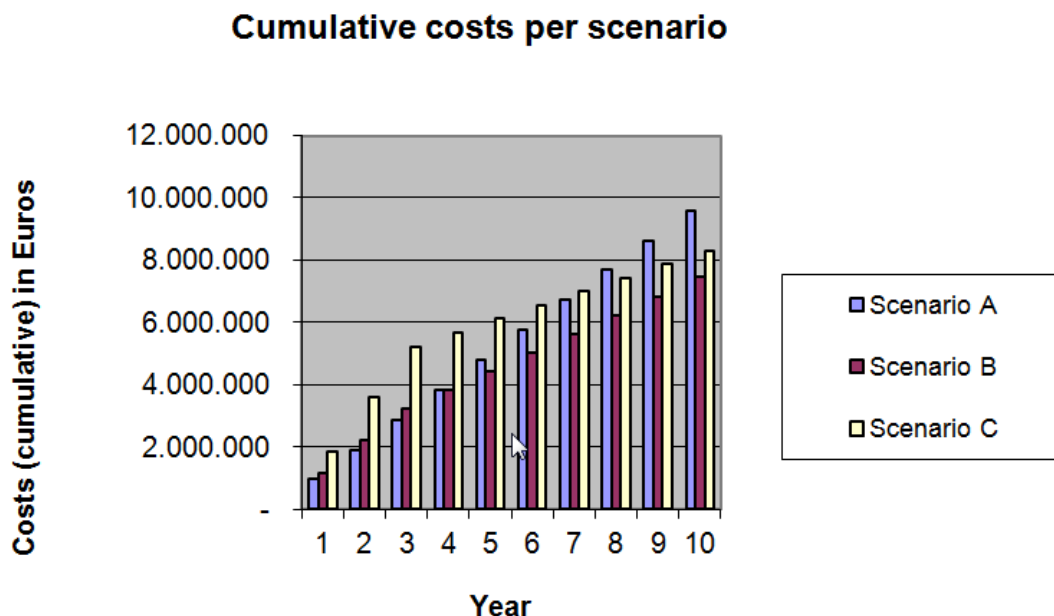


Figure 4.9 Estimated cumulative costs in Euros of TJS Scenario A, B and C in the Netherlands

In figure 4.10 the cumulative cost yields in Euros for Scenarios B and C with regard to Scenario A - the NO-TJS scenario - for the Netherlands over a period of 10 years are shown.

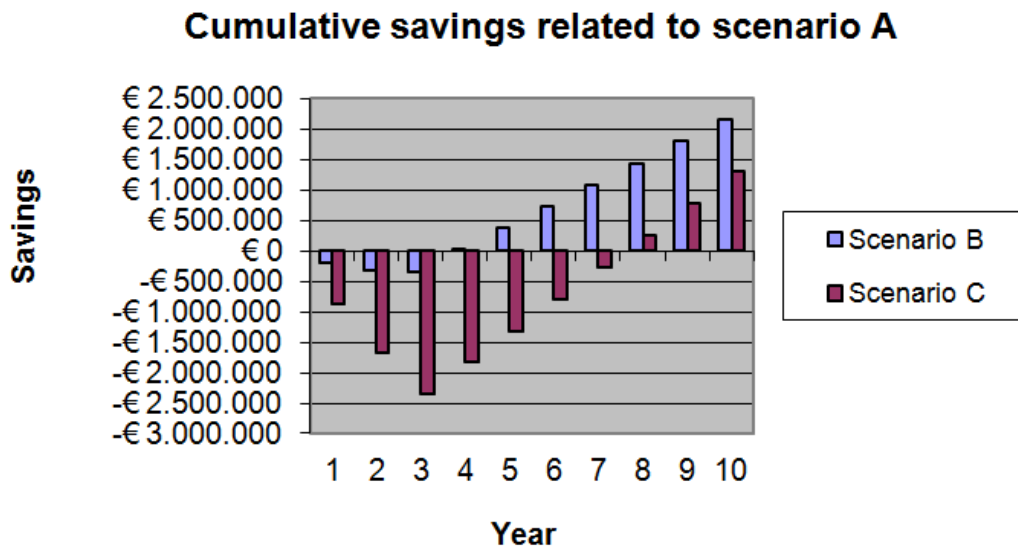


Figure 4.10 Estimated cumulative savings of TJS Scenarios B and C with regard to Scenario A

## 4.6 Benefits of TJS

The costs-benefits, as estimated in the previous paragraph, are due to savings by (in)efficiencies in the data management process in organizations. The rather labor intensive, manual data management activities and processes of joining geographic and tabular (attribute) data – often statistical data – are no longer necessary if the TJS approach is adopted. Monetary benefits that are the result of the TCO-model estimation show – although based on first rough estimations – cost savings in the case of The Netherlands in a long-term period of 10 years.

Besides these costs benefits, additional non-monetary benefits occur as well. In two expert workshops in the Netherlands (see annex B for the list of participants), several non-monetary benefits were expressed by expert opinions. This resulted in the following list of indirect and qualitative benefits that experts expect of the adoption of the TJS concept in The Netherlands:

1. Indirect efficiency benefits and more quality in output. The TJS concepts need a framework where geographic and tabular (attribute) data meet with 'keys' to join. This framework requires a common agreement between many organizations on the spatial and tabular (attribute) data offered. Such an (semantically standardized) agreement or standard approach will increase the quality of data management in several ways;
2. Societal benefits through better use of spatial statistics. Experts expect that more users will get access to spatial statistical data through TJS adoption. As an outcome, the increased use of spatial statistical data will improve policy and decision making in several application domains;
3. The use of TJS as an international open standard increases and improves the quality of the exchange of spatial statistical data between organizations. It offers faster data delivery and a better time-to-market, experts expect. Furthermore, the use of open standards for data delivery will have a positive impact on the independency for client software suppliers;
4. The use of service oriented architecture concepts of TJS will result in imago benefits for organizations. As service oriented architecture concepts offer modern data exchange and data integration mechanisms. Also, with service oriented architectures, possibilities for monitoring data access and use are available. This will offer organizations the possibilities to improve services and service delivery.

## Chapter 5 Concluding remarks

The TJS concept is developed for joining tabular data and geographic data. Joining tabular data and geographic data is necessary to show tabular data on a map or to create a new dataset for further processing and data analysis, like for mapping statistics, thematic mapping and spatial analysis.

The Table Joining Service or TJS standard offers a standardized web service interface that enables the automatic, service oriented joining of tabular and geographic data across the web, while keeping the data distributed at the data providers source location. The TJS specification - a standard of the Open Geospatial Consortium (OGC) - is relatively new and only very few ready-to-use implementations of server-side and client-side software exist. Most of them offer partial support for TJS. There are some examples though and some are under development.

The benefits of TJS follow mostly from its open, standardized interface, allowing for joining data from various, distributed data sources. There are other approaches for joining data, but none of these are standardized and (fully) web based. The concept of TJS fits the current practice of using (geospatial) web services, as is implemented for the European INSPIRE Directive (European Law) for example. TJS allows for implementations to (optionally) make the joined data directly available through commonly used OGC web service standards, as WMS and WFS. Therefore, TJS has a great potential for use in conjunction with existing spatial applications and clients used in spatial data infrastructures, like INSPIRE.

The costs-benefits of the adoption and implementation of TJS bringing tabular and geographic data together are roughly estimated for the Netherlands. The cost-benefit analysis based on a Total Cost of Ownership method (TCO), shows that –although roughly estimated – the costs of TJS adoption will have a positive business case on the long-term (after 4 years for the minimum and after 8 years) for the maximum TJS scenario) from the costs perspective. Besides the costs perspective, several potential, indirect benefits have been identified during expert workshops.

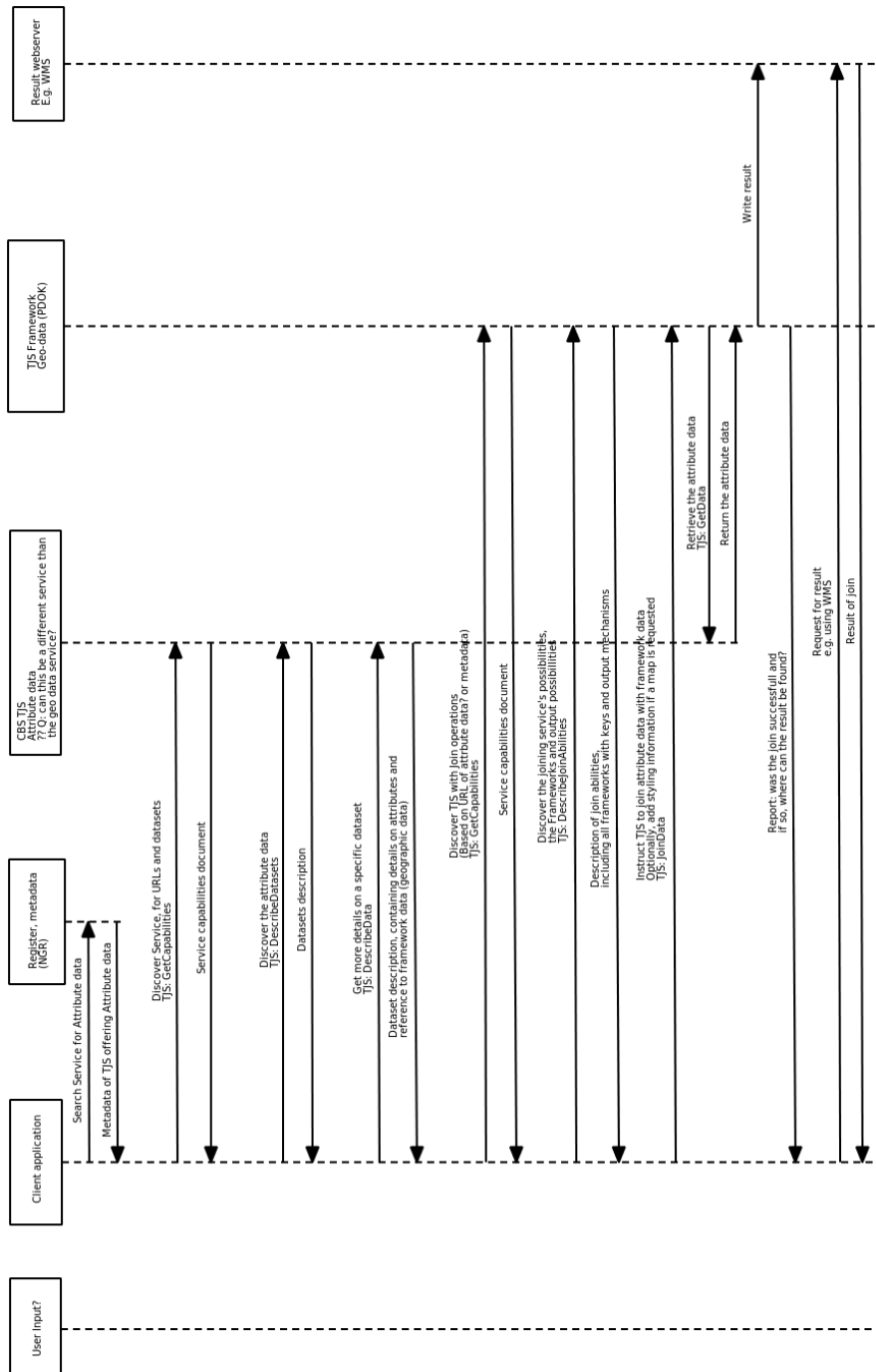
So, the major conclusion is that TJS has the potential to replace the ‘manual’ data joining operations in daily practice of data management for thematic mapping and spatial statistics. However, in order to stimulate the large scale adoption and implementation there is a need to explore the possibilities and cost benefits of the TJS concept in more detail and put TJS in practice by implementation of TJS infrastructure nodes. It is therefore recommended to:

1. Test the TJS concept in a real-world case study; e.g. to implement the minimum scenario for INSPIRE Health statistics theme in the Netherlands. This requires cooperation between several INSPIRE data providers in the Netherlands, like Statistics Netherlands and the Kadaster (“Publieke Dienstverlening op de kaart”). Also, a minimum scenario case study offers the possibility to validate the TCO-model (see point 3);
2. Establish a national spatial (identifier) framework in close cooperation and based on common agreement (data concepts and models) with public sector organizations. The spatial (identifier) framework is necessary to establish a commonly agreed situation for organizations to join their tabular data to geography, knowing that it can rely on a stable spatial (identifier) framework;
3. Further refine the presented TCO-model with real world model values based on the case study (point 1) and recalculate the TCO-model in order to have a more reliable business case for the Netherlands and as a general format for individual organizations to develop their business cases. Then, for individual organizations it is also recommended to include the characteristics of the ‘manual’ data joining practice in more detail in the TCO-model. This would lead to a better business case that supports effective decision making on TJS investments in an organization.

## Annex A – Use case scenario (UML) interaction diagram

This annex contains an UML interaction diagram that shows the interaction between a TJS client and TJS servers for joining data. The diagram is limited to one typical scenario. This scenario describes how a client joins tabular (attribute) data from organization A , with spatial framework data of organization B by:

1. Finding the service through metadata stored in a metadata registry;
2. Accessing TJS with the attribute data;
3. Finding the framework data to join the attribute data to;
4. Joining the data;
5. Fetching the result that the Joining TJS writes to the result web server.



## Annex B – List of experts consulted

The following experts were consulted for this TJS study:

Pieter Bresters	Statistics Netherlands
Niek van Leeuwen	Statistics Netherlands
Dennis Ramondt	Statistics Netherlands
Michiel Eijkhout	Statistics Netherlands
Dick Windmeijer	Statistics Netherlands
Henrico Witvliet	Statistics Netherlands
Erik Fokke	Statistics Netherlands
Caroline van Houwelingen	Statistics Netherlands
Chantal Melser	Statistics Netherlands
Johan van der Schuit	PBL
Maarten van Ham	Technical University of Delft
Jan Jaap Harts	University of Utrecht
Gerard Hazeu	Alterra
Peter Louwerse	Ministry of Infrastructure and Environment (I&M)
Tom Merkx	VSO
Matthijs van Niel	Province of Zuid-Holland
Ioulia Ossokina	CPB
Martin Ruiterman	ABF Research
Lotte Vermeij	SCP
Erik van der Zee	Publieke Dienstverlening op de kaart (PDOK)