Real-time data 2022:

Approaches to integrating real-time data sources in data.europa.eu

data.europa.eu

The official portal for European data

Publications Office of the European Union

Real-time data 2022:

Approaches to integrating real-time data sources in data.europa.eu

Luxembourg, 2022

This study has been prepared as part of data.europa.eu. Data.europa.eu is an initiative of the European Commission. The Publications Office of the European Union is responsible for contract management of data.europa.eu.

For more information about this paper, please contact the following.

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu
Email: info@data.europa.eu

Written by:
Dr Simon Jirka
Antje Kügeler

Last update: 25 April 2022
www: https://data.europa.eu/

Luxembourg: Publications Office of the European Union, 2022

ISBN 978-92-78-43132-7 doi:10.2830/182607 OA-08-22-214-EN-N

Contents

1. Introduction

The goal of data.europa.eu is to improve accessibility to and promote the reuse of public sector information. The portal provides access to open data from international, EU, national and regional sources. This is achieved by collecting the metadata of public data made available across Europe in a process called harvesting. Metadata is harvested both from open data portals and from geodata portals and is made available on data.europa.eu.

This report discusses the potential of real-time data in conjunction with data.europa.eu. Real-time data allows us to gather up-to-date information about the environment surrounding us (e.g. traffic information, weather data, measurements of environmental pollution, information about natural hazards). Consequently, the European directive on open data and the reuse of public sector information ([1]) put a specific emphasis on making such real-time data resources available.

There is a dedicated field of work in the data.europa.eu project to help users discover open real-time data sources. This comprises not only the inclusion of real-time data sources in its data index, but also the support of data providers to make their data available in a reusable and discoverable way.

This report summarises the results and findings of a webinar on real-time data held by the data.europa.eu project on 5 April 2022 ([2]). On the one hand, the report offers a short summary of the presented information and technologies (e.g. relevant standards/APIs). On the other hand, it summarises the main findings of the questions and feedback from the participants (e.g. data needs, types of data sources and required functionalities).

As a result, the report provides a set of recommendations on how to further improve the discoverability of real-time data sources via data.europa.eu.

2. What is real-time data and why is it a valuable source of information?

The European directive on open data and the reuse of public sector information refers to real-time data in a general manner:

> *'dynamic data' means documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data*

Furthermore, it states:

> *For dynamic data, meaning frequently updated data, often in real time, public sector bodies and public undertakings should make this available for reuse immediately after collection by ways of suitable APIs and, where relevant, as a bulk download, save for cases where this would impose a disproportionate effort.*

(1) https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024
(2) https://data.europa.eu/en/academy/real-time-data

For this report, we will use a compact definition that **real-time data are frequently updated data which are delivered immediately after collection**.

The nature of real-time data is quite heterogeneous. Typical types include:

- stationary measurements (e.g. time series);
- tracking data (e.g. tracking of parcels or cars);
- data measured along trajectories (e.g. floating car data);
- images (e.g. video streams from cameras, radar data).

An important added value resulting from the availability of real-time data is the enablement of live decision-making. Typical examples are dashboards that allow the monitoring of any kind of process (e.g. traffic flow, significant weather events) and real-time data analytics (e.g. detecting critical air pollutant concentrations, water level monitoring in cases of flooding events).

3. What are typical examples of real-time data?

Real-time data are relevant in many application domains, including the following:

- traffic information
- weather measurements
- rain radar
- water level measurements
- people flow information in cities
- gasoline prices.

In the following section, a few selected examples from this list are presented in more detail.

Traffic information is important for many people in their daily life. This may include information on real-time schedules of public transportation such as buses or trains. Other examples are live updates on traffic jams or the availability of parking spaces in a city. One option to provide such information is the context broker ([3]) developed by FIWARE ([4]). The aim of the Context Broker is to enable the interfacing between internet of things devices and, for example, sensor systems and users of these data. Thus, it can be considered as an efficient means to publish real-time data. To ensure an interoperable provision of such real-time data, a comprehensive set of **Smart Data Models** is available ([5]), which define how data shall be encoded. As part of the smart city domain, these data models also offer, among other aspects, ways to exchange information on public transportation networks and parking space availability ([6]).

([3]) https://www.fiware.org/developers/catalogue/
([4]) https://www.fiware.org/
([5]) https://github.com/smart-data-models
([6]) https://github.com/smart-data-models/SmartCities

In the European data portal project (the predecessor to data.europa.eu), there was already a cooperation with Context Broker developers to enable the generation of metadata, so that information about available real-time data via a Context Broker instance can be harvested.

In addition, the support of selected data models offered through the FIWARE Context Broker is already supported via the geospatial data preview functionality of data.europa.eu. Currently, this comprises a preview for bus tracking (see Figure ***1***) and parking space availability (Figure ***2***) data.

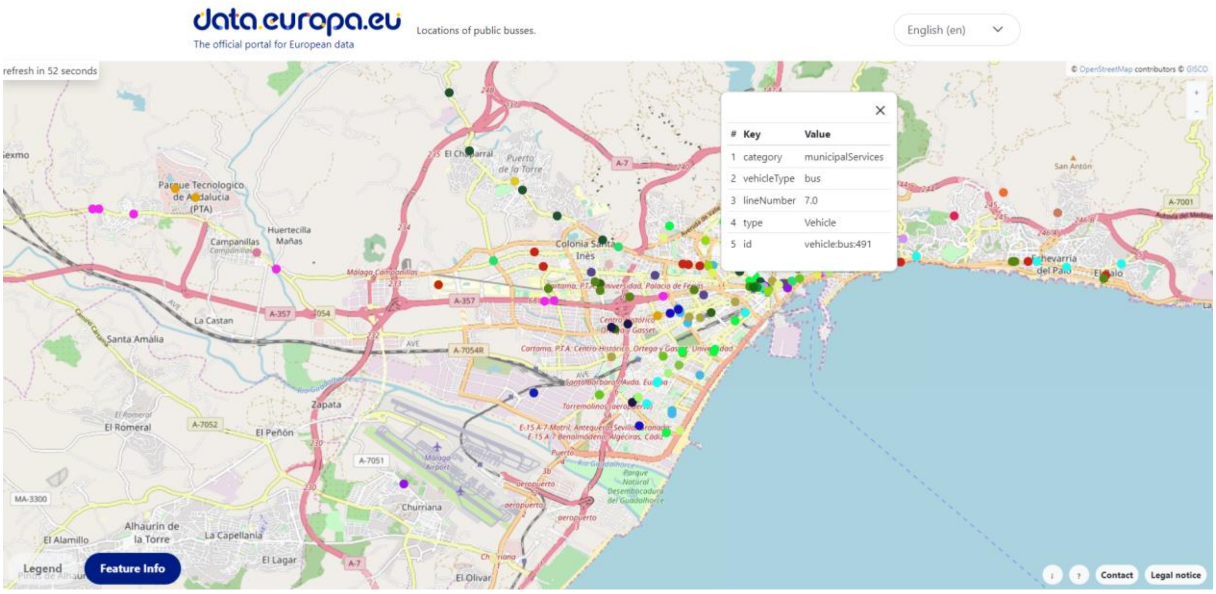

Figure 1: Preview of bus tracking data in the city of Malaga

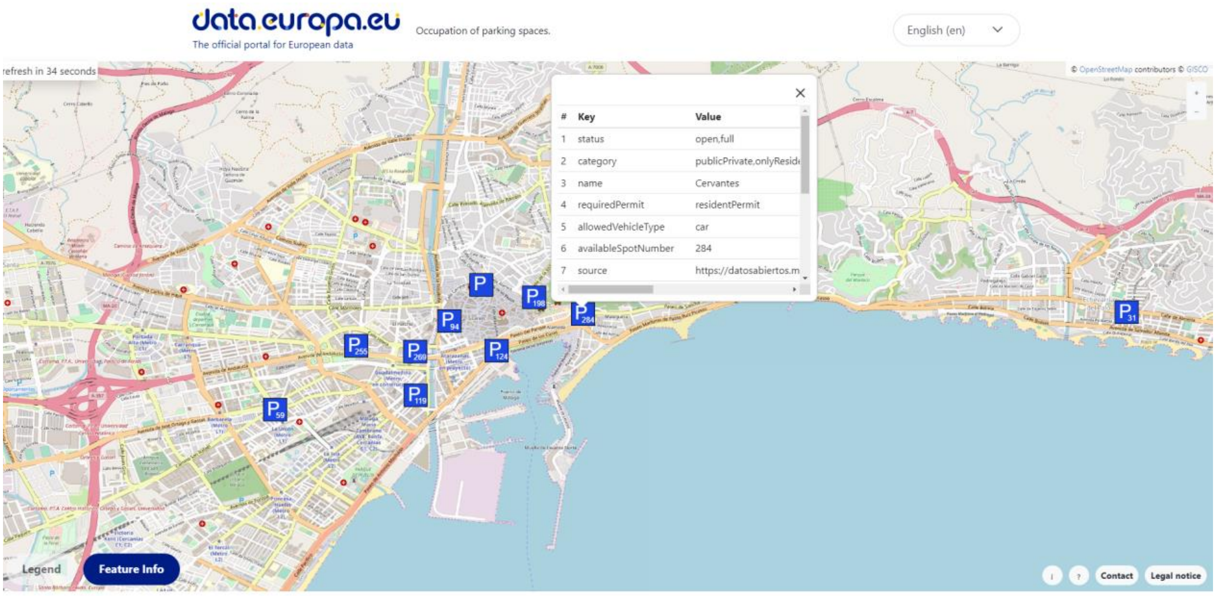

Figure 2: Preview of parking space availability data in the city of Malaga

Further examples include the provision of air quality data or hydrological measurements. In these domains, there are several examples relying on the provision of sensor data based on the standards of the Open Geospatial Consortium (OGC) (7). One of these is the e-reporting of air quality observation data by EU Member States to the European Environmental Agency (flow E2a) (8). This data flow is based on a data model derived from the ISO/OGC observations and measurements standard (9). In some cases, datasets for this flow are discoverable via data.europa.eu (10) (11).

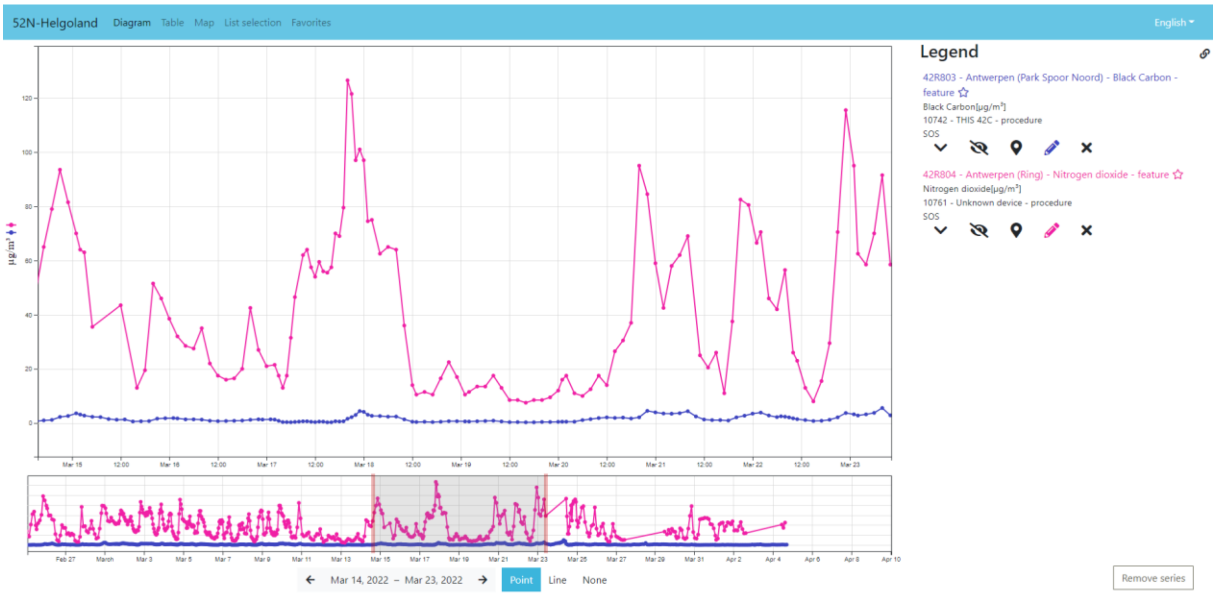

Figure 3: Exemplary visualisation of air quality measurements (external application that is not part of data.europa.eu)

Similar approaches are also used for sharing hydrological measurements. Many organisations working in hydrology collect real-time observation data for purposes such as flood monitoring, water management (e.g. ensuring a sufficient water supply for industry) and supporting recreational activities. In this case, the OGC WaterML 2.0: Part 1 – Timeseries (12) standard is worthy of mention. This standard provides guidance for interoperable encoding of hydrological measurements.

4. Which are relevant technologies to share real-time data?

When considering suitable technologies for enabling the exchange of real-time data, interoperability is a major guiding principle. In a scenario where data providers use their own data formats and interfaces, integration efforts may become huge if data from many providers are consumed by a large number of users. In this case, the link between each data user and provider might require individual implementation work. This is avoided by using a common language, namely common data formats and data access interfaces for enabling real-time data flows.

(7) https://www.ogc.org/
(8) https://www.eea.europa.eu/data-and-maps/data/aqereporting-9
(9) https://www.ogc.org/standards/om
(10) https://data.europa.eu/data/datasets/dat-3-en?locale=en
(11) https://data.europa.eu/data/datasets/d82210dc-65c4-46c6-b6fb-34f2036e2206?locale=en
(12) https://portal.ogc.org/files/?artifact_id=57222

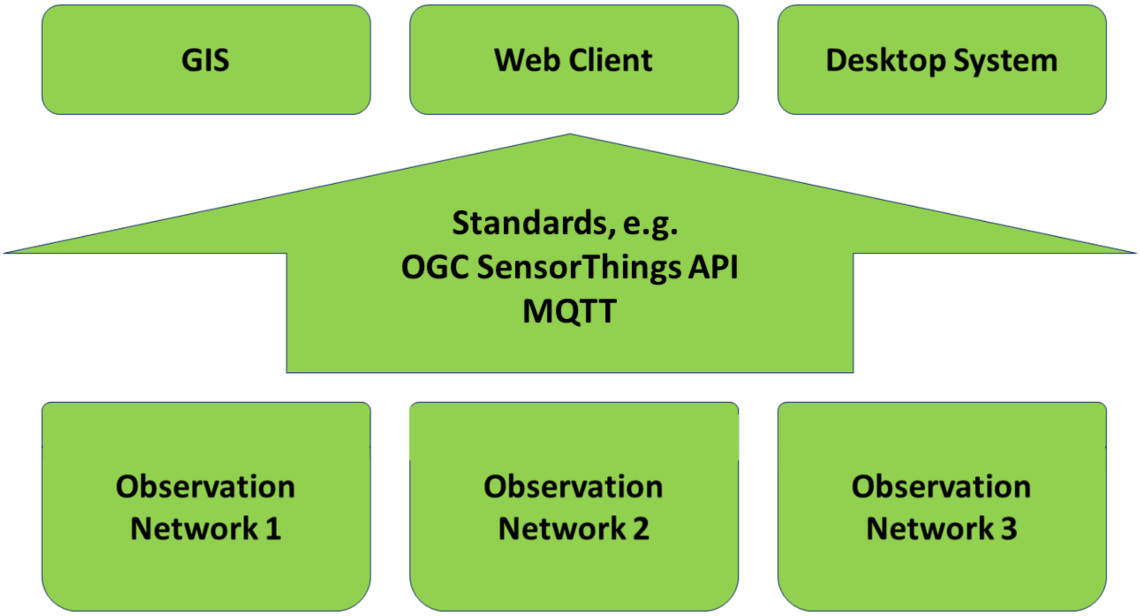

Figure 4: Basic concept of interoperability – using standards to establish a common language between data providers and users

In the following section, two selected interoperability standards are introduced. The OGC SensorThings API specification standard was selected due to its endorsement as an INSPIRE good practice for sharing observation data ([13]). The Message Queuing Telemetry Transport (MQTT) protocol is introduced as it is widely used in the internet of things community, which has the potential to become a valuable source of real-time information.

4.1. OGC SensorThings API

The OGC SensorThings API standard ([14]) specifies an interoperable way to encode and provide access to observation (data streams) generated by sensors. This standard was adopted in 2016 and offers a specification based on REST and JSON, which makes it convenient to use for web developers. More specifically, it follows the principles of the OData standard ([15]).

At its core, the OGC SensorThings API offers common functionality to create, read, update and delete sensor resources such as the following.

- **Datastreams.** Aggregations of observations from the same sensor and for the same parameters.
- **Observations.** The actual observed data.
- **Sensors.** The devices generating the offered observation data.
- **Observed properties.** The parameters which are observed.
- **Features.** The real-world objects for which the observations provide information.

([13]) https://inspire.ec.europa.eu/good-practice/ogc-sensorthings-api-inspire-download-service
([14]) https://docs.ogc.org/is/18-088/18-088.html
([15]) https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=odata

In order to specify which data a user wants to download from a SensorThings API endpoint, several query options and functionality are provided, such as:

- time filters for specifying for which time period data shall be returned;
- filters on the observed parameters;
- filters to specify which properties of a resource (e.g. observation) shall be returned (i.e. in order to reduce the transmitted data volume);
- filtering by certain result values (e.g. all observations above a certain threshold);
- ordering the content of a result by user-specified criteria;
- functionality for enabling paging within result sets.

Since its adoption by the OGC, the SensorThings API specification has gained increasing acceptance. Thus, we can observe a growing number of implementations of this standard ([16]). This has also resulted in an INSPIRE good practice document ([17]) that describes an approach for implementing an INSPIRE download service for observation data using the OGC SensorThings API standard.

In addition, the SensorThings API offers a direct integration with the MQTT standard, which is described in the next section.

4.2. Message Queuing Telemetry Transport

The MQTT protocol ([18]) enables the exchange of messages according to the publish/subscribe principle. The basic approach of the protocol is illustrated in Figure ***5***.

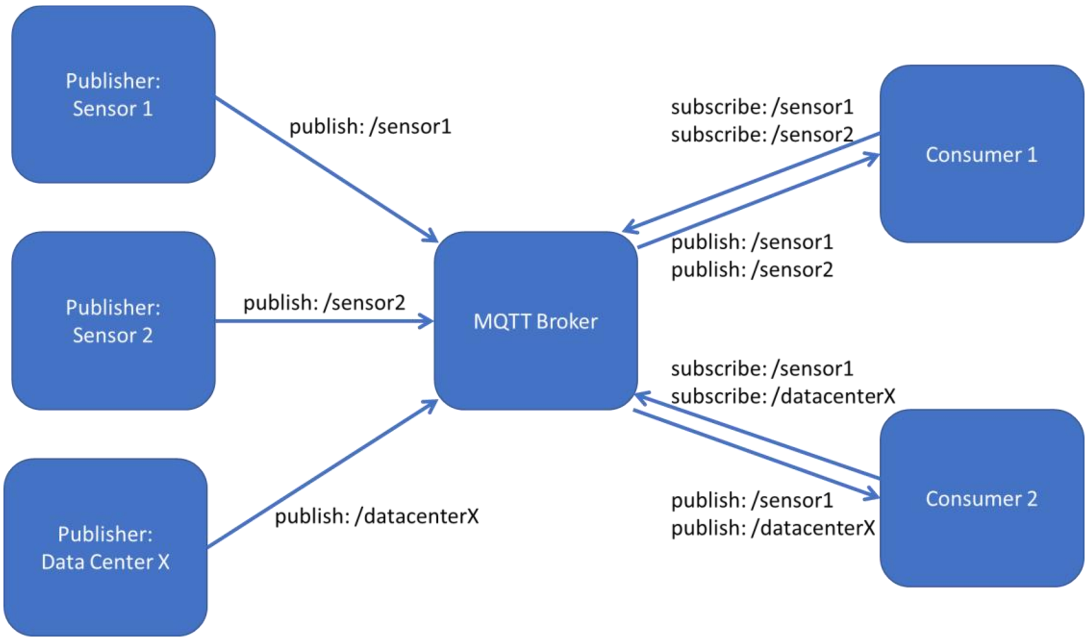

Figure 5: Schematic overview of MQTT

([16]) https://publications.jrc.ec.europa.eu/repository/handle/JRC128885
([17]) https://inspire.ec.europa.eu/good-practice/ogc-sensorthings-api-inspire-download-service
([18]) https://mqtt.org/

The central element of MQTT is the use of brokers. These brokers take incoming messages from publishers and distribute them to all users that have a subscription for the corresponding data. The data in an MQTT-based environment are organised as topics. These topics can be freely defined and allow the grouping of messages into useful thematic channels.

When new data are available, a publisher (e.g. a sensing device or a data centre) will publish them as a message on one or more suitable topics. Users that want to receive specific data subscribe to the topics they are interested in. Thus, if a broker receives a new message for a certain topic, it will check for the topic's subscribers and will then forward these messages via the corresponding topics to all subscribers.

This approach provides several advantages which have led to significant practical usage of MQTT. The push-based approach for data delivery helps to minimise latencies for data delivery ([19]). Furthermore, MQTT is a rather simple and lightweight protocol, which facilitates its implementation and use within constrained environments (e.g. limited bandwidth/connectivity).

The first MQTT-based resources of real-time data are already available via data.europa.eu. However, there is still no common approach to provide metadata about open MQTT brokers and the topics they offer. Thus, improving the discovery of open MQTT brokers will require additional work.

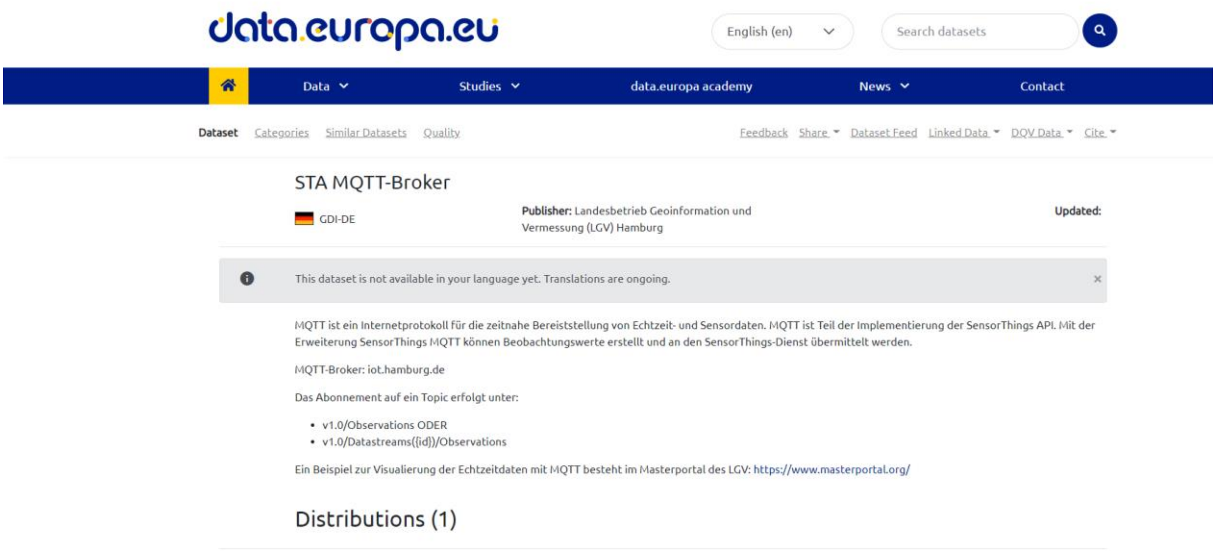

Figure 6: An MQTT broker in a search result on data.europa.eu

5. Discussion and feedback gathered during the webinar

The webinar organised on 5 April 2022 was also used to gather feedback from the participants on the usage of real-time data, current gaps in data availability and needs for future improvements.

There is already quite a broad range of use cases in which real-time data are regularly used. Several participants are already actively using real-time data, especially in the context of traffic information

(19) https://publications.jrc.ec.europa.eu/repository/handle/JRC127730

(parking space availability, live schedules of public transportation, information on traffic flow/traffic jams). The use of real-time data was also frequently mentioned for live updates on weather events. Other examples included parcel tracking, air pollution information, information on allergens (e.g. pollen), flood monitoring and stock market information.

The webinar participants also provided inputs on desired data sources for the future. Among these, mobility related data formed a significant cluster. This includes more detailed traffic information for all roads, car charger availability, availability of mobility services/taxis, taxi pooling, fuel prices and car sharing. Besides this, further information needs were mentioned regarding energy generation/consumption data and data to enable the comparison of commodity prices. When analysing these requirements, it becomes apparent that some of these data sets are already available in some countries, but the availability and discoverability of such data are still quite limited. Thus, the publication of corresponding data sets – including useful metadata – should be encouraged and supported.

This also relates to typical problems mentioned by participants to discover real-time data sets. Common issues are missing or incomplete metadata and difficulties to find good approaches to search for the needed data (e.g. lack of common search terms, lack of a dedicated search functionality). The participants also noted that there is still a general lack in the availability of real-time data and that data are often provided in a very heterogeneous manner, lacking the use of standards. Also, in several cases, it was not possible for users to differentiate between static data and dynamic real-time data. For example, real-time data may be provided as JSON files, but a suitable indication that the underlying source is continuously updated is missing. Here it would also be useful to explicitly describe the APIs through which data are offered, so that users can recognise not only the data format but also the dynamic nature of the data. Furthermore, the usability of real-time data was criticised by participants (e.g. asking for simple applications which encapsulate the access to the data).

From a technological viewpoint, several recommendations were made for technologies beyond the ones that were presented. These include Websockets ([20]) and the W3C Web of Things ([21]). A general recommendation was made to especially consider open standards for brokers and data delivery.

There were also suggestions on how to improve the exploration of real-time data, in order to enhance the preview functionality of the geo-spatial viewer component of data.europa.eu. This particularly includes the provision of context information for real-time data streams (e.g. related data streams), the provision of dashboard functionality with different visualisation options (e.g. diagrams, sensor locations) and the availability of more metadata about the data streams (e.g. periods for which data are available).

([20]) https://datatracker.ietf.org/doc/html/rfc6455
([21]) https://www.w3.org/WoT/

6. Conclusions and key points of attention

Based on the discussions and feedback gathered during the webinar, several recommendations were derived for future work requirements. These potential action items are summarised in a compact form below, ordered by the timeframe in which they can/shall be addressed.

6.1. Short-term actions (2022)

Collect use cases and derive requirements/opportunities

- Timing: 2022.
- Topics to investigate.
  - Create a systematic overview of relevant use cases (e.g. based on input from the webinar and other sources). Data sources within the focus of interest include:
    - traffic information (short-term):
      - roads,
      - car charger availability,
      - availability of mobility services/taxis,
      - taxi pooling,
      - fuel prices,
      - car sharing;
    - energy generation/consumption (medium-term);
    - data to enable the comparison of commodity prices (long-term);
  - Derive concrete needs for additional real-time data and identify potential providers.
  - Approach selected data providers and motivate them to make their data accessible.
  - Support the provision of metadata by offering best practices for metadata to relevant parties (see next action).

Support for data providers

- Timing: 2022.
- Topics to investigate.
  - How to facilitate the publication of real-time data? Which tools are especailly suitable for sharing real-time data? Create corresponding guidance documentation.
  - How can the generation of metadata for real-time data be facilitated? (e.g. suitable templates and best practices for metadata documents, see below).

Develop best practice for metadata for real-time data sources

- Timing: 2022.
- Topics to investigate.
  - How to best describe real-time data sources in DCAT-AP?
  - How to differentiate between data formats and the corresponding APIs? An approach is needed to describe both the API and the offered data format (e.g. to indicate that a JSON data source offers real-time data).

   - Which APIs/data formats should explicitly be mentioned in a metadata document in order to facilitate the consumption of data?
   - How to provide metadata for resources such as MQTT brokers and the data streams they are offering?
   - Technologies to be considered when developing best practice recommendations:
     - OGC SensorThings API,
     - MQTT,
     - FIWARE Context Broker,
     - W3C Web of Things,
     - Websockets.

6.2. Medium-term actions (2023)

Investigate approaches to facilitate the discovery and preview of real-time data sources

- Timing: 2023.
- Topics to investigate.
   - Which are the most useful APIs / data formats for real-time data which should be supported data preview? Candidates include MQTT, FIWARE Context Broker, OGC SensorThings API.
   - How to enable the preview of a (limited) history of real-time data streams (e.g. live diagrams)?
   - Are there ways to allow users to specifically search for real-time data sets? For example, is it possible to include dedicated flags in the metadata?

Increase the amount of published open real-time data streams

- Timing: 2023.
- Topics to investigate:
   - analyse which policies might help to further increase the availability of open real-time data streams (e.g. Open Data Directive);
   - how to make it easier for data providers to publish real-time data (e.g. recommendations for tools and standards to reduce the necessary efforts for data publication);
   - which success stories and use cases exist that might motivate data publishers.

7. Bibliography

Cox, S., *OGC Abstract Specification – Geographic information – Observations and measurements* (OGC 10-004r3), Open Geospatial Consortium, 2013 (https://portal.ogc.org/files/?artifact_id=41579).

European Commission and Council of the European Union (2007), Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (http://eurlex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML).

European Commission and Council of the European Union (2019), Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information (https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024).

Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., Van Den Broecke, J., Sarretta, A., Verhulst, S., Dencik, L., Oost, H., Micheli, M., Minghini, M., Kotsev, A. and Schade, S., 'Emerging approaches for data-driven innovation in Europe', EUR 30969 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-46936-0, doi:10.2760/511775, JRC127730 (https://publications.jrc.ec.europa.eu/repository/handle/JRC127730).

Liang, S., Khalafbeigi, T. and van der Schaaf, H., *OGC SensorThings API Part 1: Sensing Version 1.1 (OGC 18-088) Implementation Standard*, Open Geospatial Consortium, 2021 (https://docs.ogc.org/is/18-088/18-088.html).

Schleidt, K., Van Der Schaaf, H., Mossgraber, J., Grellet, S. and Oliveira, N., 'Evaluation of Application Programming Interfaces for INSPIRE', Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-49644-1, doi:10.2760/00811, JRC128885.

Publications Office
of the European Union

ISBN 978-92-78-43132-7