

# An Overview of Norwegian Linked Open Data

~ Applications in Regional Development and Environmentally Friendly Behavior ~

Dumitru Roman  
SINTEF  
Oslo, Norway  
Dumitru.Roman@sintef.no

David Norheim  
Computas AS  
Lysaker, Norway  
David.Norheim@computas.com

**Abstract**— With Norway being one of the few countries outside of the English speaking world with a clear governmental strategy and commitment to open data, combined with one of the highest Internet penetration and mobile access in Europe, it offers interesting opportunities for becoming a great testbed for consuming Linked Open Data (LOD). With this paper we aim at presenting potential applications consuming Norwegian LOD and showing practical benefits of aggregating open data in highly sensitive domains for governments and the general public such as regional development and environmentally friendly behaviour. At the same time, this paper will serve as an overview of the Norwegian LOD as of mid 2011. The proposed applications will not only aim at demonstrating the benefits of the current Norwegian LOD, but will also make contributions to the improvement and extension of the existing data sets.

*Keywords*-Norwegian LOD, data sets, applications

## I. INTRODUCTION

Norway is one of a handful of countries outside of the English speaking world with a clear commitment to open data. Being one of the first countries to implement the PSI-directive as a law in January 2009 [1], each new governmental project is today required to address publication of the data it creates or processes. Simultaneously the community focusing on linked open data is ever increasing, and major governmental agencies are involved in making their data available as Linked Open Data (LOD). However, as we write 2011, few if any, applications are using LOD as their data source. We expect this to change soon.

We believe that the interest in LOD, combined with the highest Internet penetration in Europe after Iceland, and second only to Sweden on mobile internet access in Europe, Norway offers interesting opportunities for becoming a great testbed for consuming LOD data. With this paper we aim at giving an overview of the current status of the Norwegian LOD, and propose applications consuming Norwegian LOD in highly sensitive domains for governments and the general public such as regional development and environmentally friendly behaviour. The proposed applications will not only aim at demonstrating the benefits of the current Norwegian LOD, but will also make contributions to the improvement and extension of the existing data sets.

In the following we provide an overview of two potential applications consuming Norwegian LOD (Section II), and then

present the current LOD in Norway, with a focus on its relevance to the proposed applications (Section III). Section III summarizes this paper.

## II. APPLICATIONS FOR NORWEGIAN LOD

### A. Regional development with LOD

*Why?* Regions are often faced with challenges such as lack of competitiveness and need for regeneration to various political challenges. The long-term trends and effects of development schemas are however often not well understood and readily available to the public. Local news agencies and the general public are often served numbers without being able to look behind the scene. There is a clear need in this domain for support in creating visualization tools over statistical data or aggregated data to follow the long term effects of regional development.

*Who are the stakeholders?* Data-journalists in Norway, both from the broadcasters and news agencies, have requested a tool allowing ad-hoc combinations of datasets in their regional footprint. A tool based on linked data will increase its value as new datasets become available. Data visualization and investigation support should also be available for public use, allowing citizens to drill into public data.

*What and how?* Inspired by Hans Roslings famous presentation tool Gapminder [2], we propose to build an application focusing on innovations and development in various sectors in regions and municipalities. Central Norwegian public sector data owners have opened their data as linked open data during 2010. The Brønnøysund Registry Center<sup>1</sup> has opened the national organization registry, Enhetsregisteret [3]. Others, notably the Norwegian Research Council,<sup>2</sup> have already used the dereferencable URIs from Enhetsregisteret while publishing their own data sets. This case study will require the need to improve the existing company registry service (exporting also NACE codes), and to include geographical information (linking to Kommuneregisteret,<sup>3</sup> the details of municipalities and counties). In addition it requires the use of visualization tools. The successful implementation of

<sup>1</sup> <http://www.brreg.no/english>

<sup>2</sup> <http://www.forskningradet.no/en>

<sup>3</sup> <http://www.kommunenokkelen.no/adresse/side2.do?side=kommuneregisteret>

this application is dependent on the quality of the data and links in the existing Norwegian LOD, and on useful and flexible analytics and visualization of the linked data which will be developed as part of this case study.

### B. Assisting Environmentally Friendly Behaviour with LOD

*Why?* The current Norwegian government has made a point that environmentally friendly behaviour should pay off. However, behaving environmentally friendly is not an easy task even if financial incentives are in place. For example, often the information about public transportation and availability of city bikes is not as easily available when a trip decision needs to be taken. In such situations, there is a need for applications that combine linked open data in the environmental domain as a decision making tool.

*Who are the stakeholders?* Citizens interested in applications assisting them in environmentally friendly behaviour.

*What and how?* In the private sector a number of applications and mobile apps have been created as a result of open data initiatives. In Norway this includes applications ranging from real-time public transportation information [4], snow and ski conditions [5], the presence of electric cars charging stations [6], real time status for city bike stands [7], weather forecasts [8], and many more. Common to all this is that they use only *one* source of open information. The proposed application in the environmental domain will combine the use of transportation (e.g. public transportation, electric cars parking lots, bikes stands) to events (e.g. concerts, art galleries) while including other decision relevant real-time information (e.g. forecast, traffic messages). The application will be made as a mobile app. The overall goal here is to invoke the citizen's own interest in environmentally friendly behaviour. The successful implementation of this application is dependent on extending the existing Norwegian LOD with new data sets and ensuring the quality of the new data sets and their proper linking to the existing Norwegian LOD, as well as on the usability of such an environmentally friendly behaviour application which will be developed as part of this case study.

### III. NORWEGIAN LOD

On January 1<sup>st</sup> 2009 the EU PSI Directive was implemented in Norwegian law. As a direct consequence of this, the Ministry of Governmental Reform and Administration in 2010 announced the development of [data.norge.no](http://data.norge.no) (currently in alpha mode), the equivalent of [data.gov](http://data.gov) and [data.gov.uk](http://data.gov.uk), to be launched during spring 2011. While the portal will be filled with PSI data sets, other data sets will be included such as those provided two central Norwegian LOD projects, Sesam4<sup>4</sup> and Semicolon II<sup>5</sup>, which opened data sets and converted them to RDF with links in-between and to other sources. Of special interest was the national registry for company information, Enhetsregisteret, and the archive for research funding, Prosjektdatabasen, from the Norwegian Research Council.

Figure 1 provides an overview of the existing data sets in the Norwegian LOD and their relationships. The blue-coloured data sets are related to the regional development case study and the green-coloured data sets are related to the environmentally friendly behaviour. The solid arrows between the data sets represent the already existing links between the data sets, whereas the dotted arrows represent the logical connections for which we plan to create the necessary links in the project (as far as the relevant data sets are concerned).

A brief description of the relevant data sets we are considering for the two applications is given below:

- All legal entities are registered in the **Central Coordinating Register for Legal Entities (Enhetsregisteret)** at Brønnøysund Registry Centre. URIs and a RESTful service returning RDF has been created to look up based on organizational number. The service needs to be made searchable by including a RESTful service based lookups on NACE codes and regional codes.
- The Norwegian Association of Local and Regional Authorities (KS) and Statistics Norway maintain a **registry of municipalities and counties (Kommunekatalogen)**. The information translated to RDF as a proof of concepts by the research project Semicolon II, but URIs haven't been made RESTful and outgoing links to e.g. dbPedia has not yet been created.
- The Norwegian Research Council maintains a **catalog of funded projects (Prosjektarkiv)**. This relates instruments and projects to organisations. This catalog has been translated to RDF and outgoing URIs have been created to the Enhetsregisteret. The catalog is hosted on a triplestore supplied by Computas through the SESAM4 research project.
- All the parties are, according to the Political Parties Act, obliged to report their income to a central register. The Ministry of Government Administration, Reform and Church Affairs, maintains a **central register of parties and their income (Partifinansiering.no)**. This data is available in RDF / SPARQL, on Computas triple store and have been assigned URIs, but has not yet been linked to the Kommunekatalogen.
- The Ministry of Local Government and Regional Development have made **election results** from elections available in structured form. These data should easily be translated to RDF with outgoing links to the party register mentioned above and then made available in a triple store (SPARQL).

<sup>4</sup> <http://sesam4.net/>

<sup>5</sup> <http://www.semicolon.no/>

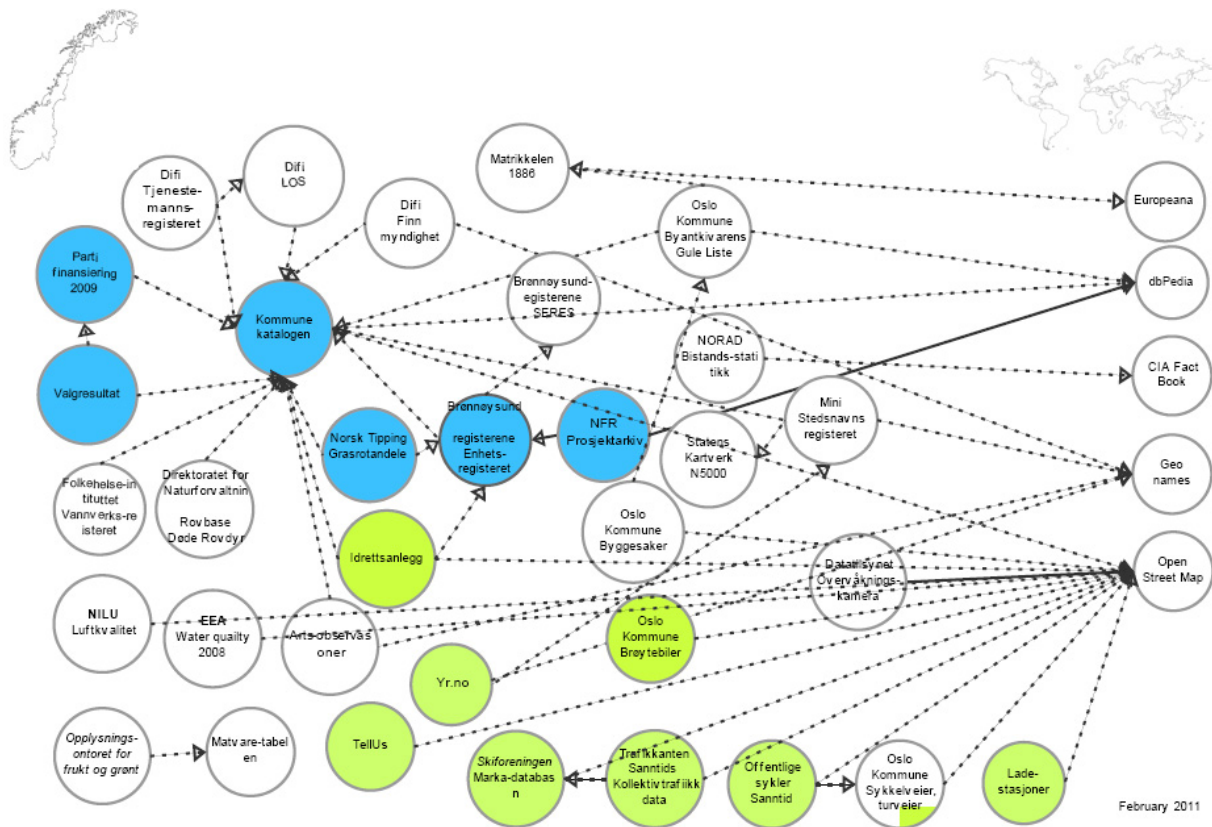


Figure 1. Norwegian LOD.

- Norsk Tipping is a wholly state-owned company under the jurisdiction of the Ministry of Culture with a special role in regulating gaming in Norway. Gaming funds benefit sporting and cultural activities as well as voluntary humanitarian organisations all over the country through the Grassrootshare (Grasrotandelen). Grasrotandelen is a **registry showing organisations that receive funding**. The registry has been translated to RDF but without the links to the organisations involved (in Enhetsregisteret).
- Trafikanten’s main purpose is to provide the public with **up to date travel information for the public transportation system** in and around the counties Oslo and Akershus. The data is about all public transportation, the current location and delays of route. Currently the data is not available to the public as raw data (only through their own app), but there are currently work going on in Oslo Municipality to make it available as RDF.
- Yr.no is the joint online **weather** service from the Norwegian Meteorological Institute (met.no) and the Norwegian Broadcasting Corporation (NRK). They have a XML service available for 7 million places in Norway for free. Work is going on in the Semicolon II project to make this available as RDF RESTful services.
- Markadatabasen is maintained by the Norwegian Ski Association. It's a public source that provides **routes and real time status of preparation of ski slopes**. The data is available in XML. It is also linked to Open Street Map. The data needs to be translated to RDF and made available as a RESTful and SPARQL endpoint. The data will also be linked to public transport routes from Trafikanten.
- Oslo is equipped with around 3000 city-bikes available for the public. A **real-time status of the bike stands** (where bike stands are and how many are available) is developed. ClearChannel is maintaining this on behalf of Oslo Kommune. The data is made available as a XML service. The service needs to be wrapped as RDF and interlinked with bike tracks and tour tracks.
- **Bike tracks and touring tracks** in Oslo is a dataset that is made available from Oslo Kommune. This should be connected to the city-bike stands and made available as a SPARQL/RDF endpoint.
- All national **electric cars charging stations** for electric-cars are made publicly available as a dataset and translated to RDF, and made available by the research project SESAM4 on a triple store. It is now being extended to include real-time availability. This real-time service will have to be interfaced to the charging station locations.

- **Snow ploughing real-time status** is another service that is made available by the Oslo Kommune in near future. This data set could relate to traffic information making it more sensible to take public transportation. The dataset should therefore link to public transportation and made available in RDF, and also connected to the weather information.
- Tellus is a national dataset of **accommodations, attractions and events** produced by the same organisation. It has been made available as a SPARQL endpoint with RESTful RDF lookup by the Sesam4 project. It should be linked to Open Street map and possibly to public transportation.
- The Ministry of Culture maintains a dataset of **all sports arenas** in Norway. This dataset has been converted to RDF and made available as SPARQL endpoint with RESTful lookup in RDF by the Sesam4 project. This should be connected to the organisation in the Enhetsregisteret.

Table 1 below provides information of the above mentioned data sets regarding the owners, the current formats they are available in, where they are made available, the estimated size, the quality of data, and the case study where they will be applied.

**Table 1. Relevant Norwegian LOD data sets for the proposed applications.**

Name	Owner	Format	Hosting	Estimated # of triples	Quality (stars)	Application
Enhetsregisteret	Brønnøysundregisterene	Restful RDF Web service	Brønnøysundregisterene	> 4.500.000	5	App #1
Kommunekatalogen	KS	XML	Univ of Oslo / Semicolon II	Ca 2.000	3	App #1
NFR prosjektarkiv	Norwegian Research Council	RDF	Computas/ Sesam4	Ca 200.000	5	App #1
Valgresultat 2005	Government	RDF	Computas/ Semicolon II	Ca 100.000	4	App #1
Partifinansiering 2009	Government	RDF	Computas/ Semicolon II	Ca 100.000	4	App #1
Grasrotandelen	Norsk Tipping	RDF	Computas/ Semicolon II	Ca 70.000	4	App #1
Trafikkanten sanntid	Oslo Kommune	XML, JSON	Oslo Kommune or Computas	Ca 50.000	3	App #2
Yr.no	Met. Inst.	XML	Univ of Oslo / Semicolon II	Ca 700.000.000	3	App #2
Markadatabasen sanntid	Skiforeningen	XML	Oslo Kommune or Computas	Ca 100.000	3	App #2
Offentlige sykler sanntid	ClearChannel	XML	Oslo Kommune or Computas	Ca 10.000	3	App #2
Sykkelveier, turveier	Oslo Kommune	XML	Oslo Kommune or Computas	Ca 10.000	3	App #2
Ladestasjoner (Sanntid)	Ladestasjoner.no	RDF	Computas/ Semicolon II	Ca 1.100	3	App #2
Brøytebiler Sanntid	Oslo Kommune	XML	Oslo Kommune or Computas	Ca 10.000	3	App #2
Tellus	Tellus	RDF	Computas/Sesam4	Ca 600.000	3	App #2
Idrettsanlegg	Ministry of Culture	RDF	Computas/Sesam4	Ca 1.000.000	3	App #2

#### IV. SUMMARY AND OUTLOOK

This paper gave an overview of the Norwegian LOD as of mid 2011, and proposed two applications consuming the Norwegian LOD. The aim of the two applications is to enhance transparency in making factual information available to the public when it comes to regional development, and help individuals to act more environmentally friendly. We outlined the data sources available for building these two applications. The two applications rely on local and national data sets, but are highly generic and could be applicable for other countries as well.

Whereas the data is available, improving its quality is of crucial importance for a successful realization of the proposed applications. In particular, the quality of links between the relevant data sets needs to be ensured. The links should be created with quality information based on the method being used, e.g. lookups in other registries, ontology reasoning, etc, and be put into a separate graph holding this provenance information. In addition, especially for the second application, available open data sets need to be published as RDF in a triple store and an infrastructure for the triple store and RESTful services needs to be available.

#### ACKNOWLEDGMENT

This work is partially funded by the nationally funded Semicolon II project, and the EU funded projects PlanetData and ENVISION.

#### REFERENCES

- [1] Implementation of the PSI-directive: <http://www.lovddata.no/all/hl-20060519-016.html#9> (In Norwegian).
- [2] Gapminder: <http://www.gapminder.org>.
- [3] Enhetsregisteret: <http://www.brreg.no/registrenehets>.
- [4] Trafikkanten sanntid: <http://itunes.apple.com/no/app/trafikkanten-sanntid/id299318111?mt=8>.
- [5] iMarka: <http://itunes.apple.com/no/app/imarka/id344496606?mt=8>.
- [6] LadeNå!: <http://itunes.apple.com/no/app/id329169428?mt=8>.
- [7] Oslo Bysyssel: <http://www.oslobysyssel.no/>, Accessed April 2011.
- [8] Yr.no: <http://itunes.apple.com/us/app/yr-no/id300709016?mt=8>.