Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud

Studies and Reports

Analysis of the State of the Art, Barriers, Needs and Opportunities for Setting up a Transport Research Cloud

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the internet (http://europa.eu).

Analysis of the State of the Art, Barriers, Needs and Opportunities for Setting up a Transport Research Cloud

Expert Group

Final Report

Martin Böhm, AustriaTech

J. Rod Franklin, Kühne Logistics University – Chairperson

Sarah Jones, Digital Curation Centre

Tatiana Kovacikova, University of Žilina

Katarzyna Nowicka, SGH Warsaw School of Economics – Rapporteur

George Yannis, National Technical University of Athens

Table of Contents

EXECUTIVE SUMMARY

The transport research community is composed of a diverse group of researchers *conducting research in academic institutions, public-private research institutions, commercial organizations, and public institutions. The community examines a* ***highly diverse set of issues*** *from urban transport operations to commercial people and freight movements. The modes of transport that they focus on include air, sea, land, rail, space, pipelines, rivers, and any other new or emerging approaches to movement and carriage (hyper loops, drones, etc.).* ***Transport researchers collect and generate large amounts of data whether from monitoring actual freight/person movements, recording sensor data*** *from vehicles and infrastructures, or capturing video of various transport related phenomena. They come from different background and apply a wide range of methods: transport research is* ***inherently cross-disciplinary*** *and so provides an ideal context in which to apply principles such as FAIR. Unfortunately,* ***most of the data that these researchers collect is used once and then stored away in locations that are inaccessible to other researchers****.*

This report focuses on the requirements for data sharing within the transport research community*. In particular, the report examines the* ***potential of a Transport Research Cloud (TRC) as a subset of the European Union's European Open Science Cloud (EOSC) initiative.*** *Six domain experts collected data based on their personal experiences, contacts, prior research and a survey sent out to other researchers in the transport domain to enable a preliminary analysis concerning the needs, barriers and potential benefits for the domain should a TRC be realized. From this work ten recommendations, grouped into five broad topic areas, have been developed that the Experts believe must be addressed if a sustainable TRC is to be realized. The five topic areas are organized in a rough precedence order that indicates a logical flow in time. However, this order is not necessarily required for the implementation of all of the recommendations as many can be performed in parallel.*

Reusable research data:

1. *The Commission should bring together researchers, research data users, and data generators to* ***define what constitutes transport research data****.*

2. *The Commission should conduct a detailed study among transport researchers in order to* ***identify the objections behind limited use of data collected by others*** *and develop recommendations on how to overcome these objections.*

Data as a public good:

3. **Any and all data collected under contracts that are paid for by tax payer funds by default should be classified as public data.**

4. All data collected under a publically funded project should include a **clear demarcation between the Intellectual Property created by the individual researchers in analyzing the data and the data itself**.

Standards:

5. The Commission should bring together members of the transport research community, governmental entities that generate transport data used in research, infrastructure operators, and commercial consumers of transport research and transport research data to **define the standards that will be necessary for the collection of transport data** by public institutions, the data formats these data should adhere to, the metadata that must be used to describe the data, and formats of this metadata so that automated search engines can easily find and characterize the data.

Infrastructure:

6. The Commission should conduct a detailed study on **the infrastructure and operating requirements for a TRC** to ensure that an appropriate level of service can be provided at a cost that is understood by all stakeholders.

7. The Commission should conduct a detailed study of **what the potential user and stakeholder communities would require from a TRC in order to make it the "go to" place for doing cutting edge transport research**.

8. The **TRC should be modelled after the EOSC and governed in a manner similar to the EOSC** taking into account the particularities of the transport domain.

Incentives, training, and education:

9. EU policies for academic promotion, training, publication, and knowledge generation at public universities (and private degree granting universities) should be examined and harmonized to ensure that **researchers are**

uniformly trained in the requirement and process of placing their research data into the EOSC (and its subsidiary cloud infrastructures), that universities provide the proper (and uniform) incentives to their faculty and researchers to ensure that their research data (if not constrained by commercial non-disclosures or other requirements for privacy) is placed into the EOSC, and that proper credit for the generation of data that is reused is given to the individuals who originally collected the data (note that where universities use journal citations as an element of promotion references to who collected the data should be looked at in a manner similar to any of the authors of the paper).*

10. *An analysis of training requirements should be conducted by the Commission and, based on these requirements, **training/education programs developed for existing researchers and future researchers, libraries and librarians, data curators, and other individuals** who will be needed to carry out the development of a mind-set of open data by default.*

The Expert Group believe that if these recommendations are followed, then the transport research community will become more open to the concept of "open data by default" and the TRC will have the greatest chance of being a sustainable and high value asset to the transport research community and European Union. At the same time the Experts take the position that open data in general, and open transport research data in particular, must be elevated on the agenda of the EU's political bodies to ensure that the EU continues to be the world's leader in transportation research.

1 Introduction

1.1 Vision of TRC

Since the turn of this century a growing number of governments and funding agencies have begun to promote the sharing of research data to make research results more widely available for other researchers, education, business, and other purposes[1]. This interest in the sharing of research data has led to policies that promote open access to data from observational research networks, governments, and other publicly funded agencies[2]. Many private foundations have also begun to encourage or require the release of data from research they fund[3]. Data sharing policies vary widely by research domain, country, and agency, but have many goals in common. Many of these policies have been analysed and resulted in grouping the arguments for data sharing in four general categories:

1. To reproduce research: Numerous commentators have expressed concern that much of the research that is published today cannot be reproduced by other scientists and, therefore, this lack of reproducibility has led to a crisis of credibility in science [4]. To address this "reproducibility crisis," research organizations, funding bodies, and researchers themselves have indicated that a more open sharing of data and research methodologies is needed to ensure that reported research outcomes can be reproduced and the credibility of the scientific endeavour maintained[5].

2. To make public assets available to the public: A strong social argument has been made that data generated by public entities, or through publicly funded research activities, should be available to the public who paid for its creation[6]. In addition, similar arguments have also been made that open government is facilitated by the open sharing of publicly developed data[7].

---

[1] https://openaccess.mpg.de/Berlin-Declaration
[2] https://ec.europa.eu/digital-single-market/open-access-scientific-information
[3] https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy
[4] "Trouble at the Lab," The Economist, October 19, 2013, pp. 23-28.; Grand, J.A., Rogelberg, S.G., Allen, T.D., Landis, R.S., Reynolds, D.H., Scott, J.C., Tonidandel, S., Truxillo, D.M., (2018). A systems-based approach to fostering robust science in industrial-organizational psychology. Industrial and Organizational Psychology, 11(1), pp.4-42; Landis, S.C., S.G. Amara, K. Asadullah, C.P. Austin, et al., (2012). "A Call for Transparent Reporting to Optimize the Predictive Value of Preclinical Research," Nature, vol. 490, pp. 187-191.
[5] McCormick, D., Moore, K., (2016). "Report on the First IEEE Workshop on The Future of Research Curation and Research Reproducibility," John Baillieul chair, IEEE, 5-6 November.
[6] Duch-Brown, N. B., Martens, B., Mueller-Langer, F., (2017). The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01.
[7] Bertot, J.C., Jaeger, P.T., Grimes, J.M., (2010). "Using ICTs to create a culture of transparency: e-Government and social media as openness and anti-corruption tools for societies," Government Information Quarterly, vol. 27, no. 3, pp. 264-271.

3. To leverage investments in research: The European Commission has made a strong argument that publicly funded research data should be made openly available to researchers and entrepreneurs to ensure that the maximum return on publicly funded research can be realized[8].

4. To advance research and innovation: In conjunction with the Commission's desire to leverage investment in publicly funded research, the Commission also believes that by opening research data for further use the research agenda of Europe can be advanced and European leadership in scientific research can be maintained[9].

Data sharing, in its simplest form, is merely releasing or posting data. The proposed benefits of sharing can be achieved, however, only if the available data are understood and used or reused by others[10]. Thus, sharing the underlying research data is only one aspect of what is required to realize the benefits of shared data. To be reused successfully, information on how the data was collected, for what purpose it was collected, when and where it was collected, etc. must be provided so that interested researchers can understand the data and properly employ it in their research.

The European Commission is interested in the areas of open science and open data to ensure that money spent by the Commission on research is leveraged in a manner that generates returns (scientific, economic, environmental and social) well after the originally funded research has concluded. The benefits are expected to go far beyond academia to the public and commercial sectors, with tangible increases in economic growth and innovation being direct results from open science strategies[11].

The European Commission is committed to Open Science and has initiated several actions over recent years to develop the policy and infrastructure required to support this. Specifically, a vision has been put forward for a EOSC that federates existing infrastructure from research disciplines and member states to remove the technical, policy and human barriers to data sharing. The EOSC intends to provide a single point of access to all European research data and the world-class data services, tools and standards needed for reuse. The proposed TRC is envisaged as a thematic pillar of the EOSC, addressing the specific needs of the transport research community.

---

[8] Council of the European Union, Draft Council Conclusions on the European Open Science Cloud, 9029/18.
[9] Ibidem.
[10] Borgman, C.L., (2015). Big data, little data, no data: Scholarship in the networked world. MIT press.
[11] David, PA, (2003). "The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A primer," Stanford Institute for Economic Policy Research, No. 02-30, March; Manyika, J., M. Chui, P. Groves, D. Farrell, S. Van Kuiken and E.A. Doshi (2013), "Open data: Unlocking innovation and performance with liquid information," McKinsey Global Institute, vol. 21.

Two High Level Expert Groups (HLEGs) have issued reports on the EOSC. The first laid out a strategic vision[12]. Recognising that most of the challenges are social rather than technical, the report highlighted the need to urgently develop adequate data stewardship capacity in European Union Member States and rethink recognition and reward structures to encourage FAIR[13] and Open data. The results of a survey conducted for this report indicate that a culture and practice of data sharing still needs to be developed in the area of transport research. The second HLEG EOSC report builds on the European Commission's Strategic Implementation Roadmap by focusing on the practical measures needed to implement the European Open Science Cloud. It proposes a business model to finance the EOSC and Rules of Participation for service providers[14]. These recommendations have helped to inform our discussions and recommendations around potential models for a TRC. Several projects funded by the EC, namely the EOSC Pilot[15] and e-Infrastructures such as eInfraCentral[16], EOSC-Hub[17] and OpenAIRE[18], are defining the governance framework and establishing initial EOSC services for 2020.

The future vision for the EOSC is to federate existing national and European infrastructure investments and assign "cloud coins" as part of grant proposals to enable researchers to purchase data services from approved suppliers. Transport research providers must engage with these initiatives so appropriate tools and services that meet the needs of the community in terms of controlled data sharing and private clouds are available to users within the EOSC. There may be some hurdles with this approach if services have been established in national interests and preclude cross-territory data sharing and reuse, but providers should be encouraged to meet the EOSC Rules of Participation so the transport research community has adequate access to the EOSC for the conduct of their research.

To enhance the use of transport research data, there is a need to establish an EU TRC in line with the EU's open science efforts and the EOSC. A primary aim for a TRC will be to provide researchers in the transport and logistics domain with access to open data sets covering topics of importance to their research. Current approaches to data access, which are ad hoc, based on personal relationships, and individual knowledge of where data resides, are insufficient to provide the broad access to properly curated data needed by researchers in this domain.

The TRC presents numerous opportunities for the domain. If data can be made open and easily accessible, there will be much greater availability of relevant datasets and wider opportunities for reuse and remixing. This will facilitate research across communities and countries, advancing the state-of-the-art in the

---

[12] EC (2016). Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. European Commission. Directorate-General for Research and Innovation. Brussels.
[13] A set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.
[14] EC (2018). Prompting an EOSC in Practice. Interim report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud. Brussels.
[15] https://eoscpilot.eu/
[16] http://einfracentral.eu/
[17] https://www.eosc-hub.eu/
[18] https://www.openaire.eu/

field more quickly. If the TRC becomes the go-to place for datasets in the field, it could also facilitate more public-private partnerships as commercial companies are encouraged to make their data available and research teams do not have to approach data "owners" individually and make separate agreements for reuse[19].

1.2   The main goals of the TRC Group

The aim of this report is to analyse the state of the art, barriers, needs and opportunities for setting up a TRC in the EU. The Expert Group was tasked with:

- Identifying the main needs, obstacles and opportunities for data sharing and the open science cloud model in the area of transport research, in particular with regard to data protection, security, FAIR data principles, governance (including public – private partnerships to exploit data), funding models, and competitiveness and IPR issues.

- Identifying the scope and the characteristics of data to be included in a possible TRC with particular attention to the distinction between research and operational data and IPR issues.

- Identifying data mining and analysis requirements and tools to maximise the usability of big data within the framework of the TRC.

- Identifying a common set of cloud-based services with application to a possible TRC.

- Assessing the relevance of international dimension of a possible TRC.

- Providing recommendations for different options of funding of a possible TRC.

1.3   Methodology

The six experts convened to advise the Commission on a TRC have drawn on their personal expertise and networks to define current practice and requirements. The report is based on their recommendations, supported by different types of empirical studies. An initial analysis of existing open science practice and support was provided on a country-by-country basis and is presented in Attachment I, the Global Approaches to Open Transport Data describing different national or local approaches is included as Attachment II. Two case studies were developed from the key research areas of Mobility as a Service and Automation in Road Transport. These case examples inform potential models to explore to better understand the characteristics of transport data – they are presented in Attachment III. Additionally, a broad-based survey was undertaken in the summer of 2018 to determine existing data documentation and sharing practices of transport researchers, and what the transport research community would expect from a cloud service. The survey was conducted among

---

[19] Janssen, M., Y. Charalabidis, A. Zuiderwijk, (2012). "Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems and Management, vol. 29, no. 4, pp. 258-268.

researchers from academic institutions, representatives from public bodies and the commercial sector.. Data collected through this survey was anonymised and used to inform recommendations and published under an open licence. The survey consisted of four parts: transport research data, cloud service requirements and expectations, opportunities and barriers and funding mechanism. A total of 87 responses were collected between June and July 2018. Respondents came from 29 countries – primarily from the United Kingdom (11%), Slovakia (9%), Greece (9%), Germany (7%), Austria (7%), Israel (6%), Poland (5%) and Spain (5%). The detailed results of the survey are presented in Attachment IV. The results of this work have led to a number of recommendations that are presented within the content of this report to illustrate and support selected arguments. Parts of the survey results are also shown in the content of the report to draw out specific needs, requirements and recommendations for future work. Additionally in the Attachment V. Minimum Metadata Elements as described by the tri-national cooperation between Germany, the Netherlands and Austria are presented.

1.4   Report Structure

This report consists of four chapters that describe the most important issues identified by the Expert Group. After the introduction, where the vision of the TRC and Group goals are presented, we explore, in the second chapter, the typology and the nature of transport data, metadata and data standards, approaches to open data support, and opportunities and barriers to transport data sharing. In the third chapter we consider the needs for data sharing, revising policy, infrastructure construction, cost coverage, and a sustainable business model. In chapter four we propose recommendations for the development of the TRC.

2 The data culture of the transport research community

2.1   Definition, sources, scope and characteristics of transport data

To identify the scope of the TRC, it is important to understand the dimensions of the transport sector, which also explains the complexity of transport related research. Transport sector dimensions consist of[20]:

- Transport modes: road, rail, air, maritime, inland waterways, space and pipelines with each mode characterized by a set of technical, operational and commercial characteristics.

- Types of "intermodals:" inter-modal transport (several modes of transportation with separate tickets/contracts), multi-modal transport (several modes of transportation with one ticket/contract), trans-modal

---

[20] Rodrigue J. P. (2017), The Geography of Transport Systems, Fourth edition, New York: Routledge. ISBN 978-1138669574

transport (e.g., ship-dockside-ship), and synchro-modal (integrated scheduling and operation between different modes of transport).

- Transport sectors: passengers and freight each having different options for transport mode and transport means, e.g., passengers using road transport can use a car, bicycle, walk, etc., similarly with freight.

- Vehicles: each mode of transport has several types of vehicles that act as the means of transport, and producers that are focused on differing propulsion systems for each means.

- Geographic area: urban transport, trans-border / cross-border transport, rural transport, international transport, etc.

- Infrastructure: road, rail, ports[21], air traffic control, etc.

- Evaluation perspectives: long-term views, decision support methodologies, the environment, economic and regional impacts, accessibility, social and equity impacts, etc.

- Policy aspects: financing, pricing and taxation, regulation, competition and public services, infrastructure and TEN-T, transport planning, climate policy and energy efficiency, security and safety, international and EU collaboration, awareness information and user rights.

- Technology: ITS, IoT, connected cars and autonomous vehicles, transport management systems, big data (machine learning, artificial intelligence),[22] propulsion systems, V2V, V2I, V2X, etc.

- Applications[23]: e.g., asset management, capacity planning, construction, design and planning, emergency response, estate management, journey planning, maintenance, navigation, etc.

- Types of data: traffic data, infrastructure data, passenger data, sensor data, tracking data, social networks data, ticketing and fare data, scheduling and asset management data, environmental data, safety data, etc.

---

[21] Note that the use of the "ports" in this document covers all port types including air, sea, inland waterway, inland, etc.
[22] Leveraging Big Data to Manage Transport Operations (LeMO project), https://www.big-data-europe.eu/wp-content/uploads/5-BDE_SC4_WS3_Arnaud-Burgess-LeMo-project.pdf
[23] INSPIRE Data Specification on Transport Networks – Guidelines: http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_TN_v3.0.pdf

Numerous stakeholders will be involved in making the TRC a success. Infrastructure hosting institutions, governments (local, regional, national and international), researchers, societal bodies such as NGOs and community organizations, foundations, commercial transport and logistics industry players, private researchers, and many others all will have an interest in either sharing data, using data, providing funds or reviewing and commenting on the success of the TRC. Managing the complex ecosystem necessary for the TRC will be a challenge due to these many partners, but one that must be met if a TRC is to be a success.

Transport related data sets will include data from governmental entities such as cities, regions, countries and federated communities, where transit, traffic, safety, and other operational data are regularly collected. Commercial data may also be included as commercial transport companies, service providers, vehicle producers, etc. gain interest in the TRC. In addition, data from governmentally funded research projects in the area of transport and logistics will be deposited in the TRC. These data sets should be provided by the TRC to interested researchers for reuse in a curated and open manner to address their need for access to broad detailed data on transport operations in order to tackle the increasingly difficult problems of efficiency, effectiveness, safety, environment and social impacts that face the domain.

Data owners and data users

It should be noted that for all three types of transport research data (original research data, operational data and publications) there are different types of **data owners** and **data users**. There are owners and users from the public sector (European, governmental, regional, local authorities, research and HE institutions), private sector (companies, industry, research and HE institutions), and citizens. Taking into account that transport data is often ethically or commercially sensitive, it is evident that tight controls are needed around access to the data, with appropriate adjustments with regards to data ownership and use.

A fundamental issue for the development of the TRC is to identify what data is considered to be "transport research data" in contrast to transport data in general. This identification will require more detailed analysis of the types of research problems transport researchers are working on today so that a proper taxonomy of "transport research data" can be established. Currently, defining what classifies as research data is quite challenging, since there is no consensus on the definition and the definition may vary according to the scientific discipline or the research funder. Definitions of research data commonly used are[24]:

- Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.

---

[24] University of Leicester, (2018). Research Data. Retrieved May 30, 2018, from https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data

- Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.

Transport data are characterized by a complex, multilevel topology corresponding to the various aspects of transport research, planning, design and operation. The potential outline of transport research data classification must refer to the different transport modes (e.g., road, rail, marine, air, space, pipeline, and multi-modal) and types (e.g., persons/freight, urban/interurban/rural, domestic/international transport, commuting/school/recreational, etc.). Focusing only on the means of transport, however, will miss other areas of transport related research that enable and define the efficiency, safety, cost/value, environmental impacts, and security of transport operations. Finally, any classification will need to cover all phases of transport project lifecycles (planning, design, implementation, operation and management) and concern all categories of content (raw and processed data, research outputs and publications).

Taking into account the multiple characterizations of transport data, while all transport data could potentially be used for research purposes, it is neither feasible nor advisable to attempt to include all data in the TRC. It is therefore suggested that, at least as a first stage, only data directly related to current definitions of transport research are incorporated in the cloud. Research related data do not necessarily include business related data (e.g., data from car manufacturers on vehicle technology development, from airlines, from logistic companies on freight volumes, etc.), which are often also very difficult to obtain due to proprietary issues.

Therefore, the following three main categories of data are suggested to be included in the TRC:

1. **Original research data**, e.g., data from Field Operational Tests (FOTs), Naturalistic Driving Studies (NDS), research results and research models from published papers.
   This category of data will most probably constitute the core of the TRC and, for each research project that submits its data to the TRC a minimum of information will be required. This information should include at least the following:

- Experiment context (possibly included as metadata);

- Data file description (possibly included as metadata);

- Data access instructions (possibly included as metadata);

- Raw dataset(s) used in the research;

- Transformed and aggregated datasets: cleaned-up, derived, annotated data, generalized graphs and tables, data combined with other datasets, etc.;

- Research models and research results.

For research data a number of public and private sources can be identified. Some of these sources are:

- EC funded research projects[25] - all types of H2020 Actions (probably also previous FP Programs) – Research and Innovation Actions, Innovation Actions, FOTs, NDSs (e.g., U-DRIVE), Research Infrastructures and e-Infrastructures.

- Non-governmental and government projects and initiatives such as Transforming Transport (TTT)[26], Big Data Europe (BDE)[27], NOvel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services (NOESIS)[28], DATA science for SIMulating the era of electric vehicles (DATA SIM)[29], etc.

- Industry and research projects and initiatives: "On-Road Integrated Optimization and Navigation system" (ORION)[30], Railigent – Siemens[31], etc.

- Various libraries of research activities, either operated by public entities (e.g., TRIMIS[32]), by public research entities (e.g., universities), private institutions (e.g., publishing houses or research institutes such as the Fraunhofer Institute), or in public-private operated associations (e.g., ERTICO[33]).

- Other international research projects.

- University projects and initiatives: MIT SENSEable City Lab[34], Goethe University Frankfurt Big Data Lab[35].

2. **Operational data directly related to research**, such as accident data, transport volumes data, etc.
   This category of data consists mainly of data from public authorities, either national or European/ international. Publicly available datasets could include data on public roads[36], traffic accidents, road lengths, vehicle registry,

---

[25] https://ec.europa.eu/info/research-and-innovation/projects/project-databases_en

[26] https://transformingtransport.eu

[27] https://www.big-data-europe.eu/pilot-transport

[28] http://noesis-project.eu

[29] https://cordis.europa.eu/project/rcn/100232_en.html

[30] https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/UPS-On-Road-Integrated-Optimization-and-Navigation-ORION-Project

[31] https://www.siemens.com/global/en/home/products/mobility/rail-solutions/services/digital-services/railigent.html

[32] TRIMIS, (https://trimis.ec.europa.eu/) is coordinated by the European Commission

[33] ERTICO (www.ertico.com) is operating the ITS Observatory (https://its-observatory.eu/)

[34] http://senseable.mit.edu/shareable-cities

[35] www.bigdata.uni-frankfurt.de

[36] Yin, H., Berger C., (2017). "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets", In Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference, pp. 1-8.

driving licenses registries, public transport[37], road/ rail/ port/ air traffic, etc. In this category of data, data anonymization, data security and user authorisation are of paramount importance.

For operational data, potential sources are Eurostat, national statistical agencies, national governmental bodies (e.g., ministries or general secretariats of transport, of infrastructures, etc.), cities and localities, port authorities and aviation authorities.

3. **Data from published transport research** appearing in scientific journals, delivered at conferences, workshops, etc.
Data from transport research related publications in scientific journals, conferences, workshops etc. could also constitute a significant asset of TRC. Currently, data from published research should be available (although not always for free) through the websites and portals of the various publishers. Searching for these datasets, and the publications they are associated with, is usually limited to the publisher and a comprehensive central search engine does not exist[38]. Incorporating transport research datasets from publications into the TRC will provide for advanced search services and, more importantly, will enable linking publications to the datasets, models and research results of actual research projects.

2.2   Metadata and data standards

A cloud-based service provides data archivists with the possibility to make data, which is located in different physical locations, available to researchers anywhere in the world. The main focus of such a service concerns the distributed collection and distribution of data. The core elements to enable such a service are the need to understand the content and structure of the data being managed. Concerning the content, a detailed description of the data available is key, there must be catalogues (e.g., for datasets, services, standards) based on machine readable metadata and identifiable by means of a common and persistent identification mechanism available to ensure that the data can be found, understood, and accessed by interested researchers[39]. High quality metadata (for instance the conditions under which data has been collected, for which purpose, how it has been stored, processed, and how it can be accessed) enable data reuse by

---

[37] https://www.europeandataportal.eu/data/en/dataset?tags=public-transport
[38] Note that Google Scholar (https://scholar.google.com/) attempts to perform this function. However, its U.S.A. and English focus (it does have some European research focus as well) make doing thorough international searches for published research difficult. Google's recent release of its Google Dataset Search service (https://toolbox.google.com/datasetsearch), while an interesting initial attempt at developing a comprehensive dataset search engine, still is very U.S.A. focused.
[39] EOSC Declaration, (2017). EOSC Declaration. European Open Science Cloud. New Research & Innovation Opportunities. Brussels.

providing precise information on the data itself and documenting data to a level where an assessment of the level of validity can be performed[40].

The European Commission has published the DCAT-AP[41] (application profile for data profiles in Europe) specification, which is followed by several transport data access points[42]. The DCAT-AP specification defines how metadata elements should be described, especially when it comes to:

- data elements (description of a dataset in a minimal but adequate way),

- wording and semantics,

- predefined categorisations,

- data field names,

- data value types,

- data field lengths.

The specifications for these descriptions must be followed by all data owners to ensure automatic machine reading via cloud-based services. While it remains as one of this report's recommendations to identify or develop an appropriate metadata standard for the TRC, based on the principles defined by the DCAT-AP specification, a metadata description for data generated from the transport domain should include at least the following major categories:

- metadata information (including date of creation of metadata, metadata language),

- content information (including name of dataset, description of dataset),

- temporal information (publication data and eventual date of expiry),

- geographical information,

- contact information (including data ownership),

- conditions for usage,

- access information (including data format, data structure, access URL),

- quality information (including update frequency, data collection methodology).

---

[40] Gellerman, H., Svanberg, E. & Barnard, Y. (2016), "Data sharing of transport research data," Transportation Research Procedia vol. 14, 2227 – 2236.
[41] https://joinup.ec.europa.eu/release/dcat-ap-v11
[42] E.g., the German National Mobility Data Warehouse (MDM - https://www.mdm-portal.de/) or the Dutch National Data Warehouse (http://www.ndw.nu/en/)

The national access points of Germany, the Netherlands and Austria, working in accordance with Delegated Regulation EU 2015/962 of the European Parliament, developed a minimum set of data elements based on the DCAT-AP standard for the description of data generated by real-time traffic information services. A copy of this metadata specification can be found in Attachment V.

When it comes to transport research data, information on the original purpose of data collection, population from which the sample was selected, and the structure of data is also required so that future researchers using the data can properly include it in their research methodology. Feedback from the survey of transportation researchers conducted by the expert team for the TRC confirms that the metadata discussed previously is necessary if researchers are to use the TRC to identify potentially useful data sets for their research (Figure 1).
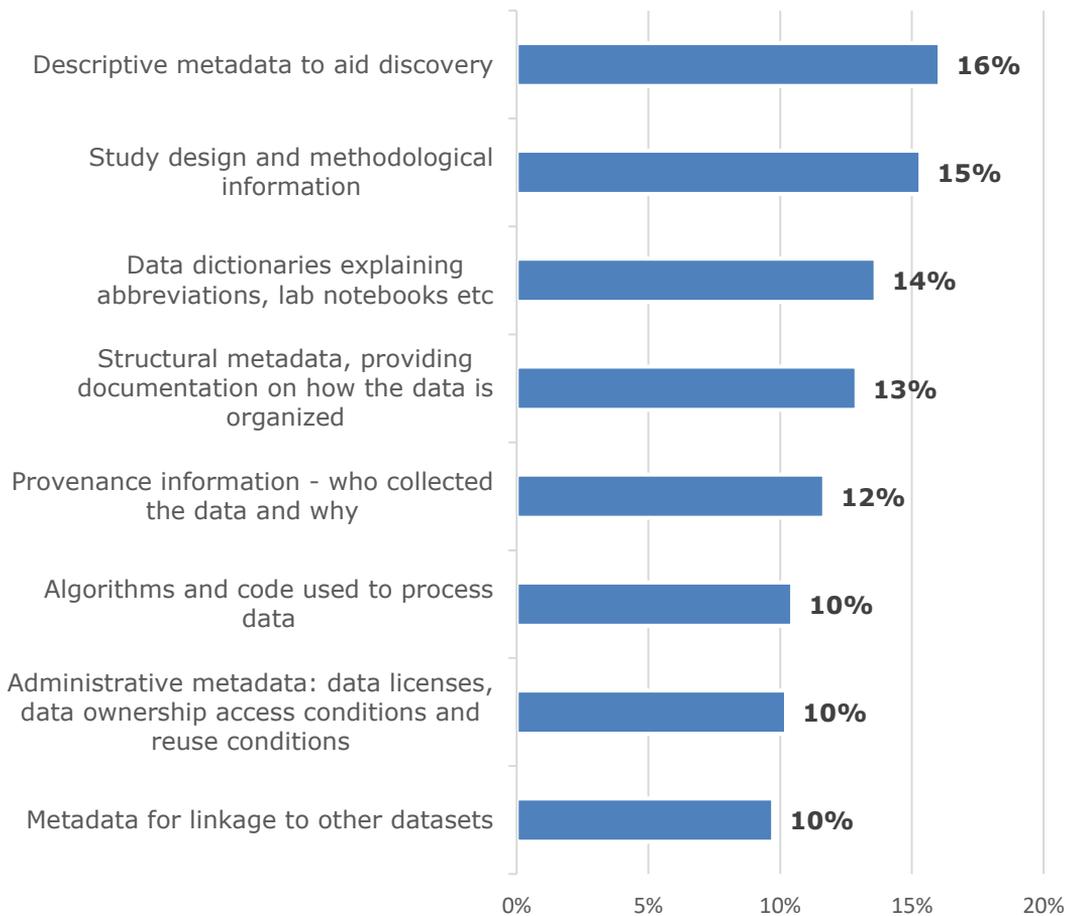
Figure 1. Metadata Requirements

An important issue for consideration, related to both research data and operational data, is the size and complexity of datasets. FOTs and NDSs, for example, may involve hundreds of hours of video footage and related data collection in controlled access facilities or on public road networks, especially when continuous data-logging is favoured over event-based data collection. These datasets and other "big data" transport research efforts require ample storage space and powerful computers to process the data and reveal patterns that may lead to knowledge extraction. Processing power and machine learning algorithms, therefore, are a vital element for big data transport research. Curating, storing, and handling these large, unstructured datasets will also require specialized databases, data management systems and infrastructure to ensure that access and reuse are made as simple as possible. These requirements for the TRC must be considered at the outset to ensure that valuable, but complex data concerning transport activities can be made available to interested transport researchers.

The re-use of transport research data may require processed, cleaned and annotated versions of original datasets. If the data reuser was not involved in the original project, does not know how the experiments were performed in detail or how the processed data were derived from the raw datasets, he/she cannot verify data validity unless the data are sufficiently described. Data re-use, therefore, requires knowledge about the data itself. This implies additional elements are required for any metadata standard used for transport research data to enable advanced search options for the proper interpretation and filtering of data based on criteria appropriate to high quality transport research.

The various elements of a transport specific metadata should also be composed based on global standards in order to ensure the broadest use and understanding for these data. Currently, the primary metadata standards that are being employed for documenting transport data are the ISO/IEC 11179 standard and the EU's DCAT-AP standard, along with the ISO 14817 standard for ITS data dictionaries and the ASTM E2468-05 standard for metadata for archived data. These standards should be evaluated in light of the needs of the transport research community and adopted, modified or replaced depending on whether they are determined to be acceptable for the broad requirements of this community.

Besides metadata describing the datasets, the format of the dataset is important. Data is often collected and stored in proprietary formats, which hinders other researchers in re-using the datasets. Even though there are regulatory requirements for the collection of certain transport related data requiring clearly defined formats, the fact that the data may be collected in incompatible formats is a barrier to making data reusable (even if the data is accessible). In accordance to the Delegated Regulations following the European ITS Directive

(2010/40/EU)[43] the following data formats are currently recommended for different ITS used in different transport domains:

- DATEX II for road transport data[44],

- NeTEx and SIRI for public transport data[45],

- TN-ITS[46] and Inspire[47] for geographical data.

While this list is currently valid, and used for the creation of traveller information services, these standards were not developed with the idea of providing researchers with access to research datasets. A more detailed analysis needs to be made of the formats before any recommendation can be made for them to become recommended data formats for transport research data. Such an analysis might lead to additional formats being recommended as standards for transport research data. In addition, there are different data formats used for different transport domains that make the linking of data across domains difficult. Without clear direction from the European Commission, an uncontrolled increase in additional data formats will create problems for realizing the vision of data reuse that the TRC is being proposed for.

Standardized data formats are key to enable the linkage between different transport related data stored on different cloud platforms. In terms of expectations to provide/access data to/in a cloud via specific data formats, participants in the survey conducted in conjunction with this report were asked to indicate which standards they expect to be used. The majority of respondents identified DATEX II as the standard to use. The following formats were also identified as being used by the respondents; NeTEx, GTFS, INSPIRE, BS7666, VICINITY and FIware. General Open Data standards was also a response indicating that, while non-specific, any standard used would need to focus on an "open" approach to the use and re-use of data.

In order to enable scientists to search, browse, review and access available data in the TRC, the development of new software tools will be required. Software and search tools will make use of the metadata that annotates and identifies the underlying data. Ideally, all research data will be available programmatically, through web APIs, so that it can be identified and accessed by search engines and automated systems[48].

---

[43] European Union Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010, "on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport," Official Journal of the European Union, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040.
[44] http://www.itsstandards.eu/rtd
[45] http://www.itsstandards.eu/pt
[46] http://tn-its.eu/
[47] http://inspire.ec.europa.eu/
[48] Science¦Business (2018), "How the Science Cloud could pay its way," Science¦Business Network's Cloud Consulation Group, Brussels.

Automated systems have become essential transport research assistants, both for data generation, data processing and analytics. Data formatting, terminology /identifier mappings, metadata and provenance must therefore be optimally organized in order to support machine processing as well as knowledge extraction. The tools supporting these two processes are fundamentally different, pattern recognition tools being mainly for machines and tools for confirmational reading and interpretation being mainly for humans. Machine actionability of whatever is published is therefore a crucial consideration in modern narrative and data publishing[49]. The FAIR principles also emphasize the need for machine understandable transport research data.

Besides the need for specific transport data formats and the ability to have machines process the data, a clear understanding of the benefit of linking cloud platforms together in a federated manner forms the basis for providing access to the data that is stored on those platforms. Linking cloud platforms together will require standardizing communications protocols between the various clouds, search approaches for accessing metadata concerning the data stored on the platforms, and upload/download mechanisms so that researchers can easily upload data and metadata to their preferred platform(s) and access cloud hosted datasets that they believe can be used in their research[50]. One possibility might be to link the platforms together using open-APIs, which would allow the platform managers to keep their systems "closed," but make data generally accessible. The specific approach to be used in federating these platforms needs to be examined more thoroughly to ensure that not only European platforms can be accessed in a seamless manner, but internationally hosted platforms as well.

2.3 Current approaches to open data support

There are practically as many operational models for open data platforms as there are platforms[51]. Domain focused platforms, governmental platforms, cross-domain platforms, research society platforms and numerous others exist. Each of these platform models have pros and cons, many of which have been discussed previously[52]. With respect to their applicability as frameworks for the TRC, they all have potential for implementation in managing open transport data. The key to which of the several models will be successful for the operation of any of the federated transport research platforms is that the platforms in question have well defined business models, a clear understanding of their stakeholder value

---

[49] EC (2016). Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. European Commission. Directorate-General for Research and Innovation. Brussels.

[50] Hey, T. Trefethen A.E., (2005), "Cyberinfrastructure for e-Science," Science, vol. 308 no. 5723, pp.817-821; Research Data Alliance (2015), "Sustainable Business Models for Brokering Middleware to support Research Interoperability," Sustainable Business Models Team Report to the Brokering Governance Working Group, Research Data Alliance.

[51] OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en.

[52] e.g. Danneels, L., S. Viaene, J. Van den Bergh (2017), "Open data platforms: Discussing alternative knowledge epistemologies," Government Information Quarterly, vol. 34, no. 3, pp.365-378.

propositions, and sufficient start-up funding to ensure that they gain traction with their stakeholders so that they can become relatively self-sufficient[53].

It is worth analysing existing models of international research collaboration in terms of their characteristics, operations, management and sustainability. Current international open data platforms have generally formed around a specific "big science" project in which data requirements are not confined to the borders of nations[54]. Many of these existing platforms have been in existence for a number of years, driven by the needs of the particular domain and research topic (e.g., CERN or SSRN). Recently, more domain agnostic platforms have begun to appear, particularly in the area of social sciences, where multiple disciplines have combined their data to generate integrated platforms for social scientific research[55]. Supra national governmental bodies, such as the United Nations[56], World Trade Organization[57], OECD[58], the World Bank[59], etc. have also contributed to the cross border, international open data initiative. Through their collection, curation and online provision of trade statistics, demographic data, and other searchable international data, these organizations provide examples of successful international open data portals.

In terms of operations, just as with national and local open data platforms, there are as many operating models for international open data platforms as there are platforms. Each platform has evolved operating models that address their particular user community and funding agencies. Examples vary from comprehensively curated data that is searchable with strong query and segmentation tools to simple data repositories that provide little by way of curation or tools. Most platforms reporting from recent studies indicate that, because of the reasons they were originally established, they are not prepared for the scalability and value added services that are implicitly envisioned when governments either require or promote open data services. This indicates that operational issues are potentially serious inhibitors in moving from these ad hoc formulated platform communities to the integrated/federated collection of open data platforms of the future[60].

When it comes to the management aspects of the various open data platforms, it can be stated that platform management varies by how and for what purpose the platforms were established. Platforms that were formed via lead universities tend to be managed through the library function within the lead university. Satellite platforms connected to these centralized university platforms also tend

---

[53] OECD (2017), "Co-ordination and support of international research data networks", OECD Science, Technology and Industry Policy Papers, No. 51, OECD Publishing, Paris, https://doi.org/10.1787/e92fa89e-en.
[54] Ibidem.
[55] https://www.icpsr.umich.edu/icpsrweb/
[56] http://data.un.org/
[57] https://www.wto.org/english/res_e/res_e.htm
[58] http://www.oecd.org/gov/digital-government/open-government-data.htm
[59] https://data.worldbank.org/
[60] OECD (2017), "Co-ordination and support of international research data networks", OECD Science, Technology and Industry Policy Papers, No. 51, OECD Publishing, Paris, https://doi.org/10.1787/e92fa89e-en.

to be managed by university libraries. Domain specific platforms, such as the CERN platform[61], have dedicated management structures that are funded through the platform partners. These dedicated management structures facilitate a more strategic/business focused understanding of the platform and the services it provides to its stakeholder community. Governmental platforms, and platforms developed based on ad hoc governmental requirements, are managed through either pure governmental departments or public/private partnerships. These types of platforms can have dedicated staff that manage the platform or temporarily assigned staff and volunteers managing the platform. Finally, pure ad hoc platforms are supported by limited dedicated staff and a significant number of volunteers. These types of platforms, while providing valuable services in the short term, have difficulties in maintaining continuity over the long term[62].

Finally, platform sustainability requires a business model that generates value for stakeholders of the

Open Data Cases

In order to provide a practical indication of potential scope and characteristics of transport data covered by the TRC, it is considered useful to examine a few indicative examples of open transport initiatives and databases.

An important initiative of the European Commission is the launch of the Transport Research and Innovation Monitoring and Information System (TRIMIS) which aims at becoming the EC's tool for mapping technology trends and research and innovation capacities in the transport field. TRIMIS incorporates an open-access information and knowledge management system that includes:

- Strategic Transport Research and Innovation Agenda (STRIA) **roadmaps** of innovation and new technology in transport,
- A searchable database of a large number of transport **projects** at European (FP4, FP5, FP6, FP7 and Horizon 2020), national and international levels. Depending on the status of each project, the following information is provided:
  - For "Ongoing" and "Complete" projects: information on origin and funding of the project, STRIA Roadmap and other transport themes, available contact point and a project outline including background, objectives and methodology.
  - For "Complete with Results" projects (in addition to the above): key results of the project, final reports and other useful project deliverables for download.

This section of TRIMIS is closely related to the concept of data to be covered by the Transport Research Cloud organised, however, on a project by project basis.

- A database of transport **programmes**, i.e. activities that finance or promote transport research.
- A database on **country profiles**, providing background information on the organisation of transport research across Europe and overviews for the Member States of the European Union.

A second example of an open data portal in the European Union is the Transport section of the **European Data Portal (**https://www.europeandataportal.eu/data/en/group/transport). It currently includes 7,210 datasets on a variety of diverse transport issues, however only with elementary search capabilities (by country, catalogue, tag, format and license). The portal has limited practical use due to the lack of standards and high quality metadata to describe the actual databases. As a result, although there are very useful transport data openly available in these datasets, it is actually very difficult to identify and locate them, to understand the context of the data gathering and development process and to assess data reliability and accuracy.

---

[61] https://home.cern/

[62] Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L. B. da Silva Santos, P.E. Bourne, J. Bouwman (2016), "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3.

platform and funding sources that, recognizing this value, provide long term funding for the platform. Unfortunately, these two items, a value based business model and funding from engaged sources that appreciate the platform's value proposition, are scarce commodities in the open data world[63]. **Very few platforms that have been studied to date, save for those that are funded purely by governments or supra-governmental bodies, have a truly sustainable business model and, therefore, funding sources. This factor is one of the key issues that must be addressed if open data and open science is to gain credibility and become the long term approach to scientific research**[64].

The findings of how international open data platforms are currently supported leads one to conclude that sustainability of the TRC will be an ever present issue. While it is easy to say that a sustainable platform must create value that is seen both by its stakeholders and users, it is far more difficult to identify what a platform's value proposition is and how its stakeholders and users will respond to this value proposition in funding the platform. Current recommendations on how the EOSC might become sustainable envision various "for fee" business models including "cloud coins"[65], subscriptions and pay-as-you-go[66], and continued government funding (recognized at least as an initial stage requirement by the Commission in Council of the European Union report 9291/18). Each of these ideas has the potential to act as a means of sustainable funding for the EOSC, however they do not define a truly sustainable business model for the EOSC. The difficulty arising in defining a sustainable business model for the EOSC (one that does not require continuous public funding) indicates that considerably more work is needed to determine what business model(s) are appropriate for the TRC to ensure sustainable operations of this domain element of the EOSC.

2.4   Opportunities and barriers to transport data sharing

Enabled by recent developments in Information and Communication Technologies (ICT), cloud computing, artificial intelligence, machine learning and the Internet of Things (IoT), a large amount of data is generated, collected, processed and used in research. By the end of 2020 it is estimated that there will be over 50 billion connected devices globally collecting over 2.3 zettabytes of data each year[67]. Shared and open data gathered from these devices can underpin transport solutions that support integrated, efficient and sustainable transport, transport systems and services. The data collected by these devices can also contribute to

---

[63] OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en

[64] Duch-Brown N. B., Martens, B., Mueller-Langer, F., (2017). The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01.

[65] Science¦Business (2018), "How the Science Cloud could pay its way," Science¦Business Network's Cloud Consultation Group, Brussels.

[66] EC, (2018). Prompting an EOSC in Practice. Interim report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud. Brussels.

[67] Statistica, https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/ and Cisco, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

positive economic outcomes as entrepreneurs and businesses leverage these data for value. However, for these data to be leveraged so that researchers, entrepreneurs and society can benefit requires that an open means of access and use be established. The TRC thus presents numerous opportunities for the domain. Open and easily accessible data will facilitate research across communities and countries, advancing the state-of-the-art in the field more quickly. It could also facilitate more public-private partnerships as commercial companies are encouraged to make their data available and research teams do not have to approach data "owners" individually and make separate agreements for reuse.

Some examples of where greater access to data could lead to advances in the area of transport research come for the increased use of FOT and NDS[68]. These approaches to understanding how vehicles and people interact create extremely large datasets that contain much more data than that which is used for the original purpose of any single study. These data can be reused by researchers to better understand autonomous vehicle operations, human/technology interactions, vehicle-to-X (infrastructure, vehicle, controller, etc.) relationships and numerous other potential interesting, but open, research questions. Reuse of data collected in these studies could also reduce the funding and effort requirements to answer new questions concerning the topics covered in the original studies without having to conduct new FOT/NDS studies.

Another example where sharing transport data provides a significant opportunity are advanced mobility solutions such as journey planners and control systems that can save businesses money, increase safety and reduce congestion. A number of organisations, including transport operators, public sector actors, industry players and academicians, are taking advantage of shared transport data and advanced analytical techniques to develop new insights, products and services that seek to manage journeys in an efficient, affordable and sustainable manner.

One final area where significant research benefit could be realized through sharing of data is in the integration of cross modal, multi-modal and synchro-modal transport operations. The transport research community is generally a single mode focused community. Unfortunately, this single mode focus in research ignores the fact that all transport is ultimately multi-modal. As requirements for emissions become more stringent, community impacts due to congestion and "last mile" deliveries grow, and operational requirements for end-to-end efficiencies through synchro-modal transport increase (both personal transport and freight transport) the need for cross modal data analysis will become critical. The ability of researchers to obtain, analyse and integrate data from multiple modes of operation will facilitate their efforts to develop realistic models of how transport actually occurs and lead to improvements in the actual delivery of transport services across Europe. These benefits were supported by the individual researchers that responded to the TRC survey (Figure 2).

---

[68] www.fot-net.eu
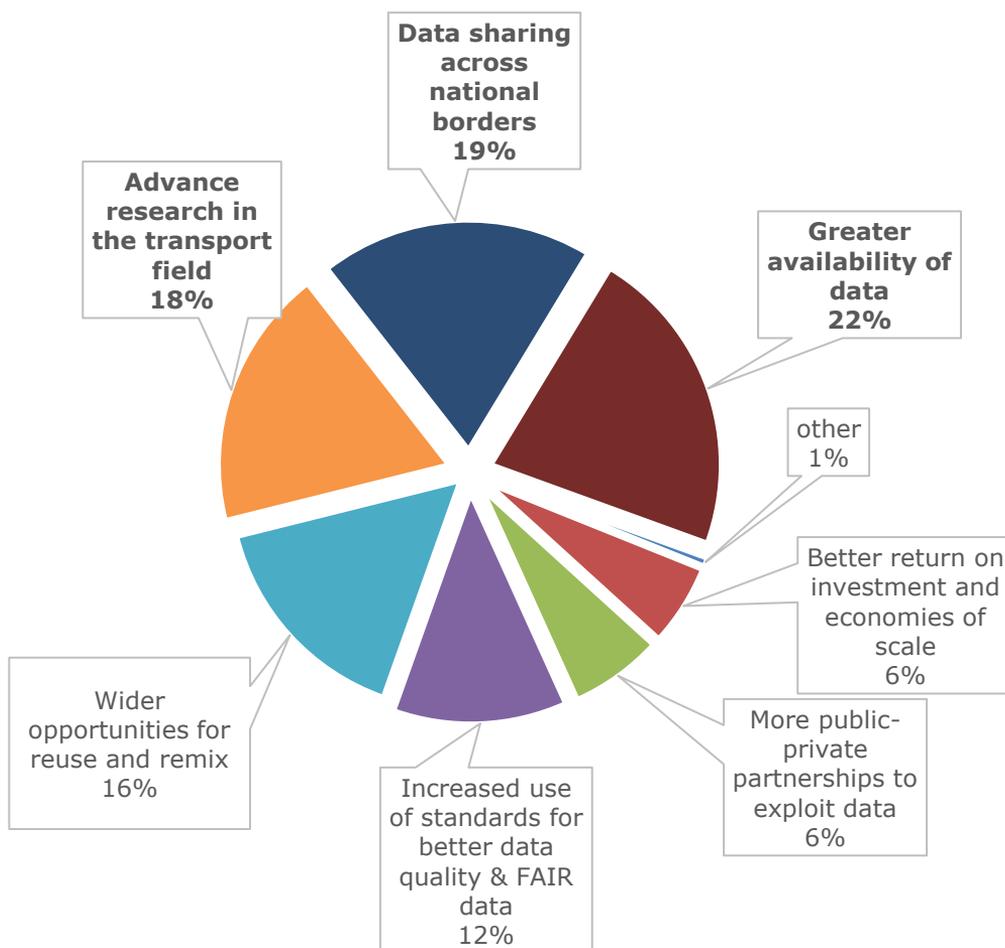
Figure 2. Research Benefits of a TRC

However, numerous challenges still hinder the reuse of transport related data[69]:

- **Data silos**: Transport data is often stored in distributed data silos (distributed data catalogues, individual data catalogues, consolidated big data platforms, distributed data aggregation platforms, etc.), which makes data analysis difficult and causes problems with many analytical models. Another obstacle to the efficient exploitation of these data assets is the fragmentation of data ownership and a lack of interoperability between datasets and platforms.

- **Different stakeholders**: Governments, public institutions, and private technology, service and vehicle manufacturers generate and use an increasing

---

[69] Janssen M., Y. Charalabidis, A. Zuiderwijk, (2012). "Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems and Management, vol. 29, no. 4, pp. 258-268.

amount of data for traffic control, vehicle management and safety, mobility planning, transport management, environmental impact management, air traffic control, etc. The private sector exploits increasing amounts of data for route planning and revenue management to gain competitive advantages, save time, and increase operational efficiency. Individuals increasingly use data via websites, mobile device applications, and GPS information for route planning to increase efficiency and save travel time. All of these stakeholders have interests in transport data, but their interests vary and create differing requirements for data access and understanding.

- **Different data ownership**: Data ownership varies by who generates and collects the data. Transportation system operators (both public and private), various transport focused agencies (e.g., IATA or the German Federal Ministry of Transport and Digital Infrastructure), cities and local governments, transportation researchers and other generators of transport data may not be willing to share their data due to privacy, legal liability, IP, competition, or cost related issues. In many instances when proprietary data is shared with researchers, non-disclosure agreements are required that limit or prohibit the sharing of research findings resulting from the use of the data.

- **Commercial data:** When working with commercial partners, the sensitivity of the data may preclude the use of cloud services, particularly when it represents trade secrets, evidence central to forthcoming patents and other data in commercial interests. The data owners may be unwilling to use cloud services for fear of data breaches or unauthorised access. Data ownership may itself be unclear as well, raising questions concerning who has authority to determine where data can be stored and how it can be made available in the future. These challenges require the original researchers to obtain adequate support in negotiating consent and partnership agreements prior to initiating their research projects. This support is required so that platform data management and data sharing decisions can be made without issue when future researchers ask to reuse the data in their research.

- **Legal concerns:** Privacy issues, particularly associated with data sets that are granular in nature and contain time and date information on individual trips, cause many data owners to restrict access to their data sets. In addition, issues with how data is used by external researchers can create legal problems if the data set is employed in a manner that places individuals at risk or that yields results that are not valid.

- **Data quality:** As result of the diversity of data sources, all developed with a particular problem definition in mind, numerous data types and data sets of differing quality are produced in the transport domain. This creates problems for those wishing to integrate the data to arrive at new insights. Poor data quality can arise when sensors collecting the data are inaccurate or have faults, from differing collection protocols, from manual processing of data, and numerous other data collection problems. Lack of assurance that data is of a suitable quality for research purposes inhibits researchers from reusing data sets that others have collected, thus creating a significant barrier to data reuse. Data volumes can be extreme particularly when data sets from IoT devices, connected vehicles, field operational tests, etc. are developed.

Lacking knowledge of all the particulars in how the data was collected and for what use, researchers can find it very difficult to assess the quality of the data in any reasonable amount of time. Data aging is also a problem as data change rapidly requiring significant amounts of curation to ensure that the data remains relevant for ongoing research purposes.

- **Lack of standards:** Transport data is collected for numerous reasons by various agencies and researchers. For the most part, data is collected without consideration for any potential reuse and, therefore, usually lacks any standards based collection structure. Variations in hardware and software used for collecting the data also create problems for reuse, as hardware and software vendors generally utilize proprietary standards in their products. Finally, basic issues of documenting any special circumstances or potential biases in the data collected are quite often ignored leading to data sets with anomalous elements that are not well understood by non-original researchers.

- **Lack of expertise**: A critical challenge for Europe is ensuring the availability of skilled experts in the data ecosystem. An effective ecosystem requires data scientists and engineers who have expertise in analytics, statistics, machine learning, data mining, and data management. It also requires technical experts knowledgeable in the domain being studied so that the nuances of the data being examined are understood and research value is created through the reuse of existing data sets.

- **Cultural barriers:** Researchers within the transport domain tend to be modal researchers, understanding their particular mode of transport, but lacking interest in, or knowledge of, other transport modes. This silo approach to what is an integrated system of movement creates significant barriers to data reuse and leverage as the research data collected may not be in a form that researchers in other modal silos can use.

- **Data sensitiveness:** Transport data is often ethically or commercially sensitive requiring tight controls concerning access to the data and ensuring it is stored within geographic areas where the legislative frameworks match European data requirements. Both of these elements pose significant obstacles in using distributed cloud services. The researchers who originally collected the data need to be assured that it will not be used in a manner that ethically compromises the subjects of the original research. In addition, these researchers need to know that the platform storing their data will be bound by European legal requirements covering security and privacy. Researchers attempting to access this data for reuse need to be informed of the constraints that are placed on the use of this data as well so that they do not inadvertently compromise the original researchers' commitments to research subjects and authorities.

- **Data control:** Transport research may include human participants, including vulnerable groups such as disabled travellers or children, so controls are needed around access to the data and how it can be reused. Users of this data need to be informed of the constraints being placed on the data and agree to these restrictions. Anonymised or aggregated datasets can be administered in the cloud environment, but data security and authorisation processes are

definite requirements if the original data collectors and their subjects are to be protected from unauthorized use.

The survey responses from transportation researchers surveyed for this report support the fact that the barriers identified above are real, and they identify other barriers that might also inhibit the use or reuse of data collected by other researchers or organizations (Figure 3). **It should also be noted that only 5% of respondents declared that they shared data openly, 28% shared under restrictions and 20% only share with collaborators on request.**
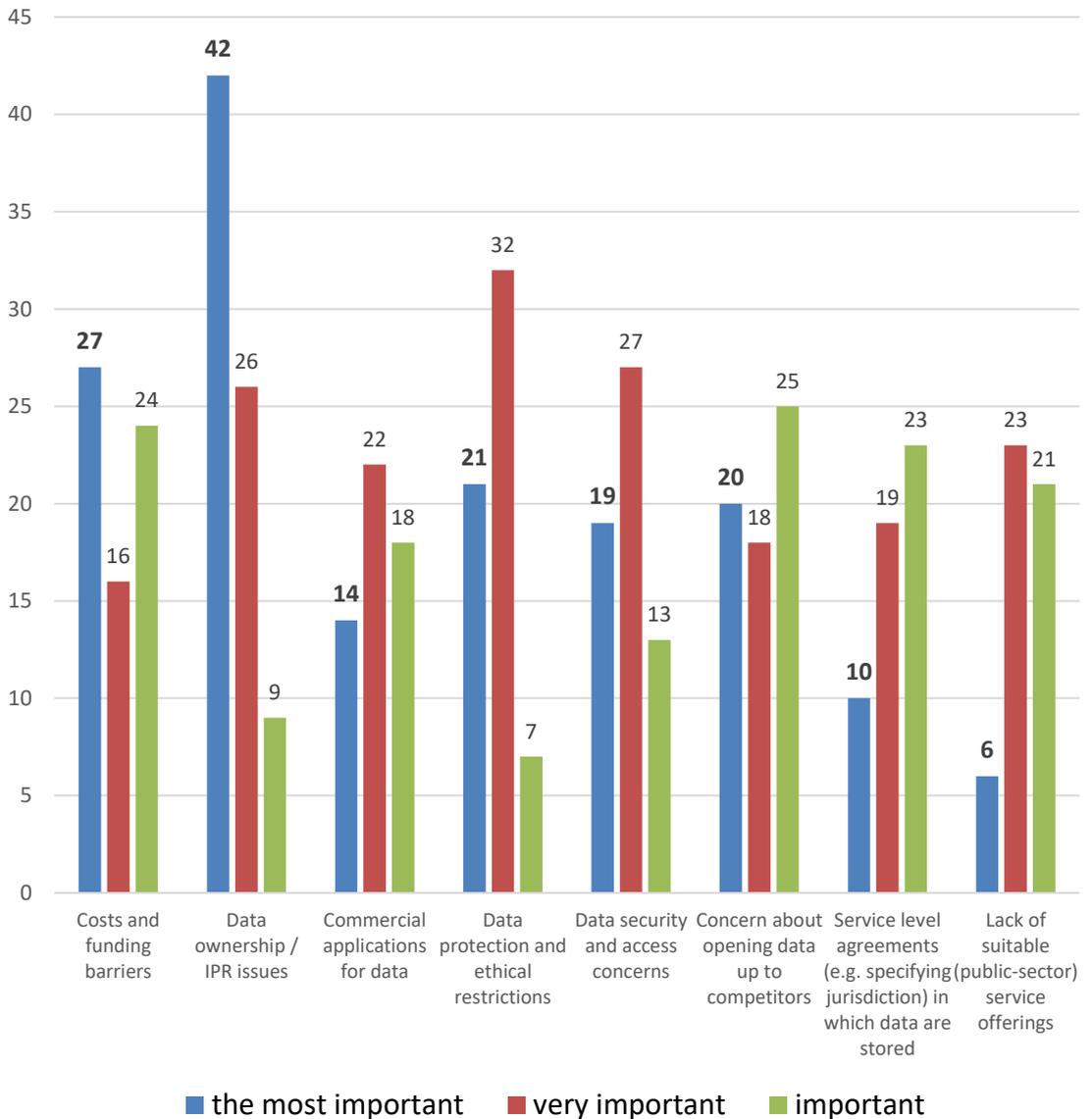

Figure3. Barriers to Data Reuse

While the barriers described are real and act as inhibitors to the free sharing of data amongst researchers, many of the assumptions underlying the barriers are more myth than reality. Similarly, a number of assumptions that underlie the benefits of open data sharing also are based on a number of poor premises. For example, it is often assumed that by simply providing access to data yields positive benefits. However, this assumption fails to understand the user side of the reuse question, which is driven by considerations such as access, understanding, quality, trust, legality and liability. Similarly, the assumption that open access to data is equivalent to open science or open government is fatally flawed. Open access to public, and even private, data may be necessary for open science and open government, but it is far from sufficient for these outcomes to be fully realized. Data without context, in the hands of users who don't understand its limitations, or that is selectively released for reuse can lead to faulty and erroneous outcomes that generate distrust in the community of potential users of the data. Finally, the idea that all data should be subject to open access is also a common misconception. Where data sharing could potentially lead to privacy problems, liability, security or legal issues, openness is not something that should necessarily be considered.

The TRC, in harmony with the EOSC, intends to support the provision and uptake of FAIR data. Like open data, FAIR is an approach to providing researchers access to data collected by other researchers or organizations. However, data that is provided in an open manner is not synonymous to data provided according to FAIR principles. Data can be Findable, Accessible, Interoperable and Reusable without being open. This is critical for transport data given the sensitivities of many transport data producers, so FAIR presents a useful set of principles to adopt for the TRC.

3 Needs for transport data and the TRC

The types of cloud services required in a Transport Research Cloud mirror the needs of researchers observed in other domains, though priorities will differ. These include:

- access to datasets,

- search tools,

- data analysis,

- storage,

- data sharing,

- preservation.

What emerged most strongly from our survey was the need for access to datasets and advanced search tools to help research communities assess the relevance of

content for their work. Since research is global and transport data inherently crosses geographic boundaries, having a single point of access to discover relevant datasets and broker (ideally open) licenses that allow access to researchers from different countries is critical. The quality of the datasets that a researcher finds in the TRC is also critical. Data sharing is not yet common practice in transport research so increasing the amount of high quality data made available is a priority to not only engage researchers, but build the reputation of the TRC.

Discovery and access mechanisms should provide a coherent catalogue with advanced search options to enable filtering based on criteria appropriate to the particular transport research question being asked. Data should be automatically indexed so shared datasets are discoverable via external catalogues and search engines, not just via the TRC. As seen in Figure 4 following, access to datasets and the ability to search effectively were significantly prioritised over other services of a TRC by the researchers we surveyed.

The next category of priorities for a TRC pertain to the actual reuse of data, with demand seen for open data sharing, data analysis, text mining and data integration tools. Large datasets and unstructured data require different management, indexing, and curation from structured data that researchers have traditionally worked with. Data analysis, text mining tools and access to High Performance Computing will be useful in this context. To encourage the broadest reuse of the stored data, the TRC should provide facilities for collaborative data access so that consortia can access the complex data available applying their combined skills to analyse the data.

Data storage and preservation was a lesser concern for the survey respondents. Perhaps this indicates that existing provision for storing data during projects is sufficient, or it may reflect a culture where data sharing is still uncommon and restrictions or barriers often apply.

TRC survey respondents were asked if they would use the TRC open data service. Almost half of the answers (48%) stated "yes, definitely" and a number of clear priorities were given for the most in-demand cloud services. The following examples of why the respondents would use the TRC demonstrate a broad set of reasons for wanting access to open data:

- "I use transport data in different fields - education and research. In education, I teach modelling and simulation course and GIS course. I teach at university with tradition in transport oriented education. In research, I work on decision support tools for transportation"

- "Open data sets can significantly escalate the development of new technologies, search for interdependence and multidisciplinary. There is an opportunity to create community development teams that will work together to find ways to optimize the transport system. It is also possible to bring new developers from other science disciplines"

- "Quality, accuracy and integrity of data are most important, else unusable!"

However, 46% of the respondents were less enthusiastic indicating "perhaps – depends on service", arguing as follows:

- "Would that come at a cost? Furthermore, generally platforms developed at a European level are poor in terms of security, user friendliness, accessibility and the requirements to access them"

- "I´m dealing with cross country comparisons and I expect that TRC can provide a common framework for comparable data. At the moment, even available data are mostly distracted, in some part - only in national languages and with significant differences in terms of their coverage, aggregation and available variables."

The short list of requirements mentioned above were supported by the transportation researchers that responded to the TRC survey (Figure 4). In addition, access to analytical tools, private workspaces, integration tools, and data use definition services were all deemed to be important services for the researchers if they were to use the TRC.
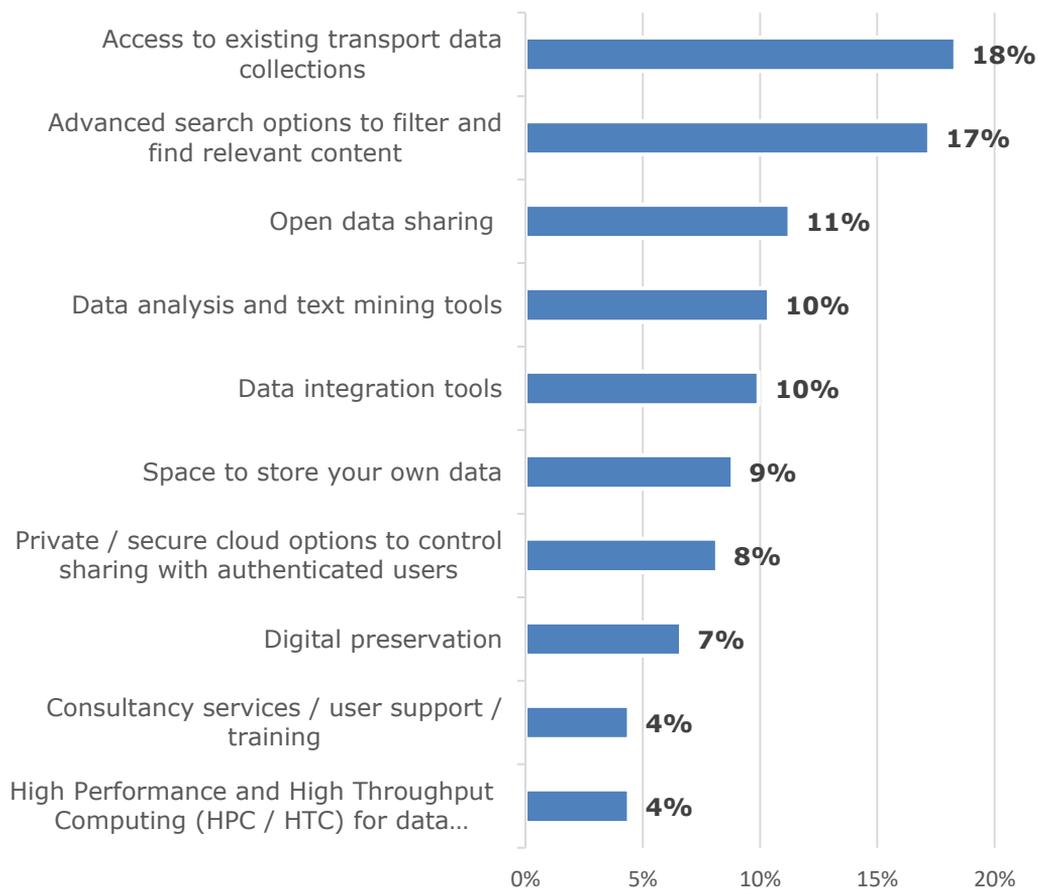

Figure 4. TRC Service Requirements

To make the TRC a viable long term project requires that it be easy to use and that it provides value to the various stakeholders that will use, supply, operate and fund it. Each of these stakeholders, due to their different needs, will have a different view of value based on their role in the ecosystem of the TRC:

- **Data suppliers** will need to know that they have a simple process for uploading their data to the cloud, that the data will be properly secured, that their contribution to any future research based on the data is clearly documented, and that any intellectual property that they have in the data is properly managed.

- **Data users** will need to be able to easily find and access data that they need, know that the data they have access to is reliable and valid, that the terms of its use are clearly understood, that the source of the data is known, and if there are any potential intellectual property restrictions based on their use of this data in their research exist.

- **Platform operators** will need to be comfortable that their costs for operations, curation, marketing, management, tools development, training, etc. are covered by the various funding programs that the platform operates under.

- **TRC funders** will want to see their funds put to productive uses in the generation of new knowledge. In addition, funders will also want to know that their investment in the TRC will generate an increasing flow of inbound data for deposit and outbound data for research.

- Finally, **society** will need to see that by using the data hosted on the Transportation Research Cloud researchers are achieving breakthroughs that will lead to addressing the critical issues that transport operations generate for society as a whole.

These primary value areas will require fine tuning for the various platform business models so that all stakeholders perceive the value they receive from participating in this open data endeavour.

Additionally, there is a need to analyse particular issues that will have an impact on the development of the TRC. The following section describes some of the most important issues.

3.1   Policy issues

Governmental policy concerning publicly funded transport research should ensure that data generated during the research is made available for general public reuse. This is a necessary condition, but it is not sufficient. Simply requiring that data be made available for reuse does not ensure that it can be reused, or that it even can be accessed. Policy must specify the condition in which the data is provided, curated, maintained and accessed. Policy also must specify how these data are to be served and how this service function is to be funded (e.g., the standards that will be necessary for all EOSC infrastructures).

Governments do not have to fund the platforms hosting and serving transport research data in perpetuity. Some government funding will be required to setup and begin operations of the platforms however. It also implies that governments need to specifically seek business plans from the platform managers that demonstrate how the platform will generate its own funds and when it will be able to operate without governmental funding. These plans will require periodic review and updating to determine progress towards self-sufficiency with decisions being made along the way to either continue funding or merge the failing platform with other, more successful, platforms[70].

3.2 Network development and infrastructure construction

Governmental mandates can act as the trigger for the development of stakeholder communities and the creation of infrastructure to support the open data platforms. However, sound business models, a clear value proposition, and a market requirement for open data are all elements that are necessary for the development of user and supplier networks, infrastructure development and funding sources. In examining the needs of the potential users of a TRC, it was found that the keys to success for the TRC revolve around the factors of ease of use and value added for the research community.

By ease of use, the researchers, both those that would contribute to the TRC and those who would use its services, felt strongly that it must not add to the workload that they have today. For data providers, the process of uploading data, along with the metadata required to document the data, should not be something that is an additional burden on their research efforts. Redocumenting what the researchers have already documented in the reports or articles developed from the data, additional documentation requirements, etc. for the data, as well as cumbersome legal rights requirements, while potentially understandable, were all deemed as overhead making the uploading and archiving process of the researcher much more demanding than desired.

For the data user, searching for data, understanding what the data means, whether it is valid and of a quality for further study, cumbersome rights requirements, etc. were also noted as problematic. As one researcher put it, "If I have to work as hard to find and understand the data set as I would to create my own data, I'd rather create my own data." Therefore, given the inherently cross-disciplinary nature of the transport research community, there is significant potential benefit to be gained by agreeing on standard data formats, metadata, sharing practices and access mechanisms so that cross-discipline data access can be facilitated.

---

[70] OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en; Duch-Brown N. B., Martens, B., Mueller-Langer, F., (2017). The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01.

3.3   The costs, funding and long-term sustainability to operate the TRC

While it is not possible to estimate costs at this time for a federation of platforms that have not yet been designed, it is never too early to point out that there will be costs associated with the operation of the TRC.   Management, support, operations, storage, marketing, education, engineering, integration, and other ongoing costs will need to be foreseen for any TRC proposal and offset against potential funding sources.   These costs will determine the viability of the TRC, both in the short term and over the long term. It should also be noted that costs change over the lifecycle of an entity such as the TRC.   Start-up costs will be incurred in developing and implementing the TRC, marketing costs will be substantial in the early years of the TRC as its services are explained and knowledge concerning its existence is developed.   Educational costs will be an ongoing cost function as new stakeholders become interested in the TRC. Growth will also generate costs as new infrastructures will need to be added, old infrastructures abandoned and more robust systems developed to handle the increased loads, data types and uses that come with growth.

The sources of revenue for the TRC and its affiliated data platforms will need to be clearly identified. Revenue sources for the TRC and affiliated platforms are conceptually numerous. Supply side revenue may be generated by charging data owners for hosting their data. Buyer side revenue may be developed by charging users for data access, providing users with value added services, and performing data consulting services (see Figure 5 for survey respondent preferences on charging models for use). Governmental funding, either directly or via individual research programs, is another potential source of revenue. Private funding from foundations or non-profit organizations is also another potential source. Finally, IP related funds may be generated through the licensing of hosted data to industry. While it might appear that numerous sources exist, it should be noted that open data platforms today are not yet self-funding entities. Public funding in one manner or another is still required for all platforms studied thus far. This provides a question for entities wishing to see the establishment of open data clouds. This question concerns how much governments are willing to pay to see these open clouds develop and become established tools for research.
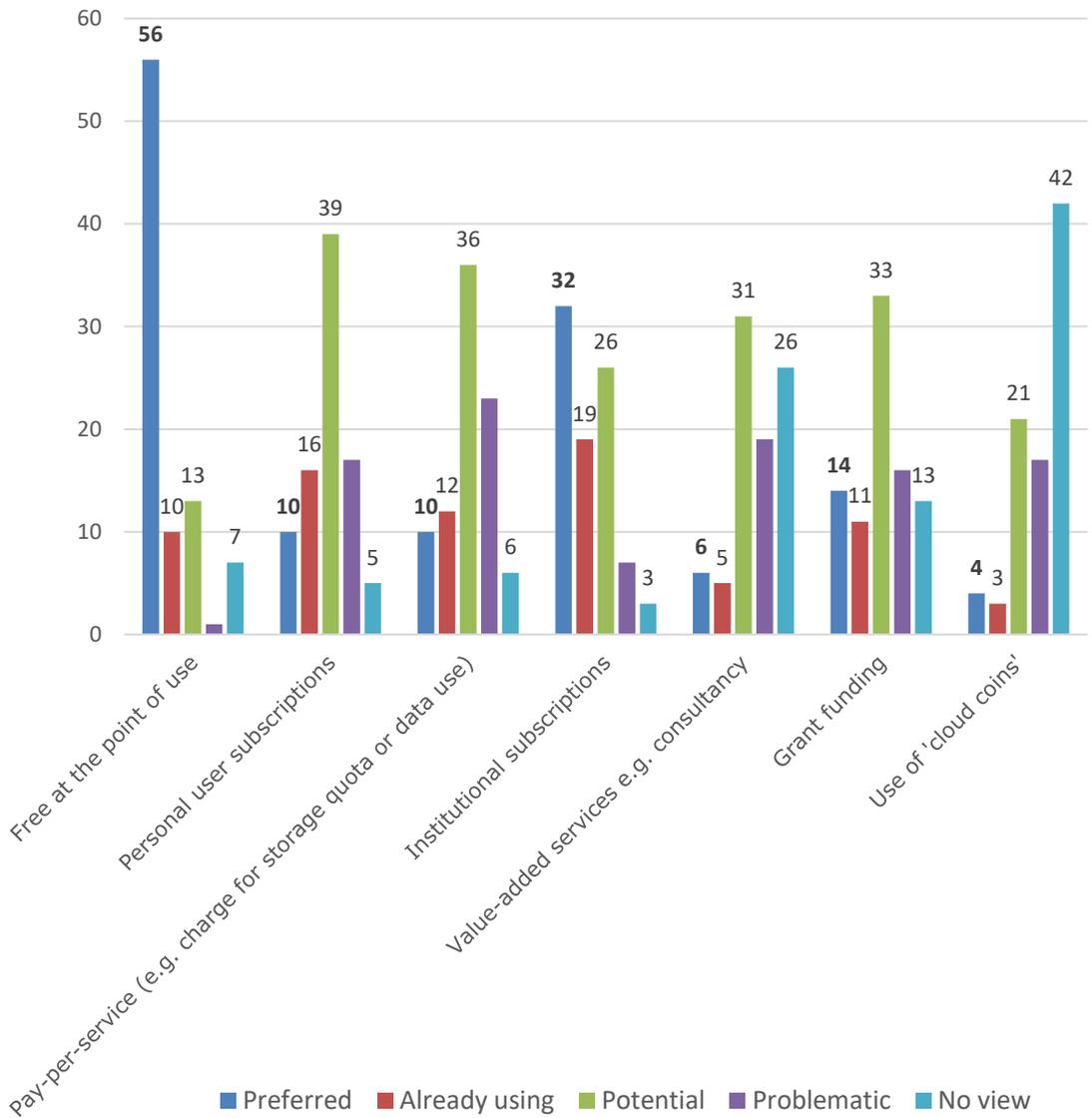
Figure 5. Preferred Charging Models for Use

Operation of a cloud based platform providing data management, storage, curation, access control, distribution, induction, and other valuable services is a nontrivial activity. Significant infrastructure (storage, computing systems, web services, network access, security systems, etc.) will be required to ensure the proper management of hosted data. In addition, since the TRC will need to be operated in a business like manner, skilled managers, operators, marketers, engineers, developers and support personnel will be needed so that the platform is perceived by its stakeholder community as professional generating high value for all parties involved.

Numerous service delivery channels will be necessary for the TRC. Marketing, education and direct outreach programs will be necessary to inform data providers and users of the existence of the TRC and to educate them in its use. Marketing and outreach programs for potential funders will also be necessary so that a sufficient community of interested funding sources is developed to sustain the TRC. Once educated, the TRC will need to perform its services in a value added manner for all parties. Data providers will need easy upload services to place their data on the platform. Curators will need tools to properly validate, curate and normalize the provided metadata to facilitate querying the data. Users will require the TRC to have robust search tools, validation processes, and download capabilities to make access to the stored data easy. In addition, data enhancement/analysis tools may be needed to ensure that data users have the option to either manipulate the data on the platform or download it for local storage and manipulation. Actual service delivery processes will depend on the specifics of the platform and its stakeholder community, but those described will most certainly have to be covered.

Given the examples of open data platforms studied to date, it is not at all clear that long term viability or sustainability is achievable without ongoing public support. All current international platforms operate to some degree using governmental funds. Supply side fees, buyer side fees, fees for value added services, etc. are all being tried to generate funds in excess of costs and to migrate the platforms away from governmental subsidies. Unfortunately, with research budgets being constrained, industry generally not encouraged to use, or knowledgeable of, the data residing on these platforms, third party funders not perceiving funding of platforms as a long term commitment and few platforms demonstrating a strong business model, there is every reason to believe that existing and future platforms will not be viable over the long term. This does not mean that sustainability is out of the question. However, it does mean that a more pragmatic and business focused view of open data and its provisioning is required if governments are going to be able to assure their constituents that they won't be obligated to fund open data in perpetuity and for open data stakeholders to actively provide financial support for open data.[71]

One potential way to address the sustainable funding issue is to rethink the purpose of the TRC. Discussion generally concerning open data platforms take a "warehousing" approach to the function of the platform. Platforms store, curate, provide information about, and may do some training and consulting concerning the data they are managing. Perhaps if the platform was looked at more as a value adding or "production" centre rather than a warehouse platform users might be more willing to pay for the services received. This reconceptualization would require the TRC to either assist users in collecting data, analysing data, integrating data, publishing data, or gaining access to new data. The TRC could also act as an IP clearing house where data used in the development of new business services could generate revenue for the platform through license fees. These services, of course, would have to be implemented over time, but changing

---

[71] Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman (2016), "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3.

the perception of the users of the TRC from one in which the TRC is simply a place to store or search for data, to one in which the TRC produces potentially new and exciting data sources, business opportunities and research would definitely lead to a more value focused perception and, potentially, a greater willingness to pay.

4 Recommendations

Setting up the TRC will not be an easy task. As has been pointed out in the previous sections, many barriers and road blocks exist to successful delivery of open data over a federated network of open data platforms. In the transport domain, the clear lack of current practice in the research community in providing "their" data to others for reuse may be the greatest impediment of all. Bearing in mind what has been found in the development of this report, the following recommendations provide a set of minimum requirements that must be addressed if a TRC is to be established[72].

4.1 Reusable research data

1. <u>Not all data is research data:</u> The TRC does not have to catalogue every data set that is developed by transport researchers, governmental agencies, and private entities. Research data is data that is of value to the research community. While it may be difficult to delineate this value, it should be understood that data that resides in the TRC is data that has the potential to create research related value. **We recommend that the Commission bring together researchers, research data users, and data generators to define what constitutes transport research data so that recommendations made by this Expert Group can be implemented with a clear vision what types of data that should be part of the TRC.**

2. <u>Motivate researchers to use existing data sets:</u> Today's researchers are used to making an analysis from the beginning to the end. This means that researchers start with data collection and afterwards analyze the collected data. This general principle needs to be overcome as it is limiting the scope of today's research. Ways need to be identified so that reusing someone else's data in a research project is rewarded and accepted as good practice in the research community. **We recommend that the Commission conducts a detailed study among transport researchers in order to identify the objections behind limited use of data collected by others and develop recommendations on how to overcome these objections. In addition, training should be developed to inform and encourage best practices in the collection and reuse of data so that the concerns of the researchers can be overcome.**

---

[72] Note that in this section our recommendations, while focused on the TRC, reference the EOSC on several occasions. This is because the TRC, as a subsidiary domain structure under the EOSC, should benefit from all EOSC required standards and requirements.

3. <u>Research data is a public good:</u> In some parts of the world (e.g., the U.S.A.) it is accepted that if research is paid for by public entities, then not only are the results a public good, but also the data collected is a public asset. In principle, scientists are paid for collection, processing and analysis of data, not for data ownership. Therefore, it should be made clear to researchers that data collected by public money needs to be made available to the public as an asset that can be reused by others. **We therefore recommend that any and all data collected under contracts that are paid for by tax payer funds by default be classified as public data. This data should be placed into the EOSC and/or its subsidiary cloud infrastructures (e.g., the TRC) as a requirement of all publically funded research (we recognize that certain data, due to privacy or secrecy requirements would need to be excepted from this requirement, but by having the requirement as a default these exceptions could be made consciously and transparently).**

4. <u>Public goods versus intellectual property</u>: Researchers may develop unique tools or algorithms in their research to collect and analyse data that is being paid for by public institutions. To encourage the researchers to release the public data, without infringing on the intellectual property that the researcher has developed, there should be clear guidelines as to where the public asset ends, and the private asset begins. **We recommend that in the development of the approach that all data collected under a publically funded project that a clear demarcation be made between the Intellectual Property created by the individual researchers in analysing the data and the data itself.**

4.3 Standards

5. <u>Make research data available in truly standardized form:</u> Without data, the TRC will fail before it starts. Numerous public websites provide various kinds of data for cities, regions and countries. In addition, transportation researchers independently collect and analyse data for many of their studies. Unfortunately, the data available from public entities or individual researchers is not generally findable via a single portal, not well documented, of differing formats, collected for various reasons and, in the minds of most researchers, unusable for their scientific research. We believe that EU wide standards for all data collected in the transport domain is required or researchers will continue to find the search costs for usable research data too high. **We would recommend that the Commission bring together members of the transport research community (all modes), governmental entities that generate transport data used in research, infrastructure operators, and commercial consumers of transport research and transport research data to define the standards that will be necessary for the collection of transport data by public institutions, the data formats these data should adhere to, the metadata that must be used to describe the data, and formats of this metadata so that automated search engines can easily find and characterize the data.**

4.4   Infrastructure

6. <u>The TRC must be fit for purpose:</u> Transport research datasets can be extremely large, composed of unstructured data requiring sophisticated data management and analysis technologies to store, curate and add value to. The TRC must be able to handle these complex datasets in a seamless manner if it is to be viewed as a valuable tool for transport researchers.   This requirement means that the TRC will need to incorporate sophisticated data handling capabilities, requiring sophisticated data managers and curators. The infrastructure to handle this requirement, along with the personnel needed to operate the infrastructure, will not be inexpensive. **Since the TRC may be a costly service to operate we recommend that the Commission conduct a detailed study on the infrastructure and operating requirements for a TRC to ensure that an appropriate level of service can be provided at a cost that is understood by all stakeholders.**

7. <u>The TRC must provide value to its users:</u> Rather than acting as a passive data warehouse taking custody of transport research data and then simply waiting for users to come, the TRC must actively work to create value for its users and stakeholders. Services such as easy upload/download, quick search and query, rapid data extraction, etc. are expected of any good online platform. However, these services are simply the price of entry, they are not things that will excite and encourage researchers to continually come back and use the TRC. For the TRC to "delight" its stakeholders and users it must create distinctive value for each of them.   This means building out sustainable business models and ensuring that the TRC's value proposition is "evergreen." **We therefore recommend that the Commission conduct a detailed study of what the potential user and stakeholder communities would require from a TRC in order to make it the "go to" place for doing cutting edge transport research.   From this value proposition, business model development and actual pilot implementations should be performed to test the findings concerning value and use.**

8. <u>The TRC must mirror the EOSC as a subsidiary:</u> While this document examines the need for a transport research cloud, it does not do so in a vacuum. The TRC will be a supporting pillar of the EOSC and as such must conform to processes and procedures established for that overarching cloud infrastructure to avoid confusion and conflict in the future. For this reason, while we believe that there are unique characteristics to transport data that need to be recognized by the TRC, we believe the TRC should structure itself as closely as possible to how the EOSC is structured. **We therefore recommend that in establishing the TRC it should be modelled after the EOSC and governed in a manner similar to the EOSC taking into account the particularities of the transport domain.   We also recommend that the TRC conform to all standards established for and by the EOSC so that confusion and conflict in accessing and searching for data are avoided.**

4.5 Incentives, education, and training

9. <u>Encourage researchers to place their research data in the TRC:</u> Researchers need to be motivated to publish not only the results of their research (which in most cases is the core element of their research contracts), but to make the data used available to other researchers. Sharing data is often hindered by the fact that researchers like to use the same datasets for several publications on an exclusive basis by looking at the data from several perspectives. Only when all analyses are finalised does the researcher consider making the data available to others, if at all. Incentives need to be changed so that researchers who collect data make it available early for others to work with. Here, credit for having collected the data should be made and rewards based on this credit should be developed so that the researcher does not hoard the data. **We therefore recommend that EU policies for academic promotion, training, publication, and knowledge generation at public universities (and private degree granting universities) be examined and harmonized to ensure that researchers are uniformly trained in the requirement and process of placing their research data into the EOSC (and its subsidiary cloud infrastructures), that universities provide the proper (and uniform) incentives to their faculty and researchers to ensure that their research data (if not constrained by commercial non-disclosures or other requirements for privacy) is placed into the EOSC, and that proper credit for the generation of data that is reused is given to the individuals who originally collected the data (note that where universities use journal citations as an element of promotion references to who collected the data should be looked at in a manner similar to any of the authors of the paper).**

10. <u>Outreach and training:</u> Developing the TRC will be a difficult technical endeavour. However, the real challenge will be in changing the minds of transport researchers so that they use the TRC for data storage, sharing and reuse. This will not be an easy task and will require that TRC infrastructure operators, educational institutions, and governmental data providers work with the research community (current and future) to let them know about the TRC, the value that the TRC brings to their research, and the value that they can bring to the transport research community by adhering to some simple rules. **We recommend that an analysis of training requirements be conducted by the Commission and, based on these requirements, training/education programs developed for existing researchers and future researchers, libraries and librarians, data curators, and other individuals who will be needed to carry out the development of a mind-set of open data by default.**

One final point needs to be emphasized if the TRC is to become a reality. This point is that the TRC needs to be looked upon as a sustainable long-term project. This means that the TRC must focus on understanding how it can provide value to its users, educate the users in the fact that this value exists and continue to add new value so that it becomes the go to place for doing research. If this final point is not taken to heart, then the TRC, no matter how well it is constructed, will not be used by its customers and will not be the success that it could be.

The recommendations made for the TRC to become a reality are, by nature, broad. This is because much more work needs to be done in defining standards, understanding stakeholder value propositions and needs, identifying the infrastructures that are/could be used and their requirements, and learning from the trail blazing work that is going in establishing the EOSC. This means that considerable additional digging is required before a TRC can be planted and much can be learned as the EOSC moves forward, so that the planting of the TRC leads to a growing service and not one that is interesting, but not used.

GLOSSARY

Term	Definition
EOSC	European Open Science Cloud An initiative to federate existing infrastructure from member states and research disciplines to provide a single point of access to European data and data services
FAIR	Findable, Accessible, Interoperable and Reusable A set of 15 high-level principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines
IPR	Intellectual Property Rights The rights owned in a creative work which can be assigned through patents, copyrights and trademarks allowing the holder to exercise a monopoly on the use of the item for a specified period of time. Licenses for datasets and software clarify the permissions a copyright holders grants to others to copy, publish, modify and reuse a work. Restrictions can be placed on the community (i.e. non-commercial) or the type of remixing allowed (i.e. no derivatives) and conditions may have to be met (i.e. attribution required).
Open access	Published research outputs that are made available online, free of charge, and free of most copyright and licensing restrictions.
Open data	Data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.
Open science	The practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.
Transport data	Data related to the broad field of transportation.
Transport research	Research conducted by researchers focused on understanding, characterizing, optimizing, improving, etc. the domain of transportation.

Transport research data	Data relevant to conducting research in the transport domain.
TRC	Transport Research Cloud An umbrella term for the federated relationship of online transport data infrastructures allowing users to upload, search, access and reuse transport related datasets.

REFERENCES

"Trouble at the Lab," The Economist, October 19, 2013, pp. 23-28.

ACEA Strategy Paper on Connectivity, (2016).
http://www.acea.be/publications/article/strategy-paper-on-connectivity

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, Max Planck Gesellschaft, (2003).

Bertot, J.C., Jaeger, P.T., Grimes, J.M., (2010). "Using ICTs to create a culture of transparency: e-Government and social media as openness and anti-corruption tools for societies," Government Information Quarterly, vol. 27, no. 3, pp. 264-271.

Borgman, C.L., (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Buschmann, K., Kasberger, S., Kraker, P., Mayer, K., Reckling, F., Rieck, K., & Vignoli, M., (2015). Open Science in Österreich: Ansätze und Status. Information. Wissenschaft Und Praxis, 66(2-3), 1–9.

Copyright Act of 9 September 1965 (Federal Law Gazette I, p. 1273), as last amended by Article 1 of the Act of 1 September 2017 (Federal Law Gazette I p. 3346).

Council of the European Union, Draft Council Conclusions on the European Open Science Cloud, 9029/18.

Danneels, L., S. Viaene, Van den Bergh, J., (2017), "Open data platforms: Discussing alternative knowledge epistemologies," *Government Information Quarterly*, vol. 34, no. 3, pp.365-378.

David, P.A., (2003). "The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A primer," Stanford Institute for Economic Policy Research, No. 02-30, March 2003.

Duch-Brown, N. B., Martens, B., Mueller-Langer, F., (2017). The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01.

ECa, (2016). Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. European Commission. Directorate-General for Research and Innovation. Brussels.

ECb, (2016). Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European

Open Science Cloud. European Commission. Directorate-General for Research and Innovation. Brussels.

EC, (2018). Prompting an EOSC in Practice. Interim report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud. Brussels.

EOSC Declaration, (2017). EOSC Declaration. European Open Science Cloud. New Research & Innovation Opportunities. Brussels.

ERTICO (www.ertico.com) is operating the ITS Observatory (https://its-observatory.eu/)

European Union Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010, "on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport," Official Journal of the European Union, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040

Gellerman, H., Svanberg, E., Barnard, Y., (2016). "Data sharing of transport research data," Transportation Research Procedia vol. 14, pp. 2227 – 2236.

German National Mobility Data Warehouse (MDM - https://www.mdm-portal.de/) or the Dutch National Data Warehouse (http://www.ndw.nu/en/).

Grand, J.A., Rogelberg, S.G., Allen, T.D., Landis, R.S., Reynolds, D.H., Scott, J.C., Tonidandel, S., Truxillo, D.M., (2018). A systems-based approach to fostering robust science in industrial-organizational psychology. *Industrial and Organizational Psychology*, *11*(1), pp. 4-42.

Hey, T., Trefethen, A.E., (2005), "Cyberinfrastructure for e-Science," *Science*, vol. 308 no. 5723, pp. 817-821.

https://africaopendata.org/dataset

http://archaeologydataservice.ac.uk/advice/chargingPolicy.xhtml

http://dados.gov.br/

http://data.stats.gov.cn/

http://data.un.org/

http://dataportal.opendataforafrica.org/

http://datos.gob.ar/

http://einfracentral.eu/

http://inspire.ec.europa.eu/

http://noesis-project.eu

http://pon.edu.pl/images/plon_publications/files/19_Towards%20Open%20Research%20Data%20in%20Poland.pdf accessed 18 March 2018

http://senseable.mit.edu/shareable-cities

http://tn-its.eu/

http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

http://www.data.go.jp/?lang=english

http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps

http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies

http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies

http://www.itsstandards.eu/pt

http://www.itsstandards.eu/rtd

http://www.oecd.org/gov/digital-government/open-government-data.htm

http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/monitoring-transition-open-access-2017.aspx

https://blogs.rcuk.ac.uk/2012/10/24/rcuk-open-access-policy-our-preference-for-gold/

https://cordis.europa.eu/project/rcn/100232_en.html

https://data.gov.in/

https://data.gov.uk

https://data.govt.nz/

https://data.worldbank.org/

https://dini.de/dini-zertifikat/liste-der-repositorien/, accessed 12 April 2018.

https://doaj.org/, accessed 10 April 2018.

https://ec.europa.eu/digital-single-market/en/policies/76011/3502

https://ec.europa.eu/digital-single-market/open-access-scientific-information

https://ec.europa.eu/info/research-and-innovation/projects/project-databases_en

https://eoscpilot.eu/

https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:32003L0098

https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32013L0037

https://home.cern/

https://idatosabiertos.org/en/acerca-de-nosotros/

https://joinup.ec.europa.eu/release/dcat-ap-v11

https://open.canada.ca/en/open-data.

https://openaccess.mpg.de/319790/Signatories, accessed 10 April 2018.

https://openaccess.mpg.de/Berlin-Declaration

https://opendataamerica.opendatasoft.com/pages/home/

https://opendatabarometer.org/4thedition/regional-snapshot/middle-east-north-africa/#key_finding

https://researchdata.ands.org.au/

https://scholar.google.com/

https://search.data.gov.au/

https://toolbox.google.com/datasetsearch

https://transformingtransport.eu

https://www.big-data-europe.eu/pilot-transport

https://www.data.cam.ac.uk/repository

https://www.data.go.kr/main.do?lang=en

https://www.data.gov/

https://www.datos.gob.mx/

https://www.ebi.ac.uk/arrayexpress

https://www.education.gov.au/2016-national-research-infrastructure-roadmap

https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris

https://www.eosc-hub.eu/

https://www.europeandataportal.eu/data/en/dataset?tags=public-transport

https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

https://www.icpsr.umich.edu/icpsrweb/

https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/UPS-On-Road-Integrated-Optimization-and-Navigation-ORION-Project

https://www.nature.com/sdata/policies/repositories

https://www.openaire.eu/

https://www.projekt-deal.de/about-deal/, accessed 12 April 2018.

https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation

https://www.siemens.com/global/en/home/products/mobility/rail-solutions/services/digital-services/railigent.html

https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

https://www.timeshighereducation.com/sites/default/files/breaking_news_files/concordat_on_open_research_data.pdf

https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf/

https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

https://www.wto.org/english/res_e/res_e.htm

INSPIRE Data Specification on Transport Networks – Guidelines: http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_TN_v3.0.pdf

Janssen, M., Charalabidis, Y., Zuiderwijk, A., (2012). "Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems and Management, vol. 29, no. 4, pp. 258-268.

Koznov, D., Andreeva, O., Nikula, U., Maglyas, A., Muromtsev, D., Radchenko I., (2016), "A Survey of Open Government Data in Russian Federation," KMIS, November 2016, pp. 173-180.

Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., et al., (2012). "A Call for Transparent Reporting to Optimize the Predictive Value of Preclinical Research," Nature, vol. 490, pp. 187-191.

Leveraging Big Data to Manage Transport Operations (LeMO project), https://www.big-data-europe.eu/wp-content/uploads/5-BDE_SC4_WS3_Arnaud-Burgess-LeMo-project.pdf

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., Doshi, E.A., (2013). "Open data: Unlocking innovation and performance with liquid information," *McKinsey Global Institute*, vol. 21.

McCormick, D., Moore, K., (2016). "Report on the First IEEE Workshop on The Future of Research Curation and Research Reproducibility," John Baillieul chair, IEEE, 5-6 November.

OECDa, (2017). "Business models for sustainable research data repositories", *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en.

OECDb, (2017). "Co-ordination and support of international research data networks", *OECD Science, Technology and Industry Policy Papers*, No. 51, OECD Publishing, Paris, https://doi.org/10.1787/e92fa89e-en.

Open Access in Deutschland: Die Strategie des Bundesministeriums für Bildung und Forschung, Bundesministeriums für Bildung und Forschung, 2017.

Planning for Connected and Automated Vehicles, (2017). http://www.cargroup.org/wp-content/uploads/2017/03/Planning-for-Connected-and-Automated-Vehicles-Report.pdf

Priority Initiative "Digital Information" of the Alliance Partner Organizations, Alliance of Science Organizations in Germany, (2013).

Research Data Alliance, (2015), "Sustainable Business Models for Brokering Middleware to support Research Interoperability," Sustainable Business Models Team Report to the Brokering Governance Working Group, Research Data Alliance.

Roadmap, Automation in Road Transport, iMobility Forum, Version 1.0, May 2013, http://vra-net.eu/wp-content/uploads/2013/12/Automation-Roadmap-final.pdf

Rodrigue J. P. (2017), The Geography of Transport Systems, Fourth edition, New York: Routledge. ISBN 978-1138669574.

SAE International, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, J3016_201609.

Science¦Business (2018), "How the Science Cloud could pay its way," Science¦Business Network's Cloud Consulation Group, Brussels.

Shaping digital transformation in science, "Digital Information" Initiative by the Alliance of Science Organizations in Germany, Mission statement 2018 – 2022, Steering Committee for the "Digital Information" Initiative of the Alliance of Science Organizations in Germany, 2017, DOI: http://doi.org/10.2312/allianzoa.016

TRIP, https://ec.europa.eu/transport/themes/research/trip_en

U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA).

University of Leicester, (2018). Research Data. Retrieved May 30, 2018, from https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data.

Vogel, G., German researchers resign from Elsevier journals in push for nationwide open access, http://www.sciencemag.org/news/2017/10/german-researchers-resign-elsevier-journals-push-nationwide-open-access, accessed 12 April 2018.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman (2016), "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol.3.

www.bigdata.uni-frankfurt.de

www.fot-net.eu

www.mobilitydata.gv.at

Yin, H., Berger C., (2017), "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets,"
In *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on* (pp. 1-8).

ATTACHMENT I. OPEN ACCESS STATUS IN CHOSEN EU COUNTRIES

AUSTRIA

Recent trends

Since December 2017 Austria has a new government which formulated a new government-program which has new attitudes as cornerstones. One of them affects open access to data very strong.

The new government program speaks opening all kind of data that is with authorities, for research and innovation activities. The only sectors that are excluded are internal affairs and justice. All other sector ministries are urged to prepare steps to make data which are belonging to public authorities available in an anonymised format for R&I activities. This includes health, education, environment, and of course transport.

Based on this new paradigm, some of the later mentioned aspects are currently valid, but might change or being adopted in near future.

R&I Programs in Austria

Today the funding agencies are increasingly incorporating the handling of research data in their specifications for funded projects. The FWF, Austria's largest funding agency, sees "Open Science" as one of the central aspects for participation in future funding programs of the FWF.

1. FWF (www.fwf.ac.at): The Austrian Science Fund (FWF) is Austria's central funding organization for basic research. In 2018 the Austrian Science Fund FWF is celebrating its 50th birthday. In this respect they are organising Austria's biggest open-air science festival in Vienna.

2. FFG (www.ffg.at): The Austrian Research Promotion Agency (FFG) is the national funding agency for industrial research and development in Austria. The FFG was founded in 2004. The FFG is wholly owned by the Republic of Austria, represented by the Federal Ministry for Transport, Innovation and Technology (bmvit) and the Federal Ministry for Digital and Economic Affairs (BMDW). In this respect most transport related research activities are funded by the FFG.

History of Open Science/Open Data in Austria

From 2014 on Austria is making considerable progress in the area of Open Science, especially in the areas of Open Access and Open Data. The foundation of the Open Access Network Austria (OANA) and the launch of the E-Infrastructures Austria project early 2014 can be seen as important cornerstones of a developing Austrian Open Science landscape. The Austrian Chapter of the Open Knowledge Foundation substantially contributes to Open Science practice- and awareness building as well. Among others, these

initiatives are the basis for setting up a national Open Access strategy as well as a nationwide Open Access and Open (Research) Data infrastructure. This paper describes these and similar national as well.[73]

Open science community in Austria

1. Open Knowledge Österreich (www.okfn.at) is a community, where everybody can as a person participate. The goal is the promotion and establishment of open science within the Austrian research community. It is operated by the association Wikimedia Austria.

2. Open Access Network Austria (www.oana.at) is operated by the Austrian Science Fund together with the Platform of Austrian Universities. The network develops specific recommendations for the implementation of Open Access in Austria:

- Coordination of and recommendations for the Austrian OA-task/ activities of research institutions, funding organisations and research policies (including international developments)

- Positioning towards the information providers (mainly publishing houses)

- Contact persons and resource of information for scientists, research institutions and (research-) policies

3. E.infra Austria (www.e-infrastructures.at) is the homepage of the E-Infrastructures Austria projects, which started in 2014. The current project, called "e-Infrastructures Austria Plus", is a project carried out by nine Austrian universities and is running until December 2019. The goal is to implement infrastructure for „eScience". The developments are following the FAIR principle.

GERMANY

The German research community has been active in the open science movement since at least the beginning of the 21st century. In 2003 the Max Planck Gesellschaft sponsored the Berlin Declaration on Open Science in the Sciences and Humanities[74]. This declaration on open science was drafted "to promote the Internet as a functional instrument for a global scientific knowledge base and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives and museums need to consider."[75] This document was signed by representatives of 19 research organizations, including all of the primary German research organizations. The Berlin Declaration acted as the foundation for a number of follow on activities in

---

[73] Buschmann, K., Kasberger, S., Kraker, P., Mayer, K., Reckling, F., Rieck, K., & Vignoli, M. (2015). Open Science in Österreich: Ansätze und Status. Information. Wissenschaft Und Praxis, 66(2-3), 1–9.
[74] "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," Max Planck Gesellschaft, (2003).
[75] Ibidem.

Germany concerning open science and remains one of the cornerstones of Germany's open science initiatives. Today there are over 600 signatories to this initiative and the list continues to grow[76].

In 2008 the Alliance of Science Organizations in Germany launched its Priority Initiative "Digital Information"[77]. This initiative, supported by the major research organizations in Germany, lays out a long term plan for the development of the infrastructure and processes necessary to facilitate the broadest access to research and data for not only German scientists, but international researchers as well. This initiative was recently updated and the Alliance has reiterated its mission statement and extended its remit through the year 2022[78].

The Berlin Declaration and the Priority Initiative of the Alliance have influenced the development of the German government's position on open science. Both the German Research Foundation (DFG) and the Federal Ministry of Education and Research (BMBF) have established open access clauses for all research funded by either organization. In addition, the German government has adopted a secondary publication clause that has been incorporated into German copyright law. This law gives German researchers the legal right to publish their research on the Internet whether they have agreed to transfer exploitation rights to a publisher or not[79].

The many initiatives currently underway by the public research organizations in Germany are extensive. Approximately 300 open access journals are currently being published in Germany[80]. All public universities are required to have an open access policy by the Federal Ministry of Education and Research[81]. The primary research institutes (Fraunhofer, Max Planck, Leibniz, Helmhotz Association, DFG, German Academic Exchange, Alexander von Humboldt Foundation, German Rectors' Conference, German Council of Science and Humanities and German National Academy of Sciences Leopoldina) have developed detailed open science publication plans. Finally, the German standards organization, the DINI, has developed a set of standards for open access repositories and any services that will use thee repositories[82].

---

[76] https://openaccess.mpg.de/319790/Signatories, accessed 10 April 2018.
[77] "Priority Initiative "Digital Information" of the Alliance Partner Organizations," Alliance of Science Organizations in Germany (2013).
[78] "Shaping digital transformation in science, "Digital Information" Initiative by the Alliance of Science Organizations in Germany, Mission statement 2018 – 2022," Steering Committee for the "Digital Information" Initiative of the Alliance of Science Organizations in Germany, 2017. DOI: http://doi.org/10.2312/allianzoa.016
[79] Copyright Act of 9 September 1965 (Federal Law Gazette I, p. 1273), as last amended by Article 1 of the Act of 1 September 2017 (Federal Law Gazette I p. 3346).

[80] https://doaj.org/, accessed 10 April 2018.
[81] "Open Access in Deutschland: Die Strategie des Bundesministeriums für Bildung und Forschung," Bundesministeriums für Bildung und Forschung, 2017.
[82] https://dini.de/dini-zertifikat/liste-der-repositorien/, accessed 12 April 2018.

In an attempt to put teeth into the open access process, the Alliance for Science Organizations in Germany has established Projekt-DEAL[83]. This initiative is an attempt by the research community in Germany to step away from individual library relationships with the major scientific publishers and establish national relationships with these publishers. This initiative has set the German scientific community at odds with major academic journal houses such as Elsevier and Springer although these publishers are attempting to see how they might support the open access initiative in other, less economically damaging ways[84].

These numerous efforts underway in Germany provide clear evidence that open access is a serious objective of the German science community. While these efforts demonstrate a clear desire for research to be accessible and reusable, there remain a number of issues. The primary issue, as in any broadly defined, but locally implemented, initiative is the integration of access and search so that researchers do not have to spend time learning about, and accessing, multiple open access sites to obtain the data they are looking for. The DINI standard, along with the efforts of the Alliance of Science Organizations in Germany, are a good first step towards the goal of integration. What remains to be done will be the actual linking of the various open access databases so that a seamless access process can be realized.

GREECE

Research environment in Greece

The main governmental body responsible for research and technology in Greece is the General Secretariat For Research and Technology (GSRT - ΓΓΕΤ - www.gsrt.gr/), under the Ministry of Education, Research and Religion. GSRT supervises several research centers, e.g. ATHENA (www.athenarc.gr), CERTH (www.certh.gr), EIE (www.eie.gr), NOA (www.noa.gr), ITE (www.forth.gr) etc. and technological agencies. Besides the research centers supervised by GSRT, significant research work is conducted in state operated universities (e.g. NTUA, AUT etc.) and private research institutes.

Greek public funders in research have not yet fully imposed open access policies. A first step was performed in a 2014 call for academic research projects on the topic "Diversity, inequality and social inclusion" funded by the EU and operated by GSRT, and future calls by GSRT are expected to include policies for open access to publications and research data.

Open Science Initiatives

The Greek research community is aware of Open Access benefits and of related developments in Europe and internationally; however, publications and research

---

[83] https://www.projekt-deal.de/about-deal/, accessed 12 April 2018.
[84] Vogel, G., German researchers resign from Elsevier journals in push for nationwide open access, http://www.sciencemag.org/news/2017/10/german-researchers-resign-elsevier-journals-push-nationwide-open-access, accessed 12 April 2018.

results are normally not openly and freely available. The following Open Access initiatives are worth mentioning:

A web portal to promote Open Science has been developed by the National Documentation Centre and can be accessed at www.openaccess.gr/el. The portal provides information on Open Access benefits and current situation in Greece along with guidance to researchers, publishers, students, university managers and libraries/ repositories managers for Open Access promotion. No research data or publications are publicly available in the portal.

The Organisation For Open Technologies (https://ellak.gr/) supports a campaign for the promotion of Open Technologies and Science in the following five thematic fields:

- Open Software-Open Technologies and Open Standards,

- Open Governance and Open Data,

- Open Design and Open Hardware,

- Information Systems Security and Protection of Personal Information,

- Open Business.

Recently a call was addressed to academic institutions to support this initiative and participate in work groups with the aim to promote their use within the institutions.

A Conference "Open Science: Critical Issues and Future Prospects" organized by the National Documentation Centre, was held in Athens on June 15th, 2017 (http://www.ekt.gr/en/news/21001). The conference provided a forum for the discussion and exchange of ideas and best practices between the key stakeholders in the transition to the open science/ open access paradigm. Presentations focused mainly on European initiatives and on available resources, tools and training to support researchers in complying with the Horizon 2020 open access mandate.

A Conference "Open Science Fair 2017" (www.opensciencefair.eu/) was held in Athens on September 6-8, 2017, locally curated by the National Kapodestrian University of Athens and the "Athena" Research and Innovation Center and co-sponsored by the National Library of Greece, and served as an emblematic initiative of four EU projects in the area of Open Science: OpenAIRE, OpenUP, FOSTER and OpenMinTeD.

36 Greek Open Access Repositories are listed in the Directory of Open Access Repositories (OpenDOAR). Most of these repositories are part of the academic information system of universities and other educational institutions and provide access to theses, dissertations and (few) publications. Regarding their subjects, most are multidisciplinary, there are some with historical/ archaeological content, and others with geographical or environmental content. None is transport related.

NTUA Examples of Open Science

A list of Open Science Infrastructures in the transport sector (databases, observatories, decision support systems), developed with the participation of NTUA, is provided below:

SafeFITS - The Global Road Safety Model

Published on March 2018

(https://unecetrans.shinyapps.io/safefits/)

SafeFITS aims to facilitate knowledge-based transport policy decision-making related to reducing road traffic injuries. The model is based on historical road safety data and relationships between several road safety parameters, and provides information on different road safety scenarios. The SafeFITS tool includes three complementary modules: an intervention analysis module, to allow the user to forecast the safety effects of a specific road safety measure or intervention for a given country and time period, a forecasting module, to allow the testing of combined scenarios of interventions at national level, and a benchmarking module, to allow the user to benchmark a country against other countries.

SafetyCube - The European Road Safety Decision Support System

Published on 5 October 2017

(https://www.roadsafety-dss.eu/)

The SafetyCube DSS is the first global system with knowledge of both road safety risks and measures. It provides detailed interactive information, based on European and international evidence, on a large list of road accident risk factors and related road safety countermeasures.

Africa RSO - The African Road Safety Observatory

Published: not yet (1 October 2018)

(information in: http://www.saferafrica.eu/road-safety-knowledge-and-data/)

Africa RSO aims is to support policy makers and stakeholders with evidence of critical risk factors and identification of related actions and good practices on the basis of high-quality data and knowledge. It will incorporate knowledge and management tools, such as statistics, reports, fact sheets, knowledge resources and links, which will be openly available to stakeholders through a web portal, and provide a networking platform through which stakeholders and end users may contact experts, submit their own questions or data and exchange knowledge and experiences.

PRACT - The Repository of Accident Prediction Models and Crash Modification Factors

Published on 1 April 2015

(http://www.pract-repository.eu/)

The PRACT Repository incorporates an open online searchable database that contains the most recent and high quality Accident Prediction Models and Crash Modification Factors, highlighting effectiveness of road safety measures worldwide, for use by road safety decision makers and practitioners worldwide.

DaCoTA - European Road Safety Knowledge System

Published on 1 October 2011

(http://safetyknowsys.swov.nl/Home/about.html)

The DaCoTA Road Safety Knowledge System (update of the European Road Safety Observatory) includes road safety data on a wide range of topics and also brings together research from other projects. The system's contents include: Safety issues by age, road users, driver behaviour and more; Country overviews and forecasts across Europe; Statistics fact sheets and causation information; Methods used to gather high quality data; Important links from A-Z on road safety organisations; etc.

ERSO - The European Road Safety Observatory

Published on 1 October 2005

(http://ec.europa.eu/transport/road_safety/specialist/index_en.htm)

The European Road Safety Observatory (ERSO) gathers harmonised specialist information on road safety practices and policy in European countries. ERSO collects a range of information types including a series of data protocols and collection methodologies, national and in-depth accident data, exposure data and safety performance indicators.

NRSO - National Technical University of Athens Road Safety Observatory

Published on 12 July 2004

(www.nrso.ntua.gr)

The mission of the NTUA Road Safety Observatory is to support the Greek and the International Road Safety Community with current key road safety knowledge and data, which are gathered, analysed and organised within the research activities of the Department of Transportation Planning and Engineering of the School of Civil Engineering of the National Technical University of Athens, as well as within co-operations with various national and international road safety organisations.

POLAND

The Ministry of Science and Higher Education (MSHE) adopted on October 23, 2015 an open access (OA) policy included in the document "Directions of development of open access to publications and results of scientific research in Poland". The document is in the form of recommendations regarding the introduction of OA by entities financing research (the National Centre for

Research and Development [NCRD] and National Centre of Science [NCS], MSHE), scientific units, universities and publishers, including publishers of scientific journals. The most important recommendations included in the Directions for the development of open access to publications and the results of scientific research in Poland are as follows:

1. Development and adoption by individual universities, research institutes and institutes of the Polish Academy of Sciences (PAS) and NCRD and NCS of their own institutional policies in the field of OA, which will determine the rules of publishing the results of research in the OA (mainly applies to articles in peer-reviewed journals, but also e.g. peer-reviewed conference materials, possibly research data).

2. Designation of the plenipotentiaries for the OA by heads of scientific units and universities.

3. Transition of scientific journals to open models.

4. Providing doctoral dissertations in open repositories.

5. Monitoring and reporting to the MSHE on the progress in the implementation of the OA, including systematic analysis of the number of publications produced in a given scientific unit or university in order to determine the proportion of publications in the OA in relation to all publications.

6. Organizing training in the scope of the OA for all scientific employees and doctoral students of a given scientific unit or university.

7. Taking into account the experience and potential of scientific libraries, which often coordinate the process of editing and depositing scientific publications in repositories.

Initiatives concerning OA in Poland:

- ~1000 scientific journals in the OA, including the Reading Room of the PAS.

- Centre of Open Science - CeON (otwartanauka.pl) – run by the Interdisciplinary Centre for Mathematical and Computational Modelling at University of Warsaw.

- Open Science Platform (pon.edu.pl) – partonage of the Ministry of Science and Higher Education, CRASP and the Polish Academy of Sciences.

- Federation of Digital Libraries (fbc.pionier.net.pl).

- Different actors and projects for open science: Open Education Coalition, Open Science Library, Portal Unleash Science, Citizens of Science, Young Science Foundation.

- Scientific PlatonTV (tv.pionier.net.pl) – implements and publishes professional video materials in public in the field of results of scientific research, conferences, defence of doctoral theses, etc.

- Open Science Day – organized every autumn.

The main results of the study conducted on academic data sharing in Poland

Main results of the empirical study conducted to explore the phenomenon of academic data sharing in Poland (experiences with and attitudes towards the issue, enablers and obstacles, level of knowledge about legal and technical aspects) are described below. The research was conducted among 630 researchers from universities, institutes of the Polish Academy of Sciences, and research institutes in the second half of 2015.[85]

The vast majority of respondents turned out to **be in favour of sharing data**, as well as of **using data shared by others**. However, when these general statements are replaced by more refined questions, the respondents become more cautious and selective in terms of with whom to share and to what purpose the data could be used, so that general support for sharing should not be easily equalised with enthusiasm for open or public sharing.

The majority of participants also claimed that **they know where on the Internet they could share their own data with others and where they can find data shared by others**. On the other hand, answers to basic questions concerning legal issues revealed that in this area a large fraction of respondents wrongly assessed the correctness of the respective statements testing their knowledge. At the same time, **legal issues were among the most important factors when it comes to making a decision whether to use a dataset shared by others or not**. This may mean that a large fraction of participants is aware of the importance of these issues, but at the same time a lot of them lacks the necessary knowledge.

Career-related factors were most frequently chosen as very important when making a decision about sharing or not sharing data. These were factors such as citations or having enough time to finish all planned publications before a dataset becomes available to others. Direct financial benefits were considered important by a far smaller fraction of respondents.

At the same time, the majority of researchers considered it **important that they would be able to decide who and for what purpose will be allowed to use their data**. This may mean that some ways of sharing data (in terms of the scope of potential users and potential ways of using a dataset) are more acceptable than others.

Also, among the factors that would **prevent a researcher from sharing data the most important were those that could impede a scholarly career**, such as being outrun with future publication by other researchers benefiting from the data.

The survey participants wanted **sharing to be simple and effortless** and a large fraction of them would probably resign from sharing if it would require a

---

[85]

http://pon.edu.pl/images/plon_publications/files/19_Towards%20Open%20Research%20Data%20in%20Poland.pdf (18.03.2018)

significant effort. Interestingly, when it comes to the effort already made when producing the data, the fraction becomes much lower, so it is rather an additional effort that could prevent them from sharing.

The respondents had no objections towards sharing data with researchers whom they know personally, who work in the same institution, or even who conduct noncommercial research in general. But when it comes to researchers who perform commercial research, the situation changes: here more than a half of the respondents are undecided or even oppose sharing.

SLOVAKIA

National research environment

The Ministry of Education, Science, Research and Sport of the Slovak Republic (MESRS SR) is the main body supporting research and development funded from public resources. There are public, private and state operated universities, research organisations (e.g. Slovak Academy of Science - SAS), research institutes and private companies. MESRS SR is responsible for the process of evaluation of a research organisation to perform Research and Development (R&D) activities. The certificate for research organisation as result of the evaluation process allows to use public resources for R&D.

Slovak Centre of Scientific and Technical Information (SCSTI) is the national information centre and specialised scientific library of the Slovak Republic focused on natural, technical, economic and social sciences. The SCSTI provides several information systems supporting R&D on national level funded by the ministry, i.e. Central Registry of Publication Activities, Central Registry of Theses and Dissertations, Central Information Portal for Research, Development and Innovation and Slovak Current Research Information System (SK CRIS).

Since 2013 SCSTI serves as the National point of reference for the policy of "Access and preservation of scientific information".

Major research funders

The major funder of all scientific projects in the country is the government. Funding provided through grant agencies: Slovak Research and Development Agency (SRDA), Scientific Grant Agency (VEGA), Cultural and Educational Grant Agency (KEGA) and the system of Incentives for R&D. Support of the EU for Research and Innovation (R&I) process is provided through Operational Program R&I funded by European Structural and Innovation Funds (ESIF).

Open Access stakeholders

The Slovak research community is aware of Open Access benefits but is still reluctant to provide Open Access to their publications and research results. Only a small fraction of researchers deposits their research papers - this is often a result of proactive advocacy and training provided by the library staff.

The main Open Access stakeholders in Slovakia are:

- Slovak Centre of Scientific and Technical Information (the National point of reference for the policy of "Access and preservation of scientific information", partner of OpenAIRE2020 project, LIBER member)

- University Library of Bratislava

- Slovak National Library (partner of Europeana libraries)

- Central Library of SAS

- The Slovak Library Association (International Federation of Library Associations and Institutions (IFLA) member)

- The most important Slovak universities and their libraries.

Since 2007, SCSTI as provider of SK CRIS system, has been a member of EuroCRIS, the not-for-profit organisation developing interoperable Research Information Systems.

In October 2016, the Open Access Working Group was established in Slovakia. It has 21 members from libraries, ministries and Slovak Academy of Sciences (SAS).

"The Action Plan for Open Government in Slovak Republic, 2017 - 2019" was approved by Government of the Slovak Republic in March 2017. There is a special section on Open Access to results of Science and Research with the Action Plan that includes:

- Establishment a Contact office for OA

- Introduction of OA basic principles

- Implementation of public license Creative Commons Attribution (CC BY)

- Establishment and operation of a repository.

Open Access repositories and publishing

There are no Open Access repositories in Slovakia. Some projects on setting up Open Access repositories are ongoing.

The Central Registry of Theses and Dissertations provides full text access to some theses and dissertations but does not meet all requirements for an institutional repository. Most of universities provide their repositories as a part of academic information system. However, these repositories do not have characteristics of an institutional repository.

There are 42 open access journals published in Slovakia. There are some open access journals available only from the web pages of the institutions, which do

not have any long-term preservation strategies so they often disappear from institutional websites when a project or funding is finished.

UNITED KINGDOM

Policy

There are seven research councils in the UK, previously known collectively as RCUK and now part of UK Research and Innovation. In 2005 a joint statement was released by RCUK on open access to publications. This has been revised over the years to provide further guidelines on implementation, specifically around Article Processing Charges, acknowledgement of funding, licenses and repositories, as in the current 2013 version[86].

Each of the seven research councils also has a data policy[87] and efforts have been made to harmonise requirements. A set of Common Principles on Data Policy were released in 2011[88]. Subsequently, a multi-stakeholder group representing funders, universities and others released a UK Concordat on Open Research Data in 2016[89]. RCUK, HEFCE, Universities UK and the Welcome Trust are all signatories. The associated Open Research Data Task Force is now investigating implementation of this across the Higher Education sector.

Many UK universities also have research data management policies with statements encouraging open science practices. The DCC collates a listing of institutional data policies[90] and roadmaps for implementation[91]. The Engineering and Physical Sciences Research Council (EPSRC) was particularly instrumental in the responses of UK universities as it made explicit their responsibilities to provide an environment in which researchers were supported to manage and share data.

Many journals have policies on the deposit of associated research data and some such as Scientific Data recommend specific repositories to use[92]. A Research Data Alliance interest group is investigating the harmonisation of journal data policy based on work by Springer Nature93. The Directory of Open Access Journals list 1330 titles in the UK.

The UK is also pursuing and agenda towards open government data to ensure public sector information is made widely available for broad reuse. The current

---

[86] https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf/
[87] See summaries on the DCC website:http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies
[88] See UKRI website: https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/
[89] Available at: https://www.timeshighereducation.com/sites/default/files/breaking_news_files/concordat_on_open_research_data.pdf
[90] http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies
[91] http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps
[92] https://www.nature.com/sdata/policies/repositories
[93] See https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation

National Action Plan (2016-2018) is the third edition, and an increasing quantity of data is available via the Open Data portal[94].

Key infrastructure and initiatives

Repositories and data centres

The UK has an excellent network of publication repositories. Many UK universities offer a repository service and were supported to develop these by the former Repository Support Project. There are also a number of well-established subject repositories such as arXiv and Europe PubMed Central, offering researchers many routes to green open access publishing despite the RCUK policy preference for Gold Open Access[95]. OpenDOAR, the directory of open access repositories, lists 258 repositories in the UK.

There are also many institutional and disciplinary data centres in the UK. Some are supported by research funders such as the NERC data centres and the UK Data Archive, while others such as the Archaeology Data Service or institutional services like Cambridge may charge for their services[96]. Some disciplinary data centres are based in the UK but have an international remit e.g. ArrayExpress at Hinxton[97]. The Re3data registry of research data repositories lists 286 data repositories in the UK.

UUK Open Access Coordination Group

Universities UK launched a group to bring together funders, institutions, publishers and other stakeholders on Open Access. The Group works to ensure that the activities to support the transition towards open access in the UK can be effectively coordinated and that progress can be monitored. The Group publishes periodic reports monitoring the transition to open access[98].

Open Research Data Task Force

The UUK Task Force responds to a recommendation that a 'roadmap' for national open research data infrastructure be produced. The Group will lead and coordinate this work.

Jisc Research Data Shared Service

Jisc is currently brokering deals with providers and piloting a range of data services and to offer the UK higher education sector a shared service for storing, publishing and preserving research data. Universities will be able to buy into some or all of the services as required.

---

[94] See: https://data.gov.uk
[95] See RCUK blog post: https://blogs.rcuk.ac.uk/2012/10/24/rcuk-open-access-policy-our-preference-for-gold/
[96] See for example ADS charging policy at:
http://archaeologydataservice.ac.uk/advice/chargingPolicy.xhtml or Cambridge policy at:
https://www.data.cam.ac.uk/repository
[97] ArrayExpress: https://www.ebi.ac.uk/arrayexpress
[98] See for example: http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/monitoring-transition-open-access-2017.aspx

Digital Curation Centre

The DCC provides guidance and expertise on research data management and curation. It tracks funder policy, provides briefing papers and practical how-to guides, and runs services such as the DMPonline tool for Data Management Planning. It also runs an annual International Digital Curation Conference and a Research Data Managers Forum (RDMF) to promote networking and skills sharing.

Open Knowledge International

Open Knowledge is a global non-profit network that focuses on realising open data's value to society by helping civil society groups access and use data to take action on social problems. It was founded by Rufus Pollock on 24 May 2004 in Cambridge. As of 2018, Open Knowledge International has 11 official chapters and 38 groups in different countries.

Open Data Institute

The Open Data Institute was co-founded in 2012 to advocate for the innovative use of open data to affect positive change across the globe. Based in London, the ODI works with companies and governments to build an open, trustworthy data ecosystem, where people can make better decisions using data and manage its harmful impacts.

ATTACHMENT II. GLOBAL APPROACHES TO OPEN TRANSPORT DATA

Numerous studies and authors have noted that the increase of data required to address today's emerging scientific questions no longer allow individuals to develop their own private data sets. In addition, the international nature of the problems that science must address today also implies that data must be accessed across international boundaries. These two facts combine to demonstrate that modern science can no longer take a parochial and private view to research, but must look to working in international cooperation to advance science and address the problems facing society today. Adding to the need for international cooperation is the variability in scientific capabilities between countries. If science is to act as a leveller of inequality and create social harmony, then all nations must be able to access and learn from the global community of scientists.

Many arguments have been put forward as impediments to international scientific collaboration. These arguments can be summarized using the terms exploitation, culture, regulations, and national security. Each of these arguments have merit, but are not actual inhibitors to scientific data sharing. Successful international platforms for data sharing exist (e.g., CERN) and provide examples of where careful examination of these potential inhibitors, proper safe guards and governance, and sensitivity to cultural differences can overcome the arguments.

The key challenge to any cooperative endeavour, whether local, regional or global, is the development of trust between the collaborators. Trust requires significant time to develop, transparency in the relationship, and the understanding of the value of the relationship. Additional challenges arise in international collaboration due to language differences, differing country approaches to science, non-standard data structures, collection processes, quality assurance processes, documentation and other technical elements that can create roadblocks to the actual sharing of data.

Because international cooperation in the realm of transport research is a normal aspect of current research problems, it would be wise to understand how mature the major regions of the world are in their journey to open data and, most importantly, in their approach to open transport data.

North America (USA & Canada)

Significant work is being done in the area of open data in both the United States and Canada. In the United States national governmental entities, state governments, regional and local governments have all embraced open data as a means of creating transparency and trust[99]. The implementation of open data portals has become commonplace (e.g., catalog.data.gov, opendata.cityofnewyork.us, data.ca.gov). Unfortunately, these implementations

---

[99] https://opendataamerica.opendatasoft.com/pages/home/;https://www.data.gov/; https://open.canada.ca/en/open-data.

are ad hoc without adherence to any formal standards, which causes problems when someone attempts to search for data and then use it. In addition, there is no single portal that one can access that links the current and planned portals together so that a universal search for data can be accomplished. More structured data portals have been developed for specific scientific disciplines, but cross discipline sharing of data is inhibited due to a lack of standards for formatting, searching, and curating the data.

Canada has followed a similar path to that in the US with governmental entities embracing open data, but lacking any standards for implementing their open data platforms (e.g., open.canada.ca, oae.edmonton.ca). Collaborative scientific platforms provide domain partners with access to shared data, but once more cross domain sharing is difficult due to differing standards of practice (e.g., conp.ca, codx.ca).

In both Canada and the US transportation related data are provided at the federal, state/province, and local city levels. The US Department of Transportation, in conformance with the US Open Government initiative, maintains a catalogue of open data sets on the US data.gov website (https://catalog.data.gov/dataset). In a similar manner, the Canadian Department of Transportation maintains a catalogue of its open data sets on the open.canada.ca website (https://open.canada.ca/data/en/dataset). State, province, various city websites can all be queried to obtain various levels of open data of interest to transport researchers.

Since no single portal in the US or Canada has succeeded in acting as the window into all open data repositories, commercial entities have entered the open data access game (e.g., socrata.com, usadata.com, opengovca.com, datagovca.com). These commercial endeavours provide integration services to governmental data owners (socrata.com, opengovca.com, datagovca.com) or integration and curation services for public and commercial users by finding and curating data from governmental open data sources (usadata.com). As non-profit commercial entities these organizations charge for their services or are independently funded by foundations, in theory releasing governmental agencies from the burden of having to fund their own portals (note that this is "in theory" only as many governmental agencies are required by law to provide their own portal for access to the data that they generate).

Latin America

Latin American countries have traditionally been suspicious of openly sharing governmental data. Several countries (e.g., Brazil[100], Mexico[101], Argentina[102]) have begun to embrace a more open approach to government generated data. These centralized governmental data portals provide researchers with the ability to access statistics similar to those collected in most Western countries. Transport data is included in all of these portals on a macro and regional basis.

---

[100] http://dados.gov.br/
[101] https://www.datos.gob.mx/
[102] http://datos.gob.ar/

Local data from municipalities or regions is less well developed with only a few major cities actually providing open data portals. Recently, the Latin America Open Data Initiative[103] has begun to work on various projects to more broadly encourage governmental sharing of data and the opening of data for use in various research endeavours.

Europe

As this document has demonstrated, Europe has been a leader in the open data discussion establishing a number of open data portals (e.g., www.europeandataportal.eu). In addition, Europe has sought to include open data requirements into its research funded projects ensuring that data generated through public funds to researchers is made available[104]. Open data in the transport area within Europe is, like in North America, maintained at the EU level, state level, region within state level, and city level. The work of the EU in developing the EOSC, and in examining the requirements for a TRC, position it as a leader in attempting to integrate the numerous separate transport and transport research data repositories so that transport researchers can access the data they need in a more streamlined manner.

Russia and the CIS

Russia and the former Soviet satellite countries have been very hesitant to adopt the open science concept. In 2014 the Russian government began an open data initiative that resulted in the establishment of the national open data portal http://data.gov.ru. With the creation of this open data portal the Russian Federation has begun the journey to opening up at least a portion of its national statistics for examination by researchers. This portal provides access to transport researchers interested in macro-data concerning transport statistics for Russia. However, limited granular data from the many cities and regions in Russia is available for research as these governmental entities are still very hesitant to open their data to external users.[105]

Africa

The open data initiative in Africa is fragmented and varies in maturity by country. Many countries, due to either economic or political issues, do not make country statistics or data available to the general public. Other, more politically and economically stable countries have begun to provide country statistics and their data sets to the public. Initiatives, such as those of the African Development Bank[106] and the openAFRICA[107] project, are attempting to provide an EOSC like single access point for finding and accessing open data sets from various African nations. In addition, there is a growing awareness on the African continent in the importance of open government and open data. While most African countries still have some ways to go to provide the comprehensive data that Europe

---

[103] https://idatosabiertos.org/en/acerca-de-nosotros/
[104] https://ec.europa.eu/digital-single-market/en/policies/76011/3502
[105] Koznov, D., O. Andreeva, U. Nikula, A. Maglyas, D. Muromtsev, I. Radchenko (2016), "A Survey of Open Government Data in Russian Federation," KMIS, November 2016, pp. 173-180.
[106] http://dataportal.opendataforafrica.org/
[107] https://africaopendata.org/dataset

provides through its open data efforts, there are positive signs that most African nations understand the need to provide at least general country statistics to the public.

With respect to transport data, the data sets that are available are still quite incomplete. Few comprehensive city, region and state transport data sets have been made public. Current efforts are focused primarily on health and food related data.

Middle East and North Africa

The Middle East and North African region has been classified by the Open Data Barometer as an area that is regressing in its open data efforts. In the most recent review available the Open Data Barometer indicates that there were no truly open data sets available in the region.[108] The reason given by the Open Data Barometer to this unfortunate observation was that there was little pressure from the public for governments to open their data so there was no incentive for the governments to work towards the public sharing of statistics. This is unfortunate for transport related researchers as significant growth is occurring in this region and access to reliable data concerning transportation activities would be beneficial to researchers.

Far East

Attempting to classify the vast territory covered by the category "Far East" is very difficult. Countries as diverse as India, Pakistan, China, Singapore, Indonesia, Malaysia, Japan, Korea, Vietnam, Brunei, Thailand, Myanmar, and Cambodia make one very hesitant to make a statement on how well the region is embracing open data. The most one can say is that for most of the region there is an increasing recognition that open data is important. Countries such as India[109], Singapore[110], Japan[111], and South Korea[112] have government open data sites very similar in concept and structure to typical Western countries. Countries such as China[113], while providing some open data, are just beginning their open data journeys.[114]

Oceania

Australia[115] and New Zealand[116] have been at the forefront of the open data initiative. Australia in particular has driven significant advances in global thinking

---

[108] https://opendatabarometer.org/4thedition/regional-snapshot/middle-east-north-africa/#key_finding
[109] https://data.gov.in/
[110] https://data.gov.sg/
[111] http://www.data.go.jp/?lang=english
[112] https://www.data.go.kr/main.do?lang=en
[113] http://data.stats.gov.cn/
[114] Liu, X., Y. Song, K. Wu, J. Wang, D. Li, Y. Long (2015), "Understanding urban China with open data," Cities, vol. 47, pp. 53-61.
[115] https://search.data.gov.au/
[116] https://data.govt.nz/

concerning open data in general and research data in particular[117]. Australia's open research data portal is an example of what one would hope the EOSC will become. The open research portal links together a federated family of research infrastructures that are overseen by the National Research Infrastructure for Australia[118], a strategic program of the Australian government. The National Research Infrastructure Roadmap identifies nine focus areas where Australia is attempting to establish a strong research footprint[119]. Unfortunately for the transport sector, transportation research is not one of these nine focus areas.

---

[117] https://researchdata.ands.org.au/
[118] https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris
[119] https://www.education.gov.au/2016-national-research-infrastructure-roadmap

ATTACHMENT III. EXAMPLES OF OPEN ACCESS USE CASES

Automation in Road Transport

Definition

Automation in road transport refers to the transport system and all its components (vehicles, drivers, users, road infrastructure, information systems and applications to name the most important)[120]. Automation is often used to define something to be "smart", in which automation takes over control from humans to do the right thing in complex events or circumstances. Automation can in addition also prove valuable in non-complex circumstances and it is not necessarily the one or the other.

The term "connected and automated vehicle" can refer to a variety of vehicle technologies currently being implemented to improve travel. These technologies may work at the level of the vehicle, the road infrastructure, or both[121]. Many types of connectivity and automation are feasible, as are many ways to combine them. For example, some vehicles could be connected without being automated, and possibly others could be automated without being connected (though increasingly, vehicles are connected one way or the other, even if only via a 4G LTE device inside the vehicle). Meanwhile, an automated vehicle could theoretically only rely on information from its sensors (camera, radar, etc.) to perceive the external environment, and human-operated vehicles can have connectivity applications (telematics, GPS, etc.). Moreover, both connected and automated systems are often conflated with Intelligent Transport Systems (ITS). ITS may include connected and automated vehicle systems, but is a much broader concept involving a variety of advanced ICT applications that go beyond vehicle systems. For example, connected and automated vehicle technologies may or may not be integrated into ITS, depending on the specific application.

At the road infrastructure level, some systems are already automated (e.g., Automated Incident Detection (AID), but often a human operator is in full control due to safety requirements. At the most advanced vehicle level, both in Europe and the U.S. there are prototype vehicles driving automatically both in urban and in highway environment. Hence from a technical point of view, current technology for highly automated driving in controlled environments is quite mature. The vehicles use state-of-the-art sensors (radars, lidars, GPS and camera vision systems) combined with high accuracy maps, so the on-board systems can interpret the information to identify appropriate navigation paths, as well as obstacles and relevant signage. Still, for these prototypes the driver must always be ready to take over and these systems are partial automation systems only. The pace of further developments and deployment of new advanced technologies

---

[120] Roadmap, Automation in Road Transport, iMobility Forum, Version 1.0, May 2013, http://vra-net.eu/wp-content/uploads/2013/12/Automation-Roadmap-final.pdf

[121] Planning for Connected and Automated Vehicles, (2017). http://www.cargroup.org/wp-content/uploads/2017/03/Planning-for-Connected-and-Automated-Vehicles-Report.pdf

(such as Artificial Intelligence) which aim for full automation of the human intervention or control is the highest at the individual vehicle level.

Other applications of partial automation like platooning or cooperative traffic systems are also quite advanced and mature from a technical point of view and are being tested and assessed on their impact in various field operational tests across Europe, the US and Japan.

Further research and enhancements of existing prototypes and systems are needed to succeed real driving conditions and allowing a specific time buffer for driver take-over. This would mean a transition takes place from partial automated driving to highly automated driving.

SAE automation levels

The technologies for autonomous cars, connected cars, and advanced driver assistance systems are often mixed. Fully automated, autonomous, or "self-driving" vehicles are defined as "those in which operation of the vehicle occurs without direct driver input to control the steering, acceleration, and braking and are designed so that the driver is not expected to constantly monitor the roadway while operating in self-driving mode."[122] There have been multiple definitions for various levels of automation, however the SAE International definitions for 6 levels of automation[123] has been adopted by the most of relevant stakeholders in Europe and U.S. These definitions divide vehicles into levels based on "who does what, when." SAE J3016_201609 provides a taxonomy for motor vehicle driving automation systems that perform part or all the dynamic driving task on a sustained basis and that range in level from no driving automation (level 0) to full driving automation (level 5). SAE J3016_201609 also refers to three primary actors in driving: the (human) driver, the driving automation system, and other vehicle systems and components (table AIII-1). These other vehicle systems (or the vehicle in general terms) do not include the driving automation system in this model, even though as a practical matter a driving automation system may actually share hardware and software components with other vehicle systems, such as a processing module(s) or operating code.

Within the SAE taxonomy, the upper levels (three through five) are distinguished from lower levels by the fact that the automation system is performing the entire dynamic driving task. Levels three through five are concerning to policymakers because this implies that no human is controlling the motion of the vehicle in real time.

The link to Open Science and Open Access

The advancement of connected and automated vehicle technologies, as well as the significant increase in the number of personal mobile devices and amount of "crowdsourced" information, will create new dynamics of transport data

---

[122] U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA)
[123] SAE International, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, J3016_201609, https://www.smmt.co.uk/wp-content/uploads/sites/2/automated_driving.pdf

environment. The new data environment will include various actors (individual, business, public sector, third parties), and it can be characterized by the following key transformations: big data in transport, multiple data sources (vehicle itself through sensors, controllers and processors, other vehicles in the surrounding, road infrastructure, ITS) and application platforms, and mobile data communications. A huge amount of the data has been and will be generated within research projects, publications and other research activities. Considering a complex transport data environment in terms of actors, data sources, etc., a strategy for Open Science and Open Access to transport data should be developed hand in hand with clear legislation and a legal framework. Some important elements are described in the next section.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	System	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	System	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	All driving modes

Table AIII-1. SAE International J3016 Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles summary, https://www.smmt.co.uk/wp-content/uploads/sites/2/automated_driving.pdf

A legal Framework as basis for Open Access

There is considerable need for clear legislation and a legal framework that will support data security, data protection and privacy before a full rollout. Policy matters that need addressing include also fair competition, cybersecurity, road safety and liability. The interest of third-party service providers and new competitors in accessing vehicle data and using them for commercial purposes is an issue that requires particular attention. The EU should establish a regulatory framework for access to vehicle data that will also take into account the fact that vehicle manufacturers invest heavily in the ability of vehicles to generate data

and are ultimately responsible for ensuring the vehicle's safety and integrity as well as the protection of the user's personal data and privacy[124].

Mobility as a Service (MaaS)

Definition

One of the most used terms when speaking about new multimodal services is "Mobility as a Service". With this end-user-driven approach the goal is to ensure that an end-user is reaching his destination by what means ever. In the MaaS concept a MaaS Service Operator is providing tailor-made mobility services including all kinds of transport modes and covering all transport related issues from planning over booking to payments. In a final stage of MaaS it is expected, that travellers are using monthly mobility packages, which allow them to use transport modes in accordance to their individual needs. In such a MaaS environment it is not expected, that a traveller will get tickets for single transport modes, but that a combined MaaS ticket will allow him to use whatever mode, if it is public transport, individual car, a shared bike or a taxi service. All related actions concerning information, booking and payment of single trips is organised centrally by the MaaS Service Operator.

One enabling technology hereby is with Big Data management and Artificial Intelligence (AI). The MaaS Service Operator needs to get access to static as well as real time data of the single transport modes (if private or public organised). Hereby static data (like timetables, stations, geographical information (GIS) on the transport network itself, fare prices, etc.) are combined with real time data, including traffic status of single modes, delays of public transport, occupation rate of taxis, etc. This combination of big and small data and a real time data processing enables a MaaS Service Operator to provide individual tailored services for individuals.

Hereby it can be expected, that in future not only the combination of single data, but also the linking of services (e.g. information or ticketing services provided by single transport operators) is needed to provide a sufficient and sustainable MaaS service.

A possible MaaS Architecture – several interfaces enabling individualised end-user services

A possible MaaS ecosystem will include data access at different levels: at the public-private MaaS service level, where e.g. routing services or payment services are generated as well as on the public/private MaaS Service Operator level, where data and services are integrated into one tailored MaaS end-user service (Figure AIII-1.).

---

[124] ACEA Strategy Paper on Connectivity, (2016).
http://www.acea.be/publications/article/strategy-paper-on-connectivity
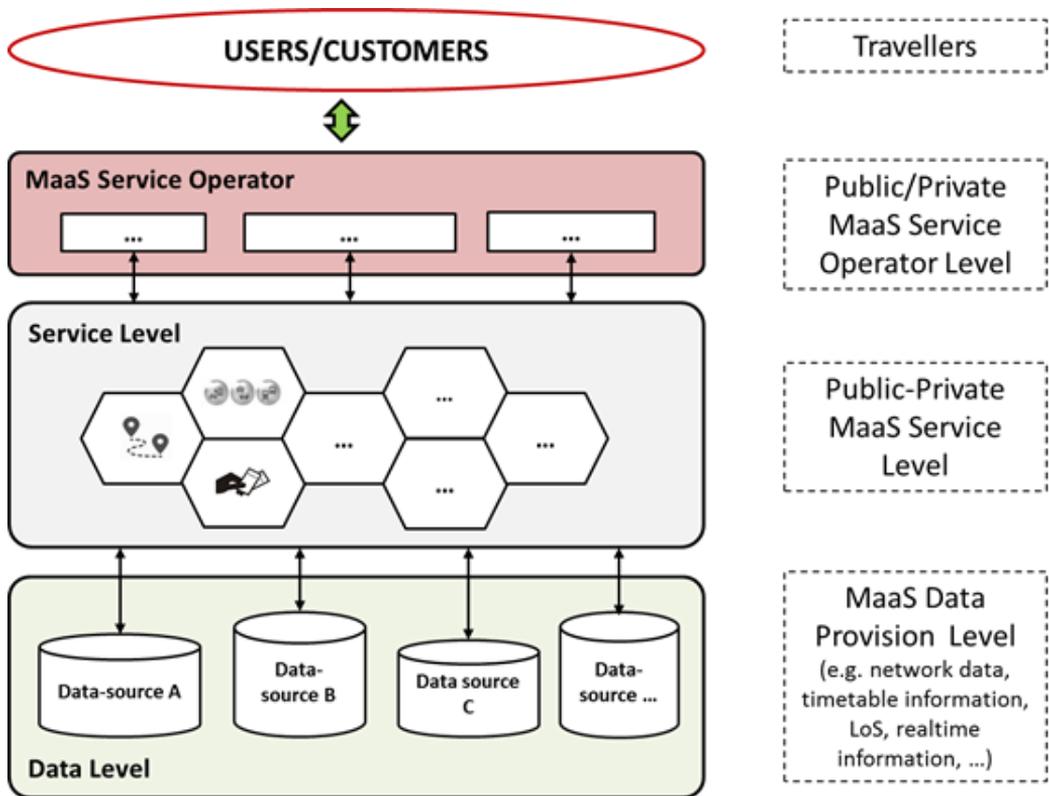
Figure AIII-1. Possible MaaS Architecture

The link to Open Science and Open Access

Even MaaS is a possible end-product of the current deployment of multimodal traveller information services this might have some consequences for Open Science as well. This especially as access to data and services, which will be the backbone for future MaaS services, will not only be given within the business value chains, but might be available as well for research and innovation actions.

Therefore a short analysis of the actions done to make data available will be given. This process might helpful as well for the establishment of a European Open Science Cloud.

A legal Framework as basis for Open Access

The whole process of giving access to data from the public sector started with the publication of the PSI[125] Directive (2003/98/EG)[126]. Here the re-use of public sector information is regulated. What came out was that this was seen more as an advice, than as an obligation. Therefore the PSI Directive was improved in 2013 (2013/37/EU)[127] making the opening of public sector information an

---

[125] Public Sector Information
[126] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:32003L0098
[127] https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32013L0037

obligation and having a clear indication, that digital information and data is a part of public sector information.

Also in the transport sector the digitalisation became more and more important. Especially providing access to data was becoming a central request. Therefor the European Commission published the ITS Directive (2010/40/EU)[128] as a framework directive, where later on specifications of specific datasets to be opened were published as Delegated Regulations.

The datasets mentioned in the afterwards following Delegated Regulations deal with parking data (2013/885), road safety data (2013/886), traffic information data which even include historical data (2015/962), and finally travel information data covering all modes including sharing, public transport, walking, etc. (2018/1926). Insofar the WHAT on data provision is clearly regulated including a roadmap until when all data shall be made available (until 2023).

Aside the "what" also the WHERE is defined. Hereby Member States are requested to setup National Access Points (NAPs) to transport related data. These National Access Points need to provide access to the datasets defined under the ITS Directive. Hereby it is not necessary to have all data physically stored at the NAP, but it is fine to set them up as a registry services including search possibilities based on Metadata description. Here Member States discussed a common Metadata-schema, which is now forming the basis for several NAP implementations, e.g. in the Netherlands (National Data Warehouse NDW), in Germany (Mobility Data Marketplace MDM), or in Austria (www.mobilitydata.gv.at). This Metadata scheme gives a good description of the single data sets including quality data or actuality indicators.

Finally also the HOW is defined in the Delegated Regulations. The interfaces to datasets have to follow specific Interface standards, in detail these are DATEX II (for dynamic road data), Inspire (for geographical data), NeTEx (for static public transport data), or SIRI (for dynamic public transport data). In addition the possibility of linking services is described by using the OPEN_API standard.

Even someone might say, that for such an opening of transport related data a Delegated Regulation might have been the wrong way forward and that instead of a regulatory framework more freedom should have given to the market itself, these specifications are of high importance for public authorities and public transport infrastructure operators. This especially for following reasons:

- The Delegated Regulations try to ease access to data in a common format, which aims of a quick uptake by the private market

- The Delegated Regulations give a clear guidance to Member States on what to do. Such a guidance eases the process of implementing an open access

---

[128] European Union Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010, "on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport," Official Journal of the European Union, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040.

infrastructure, as it insures public authorities will have no stranded investments (all Member States are providing access in a similar way)

- In addition the Delegated Regulations provide the ground for addressing directly data holders for opening their data sets

At the end it is to hope, that the now accessible data sets will not only be used for improving existing services, but that new services as foreseen in the MaaS environment will be enabled. In parallel these opened data sets should as well be available for other research activities

Current state of discussion on Open Access

- Aside national access points it is currently discussed to implement as well European wide cloud services, where Europe-wide acting stakeholders (e.g. car manufactures) can make their data accessible. For that a specific cloud environment needs to be setup accordingly. Ideally metadata services are linking to content provided by data owners and not all data is stored in that central cloud.

- But if data is stored on a central place, it is important to identify a trusted environment where data is accessible on a discrimination free basis. And here the big question is with the "who" can operate such a trusted environment. Ideally it is an organisation that has no business interest in the transport sector.

- The formats of the interfaces for data access are defined. It can be expected, that all stakeholders will follow the defined Interface standards. That of course means, that additional efforts are needed to transfer data stored and used in proprietary systems in the respective exchange formats. Here a possible solution could be the definition of a kind of Middleware which is linking datasets of different formats.

- For possible future MaaS operators it is in addition important to understand, what kind of input is needed. This might be services linked via OPEN-API, it might be interpreted data (information) in accordance to the Delegated Regulations of the ITS Directive, or it might be raw data that might be accessible (in accordance to the PSI Directive) only in sometimes proprietary formats.

ATTACHMENT IV. RESULTS OF THE SURVEY CONDUCTED ON TRC NEEDS, OBSTACLES AND EXPECTATIONS

The goal of the survey was to identify the main needs, obstacles and opportunities for open data sharing and the open science model in the area of transport research. The survey was conducted among researchers from academic institutions, representatives from public body (e.g. State Department of Transport), commercial sector and general public. Data collected through this survey was anonymised, used to inform recommendations and published under an open licence. The survey consisted of four parts: transport research data, cloud service requirements and expectations, opportunities and barriers and funding mechanism.

A total of 87 responses were collected between June and July 2018. Respondents came from 29 countries (11% from United Kingdom, 9% from Slovakia and Greece each, 7% from Germany and from Austria, 6% from Israel and 5% from Poland and from Spain, the single replies also came from Switzerland, Portugal, Italy, the Netherlands, Sweden, Serbia, Macedonia, France, Finland, Cyprus, USA, Romania, Norway, Malta, Latvia, Iceland, Hungary, Czech Republic, Canada, Bulgaria, and Bosnia and Herzegovina).

The largest group of respondents were researchers working for academic institutions – 85%, followed by representatives of commercial sector – 6%, public bodies – 4% and others. In terms of relationship with transport data respondents, in most of the cases, "analyse transport data" (28%), "use transport data" (25%), and "process data" (20%). The distribution of answers is shown in Figure AIV-1.
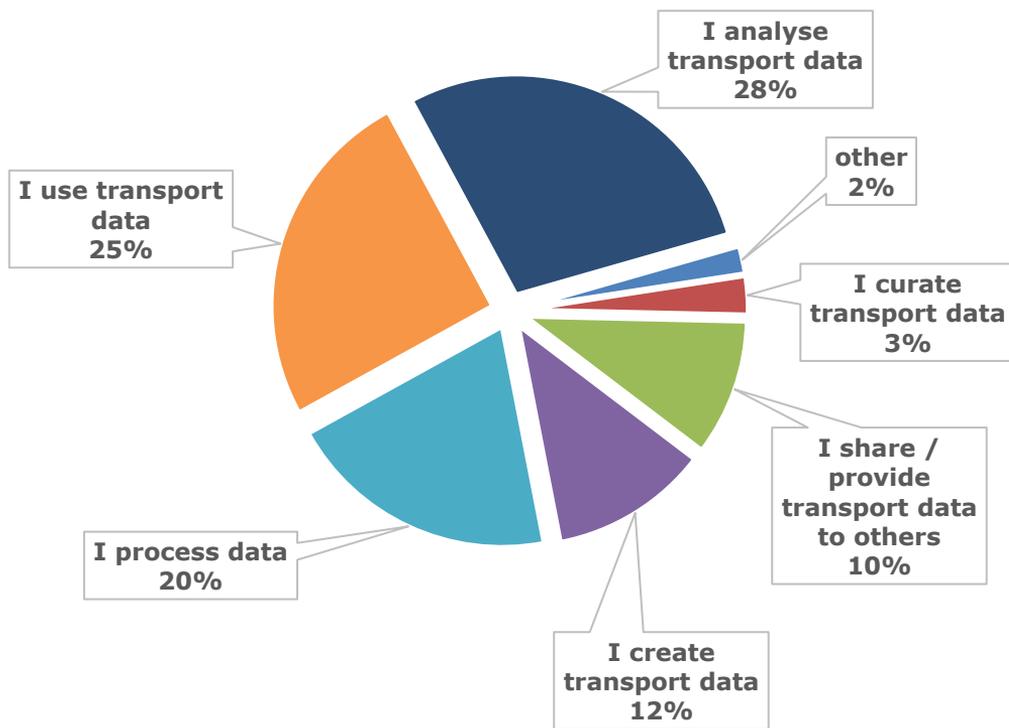
Q1: What is your relationship with transport data?

Figure AIV-1. Respondents' relationships with data transport

Respondents were asked if they share data whenever it is created by them. 28% admitted that they share data but mostly under restrictions to control access and reuse, 20% publish data which underpin research results, 20% only share data with collaborators on request and 5% share data publicly under an open licence whenever possible.

Part I. Transport research data

In the first part of the survey respondents were asked for the kinds of transport data that is mostly needed for their work. The majority of participants claimed that traffic flow and urban management data is the most needed. Further they need "accident data" and "sensor data from operations", and additionally "timetable data" was valuable source of information (Figure AIV-2).

Traffic flow / urban management data	17%
Road/ Rail/ Air/ Maritime accident data	11%
Sensor data from operations	10%
Timetable data	10%
Road/ Rail geometry & design data	8%
Logistics and delivery data	8%
Vehicle/ Ships/ Airplane emissions data	8%
Vehicle/ Ships/ Airplane fuel consumption data	7%
Vehicle test data	6%
Accessibility data from studies with disabled users	4%
Air traffic data	3%
Port operations data	3%
other	5%

Q2: What kinds of transport data are needed for your work?

Figure AIV-2. Kinds of transport data needed for work purposes

At the same time respondents were asked what kind of data would be important for their work if open data on transport operation were to be available. Traffic flow / urban management data, accident data and data from sensors are "top three" data typew that were rated the most often as "the most important" and "very important" by researchers. Also, the timetable data and road or rail geometry and design data were underlined (Figure AIV-3).
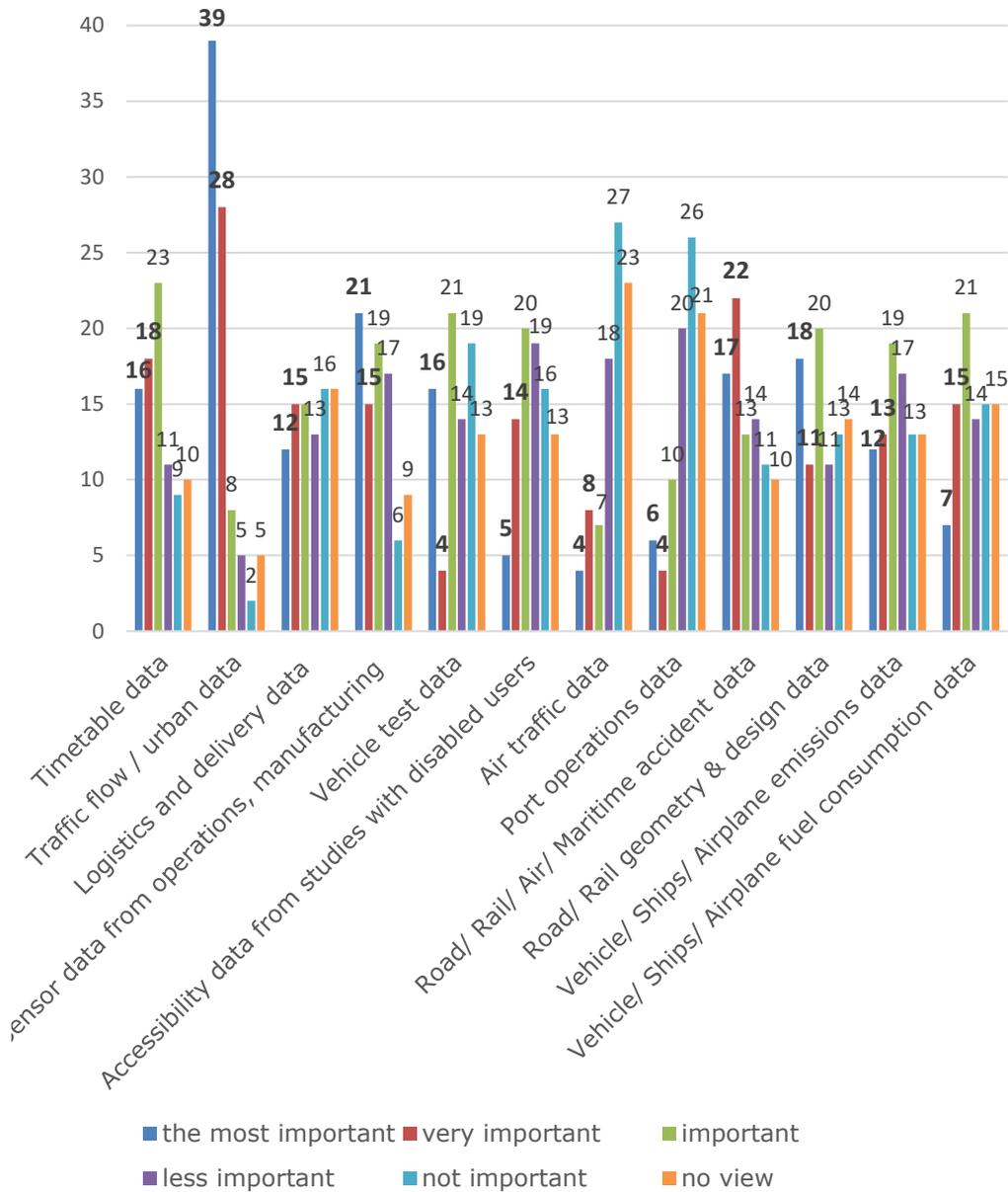
Q3: If open data on transport operations were to be available, what kind of data would be the most important for your work? Please prioritize the data types.

Figure AIV-3. The most important transport data

Respondents were also asked if some data should not be accessible via open data model. The vast majority of participants admitted that personal or privacy sensitive data should not be accessible via open access model. On the other hand

opinions such as : "making more data accessible will help researchers/companies improve peoples' lives" were also noticed. A consensus between most respondents was reached that data should be available primarily for research purposes; however some survey data might be more sensitive (due to data privacy issues) "so maybe in such cases restrictions should exist and only aggregated form should be accessible". At the same time, 95% of respondents expect data to be described (supported with documentation and metadata information) if it is made openly available.

In terms of the standards followed when publishing metadata the majority of respondents reply that they do not follow any standards. Part of them do not publish metadata or they provide the meaning of each field and name, a glossary of definitions, and a description of data collection process, etc. Participants who follow standards use: EIP recommendation according the national access point issue (delegated acts 885, 962, 886, 1926) or M3 (standard for exchanging structured information between nodes).

In terms of expectations to transport metadata and documentation, respondents would like to get information concerning descriptive metadata to aid discovery (16%), study design and methodological information (15%), and data dictionaries explaining abbreviations, lab notes etc. (14%). More information on metadata and documentation expectations are shown on Figure AIV-4.
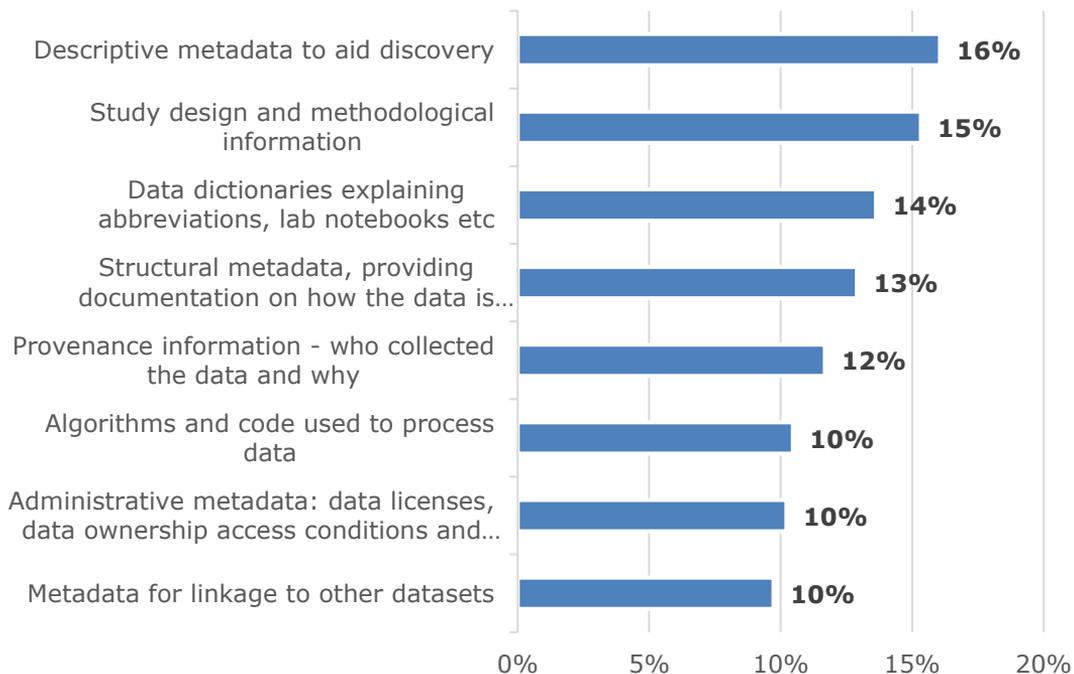
Q4: What metadata and documentation would you expect to see with transport data?

Figure AIV-4. Metadata and documentation expectations

Part II. Cloud service requirements and expectations

In the second part of the survey respondents' opinions on requirements and expectations to a TRC service potentially offered by European Commission were examined. First of all participants were asked if they would use this kind of service. Almost half of the answers (48%) were "yes, definitely". However, not much less (46%) stated "perhaps – depends on service". In general, researchers expressed their great interest in accessing data in open access model. Mainly in terms of possibilities to use it in the field of education and conducted researches. They underlined lack of access to wider range of different types of transport data. Below are selected opinions on access to TRC solution:

- "I use transport data in different fields - education and research. In education, I teach modelling and simulation course and GIS course. I teach at university with tradition in transport oriented education. In research, I work on decision support tools for transportation."

- "Open data sets can significantly escalate the development of new technologies, search for interdependence and multidisciplinary. There is an opportunity to create community development teams that will work together to find ways to optimize the transport system. It is also possible to bring new developers from other science disciplines."

- "Big data processing requires service not available on the local base."

- "Would that come at a cost? Furthermore, generally platforms developed at a European level are poor in terms of security, user friendliness, accessibility and the requirements to access them."

- "Quality, accuracy and integrity of data are most important, else unusable!"

- "A central system/repository will certainly be very useful, as a reference place for all data we need but also for all data we want to publish and promote."

- "Availability of large amount of high quality data will create significant potential for research of higher quality and research in other transport sectors, different from the ones I am already familiar with (and have access to data)."

- "Lack of data is a strong limitation to research in my field, especially as regards emerging issues in safety, human factors and automation in transport."

- "Data is the infrastructure of the future. The EU should invest in this much more, in all domains. It will be beneficial for society, just like classical physical infrastructure."

- "I am hoping for getting access to the data that is currently difficult to get an access to by small research units or to the data that takes very long to obtain due to bureaucracy."

- "I´m dealing with cross country comparisons and I expect that TRC can provide a common framework for comparable data. At the moment, even

available data are mostly distracted, in some part - only in national languages and with significant differences in terms of their coverage, aggregation and available variables."

- "Interpretability, interoperability and long-term reliability."

In terms of functionality of TRC, respondents would expect to have access to existing transport data collections (18%), advanced search options to filter and fine relevant content (17%) and open data sharing (11%). Additional expected functionality is presented in Figure AIV-5.
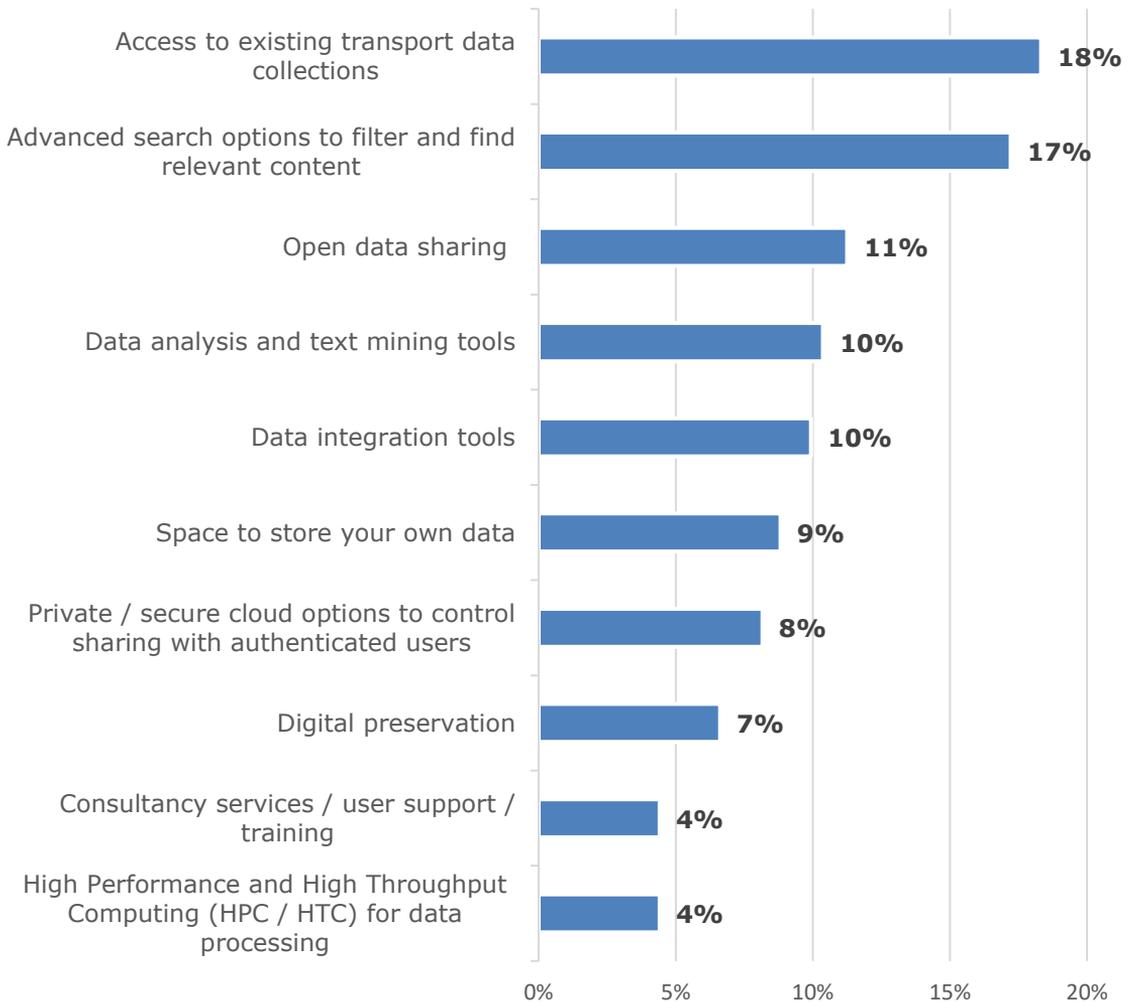
Q5: What functionality would you expect from a TRC?

Figure AIV-5. Functionality expected from TRC

82

According to respondents, the data base should have search functions and criteria for filters i.e. by transport mode (road, rail, air, maritime, inland waterways) – 19% of respondents marked this option, by country (18%), by transport sectors (passengers, freight) – 16%, by vehicle type – 14%, by technology (ITS, IoT, MaaS, CAVs, …) – 11%, by industry – 11% and by application (asset management, construction, design & planning, navigation, journey planning, etc.) – 9%.

Survey participants were asked if they currently have access to all the cloud services they desire. 48 answers were negative and 30 – positive. Besides respondents underlined that "*availability of cloud services does not seem to be an issue, but fragmentation and different levels of functionalities yes. A single access point with standard functionalities meeting the needs of the community would be very useful*" and "*companies such as Amazon and Google provide adequate computing resources, but access to useful data is lacking.*"

The vast majority of respondents reported using Dropbox (32%) or Google Drive (28%) as cloud service. Other services mentioned were: One Drive (16%), SharePoint (9%), and Amazon Web Services (6%).

In terms of open data services in the field of transport research (e.g., analysis tools used, data collections, national initiatives) awareness, respondents mentioned several, e.g.: PRACT repository (http://www.pract-repository.eu/); DaCoTA; European Road Safety Knowledge System (http://safetyknowsys.swov.nl/Home/about.html), - ERSO - The European Road Safety Observatory (http://ec.europa.eu/transport/road_safety/specialist/index_en.htm); NRSO - National Technical University of Athens Road Safety Observatory (www.nrso.ntua.gr); "Otwarte dane po warszawsku" (open data from Warsaw); Eurostat; UN-ECE; OECD/ITF; Quite a few data form national data files, and some data from tee Industry. A lot of data from EU projects (Horizons, etc.); Clearingstelle from dlr; Data distribution interface (National traffic information centre of Cz); National accident data datasets; Global network data (spatial data); datasets for the public sector licences, CIS (central informational system of timetables in Public transport); - Google Maps/Traffic, Waze, OpenStreetMap; GUS (General Statistics Office in Poland); http://transportnetworks.cs.aalto.fi/; Iowa DOT Open Data Portal (http://data.iowadot.gov/); KITTI; NAPs, specific research groups' data repositories, opendata.imet.gr; Open data portals at various levels of government, global urban data repository (GUDR) in progress; Ordnance Survey Open Data, data.gov.uk, Network Rail open data, TfL open data; PNMV and Ecologiegistique in France; Quantum GIS – AequilibraE; TT Transforming Transport data portal; UK government data.gov.uk website; www.mobilitaetsdaten.gv.at.

In terms of expectations to provide/access data to/in a cloud via specific data formats, participants were asked to indicate which standards they expect to be used. Majority of them pointed out DATEX II, then the following formats were indicated: NeTEx, GTFS, INSPIRE, BS7666, open data formats (i.e. the specifications are openly available and can be easily be re-implemented), VICINITY and FIware.

Part III. Opportunities and barriers

In the third part of the survey participants were asked to underline benefits and barriers in using open access data. The most important benefit from TRC would be "greater availability of data". This answer was pointed out by 81 respondents (22% of replies). Next "data sharing across national borders" was important by 71 respondents. The third answer was "advance research in the transport field" – 68 participants marked this answer. Figure AIV-6 presents more detailed answers to the question on envisaged benefits from TRC.
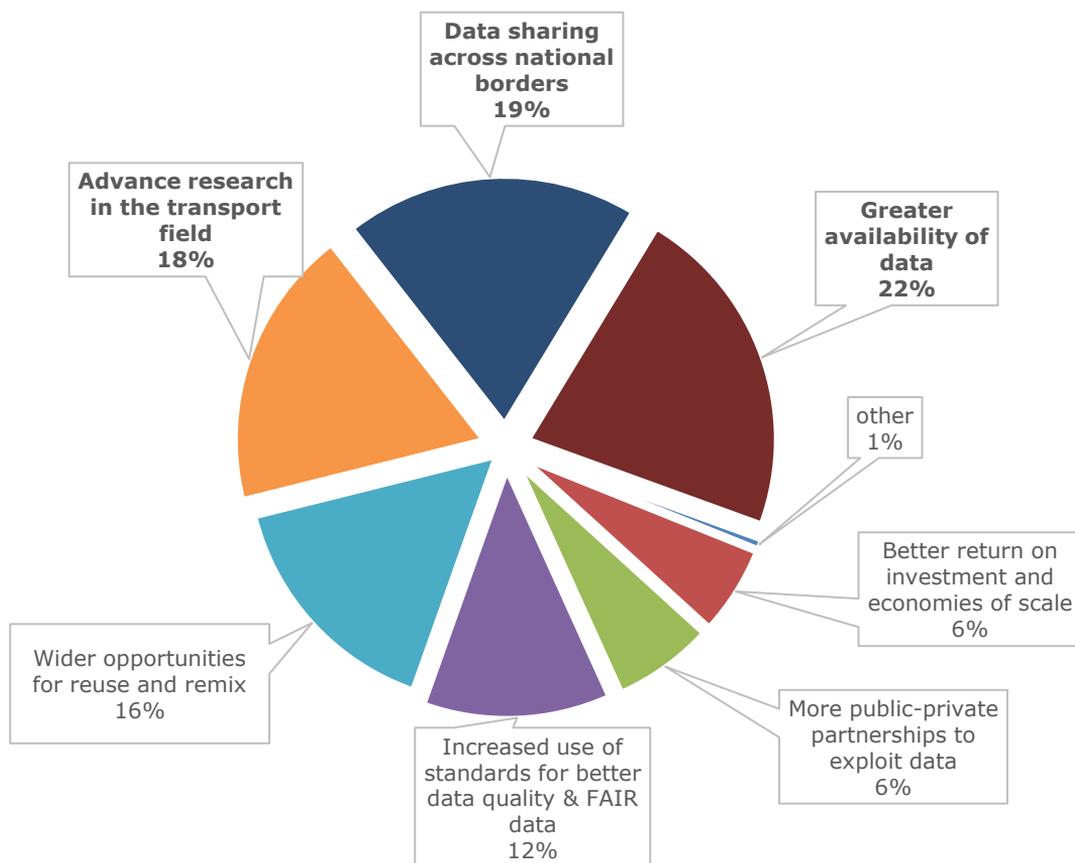
Q6: What benefits do you envisage from a TRC?

Figure AIV-6. Benefits envisaged from TRC

The respondents were also asked if a European Transport Research Cloud should connect with international initiatives e.g., TRANSPOREON for intelligent transport logistics, Transportation Mobility Cloud (TMC) for Smart Cities by Ford, etc. The vast majority (40%) agreed, that probably it would be a good idea. 33% strongly agreed with it, and 22% was unsure. Uncertainty resulted mainly from lack of

familiarity with such initiatives. Others underlined that collaboration and synergy will assist in the advance of knowledge and integration of data can provide added value to the community. Connected initiatives concerning transport activities "should be open and led by the EU. Private companies can join if they accept the conditions of the EU. EU should copy the model of classical infrastructure."

As the three biggest barriers that can be foreseen in using open data services, respondents pointed out: (1) "data ownership / IPR issues", (2) "cost and founding barriers" and (3) "concern about opening data out to competitors." All the barriers and their evaluation are presented in Figure AIV-7.
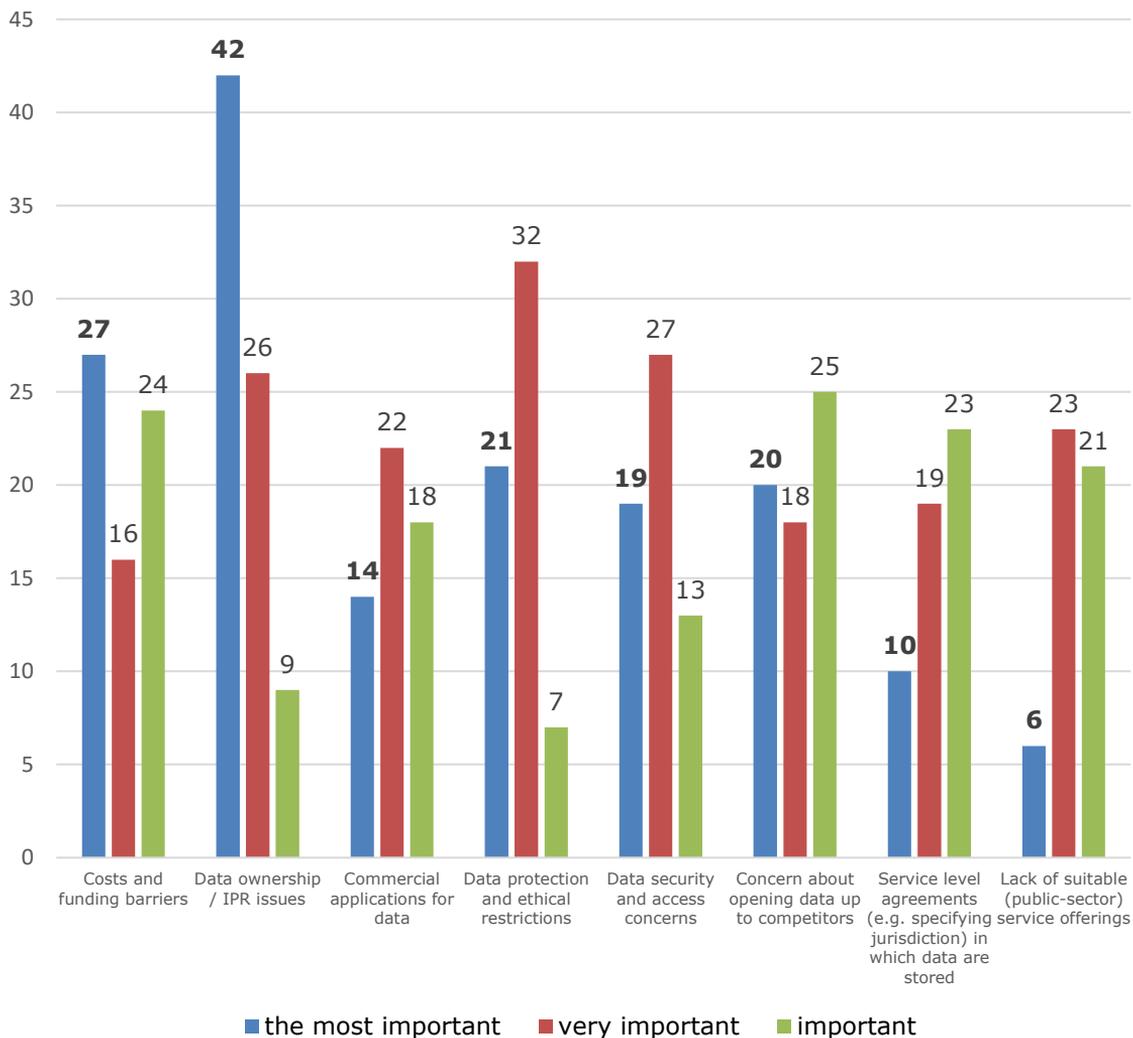
Q7: What are the three biggest barriers that you see in using open data services?

Figure AIV-7. Barriers in using open data services

Part IV. Funding mechanism

In the last part of the survey participants were requested to evaluate potential funding mechanisms for the TRC. Currently 76% of respondents does not pay for storing data in an open data service and 21% does so. In terms of charging models that would allow to use open data storage services most effectively, respondents marked in the first place "free at the point of use", next "institutional subscriptions", and then "grant founding". Charging models preferred by respondents are shown in Figure AIV-8.
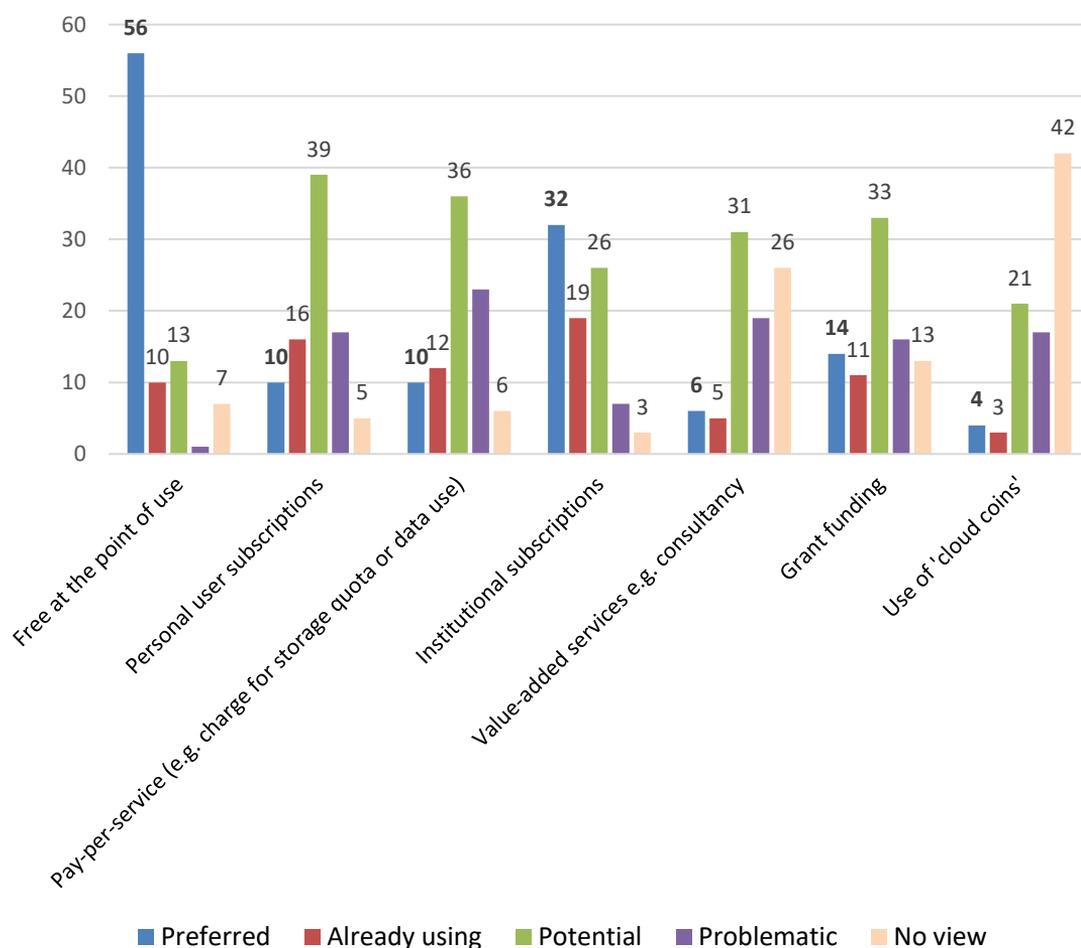
Q8: What charging models would allow you to use open data storage services most effectively?

Figure AIV-8. Charging models that allow to use open data storage services

As justification for the chosen option in the previous question the following answers were given:

- "Data should be free, just like road space is provided for free. Data are much cheaper to provide than roads, yet we do not want to provide them for free. This does not make any sense."

- "Free when a data interchange between producers is made (i.e., mutual benefit)"

- "By contribution. If you ever made any contribution, you can use other data sources for free."

- "I would prefer free at the point of use, but this involves significant up front funding, and continued central funding to ensure data remains accurate and up-to-date."

ATTACHMENT V. MINIMUM METADATA ELEMENTS AS DESCRIBED BY THE TRINATIONAL COOPERATION BETWEEN GERMANY, THE NETHERLANDS AND AUSTRIA

	Name of Metadata element	Mandatory for Nation	Field name (proposal)	Type of value	Field length (proposal)	Technical description	Example
Metadata information	Metadata Date	True	metadata_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NOT NULL	2015-10-23T09:00:00+01:00
	Metadata language	True	md_language	Predefined Text	-	Predefined EU24 Language set ISO 639-2 conform; multiple choice; NOT NULL	ger; eng;
	Contact point for metadata	False	cp_name	Free text	50	Text; utf8; NULL	Hans Maier
		True	cp_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		False	cp_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		True	cp_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		False	cp_website	Free text	50	Text; utf8; NULL	http://data.at
		False	cp_tel	Free text	50	Text; utf8; NULL	-
Content Information	Name of dataset	True	d_name	Free text	250	Text; utf8; NOT NULL	Highway network Austria
	Description of dataset	True	d_description	Free text	1000	Text; utf8; NOT NULL	Contains static high priority network of Austria, Link information: Speed, lanes, direction
	Dataset type category	True	data_agr_type	Predefined Text	-	Predefined 15 data categories; Lookup Table; multiple choice; NOT NULL	
	Dataset detailed type	True for self-declaration	data_org_type	Predefined Text	-	Predefined 50 data types;	

						Lookup Table; multiple choice; NULL	
	Dataset language	True	ds_language	Predefined Text	-	Predefined EU24 Language set; single choice; NOT NULL	ger;
Temporal information	Start date of publication	True	p_start_date	DateTime	-	YYYY-MM-DD'T'hh:mm:sTZD; NOT NULL	2015-10-23T09:00:00+01:00
	End date of publication	False	p_end_date	DateTime	-	YYYY-MM-DD'T'hh:mm:sTZD; NULL	2015-10-23T09:00:00+01:00
Geographical coverage	Area covered by publication	True	val_area	Predefined Text	-	Predefined NUTS 0 – 3 Codes; UTF8; multiple choice; NOT NULL	AUT11; AUT12;AUT13;
	Network coverage	True	net_category	Predefined Text	-	Predefined; UTF8; multiple choice; NOT NULL	Motorway
	Network coverage description	False	net_description	Free text	1000	Text; utf8; NULL	structural separated bidirectional lanes, 2 to 4 lanes, minimum speed 80, use condition
Responsibilities/ Contact Information	Publisher	False	p_name	Free text	50	Text; utf8; NULL	Hans Maier
		True	p_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		False	p_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		True	p_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		False	p_website	Free text	50	Text; utf8; NULL	http://data.at
		False	p_tel	Free text	50	Text; utf8; NULL	-
	Data owner	False	do_name	Free text	50	Text; utf8; NULL	Hans Maier
		True	do_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		False	do_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		True	do_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		False	do_website	Free text	50	Text; utf8; NULL	http://data.at
		False	do_tel	Free text	50	Text; utf8; NULL	-

Condition for use	Contract or licence	True	con_lic	Predefined Text	-	Predefined; UTF8; single choice; NOT NULL	Licence
	Condition for use	True if con_lic is used	con_description	Free text	1000	Text; utf8; NULL	http://data.at/terms.pdf
Access information	Structure of publication	True	tech_structure	Predefined Text	-	Predefined; single choice; utf8; NOT NULL	Datex II XML
	Publication structure description	True if con_lic is used	p_structure_description	Free text	1000	Text; utf8; NULL	http://data.at/tech_description.pdf
	Access interface	True	interface	Predefined Text	-	Predefined; single choice; utf8; NOT NULL	SOAP;
	Communication method	True	com_method	Predefined Text	-	Predefined; Multiple choice; utf8; NOT NULL	push;
	Access URL	True	access_url	Free text	250	Text; utf8; NOT NULL	http://data.at/access.csv
Quality information	Update frequency	True	update_freq	Predefined Text	-	Predefined; Single choice; utf8; NOT NULL	yearly
	Quality Indicator	True	qm_indicator	Free text	1000	Text; UTF8;NOT NULL	According to the EIP+ quality measures
	National Body assessment date	False	assessment_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NULL	2015-10-23T09:00:00+01:00

This report focuses on the requirements for data sharing within the transport research community. In particular, the report examines the potential of a Transport Research Cloud (TRC) as a subset of the European Union's European Open Science Cloud (EOSC) initiative. Six domain experts collected data based on their personal experiences, contacts, prior research and a survey sent out to other researchers in the transport domain to enable a preliminary analysis concerning the needs, barriers and potential benefits for the domain should a TRC be realized. From this work ten recommendations, grouped into five broad topic areas, have been developed that the Experts believe must be addressed if a sustainable TRC is to be realized.

Studies and reports

Publications Office