Marie Senumstad Sagedal

What are the Inconviniences of Storing Data Centrally?

Information on Central Geospatial Databases in Norway

Trondheim, December 2017

PROJECT THESIS: TBA4560

Main supervisor: Professor Terje Midtbø

Co-supervisor: Lars Eggan, Norconsult Informasjonssystemer

Faculty of Engineering Science and Technology

Department of Civil and Transport Engineering

Norwegian University of Science and Technology (NTNU)

NTNU – Trondheim
Norwegian University of
Science and Technology

Preface

TODO

This project assignment is written for the division of Geomatics at the Norwegian University of Science and Technology (NTNU). It is part of the study program Engineering and ICT, and was written in the autumn of 2017.

I would like to thank my advisers Lars Eggan (Norconsult Informasjonssystemer) and Terje Midtbø (NTNU).

Trondheim, 2013-12-20

Marie Senumstad Sagedal

Summary and Conclusions evt Abstract

TODO

Contents

Chapter 1

Introduction

For nearly twenty years the Norwegian cadastre, *Matrikkelen*, has had a centralised production system - will this work for other geospatial systems in Norway as well? Today the management system of the basic map data in Norway is being transferred from a local production at the municipalities - where the updated map data were sent to the central data store once or twice a year - to a system where the municipalities directly update the map data to the data store managed by the Norwegian Mapping Authority, *Kartverket*.

(Litt mer funfacts her)

The organisation of the paper is as follows: Initially there are two theoretical chapters presenting the two main registers of geospatial information in Norway and the techniques for updating and synchronise to and from those. This is followed by a brief presentation of the managing systems in the neighbouring countries Sweden and Denmark. The following chapters discuss the benefits and disadvantages of updating data into a central data store, and includes as well a brief discussion on generic versus platform dependent API's. The paper ends with a conclusion whether there are some disadvantages to centralising the production of the basic map data.

The main objectives of this project are

1. This is the first objective
2. This is the second objective
3. This is the third objective
4. More objectives

The rest of the report is structured as follows. Chapter 2 gives an introduction to ...

r det de siste versjonene av figurer som brukes?

sjekke over referanser

Er det sections og subsections som bør slåes sammen, ikke stå for seg selv? og motsatt

e

Chapter 2

The Main Geospatial Data Systems in Norway

Two of the main registers of map- and real property data in Norway are the *Felles Kartdatabase* (FKB), the primary map data, and *Matrikkelen*, the cadastral data store. The production, maintainance and updating of both are done by *Kartverket*, the Norwegian Mapping Authority, in collaboration with the members of the Geovekst project. Geovekst is a geodata collaboration between different Norwegian public agencies, such as Kartverket, the municipalities and Statens vegvesen (Kartverket, 2017c). By carrying out joint mapping projects Geovekst establish and maintain a common set of map data in Norway.

2.1 *Sentral Felles Kartdatabase* - The Central Map Data Store

To enhance the efficient integration, distribution and transport of the FKB-data, the implementation of *Sentral felles kartdatabase* (SFKB), the centralised map data store, started in October 2016. This is a system that centralises the management of the primary map data (FKB) in Norway (Kartverket, 2017d). The FKB data is the most detailed map data, ranging on a scale from 1:500 to 1:30000. They are all on a vectorised form, and used in e.g. production of technical maps and geographical analysis. Some examples of the FKB data are buildings, roads and water.

Through the SFKB system public agencies, such as the municipalities, are able do directly update the FKB-map data into a central map data store. This ensures the users of the FKB-map data access to fresh and quality assured data at all times. Earlier the updated map data was stored locally at the municipalities, and sent to Kartverket only 1-2 times a year. As of November 2017 there are five municipalities (Bergen, Bærum, Oslo, Stavanger and Trondheim) that are not members of Geovekst collaboration. These five municipalities have the opportunity to update their FKB-data directly to the central registry, but it is only Bærum that have chosen to do so (Kartverket, 2017d). As a pilot Trondheim municipality, using geosynchronisation (see figure 3.1.2), works as a provider of the FKB offering map data to subscribers (Sæther, 2016; Sandal, 2016). The remaining three municipalities have chosen to keep and update local copies, and send those copies to Kartverket regularly as was the earlier standard.

2.2 *Matrikkelen* The Norwegian Cadastre

The objective of the cadastre, in Norwegian called *Matrikkelen*, is to serve as a registry of cadastal units and property in Norway, i.e buildings, parcels and addresses. There are several usages of the cadastre. For the municipalities and the local administrations the cadastre is an important tool for e.g. land use planning and collection of fees; for the government it is a tool for deriving statistic and for the private sector it gives valuable information for the property marked amongst other things (Mjøs, 2002). The server of the cadastre runs centrally at Kartverket.

2.2.1 Data Flow of the Two Systems

The management systems of the *SFKB* and the cadastre have the same dataflow as is illustrated in figure 2.1. *The municipalities* update the central data stores through an application programming interface (API), and *the central management*, with the main responsibility for the content of the database (Kartverket), controls the data stores and does periodical updates on them. With a synchronisation API (further explained in chapter 3), local copies at the municipalities and distributional copies for all end-users of the mapdata will be kept up-to-date (Kartverket, 2015).

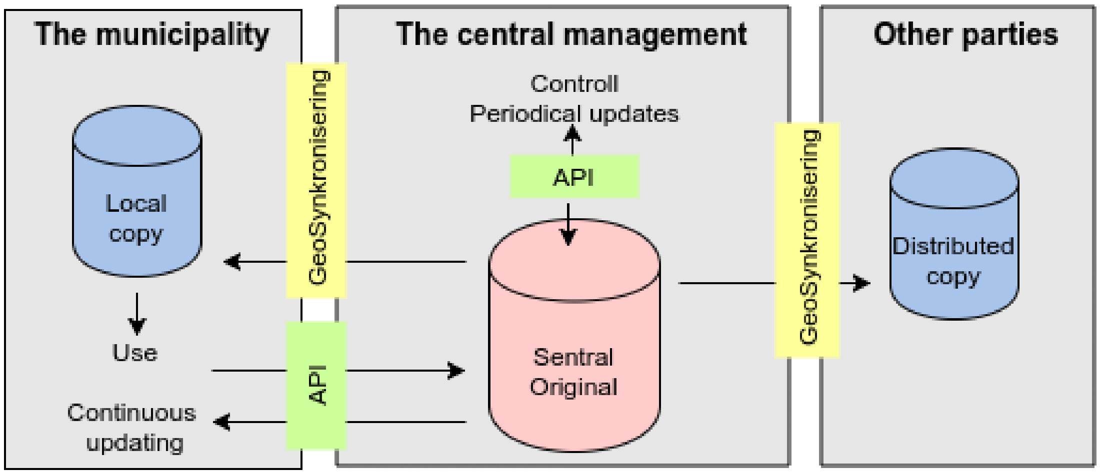

Figure 2.1: Illustration of the concept of data distribution in the systems of both SFKB and the cadastre. The figure is adopted from Kartverket (2015)

Chapter 3

Techniques for Updating and Synchronising Geodata

When submission of new data and modifications on existing data is done for the cadastre and basic map data in Norway, one of the criteria is that all the databases with the same content are updated regardless of where the update is done. What is needed is a geosynchronization service. The synchronization of the FKB data is done with *geosynchronisation*, and updating through the *NGIS-API*, and the cadastre data through the *MatrikkelAPIs*. This chapter will present these three techniques for getting access to and synchronizing geospatial data.

3.1 Quadri Map Server (QMS) and Its Synchronising Services

The *Quadri Map Server* (QMS) is the server of the of the SFKB system. QMS has a client-server architecture; the servers distribute the data and the clients are able to both fetch data and perform services on them (Norkart AS, 2010). The structure of the QMS system is as Kartverket (2017d) illustrated it in Figure 3.1: the *archives* (data stores) are used for storing the FKB-data, the *portals* defines the clients and authorizes users for different tasks, and a task is defined as access to an archive. The data stored in, and uploaded into, QMS is defined by the *object catalogue*, and the service of translating logical, humanly meaningful, names of the distributing servers into unique identifiers (UUID) is the *name server*. All the objects in the archives are given UUIDs, and no other spatial object carries the same[1], simplifying search and linking processes.

[1] For example, the number of random version 4 UUIDs which need to be generated in order to have a 50% probability of at least one collision, is 2.71 quintillion. This is computed as follows: $n \approx \frac{1}{2} + \sqrt{\frac{1}{4} + 2 \times \ln 2 \times 2^{122}} \approx 2.71 \times 10^{18}$ (Eggan and Pedersen, 2017)

The *NGIS-API* is the application programming interface for storing spatial data into QMS.

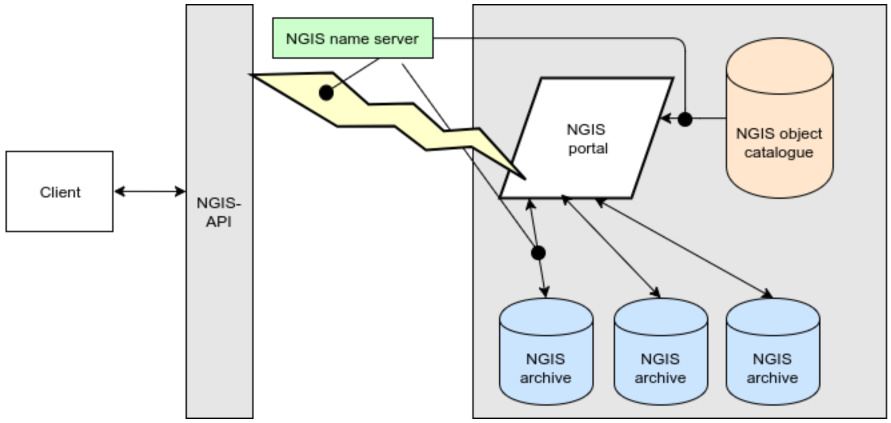

Figure 3.1: The structure of the Quadri Map Server (QMS) system. For updating the data in QMS the NGIS-API is used. The main components of the system are portals, object catalogue and archives. It is possible to use geosynchronisation (chapter 3.1.2) to manage the data in QMS, in addition to the NGIS-API. Figure adopted from (Kartverket, 2017a)

3.1.1 NGIS-API

When editing an existing object in QMS, e.g. a building or the land use of an area, the object gets locked in the archive, and the user edits the object locally. When the updating is done the user has to send the object back, together with objects that have been edided, added or deleted, to it's archive. When the objects are locked in the archives, other users can not edit the same object, but they might look at it. This process is called long transactions (Kartverket, 2017a).

3.1.2 *GeoSynkronisering* - the Geosynchronisation Standard

Geosynchronising (*GeoSynkronisering*) is a Norwegian standard for synchronising geographical information across computer systems, and is a provider-subscriber system as illustrated in figure 3.2 (Kartverket, 2013). The provider updates the database with new data, and subscribers are allowed to update their databases with the changes done by the provider. The SFKB works as a provider: if data is changed or added in QMS the features will automaticly be geosynchronised to all subscribers. Subscribers in the SFKB-system are e.g. the municipalities and GeoNorge. GeoNorge is the web page where public agencies and institutions as well as private actors can get map data and other geospatial information for Norway.

The project of making a standard interface for geosynchronisation services in Norway was carried out in 2012, and was a collaboration between Kartverket and system providers in Norway. This Norwegian standard is based on the concepts and methods from international standards, i.e. ISO 19100 Geographic Information/Geomatics and Open Geospatial Consortium (OGC) (Kartverket, 2013).

Figure 3.2: The geosynchronisation concept, adopted from (Kartverket, 2013, p. 16)

The data transfer with geosynchronising (figure 3.3) is a set of transactions of changed features on the GML-format and their respective changelogs are specified by the WFS-T. GML (Geography Markup Language) is a open interchange format for geographic transactions and a modelling language for geospatial systems (OGC, 2017).WFS-T (Transactional Web Feature Service) allows creation, deletion, and updating of geospatial features on the web (OGC Network, 2011) and will be further described in chapter 3.3.1.

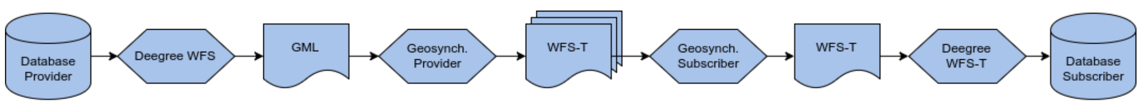

Figure 3.3: A detailed description of the Norwegian geosynchronisation process, adopted from Eggan and Pedersen (2017). MÅ FORKLARE DENNE. Deegree is a implementation of the Web Feature Service

Geosynchronising is a service that permits databases with geospatial content to synchronize across different platforms and system solutions (Kartverket, 2013), as opposed to the NGIS-API that is a platform dependent API due to its criterion of using c++ programming language and that the developer using it needs to provide the compiler Microsoft Visual Studio (Kartverket, 2017a; Norkart, 2011).

3.2 The central cadastral server and its APIs

There are several services provided by the cadastre, Matrikkelen; the *cadastral server* serves all clients, both editing and viewing (figure 3.4). The logic of the cadastre server includes management and assembling of cadastre information, as well as control and validation of the business rules of it. The only way to get access to the data on the *cadastre database* is through the *MatrikkelAPI* (chapter 3.2.1) that is on the cadastral central server (Matrikkelavdelingen, 2017, p. 338). Submissions and modifications on the cadastral data are done to the centralised data store directly, as for the new SFKB system. When wanting to keep a local cadastre copy, the Matrikkel system provides a *Changelog API* for synchronizing changes on the central data store. See figure 3.4.

The physical architecture of the central cadastral server is clustered, meaning there are several physical servers, but for the clients they will appear as one.

3.2.1 The MatrikkelAPIs

The MatrikkelAPIs is application programming interfaces for extracting data, in both small and large scale, of the cadastral data.

There are three MatrikkelAPIs. The first one is for the clients with access to updating the cadastre data - the *Editing API*, the second one the *Viewer API* includes services for inspecting and reading the data and the *Changelog API*, as stated earlier, provides services for fetching data changes - primarily for the local copies at the municipalities, but also for external registers. The Viewing API support other services as well: web services as WMS that deliver raster maps with addresses as points and buildings and areas for land parcels as well as WFS delivering vectorised map and properties for addresses, buildings and land parcels.

There are several types of clients to the cadastre system: the *updating*-, *viewing*- and *retrieval clients* as well as *other systems*. The updating clients support inspection and updating for all information on the cadastre, the viewing clients supports inspection of the data, as well as viewing the cadastral data joint with data from external systems, and lastly the retrieval clients supports updating external municipality registers (KommuneGAB). Other systems using the cadastre system are mainly systems that uses cadastral information for other software solutions, e.g. a municipal specialized system (Matrikkelavdelingen, 2017, p. 337-338).

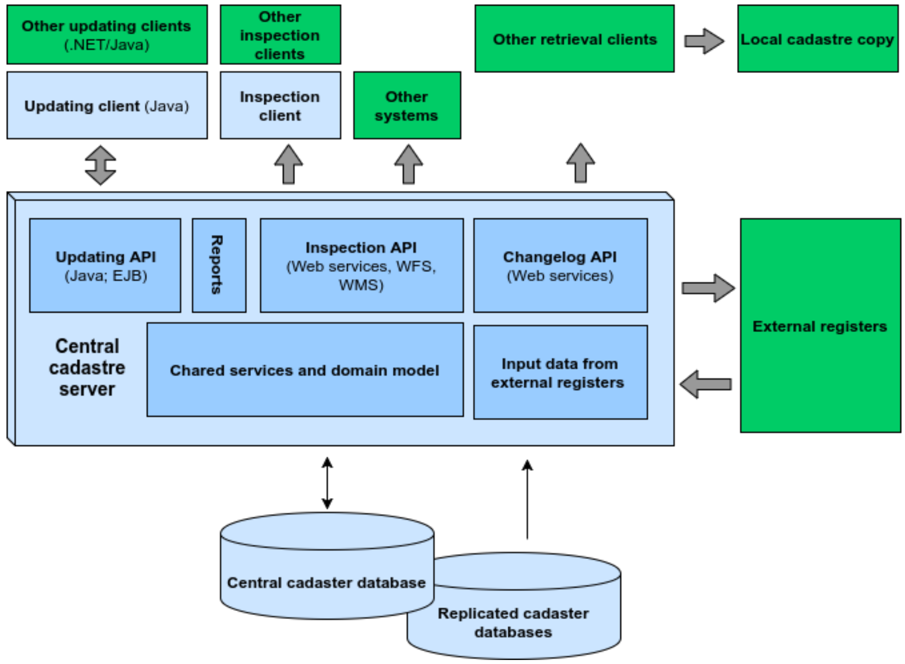

Figure 3.4: The cadastre system, *Matrikkelen.* Solutions for databases through API, clients and relations to other information systems. Adapted from (Matrikkelavdelingen, 2017, p. 337)

3.3 Other Generic Services

There are several web services, ways to access web-based geospatial data. These technologies have been created to facilitate the exchange of geospatial information and to find new ways to communicate geodata and to redefine the work flow of the geospatial analysis. In the following subchapters some of them are presented.

3.3.1 Standard Spesifications from the Open Geospatial Consortium)

To get a higher interoperability in the Geographical Information System (GIS) community the Open Geospatial Consortium (OGC) has created standard specifications for data sharing, processing, retrieval, visualisation, content and cataloguing (Giuliani et al., 2013).

The Web Feature Service (WFS)

The Web Feature Service (WFS) is a standard geodata extracting service for describing data manipulation on feature level (Peng, 2005; {Norge digitalt}, 2014). The specification defines interfaces required to support transactions and query operations on geospatial features over the internet.

The GML-format is the defacto standard transferring format, but WFS supports other formats as well (Eggan and Pedersen, 2017). By using GML for the exhange of geospatial data, interoperability between the heterogeneous system is provided (Yao and Zou, 2008).

Whereas WFS allows queering and retrieval of features, transactional Web Feature Service (WFS-T) permits the user to create, delete and update features (OGC Network, 2011),

The Web Mapping Service (WMS)

The Web Mapping Service (WMS) is another standad from OGS for exchanging geograpichal information over the web. When using the Web Mapping Service(WMS) the user is given an image of a map that cannot be edited or spatially analysed.

3.3.2 Representational State Transfer (REST)

The definitions of Representational State Transfer (REST) are many (Fielding et al., 2017; Richardson and Amundsen, 2013), this paper uses the description by Fielding and Taylor (1995): *REST is a coordinated set of architectural constraints that attempts to minimize latency and network communication while at the same time maximizing the independence and scalability of component implementations.*

REST is based on the HTTP and HTML standards, and provide simplified calls to services through HTTP. It is only data structures and the transferring of their state that is dealt with in a RESTful service **?**. The state transferring calls are POST, GET, PUT and DELETE, which receptively creates, reads, updates and deletes resources in the RESTful service. A resource is identified and resolved with a URL.

3.3.3 SOAP

Simple Object Access Protocol. Tjenestegrensesnitt med mekanismer for å hente ut objekter fra en web-basert tjeneste. Ofte omtalt som Web Service (den opprinnelige web servicen).

Chapter 4

Equivalent systems in the other Scandinavian countries

As each country normally operates with separate file formats for the geospatial data Frenvik (2017) Infrastructure for Spatial Information in the European Community (INSPIRE) is the European collaboration for a common standards for describing and sharing spatial data INSPIRE (2017). In the following sections there will be brief explanations of the geodata systems in Sweden and Denmark.

4.1 Sweden

The mapping authority in Sweden is Lantmäteriet. Lantmäteriet grant maps, aerial images and other geographical information about Sweden, together with central and local authorities. Students, scientist or employees of universities or other educational and cultural institutions can get the geodata provided by the Lantmäteriet for free.

https://www.l
and-
geographic-
information/

4.1.1 The geodata cooperation, *Geodatasamverkan* evt geodataportalen

https://www.g

geodatas

4.2 Denmark

Geodatastyrelsen "Kartverket" http://gst.dk/om-os/ Datafordeleren http://sdfe.dk/hent-data/datafordeleren/ Grunndataprogrammet Sara Bjerre på sarbj@digst.dk Matrikkelen http://gst.dk/matriklen/ejendomsdataprog udvidelse/

4.2.1 INSPIRE

Chapter 5

Directly Updating versus Local Updating

The benefits of the new SFKB sytem are numerous, as compared to local storage; freshly updated data for the end-users of map data, and the changes will carry through to *all* end-users, as opposed to local storage, which only applies to the users of that specific municipality (blir det riktig å anta at kommunene distribuerer fra sin lokale kopi mens de har de dataene?) (Dontigney, 2017). Another benefit is a more efficient and effective validation process, as the requirements are defined and validated one place and every change to the central store is done by the agency where the change actually occurred (Kartverket, 2017b).

There are several authors that agrees on the befits, the Canadian geosynchronisation project was according to Reichardt (2012) providing the most current and reliable data, consequently avoiding unnecessary versioning and keeping the duplication of data at a minimized level.

Centralisation of data management within different scientific fields where sharing and analysing data is major tasks have been important for (...), and entering data directly into databases In the paper *A Generic framework for synchronidistributed data management in archaeological related disiplines* Lohrer et al. (2016) discuss the advantages and disadvantages of centralised data management within the archaeological field

Spatial Data Infrastuctures

According to Peng (2005) sharing data on file level is causing latency in update of the data, and it is difficult to deduce a generalised standardisation for all geodata, and reaching consensus might be difficult. Spatial data infrastructure (SDI)

Spatial data infrastructures seek to facilitate the access and integration of geospatial data coming from various sources. To achieve this objective, systems must be interoperable. Giuliani et al. (2013) With the standard for exchanging geospatial data in Norway, SOSI (Samordnet Opplegg for Stedfestet Informasjon),

One of the disadvantages of updating directly into a data store is that, as was stated earlier in this chapter, that it may be difficult to obtain a generalised data format and that validation thus will be difficult. As the Norwegian geospatial data have been standardised through the SOSI-standardisation for 30 years this will not be a problem (formuler annerledes). Validation problems wil mainly be because of humaly errors.

- definitions of centralized and decentralized system
- pros and cons
- Kartverket's web page of QA
- check reference list: `https://en.wikipedia.org/wiki/Centralized_database`
- Breslow (2004) WAN
- coherency
- consistency

5.0.1 Previous work - generic API's

This chapter will present papers concerning generic API's and framwork depending API's

Giuliani et al. (2013) wanted in their paper to benchmark and evaluate the quality of the web services WFS and WCS (Web Coverage Service). They conclude that amongst other things the OGC specifications (WFS, WCS) are providing interoperable access to data in an efficient and timely manner, and that the specifications are not convenient for transferring large volumes of data.

—-Kan jeg rerferere til samtale?—– - zipped GML da gjør det ikke noe at det er store filer, smat større datamaskiner

Chapter 6

Conclusion

Fordeler og ulemper med sentral lagring styrker og svakheter

- Matrikkelen har fungert fint til nå
- oppsummere noe fra litterature review
- version-control issues as well as inefficient collaboration for storing local copies
- keeping service databases up-to-date in a multi-provider situation.

Avslutte med:

Further investigated Future work : This includes tracking latencies, bottlenecks, and errors that may negatively influence its overall quality. Giuliani et al. (2013)

The process of storing map data locally and later send those data to *Kartverket* was perhaps unnecessarily time-consuming and inefficient. As this paper has presented there are a few drawbacks of updating directly to a central data store. On the other hand, the benefits seem to be greater and thus the SFKB-system have a bright future ahead; the *SFKB*-system will reduce operational cost by lessening the import, export, copying and control costs of the data, as Kartverket (2017b) sums it up. This is in addition to improving service delivery for the public, providing it with fundamental and reliable map data freshly updated at all times.

Appendix A

Acronyms

API Application Programming Interface

FKB Felles kartdatabase (The basic map data)

SFKB Sentral Felles kartdatabase (The Central Map Data Store)

WFS Web Feature Service

WMS Web Map Service

Appendix B

Additional Information

This is an example of an Appendix. You can write an Appendix in the same way as a chapter, with sections, subsections, and so on.

B.1 Introduction

B.1.1 More Details

Bibliography

Breslow, N. (2004). Wide Area File Sharing Across the WAN. *Computer Technology Review*, 24(8):33–34.

Dontigney, E. (2017). The Advantages of Centralized Database Storage | Bizfluent.

Eggan, L. and Pedersen, J. (2017). WFS for transaksjoner WFS-T.

Fielding, R. T. and Taylor, R. N. (1995). Principled Design of the Modern Web Architecture.

Fielding, R. T., Taylor, R. N., Erenkrantz, J. R., Gorlick, M. M., Whitehead, J., Khare, R., and Oreizy, P. (2017). Reflections on the REST Architectural Style and " Principled Design of the Modern Web Architecture " (Impact Paper Award). *Proceedings of 2017 11th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering*.

Frenvik, A. (2017). Det åpne geodata-økosystemet. *Kart og Plan*, 77(5003):108–120.

Giuliani, G., Dubois, A., and Lacroix, P. (2013). Testing OGC Web Feature and Coverage Service performance: Towards efficient delivery of geospatial data. *Journal of Spatial Information Science*, 2013(7):1–23.

INSPIRE (2017). About INSPIRE.

Kartverket (2013). Standarder Geografisk Informasjon -Geosynkronisering Versjon 1.0.

Kartverket (2015). Sentral forvaltningsløsning for primaerdata.

Kartverket (2017a). Informasjon og dokumentasjon om NGIS-API | Kartverket.

Kartverket (2017b). Ofte stilte spørsmål.

Kartverket (2017c). Om Geovekst-samarbeidet.

Kartverket (2017d). Sentral Felles Kartdatabase - Systembeskrivelse.

Lohrer, J.-Y., Kaltenthaler, D., Kröger, P., van der Meijden, C., and Obermaier, H. (2016). A generic framework for synchronized distributed data management in archaeological related disciplines. *Future Generation Computer Systems*, 56(Supplement C):558–570.

Matrikkelavdelingen (2017). Samla systemspesifikasjon Matrikkelens innhold og funksjonalitet Klient Dokumentnummer: MA–nnnn.

Mjøs, L. B. (2002). New Cadastre in Norway. *TS7.12Regional Experience in the Cadastre – Europe II*.

{Norge digitalt} (2014). Veileder for Web Feature Service (WFS).

Norkart (2011). Quadri Map Server - Utviklerhåndbok.

Norkart AS (2010). Quadri Map Server - Overordnet beskrivelse.

OGC (2017). Geography Markup Language.

OGC Network (2011). WFS-Transaction (WFS-T) tutorial.

Peng, Z. (2005). A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *International Journal of Geographical Information Science*, 19(4):459–481.

Reichardt, M. (2012). New OGC standard supports "GeoSynchronization". *GeoInformatics; Emmeloord*, 15(7):1.

Richardson, L. and Amundsen, M. (2013). *RESTful Web APIs*.

Sæther, M. H. (2016). Status geosynkronisering i Trondheim kommune.

Sandal, K. (2016). Sentral lagring av FKB med ISY WinMap Kjell Sandal Norconsult Informasjonssystemer AS.

Yao, X. and Zou, L. (2008). Interoperable internet mapping–an open source approach.(Technical report). *Cartography and Geographic Information Science*, 35(4).