# HARNESSING MOBILITY DATA IN CITIES: A CASE STUDY FROM THE BERGEN REGION

Dirk Ahlers[1], Rajendra Akerkar[2], John Krogstie[1], Andreas L. Opdahl[3], Bjørnar Tessem[3], Weiqing Zhang[4]

[1] Dept. of Comp. Science, Norwegian Univ. of Science and Tech., NO-7491 Trondheim, Norway
[2] Western Norway Research Centre, P.O.Box 163, NO-6851 Sogndal, Norway
[3] Dept. of Info. Science and Media Studies, Univ. of Bergen, P.O.Box 7802, NO-5020 Bergen
[4] Telenor, Postboks 800, NO-1331 Fornebu, Norway

***Abstract:*** *Smart cities attempt to use big data, machine learning, and other topical information and communication techniques (ICTs) to improve energy-consumption, mobility, waste management, and other crucial city functions. Many international research projects have been reported but, so far, few of them have addressed mobility in Norwegian cities specifically. This paper reports on a pre-study that focusses on mobility-related data sources in the Bergen region and discusses the needs and opportunities they present. We have identified central actors and the data they own, discussed opportunities and challenges with central stakeholders, developed a taxonomy of data types, and reviewed available ontologies for data integration. We are currently exploring a big-data architecture for harvesting, integrating, and making open mobility data more ready for use through a single-entry point.*

## 1. INTRODUCTION

Smart City is a term and a practice adopted by a variety of governments, local authorities and international bodies all over the world and is often used as a tool for economic growth and global presence for an actor or organization that adopts it (Caprotti and Cowley 2016). According to (Caragliu et al. 2011), a city is smart "when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance." Other definitions also highlight a Smart City as having a better integration of information technology with the organisation of a city and better participation of its citizens and other civic stakeholders (Gil-Garcia et al. 2015, Goldsmith and Crawford 2014, Nam and Pardo 2011). Smart city services can then also be understood as a deep collaboration of citizens and technology (Ahlers et al. 2016).

In recent years, many researchers have focussed on providing the ICT infrastructure, analysis power, and services required to fulfil this vision, as we will review in the following section. Topical themes include ubiquitous computing, the Internet/Cloud of Things (IoT/ClouT), open data, social media, big data analytics, and machine learning/AI. Common goals are to improve energy-consumption, mobility, waste management, and other crucial city functions. In this paper, our focus is on mobility in a wide sense. Many pilot projects have been reported in the international research literature but few of them have targeted mobility in Norwegian cities specifically. However, many Norwegian cities have Smart City initiatives and projects[1] and, in 2017, the Norwegian government launched the Smarter Transport in Norway challenge[2] which aims to develop new, efficient, and environmentally friendly transport solutions in cities and city regions, and which can hopefully stimulate further initiatives.

1 For example: https://www.oslo.kommune.no/english/politics-and-administration/smart-oslo/ ; http://smartcitybergen.no/ ; http://trondheim2030.no/2017/11/16/trondheim-soker-om-a-bli-smart-by/ ; http://triangulum-project.eu/index.php/lighthouse-cities/city-of-stavanger-norway/ ; http://nyby.bodo.kommune.no/ ; https://www.nordicedge.org/ ; https://smartcities-infosystem.eu/search/node/norway ; https://www.ntnu.edu/smartcities ; http://www.innovasjonnorge.no/no/mulighetsomrader/smarte-samfunn-og-byer/smart-city-programs-and-events/ .

2 https://www.regjeringen.no/no/aktuelt/regjeringens-konkurranse-om-smartere-transport-i-gang-100-millioner-kroner-i-premiepotten/id2578517/

Mobility data has been defined as “information about the movement of objects, which includes, at least, location and time information” (Pelekis and Theodoridis 2014), to which we add a prescriptive dimension: *mobility data is data that are about or that can inform the movement of people and objects related to transport*. A particular type of mobility data is *movement data, which is time- and location-stamped data that describe a moving person or object*. Other types of mobility data such as timetables, maps, and information about places add context to the movement data *(mobility-context data)*. In a city, public data combined with data produced by citizens and private businesses offers new prospects. Understanding the *urban data potential* is therefore a key challenge in the field of transportation and mobility.

In the Ubiquitous Data-Driven Urban Mobility (UbiMob) project, we have worked towards a vision for harnessing such data. The aim is to, on the one hand, help citizens to make smart decisions while taking personal needs into account and, on the other hand, help service providers and operators to reach equilibrium of mobility services, supply and demand, by smarter resource planning and matchmaking. UbiMob investigated opportunities and challenges around the three biggest Norwegian cities, of which this paper will focus on the Bergen region. Our research questions for the Bergen investigation in UbiMob have been: *What are the sources of mobility-related information in the Bergen area? Which needs and opportunities do they present?* and *How can the generic smart-city concept best be instantiated in a Bergen context?* (and thus perhaps transferred to other Norwegian cities).

The rest of the paper will present the results of our study so far. We will present: examples of existing work, central creators and owners of mobility data in the Bergen region; central information users and example user stories; a taxonomy of mobility information types and their properties; information integration challenges; and preliminary work on a running architecture for mobility data. The conclusion will also outline a few possible paths for further work.

## 2. EXISTING WORK

The European Initiative on Smart Cities[3] aims to support cities and regions in taking ambitious measures to make certain progress by 2020 towards a 40% reduction of greenhouse gas emissions through sustainable use and production of energy. Similarly, the European Innovation Partnership on Smart Cities and Communities[4] (EIP-SCC) intends to develop collaborative and participatory approaches for cities, industry, and citizens to improve urban life through sustainable solutions that include a more efficient use of energy, transport and Information and Communication Technologies (ICT). It has already resulted in several European projects focussing on sectors such as *Energy*, *Transport & Mobility* and *ICT*[5]. Moreover, an abundant set of dynamic, open context-aware, ubiquitous data-driven, ITS (Intelligent Transport System) services can be seen a key step towards a so-called *smart world* — an integration of smart environments to better understand neighbourhoods so as to improve citizens’ well-being. Such services include adaptive personalised maps (Liang et al. 2008), adaptive vehicle navigation (Meng and Poslad 2009), smart fleet management and traffic monitoring (Giacobbe et al. 2010), road incident detection (Chatzigiannakis et al. 2007), congestion avoidance (Parrado and Donoso 2015), speed control via smart interaction with roadside controls (Pérez et al. 2010), context-based vehicle maintenance (Matsuzaki and Todoroki 2008), car parking aids (Ji et al. 2014), human driver monitoring (Sahayadhas et al. 2012) and better driving safety (Jang et al. 2011).

Pellicer et al. (2013) summarized the various smart mobility works or projects worldwide under different smart city areas: smart governance, smart mobility, smart environment, and smart living. The cities and the smart mobility projects that they have investigated are listed as follows: Malaga (automated meter reading, electric vehicles and charging stations, energy efficiency for public facilities, smart grid, etc.), Paris (electric charging stations, bicycles, exchange plans), Amsterdam (e-

3 https://setis.ec.europa.eu/set-plan-implementation/technology-roadmaps/european-initiative-smart-cities

4 http://ec.europa.eu/eip/smartcities/

5 https://eu-smartcities.eu/

citizen participation, electric vehicle and charging stations, energy efficient transport, etc.), Vienna (smart grid, $CO_2$ emissions reduction), Toronto (efficient metropolitan urban mobility, green mobility policy), New York (open government - open data, improving public transportation, ICT-enabled services and pedestrian spaces for citizens, start-up development for social web), Copenhagen (efficient public transport, natural resources optimization and waste management, energy efficiency), Hong Kong (online services for citizens, RFID in airports, smart cards), Barcelona (project iCity-APP to serve, electric vehicle, urban traffic management, efficient transport, smart homes), Stockholm (waste management system), London (online portal, smart card, networked public service, undergrounds optimization and management, Wi-Fi for metropolitan areas), Rio De Janeiro (traffic management, security system, smart emergency systems), and Vancouver (electric vehicles, green transport).

MK:smart[6] has aimed to collect all sorts of data relevant to the city of Milton Keynes, England, in order to deliver smart city technologies that can support sustainable economic growth. As of April 2018, their data hub has 255 datasets covering topics like transport, energy, water, business, public services information, education, etc. The project has initiated a wide range of research into smart city areas, like how to apply semantic technologies to ensure exploitability of the data hub (d'Aquin et al. 2015, Daga et al. 2016), how to enable citizens to apply smart city data for their own purposes (Wolff et al. 2017), knowledge discovery (Tiddi et al. 2015), and handling of energy consumption in smart cities (Cavero et al. 2015). Also in MK:Smart one has collected mobility data and discussed how they can be used to improve public transport (Potter et al. 2015), combining smart sensors and user provided data (Valdez et al. 2018, Valdez Juarez et al. 2015). Daga, d'Aquin, and Motta (2017) have investigated how data policies propagate through smart city infrastructures where different data with different licenses are constantly being combined, processed, and redistributed in complex and dynamic ways.

# 3. SOURCES OF MOBILITY DATA

There are four major urban mobility data stakeholders: *public actors*, such as central governments, cities, municipalities, urban communities, etc.; *private companies*, whether they are public service concession holders or not, including start-ups; *end-users,* such as citizens and tourists; and *social networks and virtual communities*. In order to gain an overview of the creators and owners of mobility data in the Bergen region, we have communicated face-to-face, by phone, and by email with more than 20 different stakeholders and additionally reviewed numerous documents, websites and APIs, which we divide broadly into: government, ideal organisations, and business/commerce. Citizen-generated and open data constitute additional categories that we will address.

## 3.1 Public actors

Several units in the City of Bergen (Bergen kommune) collect and maintain mobility data. The Agency for Urban Environment (Bymiljøetaten) in the Department of Urban Development has the main responsibility for mobility, including city roads, street parking, e-car loading spots, and bike lanes, about which they collect and maintain data. The *Parkering i Bergen* app manages street-parking payments. Public parking houses are run by a semi-private entity (Bergen parkering) that runs automatic number-plate registration (ANPR) as part of one of their payment systems. Information about available parking slots – including e-car loading stations – are made available online and through a REST API. The payment app is maintained by a subsidiary company, SesamSesam. The section also administers the city bikes, which has so far been operated by Urban Infrastructure Planner (UIP, a private company, see below). In addition, they manage the city's car-toll stations, operated by BT Signaal (below).

Other city departments also have sections that work with mobility data. The Climate Section operates four air measurement stations and sometimes, but not regularly, order surveys about perceived city climate. The Section for Digitalisation and Innovation is developing an open data lake, which will

6 http://www.mksmart.org/

increasingly include mobility-data. The Section for Societal Safety and Preparedness (beredskap) is a hub for sharing emergency- and response-relevant information among governmental, ideal, and commercial organisations, much of which is mobility-related. The Section for Information is a distributor of many information types but does not maintain mobility-related data of its own. Bergen City Event maintains a website of public events like concerts, which can impact traffic and parking conditions.

On the regional level, Skyss is an agency under Hordaland County that manages collective transport by bus, metro, speed boat, and ferry. It maintains schedules and runs automatic passenger counting (APC) systems on around half the buses in the county. GPS coordinates are collected and maintained by a subcontractor, ITS4mobility. Boats and ferries are operated by private companies we will return to later. EnTur is a public company that acts as a hub for collective transport data in all of Norway. It collects and distributes timetable as well as live data about bus, metro and train transport in the Bergen region through an open API.

On the national level, The Norwegian Public Roads Administration (Statens vegvesen) continually collects traffic data from all of Norway using a variety of methods, including digital images, much of which is made openly available in Datex II, an XML-based standard for communicating and exchanging traffic information.[7] BaneNor is a government agency that collects and maintains information about the railway network and its traffic, which is available through an open REST API and website, with SIRI as a central backend system. Avinor is a state-owned limited company that operates most of Norway's civil airports, among them Bergen Airport Flesland, making information about arrivals and departures available through open REST APIs and on the web.

Official Norwegian maps are maintained by the National Map Authority (Statens kartverk), and there are open alternatives such as OpenStreetMap and LinkedGeoData. Weather services are provided by met.no through a website and an open REST API.

## 3.2 Private companies

On the regional level, taxi companies such as Bergen taxi, alongside nationwide competitors Norgestaxi and Taxi 1, routinely collect data about taxi positions at stops. For each ride, start and stop zones, times, and travel distance are collected. Urban Infrastructure Planner (UIP) runs the Bergen city bikes. They maintain an app and an API that provides an overview of available city bikes and return slots in their stations. They provide open datasets of usage statistics including rental and return positions and times. BT Signaal operates the ring of toll stations around Bergen that collects proprietary data about numbers and categories of passing vehicles by the minute. Private parking houses run their own ticketing systems.

Regional operating companies for speed boats and ferries collect information about past, current, and anticipated future passenger counts (by route and stop in the case of buses), as well as vehicle positions, timetables and actual departure and arrival times. On the national level, long-distance air-traffic, bus, and ferry companies collect similar information. EnTur collects and disseminates data for buses, metro, and trains.

National and international mobile phone companies keep continuous track of mobile devices, identifiable through the identities of both devices (IMEIs) and subscribers (IMSIs). Increasingly, such devices go beyond mobile phones to a wider range of more or less smart, mobile-networked things. However, privacy laws strictly limit their ability to share such information – and to use it themselves. In emergency situations, UMS (Unified Messaging Systems) offers location-based citizen messaging, to which municipalities in the Bergen area have access. Internet providers can also collect similar, but less frequent and less precise, data when laptops and other mobile equipment move between network access points.

---

7 CEN TS 16157: DATEX II – The standard for ITS on European Roads, http://www.datex2.eu/sites/www.datex2.eu/files/Datex_Brochure_2011.pdf .

*Table 1. Central sources of mobility data in the Bergen region.*

| | |
|---|---|
| Movement of buses, taxis, planes, trains, boats/ferries (by GPS) | Movement of people into and out of buses with automated passenger counting |
| Passenger counts for buses/planes/trains/boats/ferries | Stops for buses/taxis/trains, quays for boats/ferries |
| Movement of vehicles, cyclists and walkers through mobility sensors or through toll stations and road counters | Travel/tourism data, including arriving cruise ships, hotel bookings, etc. |
| Zone maps for boats/taxis/boats/ferries | Weather and pollution data, e.g., about icy roads and air quality |
| Route information and timetables for buses/trains/places/boats/ferries | Information about events such as concerts, parades, protests, sports, etc. |
| Maps, including road maps and bike/walking paths, and other geographical data, including points of interest | Information about emergencies such as traffic accidents, fires, etc. |
| Availability of city bikes and pooled cars (and of return slots for bikes), and other modes of car or bike sharing (e.g. cargo bikes) | Information about other deviations such as temporarily closed roads; police investigations; delayed and suspended buses, planes, trains, boats, ferries; redirected bus/train routes, etc. |
| Movement of motorised vehicles through toll booths, into and out of parking/houses spots, e.g., through RFID or ANPR (automated number-plate recognition) | Apps (or related websites) that track personal mobility (fitness apps, and everything location-based or with a location option, including Twitter, Facebook, Instagram etc.) |
| Movement of city bikes and pooled cars between pick-up points | Apps that track vehicle mobility (parking, city bikes, pooled cars) |
| Parking spots of various types on the streets and in parking houses | Train/plane/ferry/boat ticketing systems (bus ticketing is today mostly by zone and less useful) |
| Mobile phone positions and movements (through cell towers, GPS, WiFi connections) | Ticketing systems for events such as cinema, concerts and sports |
| Open reference data from Wikipedia/DBpedia/Wikidata, OpenStreetMap, GeoNames, etc. about points of interest. | Pick-up points for city bikes and pooled cars |
| | E-vehicle charging stations |
| Population and demographic data | Survey data about perceived life quality, travel habits, etc. |

International ICT companies like Google, Apple and Microsoft harvest personal mobility data through their mobile phone operating systems and other services, which they leverage to provide map services augmented with, e.g., travel-time estimates depending on time and day. These companies may have more complete pictures of the mobility situations in Bergen and in Norway than any local or national actor, but their raw data are not open to local and national stakeholders.

## 3.3 Ideal organisations

The Bergen Car-Sharing Ring (Bildeleringen) has more than 1800 household members who collectively own several hundred cars in more than 90 locations in the Bergen city area. Online Convadis terminals in each car keep track of use and parking locations, but GPS data is not collected when a car is in use. The online car-booking system Let’s Go. A subcontractor manages member accounts, billing, and generates car-usage statistics on demand.

## 3.4 Citizen-generated data

Through their mobile phones and other gadgets (smart watches, exercise wristbands, etc.), electronic payment transactions, social media use (which may be geotagged or contain other mobility-related information), networked cars and other vehicles, most citizens leave continuous data trails as they go about with their daily lives. Coined *sousveillance*[8] (Kitchin 2014) this army of “little brothers” also generates and maintains enormous amounts of mobility-related data, which are not in general openly available (as citizens tend to give away ownership of their data to the app and service providers in return for access to free services). Some of this data is directly or indirectly collated by behemoths like Amazon, Apple, Facebook, Google/Alphabet, and Microsoft[9], the rest ends up with apps and social-

8 *Norw.: “undervåkning”.*
9 In an international context, at least Alibaba, Baidu, and Tencent should also be added to this list.

media providers. A few of them, like Twitter, at least for now share their data to some extent. Most, however, like Facebook and Google, do not share the data itself, but offer free services based on it. Although, the new European General Data Protection Directive and Regulation will return some of the control over data back to the originating person, it will work only on an individual-by-individual basis. And although mature citizen-centric/user-owned open alternatives exist (such as GNU social as an alternative to Twitter and Diaspora to Facebook), they are not yet widely used.

### 3.5 Open data

Although most of it has originated from governments and their citizens, open data has particular characteristics that make it worth discussing on its own. data.norge.no is the national portal and web hotel for open datasets. It currently contains around 1100 datasets, but only 60 of them are transport related and none of those are specific to the Bergen region. GeoNames.org and OpenStreetMap are crowdsourced datasets about geography and transport networks, which are important types of mobile context data. General datasets like Wikipedia can also be used to enrich map and other data. Of course, for many or most of these open datasets, there are also (for-pay or restricted) commercial or governmental alternatives that sometimes have higher quality (completeness, correctness, precision, timeliness, etc.), depending on the use case.

Because they use standard formats and vocabularies on the syntactic and semantic levels, semantic open datasets (or knowledge graphs) using technologies such as RDF, OWL, and SPARQL are a particularly interesting type of mobility context data. GeoNames[10] provides a simple ontology and several semantic interfaces to its data, whereas LinkedGeoData (Auer et al. 2009) is a semantic version of OpenStreetMap. DBpedia (Bizer et al. 2009) is a semantic extract from Wikipedia, whereas Wikidata[11] is a natively crowdsourced semantic fact database aiming to provide structured facts to Wikipedia projects, making reuse and maintenance easier. Table 1 presents an overview of the most central sources of regional mobility data we have uncovered.

*Table 2. Examples of user stories.*

| | |
|---|---|
| As a citizen, I want to find neighbours with similar driving patterns in order to organise car pooling to and from work. | As a taxi company owner, I want to precisely predict demands for taxis in order to lower prices and increase profits. |
| As a city official, I want to monitor greenhouse gas emissions in order to document the effects of the city's traffic regulations on the environment. | As a traffic manager, I want to detect and predict the consequences of road accidents in real time in order to proactively redirect traffic around accident sites. |
| As a city official, I want to understand policy impact on mobility in the city to better plan traffic interventions and change modal split: the types of transport people use to get into or around in the city. | As a public transport provider, I want to offer seamless single-ticket trips that may also include city bikes, car pooling, car-ring vehicles and parking slots in order to make collective travels more convenient for citizens. |
| As an emergency manager, I want a live overview of people's locations during an emergency in order to direct rescuers and medical personnel. | As a city environment official, I want to understand the interplay between weather, traffic and air pollution in order to keep the city air clean. |
| As an emergency manager, I want to predict people's movements after an emergency in order to reduce bottlenecks and better support evacuation. | As a first responder, I want to know the quickest travel path to a place of emergency, taking into account the resulting traffic congestion, in order to help citizens in need. |
| As a parking house manager, I want to offer dynamic pricing in order to make the city centre attractive for shoppers and retailers. | As a road manager, I want to monitor and predict the consequences of closing road segments for maintenance in order to avoid congestion / accidents and ensure smooth traffic. |

## 4. MOBILITY-DATA USER STORIES

By reviewing open documents, communicating with central stakeholders, and drawing on our own experience as both city residents and researchers, we have identified opportunities for using mobility

10 http://www.geonames.org/about.html

11 https://www.wikidata.org/wiki/Wikidata:Introduction

data to create a smarter city region. We have formulated each opportunity as a user story to make it as concrete and as well aligned with future development work as possible. Table 2 shows examples of user stories we have identified. In addition, many other uses of mobility data have been reported in the research literature, such as in (Benevolo et al. 2016). Along with the stories we have collected ourselves, they provide a rich starting point for further work on making Bergen and other Norwegian cities smarter.

# 5. MOBILITY-DATA CHARACTERISTICS

Developing apps and services that realise each of these user stories will in most cases require integration of several different types of mobility data originating from different sources. To facilitate large-scale data integration, we are developing a taxonomy of mobility-data types and characteristics, which may evolve into a high-level (or upper) ontology for mobility data.

## 5.1 Types of mobility data

In the introduction, we defined *mobility data* as *data that is about or that can inform the movement of people and objects*. We proceeded to distinguish between *movement data* and *mobility-context data*. Movement data is *data that describes how a person or object moves*, typically using time- and location-stamps. Hence, we reserve the term movement data for data that is both temporally and spatially located, even on a coarse/aggregate level. By live *movement data* we mean more frequently updated (many times an hour) and spatially fine-grained (at least to street level) movement data. Mobility-context data is other types of mobility data – including timetables, maps, and information about places – that add context to the movement data, but do not themselves describe how a person or object moves.

## 5.2 Specific characteristics of mobility data

With this typology in mind, we go on to investigate their detailed characteristics and how they differ on the conceptual level:

- *Temporality:* whether the data describes phenomena that change with time and, if so, the period the data covers, the delay from the mobility event until the data is available and the frequency of updates, with ranges from milliseconds to a year or longer for statistical reports (temporal granularity). We can also distinguish between different points in the lifecycle of an event, such as when it: occurs, is observed, is registered for the first time, is committed to persistent storage, becomes available to users, is covered by various types of aggregated reports, etc.
- *Spatiality:* whether the data is spatially located and, if so, which area the data covers and its spatial precision (spatial granularity).

## 5.3 General characteristics

In addition to these mobility-specific characteristics, general properties of data include:

- *Subject:* whether the data is about, e.g., individuals, vehicles, places, structures (such as buildings), events, weather conditions, etc., or about relations between two or more such phenomena.
- *Privacy:* for data about individuals, whether the person is alive and can be directly or indirectly identified through the data (i.e., whether the data are personal) and, if so, whether they are also sensitive (e.g., about criminal matters, health, political activities, race, religion, or sex life.)

Other general characteristics are availability, ownership, and provenance. Mobility data is available in different formats, publicly or not, through file downloads or web APIs. Size and bandwidth are particular concerns, i.e., how much data that is generated by the day or hour. For example, meteorological or algorithmically or socially generated data, if harvested non-selectively, may grow prohibitively large.

### 5.4 Other mobility-related ontologies

There are several relevant ontology building efforts. Datex II is a widely used XML-based format that has been "developed to provide a standardised way of communicating and exchanging traffic information between traffic centres, service providers, traffic operators and media partners" and is experimentally available as an OWL ontology, Linked Datex II. km4city (Bellini et al. 2014, p. 4) is a natively semantic vocabulary for interoperating the very large numbers of public and private mobility datasets that are available from local governments and other sources, partitioned into administration, street guides, points of interest, public transportation, sensors and time. MobiVoc is an ongoing vocabulary initiative that is currently limited to parking-related concepts. This core of dedicated mobility vocabularies are interlinked — or interlinkable — with a large number of surrounding vocabularies for domains such as geodata and maps (e.g., Auer et al. 2009), weather (e.g., WeatherOntology), sensors (e.g., Semantic Sensor Network (Compton et al. 2012)) etc. In addition, CityGML is an open standardised data model and exchange format for digital 3D models of cities and landscapes, although it is not yet defined in an ontology.

To the extent they are semantic, these vocabularies build on many of the same general vocabularies for describing general phenomena such as times (e.g., OWL-Time) and locations (geo), people (FOAF, bio), organisations (org), events (the Event Ontology), provenance (PROV-O), and data ownership (CC). The Linked Open Vocabularies site offers an overview of reusable semantic vocabularies on the web (Vandenbussche et al. 2017).

However, these mobility-related ontologies (Datex II, km4city, MobiVoc) and format (CityGML) tend to focus on the middle (domain-specific) and lower (context-specific) levels, whereas our taxonomy belongs to the upper (general) level: the subtypes and characteristics we have identified can be used to describe and systematise types and properties found in middle- and lower-level mobility ontologies and vocabularies.

## 6. DISCUSSION

### 6.1 Research contribution

The pre-study reported in this paper has been motivated by a gap in the research on smart cities and on mobility-related data that has addressed Norwegian cities specifically. We have asked: *What are the sources of mobility-related information in the Bergen area? Which needs and opportunities do they present?* and *How can the generic smart-city concept best be instantiated in a Bergen context?* The paper has contributed answers to the two first questions, largely confirming that the Bergen data sources match the sources reported in the international literature, aligned with national and regional government structure and practices.

### 6.2 Towards an architecture for mobility data

In order to approach the third question, we are currently building a historical database of open mobile data from the Bergen region. We are working to harvest, integrate, and make open mobility data ready for use through a single entry point for research and education purposes. Our still-evolving architecture is composed of: *harvesters* that download data from web and store it in a *staging area* where it can be *transformed* and *loaded* into a big data store.

*Harvesting:* We use simple scripts, mostly written in Python and some Java, to continuously download open mobility data. The data is mostly available through APIs, but also as files and by scraping web pages. So far, we have written harvesters for: Bergen bysykkel (city bikes), Bergen City Events, a selection of local newspapers, met.no, EnTur, Avinor, Tweets with Bergen-related keywords or geotagged around Bergen, and radnett.nrp.no. The harvesters (along with the transformers and loaders) run on a medium-CPU Amazon EC2 Ubuntu Linux cloud server, currently with a 512G storage. The various scripts run as cron jobs with frequencies between every minute and once an hour. While processing load is not yet heavy, the elastic storage volumes offered by EC2 makes it easy to handle growing data volumes.

*Staging:* The harvesters store the data as JSON files. The files are organised in series, typically store in the same folder in the staging area, and so that each series contain files produced by the same harvester and containing homogeneously-structured JSON objects. Each file is time-stamped and corresponds to a single harvester run. Most harvesters produce only a single JSON file series, but some produce several.

*Transformation:* Before uploading to the data store, the JSON files are prepared in several ways. Character set encoding is standardised to plain ASCII and UTF-8 if necessary, and times and dates are reformatted if necessary to the standards used by Cassandra. File series with a simple structure, i.e., whose JSON objects are regular and not deeply nested, can be uploaded directly into a database table (called column families in Cassandra). However, some harvesters generate more complex file series that must be split up and further simplified before each series can be uploaded into a database table.

*Loading:* The transformed JSON file series are uploaded to Cassandra (Lakshman and Malik 2010), which is one of the most used NOSQL DBMSs. It combines features from key-value pairs and wide-column stores. In addition to flexibility, scalability and read-orientation, Cassandra is designed to support data replication, resilience towards failure, and ease of elastically adding more machines to the database cluster. It is also optimised for retrieving and adding new data as opposed to updating existing data. We attempt to make the loading stage a streamlined as possible by doing as much of the preparation as possible in the transformation stage (where generic tools are available, like JSONPath, an XPath-like tool for JSON). The aim is to transform the harvested file series to a point where the Cassandra tables (or column families) match the JSON-document structure closely.

*Analysis:* On the usage side, we are exploring analysis tools such as Spark (Zaharia et al. 2016) to post-process the data from Cassandra, both to provide simple data services exposed as REST APIs and to extract re-combined datasets that can be used for data analytics. The user stories we have presented in Table 2 are good starting points. Semantic lifting and integration of our datasets is another high priority. In the future, we want to build more transformers and streamline the loading process. We are also developing more harvesters, and want to gain access to further data sources that are not yet open, for example by collaborating more closely with the municipality and county as well as industrial actors.

## 7. CONCLUSION AND FURTHER WORK

The paper has presented our work on a pre-study that focussed on mobility-related information sources and needs in the Bergen region. Although this is an area in rapid development, we have made progress towards our two first research goals: to better understand the sources of, needs for, and opportunities of mobility-related information in the Bergen area. In this way, we have also made some progress towards our third goal: to understand how the smart-city concept can best be instantiated in a Bergen (and thus Norwegian) context, using mobility data as an example.

As the UbiMob pre-project is over, the Bergen authors are now continuing the work as part of another project: BDEM (Big Data for Emergency Management[12]). They are still harnessing mobility data from the Bergen region, but instead emphasising uses of this data for emergency situations. Of course, this should not preclude use of the same data for other mobility-related analyses too. Independently of UbiMob, and without a research partner on the ICT side, Hordaland County and the City of Bergen have established a mobility laboratory for development of smart transport solutions ("*mobilitetslaboratorium for utvikling av smarte transportløsninger*", MUST), as a response to the government's Smarter Transport in Norway challenge. The central activities in the first phase of MUST will be (1) collecting data from different sources and providing access to and information about it; (2) establishing architectural principles and targets according to accepted standards; and (3) facilitating physical testing of solutions in the regional infrastructure (Hordaland County and City of Bergen 2017).

12 https://www.bigdata.vestforsk.no/

While more and more data is becoming openly available through the large search engines, there is a specific need to take local and regional needs and concerns into account. Building local databases or federated open data stores with analytics capabilities ultimately benefits all citizens and society as a whole by making data more accessible, findable, and interoperable. Quality assurance and processing on top of raw data increases the utility for a number of scenarios. Apart from the use cases developed here, this also includes an easier support for startups, for research, and for citizen science.

## ACKNOWLEDGEMENT

Ubiquitous Data-Driven Urban Mobility (UbiMob) was a pre-project (270785) funded by the Norwegian Research Council in 2017 through the IKTPLUSS programme. Big Data and Emergency Management (BDEM) is a project (261685) funded by the Norwegian Research Council and the Norwegian Centre for International Cooperation in Education (SiU) for 2017-2020 through the International Partnerships for Excellent Education, Research and Innovation (INTPART). The authors wish to thank Marius Falch Lillevik and Eirik Sjøvold who wrote many of the scripts used in our ongoing effort to establish a database of open mobility-related data.

## REFERENCES

Ahlers D, Driscoll P, Löfström E, Krogstie J, Wyckmans A (2016) Understanding smart cities as social machines. (International World Wide Web Conferences Steering Committee), 759–764.

d'Aquin M, Davies J, Motta E (2015) Smart cities' data: Challenges and opportunities for semantic technologies. *IEEE Internet Comput.* 19(6):66–70.

Auer S, Lehmann J, Hellmann S (2009) Linkedgeodata: Adding a spatial dimension to the web of data. (Springer), 731–746.

Bellini P, Benigni M, Billero R, Nesi P, Rauch N (2014) Km4City ontology building vs data harvesting and cleaning for smart-city services. *J. Vis. Lang. Comput.* 25(6):827–839.

Benevolo C, Dameri RP, D'Auria B (2016) Smart mobility in smart city. *Empower. Organ.* (Springer), 13–28.

Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) DBpedia-A crystallization point for the Web of Data. *Web Semant. Sci. Serv. Agents World Wide Web* 7(3):154–165.

Caprotti F, Cowley R (2016) The UK Smart Cities Survey. *Smart-Eco Cities UK 2016*. 94.

Caragliu A, Del Bo C, Nijkamp P (2011) Smart cities in Europe. *J. Urban Technol.* 18(2):65–82.

Cavero J, Kortuem G, Foell S (2015) Smart community energy initiatives in Smart Cities: learning collectively about energy.

Chatzigiannakis V, Grammatikou M, Papavassiliou S (2007) Extending driver's horizon through comprehensive incident detection in vehicular networks. *IEEE Trans. Veh. Technol.* 56(6):3256–3265.

Compton M, Barnaghi P, Bermudez L, GarcíA-Castro R, Corcho O, Cox S, Graybeal J, Hauswirth M, Henson C, Herzog A (2012) The SSN ontology of the W3C semantic sensor network incubator group. *Web Semant. Sci. Serv. Agents World Wide Web* 17:25–32.

Daga E, d'Aquin M, Adamou A, Motta E (2016) Addressing exploitability of smart city data. (IEEE), 1–6.

Daga E, d'Aquin M, Motta E (2017) Propagating Data Policies: a User Study. (ACM), 3.

Giacobbe M, Puliafito A, Villari M (2010) A service oriented system for fleet management and traffic monitoring. (IEEE), 784–786.

Gil-Garcia JR, Pardo TA, Nam T (2015) What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization. *Inf. Polity* 20(1):61–87.

Goldsmith S, Crawford S (2014) *The responsive city: Engaging communities through data-smart governance* (John Wiley & Sons).

Hordaland County, City of Bergen (2017) *Smarter Transport in Norway - Application from Hordaland County and the City of Bergen ("Smartere transport i Norge - Søknad fra Hordaland fylkeskommune i partnerskap med Bergen kommune")*

Jang JA, Kim HS, Cho HB (2011) Smart roadside system for driver assistance and safety warnings: Framework and applications. *Sensors* 11(8):7420–7436.

Ji Z, Ganchev I, O'Droma M, Zhao L, Zhang X (2014) A cloud-based car parking middleware for IoT-based smart cities: Design and implementation. *Sensors* 14(12):22372–22393.

Kitchin R (2014) *The data revolution: Big data, open data, data infrastructures and their consequences* (Sage).

Lakshman A, Malik P (2010) Cassandra: a decentralized structured storage system. *ACM SIGOPS Oper. Syst. Rev.* 44(2):35–40.

Liang Z, Poslad S, Meng D (2008) Adaptive sharable personalized spatial-aware map services for mobile users. 267–273.

Matsuzaki R, Todoroki A (2008) Wireless monitoring of automobile tires for intelligent tires. *Sensors* 8(12):8123–8138.

Meng D, Poslad S (2009) A reflective goal-based system for context-aware adaptation. (Springer), 105–110.

Nam T, Pardo TA (2011) Conceptualizing smart city with dimensions of technology, people, and institutions. (ACM), 282–291.

Parrado N, Donoso Y (2015) Congestion based mechanism for route discovery in a V2I-V2V system applying smart devices and IoT. *Sensors* 15(4):7768–7806.

Pelekis N, Theodoridis Y (2014) Chapter 1: Introduction. *Mobil. Data Manag. Explor.* (Springer).

Pellicer S, Santa G, Bleda AL, Maestre R, Jara AJ, Skarmeta AG (2013) A global perspective of smart cities: A survey. (IEEE), 439–444.

Pérez J, Seco F, Milanés V, Jiménez A, Díaz JC, De Pedro T (2010) An RFID-based intelligent vehicle speed controller using active traffic signals. *Sensors* 10(6):5872–5887.

Potter S, Valdez A, Cook M, Langendahl PA (2015) Governance in niche development for a transition to a new mobility regime.

Sahayadhas A, Sundaraj K, Murugappan M (2012) Detecting driver drowsiness based on sensors: a review. *Sensors* 12(12):16937–16953.

Tiddi I, d'Aquin M, Motta E (2015) Data patterns explained with linked data. (Springer), 271–275.

Valdez AM, Cook M, Langendahl PA, Roby H, Potter S (2018) Prototyping sustainable mobility practices: user-generated data in the smart city. *Technol. Anal. Strateg. Manag.* 30(2):144–157.

Valdez Juarez AM, Cook M, Potter S, Langendahl PA (2015) Exploring participatory visions of smart transport in Milton Keynes. Early-access.

Vandenbussche PY, Atemezing GA, Poveda-Villalón M, Vatant B (2017) Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web* 8(3):437–452.

Wolff A, Gooch D, Cavero Montaner JJ, Rashid U, Kortuem G (2017) Creating an understanding of data literacy for a data-driven society. *J. Community Inform.* 12(3):In-press.

Zaharia M, Xin RS, Wendell P, Das T, Armbrust M, Dave A, Meng X, Rosen J, Venkataraman S, Franklin MJ (2016) Apache spark: a unified engine for big data processing. *Commun. ACM* 59(11):56–65.