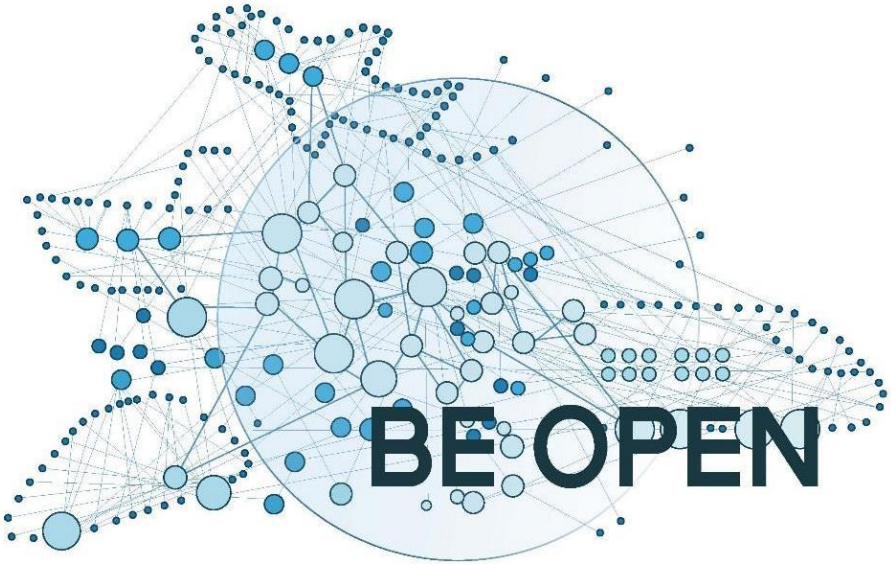

European forum and oBsErvatory for OPEN science in transport

Project Acronym:	**BE OPEN**
Project Title:	**European forum and oBsErvatory for OPEN science in transport**
Project Number:	**824323**
Topic:	**MG-4-2-2018 – Building Open Science platforms in transport research**
Type of Action:	**Coordination and support action (CSA)**

D2.3 Transport Research in the European Open Science Cloud

Final Version

Deliverable Title:	Open/FAIR data, software and infrastructure in European transport research
Work Package:	WP2
Due Date:	2019.11.30
Submission Date:	2020.09.29
Start Date of Project:	1st January 2019
Duration of Project:	30 months
Organisation Responsible of Deliverable:	ATHENA RC
Version:	Final Version
Status:	Final
Author name(s):	Natalia Manola, Afroditi Anagnostopoulou, Harry Dimitropoulos, Alessia Bardi
Reviewer(s):	Kristel Palts (DLR), Christian von Bühler (Osborn-Clarke), Anna Walek, M. Zuraska (GUT), Clara García (Scipedia), Anja Fleten Nielsen (TOI), Lucie Mendoza (HUMANIST), Michela Floretto (FIT), Ioannis Ergas (WEGEMT), Boris Hilia (UITP), Caroline Almeras (ECTRI), Rudolf Cholava (CDV), Milos Milenkovic (FTTE)
Nature:	☒ R – Report ☐ P – Prototype ☐ D – Demonstrator ☐ O – Other
Dissemination level:	☒ PU - Public ☐ CO - Confidential, only for members of the consortium (including the Commission) ☐ RE - Restricted to a group specified by the consortium (including the Commission Services)

Document history			
Version	**Date**	**Modified by (author/partner)**	**Comments**
0.1	2019.11.25	Afroditi Anagnostopoulou, Alessia Bardi, Harry Dimopoulos	EOSC sections according to EC documents. Relations to OpenAIRE.
0.2	2020.07.20	Natalia Manola	Rewrote sections to reflect new developments in EOSC governance
1.0	2020.09.20	Natalia Manola, Harry Dimitropoulos, Alessia Bardi	Integrated feedback from internal review process. Condensed Appendix describing relations to OpenAIRE

Contents

Abbreviations and Terminology

Acronym	Terminology
TRC	Transport Research Cloud
NOADs	National Open Access Desk
TDM	Text and Data Mining
UDF	User Defined Function
AAI	Authentication and Authorization Infrastructure
PID	Persistent Identifiers
MS/AC	Member States and Associate Countries
RoP	Rules of Participation
IPR	Intellectual Property Rights
DMP	Data Management Plan
SaaS	Software-as-a-Service

1 Executive summary

The objective of the BE OPEN project is to see how open science can be applied and up taken in the research transport domain: which are the different facets (open access to publications and data, FAIR data and research data management, upskilling), what are the barriers and challenges for its implementation and uptake (infrastructure, legal, cultural), who are the actors involved (public or commercial organizations, citizens), and where would potential interventions have the most impact.

Background: The transport sector plays a strategic and major role in the European single market and in a fast changing economic and societal frame. Policies, technologies and behaviours must be continually adapted to new constraints, such as climate change, the diminishing supply of fossil energy, the economic crisis, the increased demand for mobility, safety and security (with additional crucial elements in the COVID-19 pandemic) etc. Transport infrastructure, vehicles and vessels, modal share, co-modality, urban planning, energy and environment issues are the subject of extensive studies, research works and industrial innovations, conducted by universities, research institutes, companies, practitioners, and public authorities.

Moreover, with the rapid progress of the development of **intelligent transport systems (ITS)** over the last 15 years, the need for testing them in the real world and collecting data about their impact has become more and more important, and their sharing and re-use more complex. For example, operational data from government agencies and logistics data from industry is combined with user generated data and research data/metadata from neighboring scientific disciplines (e.g., energy, environment, social, health) making a rather intricate cross-sectoral ecosystem which needs explicit rules with consensus from all, and a shared infrastructure to increase the impact of data/AI driven research.

The context: The European Open Science Cloud developments are very timely as the research transport domain has already recognized a demanding need for sharing of data, publications, tools, services and knowledge, as a means for achieving data intensive science towards the emerging evolutionary leap in artificial intelligence. Already, the DG-RTD 2018 HLEG report on the **"Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud"**[1] has identified the need for a **Transport Research Cloud (TRC),** which is envisioned as a subset of the EOSC, complementing it, rather than constituting a different entity. Moreover, TRC is expected to build bridges to the European **Mobility Data Space**, one of the nine proposed data spaces in the new *European Data Strategy*[2], both on infrastructure and on the policy fronts, defining the rules and operations for sharing and re-using data and resources such as tools and services.

For TRC to become the open science hub for transport research it should address the following[1]:

1. **Reusable research data:** Identify what constitutes transport research data and investigate the barriers for the limited use of data collected by others.
2. **Data as a public good:** Properly define what *public data* is, e.g., data collected under contracts paid for by tax payer funds.

[1] DG-RTD HLEG 2018 "Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud" https://trimis.ec.europa.eu/content/analysis-state-art-barriers-needs-and-opportunities-setting-transport-research-cloud

[2] EC Communication COMM (2020) 66, A European Strategy for Data, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

3. **Standards:** Develop and use common standards among members of the transport research community in different sectors, i.e., governmental entities, infrastructure operators, and commercial consumers.
4. **Infrastructure:** Identify operating requirements for a TRC, including user communities, to design and develop infrastructure, ensuring an appropriate level of service provided at a cost that is understood by all stakeholders. Develop a proper governance model for a sustainable TRC.
5. **Incentives, training, and education:** Promote open science rewards and incentives to be adopted in the transport research discipline. Develop training/education programs for existing and future researchers and supporting human capital (e.g., librarians, data stewards, data curators).

Interventions: As EOSC is setting up its structures and prioritizes the areas of action, the transport research community recognizes the value of the emerging developments in EOSC and has identified gaps and interventions along the following six action lines which are required to fulfil the TRC vision:

Architecture: The cross-sectoral nature of transport research (research-industry-government-public) and the envisioned link to the the EC Mobility Data Space, as well as the industrial cloud platform landscape, is characterized by a plurality of architectural patterns, ranging from a high level of centralization (e.g. data lakes) to concepts promoting utmost decentralization (e.g. distributed applications using blockchain technology), or emerging *edge* computing techniques. The proposed EOSC architecture should indicate roles, processes and brokering mechanisms to manage trusted exchange of data among sectors.

Data: *As closed as necessary* is key in sharing data and EOSC needs to develop and implement safeguard measures and mechanisms to protect sensitive data. The development of a machine-readable/actionable metadata framework, which also embeds information on legal aspects, should be in the focus early on to ensure a trusted and accountable data space among the different transport sectors.

Services: User Environments provide the means to engage all actors in transport and foster a trusted open science ecosystem. It is essential for EOSC to develop a framework for community-oriented authorization policies, accompanied by privacy services embedded in data exchange and processing pipelines.

Access & Interface: A TRC dedicated portal, a marketplace for services, tools and data should be the point of reference for the transport research community and neighboring sectors. Based on the *EOSC-Core* standards and mechanisms, the TRC portal must include additional, community-focused information and have direct links to both the *EC Mobility Data Space* and the EOSC portal to actively engage the broad transport community.

Rules: Rules of Participation (RoP) currently present high-level principles which have no reference to any specifics on legal tool-kits necessary to reach the principles. To be operational and effective for transport research, EOSC RoP need to quickly move beyond the high-level view and consider elements of e.g., what constitutes *public data* and embed rules provided by the directive on open data and the re-use of Public Sector Information (PSI)[3]. Addressing the big volume of transport data generated and collected by industry, it is essential to consider the possibility to include this in early stages, defining clear rules for cooperation.

Governance: The transport research community should participate in the EOSC governance and commit to EOSC success via the TRC. Realizing the complexity of the ecosystem (i.e., different modals represented by

[3] EU Public Sector Information framework https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

different organizations, data from different sectors, different types of agencies and organizations involved) the community may engage in EOSC via multiple, non-exclusive, routes:

- Via one or more umbrella organizations, e.g., ECTRI, FEHRL, EARPA, and other academic networks like EURNEX, HUMANIST, FERSI, representing their member organizations in the EOSC Association. This umbrella organization must carry out dedicated activities to: bring requirements of transport research into EOSC; help shape emerging standards, roles, interactions, rules, architecture; align its members' policies, practices and infrastructure; develop and introduce an open science portfolio aligned with EOSC to be adopted across its members as part of TRC and solicit funding for its implementation; actively liaise with EU e-Infrastructure community as pan-European service providers addressing different layers of EOSC and especially providers of *EOSC-Core* and enablers of *EOSC-Exchange*.
- Build synergies at the national level to EOSC mandated organizations and existing or emerging national e-Infrastructures which will be linked to EOSC by design. The aim is to contemplate specific requirements from transport research local needs for policies, regulation, and infrastructure and business models in a concrete manner and provide the corresponding resources.

2 Introduction

The transport sector plays a strategic and major role in the European single market and in a fast changing economic and societal frame. Policies, technologies and behaviours must be continually adapted to new constraints, such as climate change, the diminishing supply of fossil energy, the economic crisis, the increased demand for mobility, safety and security (with additional crucial elements in the COVID-19 pandemic) etc. Transport infrastructure, vehicles and vessels, modal share, co-modality, urban planning, energy and environment issues are the subject of extensive studies, research works and industrial innovations, conducted by universities, research institutes, companies, practitioners, and public authorities.

Moreover, with the rapid progress of the development of **intelligent transport systems (ITS)** over the last 15 years, the need for testing them in the real world and collecting data about their impact has become more and more important, and their sharing and re-use more complex. For example, operational data from government agencies and logistics data from industry is combined with user generated data and research data/metadata from neighboring scientific disciplines (e.g., energy, environment, social, health) making a rather intricate cross-sectoral ecosystem which needs explicit rules with consensus from all, and a shared infrastructure to increase the impact of data/AI driven research.

This deliverable aims to position the transport research domain in relation to EOSC, related to the technical and social developments, as well as the relation to the governance as this emerges in the EOSC Partnership and EOSC Association.

In Section 3 we are showcasing the key characteristics of transport research as these have been portrayed in a series of recent Strategic Research and Innovation Agendas for Transport (STRIAs), singling out barriers, challenges and recommendations relating to data intensive research and operations, how these affect interactions among different actors in the transport cross-sector community, and proposed practices for the the way researchers and other actors manage, process and share data and knowledge.

In Section 5 we present the main objectives, principles and current implementation of EOSC, highlighting aspects to draw points of intervention and recommendations for the transport research community.

In Section 6 we discuss the **Transport Research Cloud**, first introduced in the 2018 EC High Level Expert Group report "*Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud*", and provide a list of gaps, recommendations and actions required for its implementation within and complementing EOSC.

Finally, in section 7 we present a pilot use case (detailed in *Appendix 2* – Detailed methodology for showcasing the importance of building synergies and aligning efforts with pan-European e-Infrastuctures, through the work carried out in BE OPEN jointly with OpenAIRE in building the TOPOS observatory for open science.

3 Transport research in a data driven era

3.1 Transport research characteristics

Multiple forms of knowledge are required in the planning, design, monitoring and evaluation of transport infrastructure and services, as it has multiple aspects and involves multiple actors, discourses and professional practices. Grasping the full complexity of urban conditions and potentialities is beyond the capacity of any single category of research or professionals. Thus, co-production of knowledge by multiple stakeholders, researchers, professionals rooted in different disciplines, government agencies and local citizens is one of the crucial processes for developing innovative solutions and reaching sustainability goals[4].

Key characteristics of transport research which affect interactions and the way researchers manage, process and share data are:

1. **Multi-modal:** Different modes of transportation and mobility with research topics of interest: urban logistics, intermodal transport systems, transitions towards sustainable transport, accessibility and equity analysis, technologies for improving network and vehicle efficiency, advances in integrated transport systems and intermodal transportation, traffic safety analysis, traffic flow theory and modelling, IT technologies for transport data collection and analysis, strategies for vehicle-to-vehicle communications and the transport impacts and indirect impacts of autonomous vehicles.
2. **Cross-sector:** Research data is generated, shared and consumed among research institutions, government agencies and industry. In particular, most of the data is gathered in real life situations from sensors and from individuals (public) in different settings, requiring better understanding and efficient (simple) policies to enable data crosswalks. Open and FAIR (gov) data is key into the transport research endeavor. The LEMO H2020 project (**Le**veraging Big Data to **M**anage Transport **O**perations - https://lemo-h2020.eu) groups transport related datasets into the following themes[5]: Place & Space, Environment, People, Things & Movement, Disruption and event-related data, Public Transport Services, Personal Automobility, Cargo connections, International Connections, and Consumption & transaction data. In many cases these exist in silos (e.g. automated cycle count data, Urban Traffic Management and Control traffic flow data) or are not open/available (historic passenger ticketing data). Most commonly the datasets are localised, rather than consolidated at a national level.
3. **Scattered resources:** Transport research is carried out by key national leading organizations[6], but due to its overarching and inter-disciplinary nature, research is being diluted to other research organizations. Computing/cloud resources are decentralized mostly around national service providers, sub-communities do not use dedicated or certified repositories[7], data are usually stored

[4] Löfgren S., Designing with differences, cross-disciplinary collaboration in transport infrastructure planning and design, Transportation Research Interdisciplinary Perspectives, Volume 4, March 2020, 100106, https://doi.org/10.1016/j.trip.2020.100106

[5] Lemo project (H2020): Leveraging big data for managing transport operations, Deliverable 2.2. Report on legal issues, Oct 2018, doi:10.5281/zenodo.1476424

[6] ECTRI members are European national research leading organizations https://www.ectri.org/about-ectri/members/

[7] re3data.org repositories related to transport https://www.re3data.org/search?query=transport

in different data silos of different ownerships and formats causing difficulties in cataloguing, finding, accessing and using research (meta)/data (e.g., EUROSTAT, EU open data portal, OECD iLibrary etc.), and no specific policies around open access to data and publications are imposed from these institutions. In addition, many Research Infrastructure projects or landmarks[8] cover topics in transport research (e.g., ECCSEL for energy, ICOS for CO2 emission, LifeWatch for environment) but none seem to be dedicated to transport/mobility data.

4. **Big data, AI-driven research:** Transport often faces acute challenges due to poor infrastructure, growing populations, urbanization, and in some regions rising prosperity, which increases vehicle traffic, cargo volumes, and pollution. Artificial intelligence offers new solutions to these challenges by allowing countries to reach underserved populations, creating markets and private sector investment opportunities associated with them. Emerging market start-ups and mature business in the transportation space are already starting to digitize their businesses or build new tech-enabled business models altogether (such as e-logistics and e-mobility)9. Datasets that will drive research and assist in the rise of intelligent integrated transport, on which AI applications can be further developed are: map data; weather; personal location data; network disruptions; planned events; real-time network capacity for people, vehicles & goods; public transport schedules; vehicle location data; fare and pricing data; sentiment data from service users and non-users; third party service usage data; and payment/transaction data, together with metadata in an appropriate unified standard that can be read and interpreted by AI. These data building blocks, generated from different sources (public, private, individual), constitute the seed for AI driven research and call for immediate actions that will enable open, transparent and straightforward access mechanisms.
5. **Lack of skills and education:** Demand for AI and data experts has grown over the last few years in Europe where AI investment is increasing. A lack of skilled AI/data engineering talent has been widely cited as the largest barrier to AI adoption in developed countries, with critical shortage even greater in emerging economies. It takes time for a country to effectively incorporate new technologies, particularly complex ones with economy wide impacts such as AI. This means it takes time to build a large enough capital stock to have an aggregate effect and for the complementary investments needed to take full advantage of AI investments, including access to skilled people[9]. These roles include all people involved in the process of data creation and management, not only scientists and IT, but also librarians, data stewards and university administration employees responsible for conducting research projects from the administrative side.
6. **Security and privacy:** Intelligent Transport System technologies must ensure the integrity, confidentiality and secure handling of data, including personal and financial details, and show that citizens' rights are fully protected. It is therefore important that policies, regulations and infrastructure (e.g. blockchain, anonymization) embed privacy-by-design and that all generated (big) data is shared and re-used for research purposes.
7. **Data ownership and governance:** In almost all cases technical challenges to making data available are secondary to data owners'[10] attitudes, costs of establishing and maintaining sensor networks, in-

[8] ESFRRI Roadmap 2018, http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf

[9] Maria Lopez Conde and Ian Twinn: How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets, EMCompass, Note 75, Nov 2019, www.ifc.org/thoughtleadership

[10] In this context, "data ownership" is not meant as a technically legal term describing an absolute legal right.

house skills needed to support data sharing, and data privacy concerns.[11] Transportation data governance needs to improve the security, quality and management of data assets for programmatic and stakeholder needs and to set the qualifications of the various actors involved as (joint-) controllers or processors. Few industries and governments require such strict compliance and collaborative effort across numerous functional areas where data is collected, stored and analysed. Such practices need to be strengthened and extended to research and government data to facilitate transparency in transport multi-stakeholder endeavor.

8. **Overarching initiatives:** Realizing the importance of transport and mobility, the EC has two major initiatives in place: a. JRC's Transport Research and Innovation Monitoring and Information System (TRIMIS) which gathers information on EU and national policies, projects, infrastructures, and b. the proposed *EC Mobility Common Data Space* as part of Europe's Strategy for data[12]. Both are key in establishing necessary links to all stakeholders, in addressing interactions-policies-regulatory frameworks with government agencies, industry and citizens, entirely complementary to EOSC.

3.2 Challenges and priorities related to data in transport research

TRIMIS has classified strategic areas of research in transport as follows:

1. Connected and Automated transport (CAT)
2. Electrification
3. Vehicle design and manufacturing
4. Alternative fuels
5. Network and Traffic Management Systems
6. Smart mobility and services
7. Infrastructure

It has issued over the years STRIA Roadmaps for each of these areas[13] towards a more integrated and effective transport system across Europe and a better use of innovation and new technology in transport.

The following section highlights the areas of interest for EOSC, i.e., requirements for storage, usage and sharing of valuable data, as well as new tools and models for their analysis. It is obvious in all that data sharing is needed for the designing, monitoring and evaluation phases of the research life cycle, and that data sources are from labs and monitoring facilities, but gradually increasing volumes from real-life sensors (grounded, air/water-borne, on vehicles, on people).

1. Connected and Automated transport (CAT)

Connected and automated transport (CAT) technologies and services improve traffic flows, optimise the use of infrastructure, lower noise levels, shift greater volumes of passenger traffic towards public transport, increase the efficiency of goods transport and foster the emergence of multi-modal transport solutions.

[11] Lemo project (H2020): Leveraging big data for managing transport operations, Deliverable 3.2, June 2019 Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector

[12] A European Strategy for Data, 2020, https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

[13] https://trimis.ec.europa.eu/roadmaps

Innovation towards CAT poses a multitude of challenges which include the development of technologies at hard- and software, vehicle, infrastructure, data communication and decision-making levels, as well as the validation of these technologies and the real-world testing in individual mobility, goods and passenger transportation also including the regulatory aspect.

Priorities/challenges regarding data and EOSC[14]:

Data collection and representation: Data is collected and stored in different formats and quality, and for different purposes and thus data collection is often not well synchronized nor organized. Metadata is incomplete or inconsistent with standards. Privacy, liability, IPR, security and ethics issues need to be addressed and solved before sharing and analysing data.

Big data, Artificial Intelligence and their applications: The sensors of connected and automated vehicles produce huge amounts of data. In addition, similar if not larger amounts of information are gathered from road infrastructure sensors, mobile phones, from meteorological instruments and other systems related to transport. To accelerate the development and assessment of CAD (Connected and Automated Driving) technologies and to support the optimization of road usage, the availability, quality assurance, interoperability, and the exploitation of this big traffic data must be ensured. Moreover, models for data sharing and making data accessible are needed. Also, new and innovative business services based on big data that respect the security, privacy and the highest ethical standards are generated. Therefore, 'new' AI concepts and technologies are to be developed to fulfil the challenges and responsibilities of CAD functions taking into account innovative infrastructures and new vehicle function. Operational safety via collaboration on the ways to develop, train and validate the AI and on the availability of testing sites.

Skills and training: Foster workforce and digital skills development related to CAD.

2. Electrification

The development of energy storage technologies and devices remains the cornerstone of a fully electrified transport system integrated in a clean energy network. Decreasing batteries costs while increasing their energy density and lifetime will speed up electrification of road transport. The deployment of a network of recharging points covering the whole EU road network is another key enabling condition for transport electrification.

When developing new technologies for electric vehicles, the influence of sustainability parameters must be taken into account, such as (a) primary energy savings, (b) reduction of GHG emissions, (c) reduction of noxious emissions, (d) range and speed and (e) cost of technology and constraints on raw materials. Measures of the certification of CO2 emissions and fuel consumption of vehicles and for monitoring and reporting of the certified data are well under way. Past research has been restricted by the lack of appropriate data, especially the lack of cross- country comparison of inputs, outcomes and efficiency. An effort to share data Lifetime Cycle Assessment (LCA) data on a European level has been achieved through the Joint Research Centre within the Life Cycle Data Network (LCDN)[15].

[14] Excerpts from Road, Rail and Waterborne SRIA Roadmap, 2019 https://trimis.ec.europa.eu/sites/default/files/roadmaps/stria_roadmap_2019-connected_and_automated_transport.pdf

[15] https://eplca.jrc.ec.europa.eu/LCDN/

Priorities/challenges regarding data and EOSC[16]:

Lack of data sharing standards: The most effective way for LCA researchers investigating electrification of road vehicles to improve data quality will be to use and supply data within existing networks. More importantly data gathering should focus on making production data more specific and detailed on component level instead of aggregated as now, which would require shared metadata specifications and crosswalk tools for semantic mappings. This should also include geographic information on the emissions of vehicle with a GIS based LCA.

3. Vehicle design and manufacturing

Transport vehicle design, development and manufacturing (VDM) is a collaborative, integrated and complex set of processes and tools that consider the whole vehicle life cycle and is a key element for the competitiveness of the EU transport industry. Evolutionary path of car design and manufacturing is already disrupted by the high degree of electrification, digitalization and automation. Maritime and air-transport demonstrated electric and hybrid technologies at small scale, preparing the ground for real transport vehicles in the years to come.

Transport vehicles are complex products made of millions of components. They interact between each other and the infrastructure, and they should comply with the rapidly evolving regulatory framework.

Priorities/challenges regarding data and EOSC[17]:

Embed digitalisation, big data and cybersecurity in the design and manufacturing action of next generation of transport vehicles. The average automotive manufacture uses dozens of IT systems along value chain. It creates need for research with regard to data compatibility, interchangeability and security (especially regarding digital footprints). There is necessity for research which focuses on standardization (including data / metadata standardization), as it is a key issue to allow such modular approach. This also poses the exigency for IT solutions that allow safe and efficient exchange of data between different members of the value chain.

Explore Big Data analysis, Artificial Intelligence and other methods towards linking evolutionary design and operations. Collaborative, integrated, multidisciplinary, scientific-based and validated design processes and tools, that make use of high performance computing, will allow efficient design, optimization and manufacturing of transport vehicles. Such advanced computational design tools and processes should seamlessly validate models with physical data from new and updated testing facilities. The collaborative processes should evolve from data-based to model-based exchanges.

4. Alternative fuels

The development and deployment of transport fuels such as methane, liquid petroleum gas, compressed natural gas, liquid nitrogen gas, synthetic paraffinic fuels, alcohols and ethers and esters is key to the

[16] Excerpts from Road, Rail and Waterborne SRIA Roadmap, 2016 https://trimis.ec.europa.eu/sites/default/files/stria_roadmap_-_transport_electrification_0.pdf and egvi.eu 2018 roadmap https://egvi.eu/wp-content/uploads/2018/01/electrification_roadmap_web.pdf

[17] Excerpts from 2016 roadmap https://trimis.ec.europa.eu/sites/default/files/stria_roadmap_-_vehicle_design_and_manufacturing.pdf

decarbonisation of the European transport sector and attaining a 60% reduction in greenhouse gas emissions by 2050.

The European Alternative Fuel Observatory (EAFO – https://www.eafo.eu) from the Directorate General for Mobility and Transport gathers all data (aggregate) and relevant information regarding alternative transport fuels in Europe and publishes data periodically on the European Open Data Portal[18].

Priorities/challenges regarding data and EOSC:

Data sharing principles and guidelines: Research and innovation focusses on the efficient use of advanced biofuels, fossil fuels blended with renewable fuels as well as pure renewable fuels, their match to engine characteristics for specific transport modes and development of multi-fuel combining applications. It is obvious that much of this research falls in the innovation sphere and it is not foreseen that all data will be openly available. But, as with all other research areas, this domain needs to develop and deploy guidelines for sharing and preserving research (Meta)/data, especially as risk assessment procedures are key for the uptake of any proposed solution.

5. Network and Traffic Management Systems

Network and Traffic Management (NTM) systems are used for the optimisation and management of transport networks' operation. Bottlenecks across air, rail, road and water can result in system-wide capacity constraints, traffic jams and increased pollutant emissions and environmental impacts.

Data is a major challenge for transport and network planners, including big data collection / fusion / management, floating vehicle data, data collection via social media, etc. Increasing real-time information availability can create a seamless connection, although it does not necessarily mean a clearer picture of the current traffic state. The information has to be verified, filtered, elaborated and communicated via customised interfaces to the user.

Priorities/challenges regarding data and EOSC[19]:

High quality data: Underpinning any strategic NTM system will be the provision of high quality and universally accepted data. Generated data must be shared among all interested stakeholders to facilitate the creation of an NTM system, which is appropriate for the year 2050. This will mean a change in practice for some key players, who will have to be encouraged to give others access to their data, in a manner which currently does not happen. Protocols relating to other key issues such as data privacy will also need to be introduced, to ensure data can be used productively by all relevant parties, without prejudicing the rights of the individual.

Collected data will have to allow for countries, areas and modal analytics: The availability of huge amounts of data regarding the interconnection of infrastructure, vehicles, people, goods, management and operations will offer new opportunities and challenges to better optimise traffic flows. However, it is a key issue to define and implement harmonised and standardised concepts for (cross-border) exchange of data between authorities, operators and/or users, to enable an easy and reliable synchronisation and to interconnect the national systems with (Pan-) European systems.

[18] https://data.europa.eu/euodp/en/data/dataset/eafo

[19] STRIA Roadmap 2016 https://trimis.ec.europa.eu/sites/default/files/stria_roadmap_-_network_and_traffic_management_systems_0.pdf

6. Smart mobility and services

Changes in transport behaviour and lifestyles such as the use of smart phones, mobile web applications and social media together with the trend to use rather than own a particular means of transport has opened up new pathways to sustainable mobility.

Private companies, governments and public entities should be equally encouraged to provide and share user and urban data collected on the use of public space and infrastructures wherever it is available (in such a way that protects the privacy of its citizens) so that users, governments, cities, third party apps, operators, developers and innovators can access it to inform their decisions and innovate their applications. Smart mobility and cities will combine publicly and privately developed infrastructures.

Priorities/challenges regarding data and EOSC[20]:

Equitable data sharing: Establish ubiquitous and fair-access digital public infrastructure to ensure socially inclusive digital services, to enable equitable data-sharing and evidence-based public policy and to facilitate collaborative service innovation in private and public transport offerings.

Data privacy: The flow of big and open data, while of central importance for smart mobility services and systems, will, however, require significant governance and regulatory design to ensure the interests of all stakeholders and their access to available data are equally protected. Individual data privacy rights and the ownership of mobility and city data will need to be addressed and regulated to ensure both competition and freedom from illegal governmental or commercial surveillance[21].

Large-scale public data pool with comparative formats: Provide independent, rigorous, empirical and real-time data on the environmental, social, economic and spatial performance of European mobility and transport systems to researchers and policy makers. Developing a sound database on the real-world performance of different mobility systems and services will be vital for advancing effective public policy on mobility, urban development, technology adoption, health and social equity in transport and for designing and interrogating relevant indicators. While a wealth of data is collected, collated and made available at the local and regional level, this is often insufficiently rigorous or not of adequately comparative format to allow for aggregation of transport behaviour, health, environmental, emissions or spatial effects across Europe. Creating such robust and comparative pan-European datasets will be vital in providing evidence-based information and decision support for policy and political decision-making processes and the general public.

7. Infrastructure

Transport infrastructure includes physical networks, terminals and intermodal nodes, information systems and refuelling and electrical supply networks which are necessary for the safe, secure operation of road, rail, civil aviation, inland waterways and shipping.

Advanced technologies are revolutionising the business landscape in European transport and mobility. The area needs to embrace this change and further engage in its digital transformation based on data,

[20] STRIA Roadmap 2019 https://trimis.ec.europa.eu/sites/default/files/roadmaps/stria_roadmap_2019_smart-mobility-systems-and-services.pdf

[21] Smart mobility and services HLEg`https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=34596&no=1

connectivity and automation. Moreover, optimised methods based on the use of big data, and artificial intelligence need to better take into consideration users' expectations and travel trends in planning procedures.

Priorities/challenges regarding data and EOSC[22]:

Mobility framework: Improve and create standards for transport infrastructures and data aggregation in a new mobility framework. Use artificial intelligence (including data collection and exchange protocols and standards) to better plan and deliver transport infrastructures and manage assets. Promote a platform for closer cooperation and exchange of information and data between the project owners, infrastructure operator, funding bodies, etc.

Develop skills linking infrastructures and mobility innovations: Promote ways to attract skilled professionals from the information and communications technologies to satisfy future needs for mobility, ensuring lifelong learning and training. Link education and research in transport (including ICT and data sharing) in order to prepare a resource pool of future experts.

3.3 The new European Data Strategy

The new **European Data Strategy**[23] addresses digital transformation with data-driven innovation at the centre, expecting to bring enormous benefits for citizens, for example through improved personalised medicine, new mobility and through its contribution to the European Green Deal.

The aim is to create a **single European data space** – a single market for data, open to data from across the world – where personal as well as non-personal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data, boosting growth and creating value, while minimising the human carbon and environmental footprint. This data space will accommodate B2B and B2G transactions, and will ensure that access is provided to researchers:

- *"…data generated by the public sector as well as the value created should be available for the common good by ensuring, including through preferential access, that **these data are used by researcher**s, other public institutions, SMEs or start-ups."*
- *"…sensitive data in public databases is often not made available for **research purpose**s, in the absence of capacity or mechanisms that allow specific research actions to be taken in a manner compliant with personal data protection rules."*

Building on the ongoing experience with the research community within the European Open Science Cloud, EC plans to support the establishment of nine (9) common European data spaces, including a Mobility Data Space among them:

[22] STRIA Roadmap 2019 https://trimis.ec.europa.eu/sites/default/files/roadmaps/stria_roadmap_2019_-_infrastructure.pdf

[23] EC Communication COMM (2020) 66, A European Strategy for Data, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

A Common European mobility data space to position Europe at the forefront of the development of an intelligent transport system, including connected cars as well as other modes of transport. Such data space will facilitate access, pooling and sharing of data from existing and future transport and mobility databases.

Key policy priorities listed below affect open science and how researchers will interact with the generated data:

- Open up EU approval legislation for motor vehicles (currently focused on wireless data sharing for repair and maintenance) to more car data based services, looking at data accessibility, full compliance with data protection rules and the role and rights of car owners.
- Review all necessary EU Directives to further contribute to data availability, reuse and interoperability accross countries.
- Include new provisions on data availability and market access of data service providers in order to promote the digitalisation and automation of air traffic management.
- Review the regulatory framework for interoperable data-sharing in rail transport.
- Establish common data sets in Maritime transport[24] and electronic freight transport information regulations[25] to facilitate digital exchange and data reuse between businesses and administration.

3.4 Open science in transport research

Although in the last decade, the number of open access articles in transport has been increased, data sharing practices are yet limited. The cause for the limited sharing practices lies in the fact that some researchers consider data as their property. Even if they find value in sharing unpublished research data, however, their possible lack of training on data sharing, management and reuse lead them to share or dispose very limited data, or share it in inappropriate ways without common standards. To this direction, training programs constitute effective common open data practices while universities and institutions via their libraries, data centers and emerging competence centers play a key role in the promotion of data sharing to the research community. Additionally, rewarding practices need to be applied for researchers sharing data while institutions need to implement and support open data policies.

The transport research community will aim to provide researchers in the transport and logistics sector with access to open data that cover topics of their research interest. The transport sector originates high volume of big data which require specific handling such as adequate storage space and powerful hardware. Several EU-funded projects have already studied and dealt with the use of big data in transportation. Such projects are listed in Table 1, highlighting the relevance to EOSC for data, computing, service and knowledge sharing.

In an effort to introduce intelligent, data intensive practices, research is driven towards building platforms by leveraging open data sources and open science practices: they establish repositories to share data, they build semantic layers to interlink it, and share resources such as computing, storage, services, and workflows to co-create expertise and knowhow.

In parallel, national and European policies (such as the Intelligent Transport System, the increased Open Data policies, the Automated Driving, and Smart Mobility) have been introduced with an aim to protect the privacy

[24] Regulation (EU) 2019/1239.
[25] The negotiations with the co-legislators are concluded, adoption is foreseen mid-2020.

of citizens, to promote data sharing between private and public sector and to support the digitalization of the transport sector. For instance, EU General Data Protection Regulation (GDPR) applying to personal data reshapes the way data is managed, though transparency and openness, along with confidentiality and privacy protection.

Table 1. Representative EU Big data transport research projects and their relation to open science.

EU project	Description	Open science / EOSC relevance
AutoMat https://automat-project.eu	Create an open ecosystem and marketplace for vehicle Big Data	Definition of standardized and open interfaces for unconstrained data access.
LeMO https://lemo-h2020.eu	Develop a strategy that defines the necessary research efforts for the realisation of the big data economy in the transport	Legal issues around data ownership, IPR, privacy. Use case of open data.
SafeClouds.eu **https://safeclouds.eu**	Provide input in new approaches for data mining in aviation safety. **AI** information-driven analysis of hazard identification in aviation.	Scalable, secure, on-demand **multi-side computing and data storage** platform. Provides infra for data cycle approach, improving **data management skills**.
TransformingTransport https://transformingtransport.eu	Show concrete, measurable and verifiable evidence of data value that can be achieved in mobility and logistics by leveraging big data	Building an **open data portal** which systematically catalogues data sets from industry using **open software** (CKAN) and **standard metadata** schemata.
Track and Know https://trackandknowproject.eu	Big Data for Mobility Tracking	Efficient, interoperable and scalable **toolboxes** for big data software stacks. Collection and aggregation of real-time data.
BigData Ocean http://www.bigdataocean.eu	Exploiting Oceans of data for Maritime Apps. Propose and validate maritime big data scenarios for the benefit of EU-based companies, organisations and scientists.	Largest **repository** to date containing marine data from different sources (aggregation and access). **Linked** data facilities. Extraction and **sharing of data models**.
AEGIS https://www.aegis-bigdata.eu	Create advanced Big Data Value Chain for Public Safety and Personal. Bring technologies to create a curated, semantically enhanced, interlinked and multilingual repository for "Public Safety and Personal Security" related big data.	Leverages open data sources. Framework for semantically enriching and **interlinking** data. **Sharing via blockchain** technologies.
QROWD http://qrowd-project.eu	Offer innovative solutions to improve mobility and reduce traffic congestion. Create platform to integrate human and crowd intelligence via tailor made service pipelines.	Integrate different sources of data (geographic, transport, meteorological, cross-domain and news data). Platform to **collect data, link data sources, validate algorithms, analyze data**. Use **data protection** by design and **reward data contributions fairly**.

EU project	Description	Open science / EOSC relevance
NOESIS https://noesis-project.eu	Provide a novel decision support tool for evaluating strategic big data investments in transport and intelligent mobility services.	Responsible **Code of Conduct for big data management** in transport.

In brief, open science practices are becoming all the more relevant in the transport and mobility domains and are used throughout, but not in a coordinated approach. Siloed approaches persist, specifically adhering to the different modes of transport. Furthermore, since the main data contributors are industry, citizens and public authorities, the open science terminology used in EOSC (e.g. FAIR) and its connotations may differ, which places an extra barrier in shared understanding and adoption of common mechanisms.

3.5 Main Actors

The evolution of transport requires moving away from traditional public/private sector roles, and more specifically from government, industry, mobility providers, academia, public-citizens roles and involvement. Each sector or area has a key role to play in design, development and implementation of transport systems, but the boundary between the ways in which they interact with one another will become less clearly defined over time, with each sector becoming involved with issues such as data ownership and governance, management and revenue raising.

Based on the *D1.1 – Taxonomy of Actors, terminology and experimental tools*[26] of the BE OPEN project, a classification of key actors of all transport modes in the transport domain (Figure 1) has been delivered by assessing activity of current European technology platforms in an attempt to cover the industry needs, and main influential organizations in Europe could provide an overall assessment of research scientific production against innovation uptakes. Moreover, national initiatives and actions could also foster an evidence-based discussion and cross-fertilization of ideas amongst countries acting as practical, community driven sharing of knowledge and experiences.

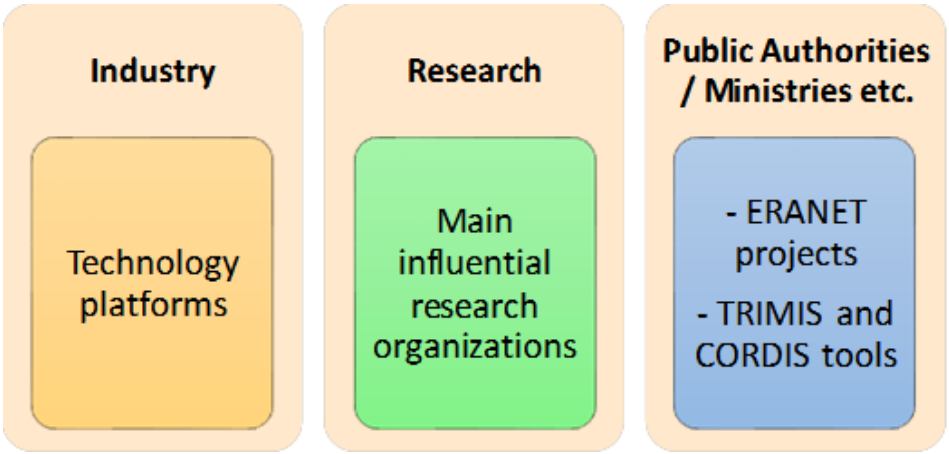

Figure 1. BE OPEN main actors' categories

In other words, a) **Technology Platforms**, b) **Influential research organizations** and c) **National initiatives and actions** could set the basis of the systematic taxonomy of key actors in order to provide a unified

[26] https://beopen-project.eu/storage/files/beopen-d11-taxonomy-of-actors-terminology-and-experimental-tools.pdf

transport terminology and related research and investigation instruments (BE OPEN, D1.1). In an attempt to consider the main areas of interest and research topics of public bodies at national, regional and local level, the ERA-NET programme was analysed. The ERA-NET scheme and TRIMIS tools aim to develop and consolidate coordination of national and regional research programmes. As such, Table 2 summarises the main research trends regarding public sector at a national and/or regional level (BE OPEN, D1.1).

Table 2. Main research trends of public authorities

Communities and Cities
Industrial Competitiveness in Transport
Clean Transport and Mobility
Smart Mobility
Energy Storage

4 The main European Technology Platforms are presented in Table 3, which also provides information about their sector of interest and their mission. Error! Reference source not found. illustrates members of the ECTRI association which brings 29 research organizations active in transport, from 20 different European countries and Table 4 in A pilot case study: OpenAIRE services for TRC

The OpenAIRE-BE OPEN liaison aims to promote and support the uptake of Open Science principles among researchers and help building and empowering the transport research community. OpenAIRE with BE OPEN partners is developing an **Open Research Gateway for Transport Research**. The gateway will offer a single entry point to discover research entities (all types of results, projects, organizations, and content providers) relevant to the domain of Transport Research. Within the gateway, researchers may find Open Science publishing tools to:

(i) Deposit research products of any type (i.e. not only scientific literature, but also datasets, software, workflows, methods, etc.) and get a persistent identifier via Zenodo.

(ii) Add details of the research context of a research product by creating links to projects and to other related research products (e.g. a link between a dataset and the software that analysed it).

(iii) Add research products to the community gateway. Via such tools, researchers populate and access an open, participatory scholarly communication graph of interlinked objects through which they can share any kind of products in their community, maximise re-use and reproducibility of science, and outreach the scholarly communication at large.

(iv) View statistics and analysis that support the monitoring of the uptake of Open Science.

The gateway will be further developed and enhanced to become the **TOPOS observatory** including organization related information. A preliminary version of the gateway is available at https://beta.beopen.openaire.eu, with work in progress for cleaning up and validating the underlying data. Once the BE OPEN gateway for Transport Research reaches a good and validated (curated) coverage and upon agreement of the BEOPEN consortium, the gateway will be registered in the EOSC Catalogue & Marketplace.

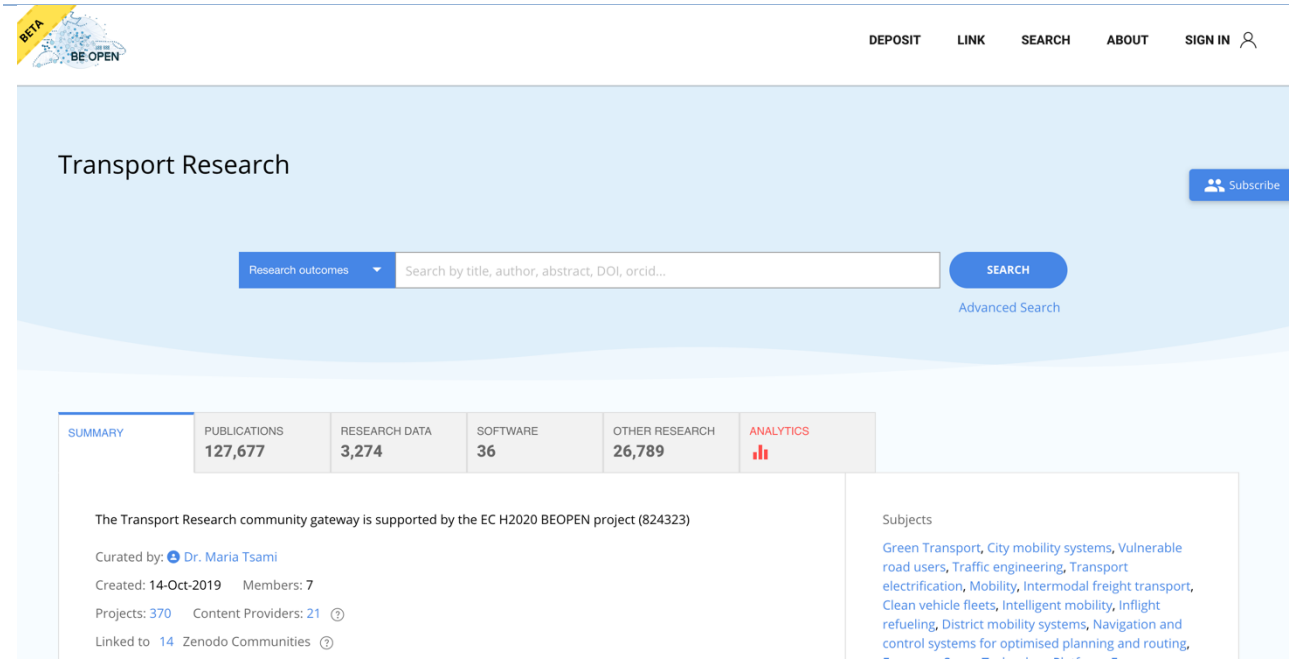

Figure 11. BE OPEN Gateway to transport open science results

Methodology: In order to proceed with the initial set-up of the gateway, the OpenAIRE team analyzed reports and BE OPEN deliverables for input to configure the gateway based on the OpenAIRE Research Graph :

- List of relevant projects – link to Cordis and TRIMIS.
- List of relevant content providers – authoritative list from repository and OA journal registries from re3data.org, OpenDOAR, DOAJ.
- List of relevant research organizations – currently identified by the consortium, with envisioned future registration process.
- Glossary for transport statistics used in text mining for classification purposes.

Detailed information on the methodology and tools is provided in Appendix 2 – Detailed methodology for BE OPEN Gateways.

Appendix describes the main influential research organizations in Europe, both research centres and associations.

Table 3. European Technology Platforms

Name	Sector	Description
ERTRAC https://www.ertrac.org	Road	ERTRAC is the European Road Transport Research Advisory Council. It is the European technology platform which brings together road transport stakeholders to develop a common vision for road transport research in Europe.

Name	Sector	Description
ERRAC *https://errac.org*	Rail	Within ERRAC, all major rail stakeholders are gathered including: railway undertakings, infrastructure managers, and manufacturers, the European Commission, the European Union Agency for Railways, EU Member States, academics and users' groups. This enables a broad coverage of all forms of rail transport: from conventional, high speed and freight applications to urban and regional services ERRAC's primary objective is to deliver the common railway sector vision of R&me to the European Institutions and important stakeholders, in order to contribute shaping a favourable funding landscape for railway research and innovation, developing innovation via calls for projects and joint undertaking like Shift2Rail. As a European Technology Platform for rail research and innovation, ERRAC is comprised of a wide range of European rail sector stakeholder organisations and their members. Their involvement in ERRAC is entirely voluntary.
WATERBORNE TP *https://www.waterborne.eu*	Mari-time	WATERBORNE has been set up as an industry-oriented Technology Platform to establish a continuous dialogue between all waterborne stakeholders, such as classification societies, shipbuilders, ship-owners, maritime equipment manufacturers, infrastructure and service providers, universities or research institutes, and with the EU Institutions, including Member States.
ACARE *https://www.acare4europe.org*	Air	ACARE is the Advisory Council for Aviation Research and innovation in Europe and provides a network for strategic research in aeronautics and air transport so that aviation satisfies the needs of society and secures global leadership for Europe in this important sector. ACARE is essential in bringing together the right stakeholders to turn the air transport vision in Europe into reality.
ALICE *https://www.etp-logistics.eu*	Multi-modal	The European Technology Platform ALICE is set-up to develop a comprehensive strategy for research, innovation and market deployment of logistics and supply chain management innovation in Europe. The platform will support and assist and advise the European Commission into the implementation of the EU Program for research: Horizon 2020 in the area of Logistics. ALICE was created in the frame of WINN project having the European Green Cars Initiative (logistics section) and EIRAC, European Intermodal Research Advisory Council, as background and supporting initiatives. ALICE was officially recognized as a European Technology Platform by the European Commission in July 2013. ALICE is based on the recognition of the need for an overarching view on logistics and supply chain planning and control, in which shippers and logistics service providers closely collaborate to reach efficient logistics and supply chain operations.

Name	Sector	Description
ECTRI *www.ectri.org*	Multi-modal	ECTRI, the European Conference of Transport Research Institutes, is an international non-profit organisation registered in Belgium that was founded in April 2003. ECTRI as the leading European research association for sustainable and multimodal mobility is committed to provide the scientifically based competence, knowledge and advice to move towards a green, safe, efficient, and inclusive transport for people and goods. It does so by: 1. Promoting transport research and enhancing its scientific quality and effectiveness, 2. Providing independent, evidence-based advice to decision makers in Europe and 3. Incorporating and representing the foremost European transport research Institutes and Universities. Its members are currently 27 major transport research institutes or universities from 19 European countries. Together, they account for more than 4000 European scientific and research staff in the field of transport.
HUMANIST *http://www.humanist-vce.eu*	Multi-modal	Research centres, universities and SMEs active in Human Machine Interface across transport modes, 22 members from 15 countries
FEHRL *http://www.fehrl.org*	Road	National research and technical institutes from across Europe, 31 members associates from non-European countries provide FEHRL with strong links to the considerable research capacity available globally (i.e. NETIVEI from Israel, CSIR from South Africa, TFHRC FHWA from USA, ARRB from Australia)
FERSI *https://fersi.org/int*	Road	The Forum of European Road Safety Research Institutes (FERSI) is a non-profit organisation that forms a flexible network of European road safety research organisations. All FERSI member institutes have a governmental mandate to perform pre-normative road safety research and are the main discussion partner of their national government for road safety issues. They analyse road safety developments, prepare solutions to be translated in legislation and guidelines, and evaluate the implementation of solutions.
EARPA *https://www.earpa.eu/earpa/home*	Road	The European Automotive Research Partners Association (EARPA) is the association of automotive R&D organisations. It brings together independent R&D providers in the automotive sector throughout Europe, ranging from large and small commercial organisations to national institutes and universities. EARPA is actively contributing to EU RTD funding programmes.
EATEO *http://eateo.eu*	Air	Common forum for European aviation training and aviation education providers, 25 members from EU and US countries
WEGEMT *http://www.wegemt.com*	Mari-time	European Association of Universities in Marine Technology and related sciences, 40 Universities in 17 countries
EURNEX *http://www.eurnex.org*	Rail	Leading rail researchers from scientific institutions and universities all over Europe. 44 members from 20 countries
EPoSS	Multi-modal	EPoSS, the European Technology Platform on Smart Systems Integration, is an industry-driven policy initiative, defining R&D and innovation needs as well as policy requirements related to Smart Systems Integration and integrated Micro- and Nanosystems. A group of major industrial companies and research organisations from more than 20 European Member States intend to co-ordinate their activities in the

Name	Sector	Description
		field of Smart Systems Integration. A main objective is to develop a vision and to set up a Strategic Research Agenda.
ECTP	Road building	The European Construction, built environment and energy efficient building Technology Platform (ECTP) is a leading membership organisation promoting and influencing the future of the Built Environment. First founded in 2004 ECTP brings together the collective vision for a leading edge European Built Environment on behalf of its Members. ECTP gathers around 140 Member organisations from across the construction sector and other sectors from the whole supply chain of the Built Environment. The membership across 26 countries, large enterprises, SME's, universities, research organisations and professional associations allows ECTP to take an integrated approach to tackling all relevant issues and connect people and organisations from across the supply chain, helping us work collectively to improve our position on many societal and industrial issues including energy, climate change, efficiency and infrastructure.

5 European Open Science Cloud description

5.1 EOSC description

The European Open Science Cloud (EOSC) aims to give Europe a global lead in scientific data infrastructures, to ensure that European scientists reap the full benefits of data-driven science. Practically, it will offer to the European research community a free and accessible virtual environment with free at the point of use services for storage, management, analysis and re-use of research data, across borders and scientific disciplines, following the FAIR data principles[27]. EOSC development is driven by scientific community needs and will offer additional open source services for education and training for digital skills and, over time, to government and business users as the technologies developed will be promoted for wider application[28].

5.2 Principles and rules of EOSC

After more than three (3) years of deliberation with all relevant stakeholders, EOSC is being shaped up as a *Research Commons* with primary focus on sharing, via FAIR and open principles, data and related research (e.g., publications and software), as well as services and tools that are used to generate, process, analyze, publish and access them.

The OpenAIRE vision paper[29] describes *EOSC as a **co-created space,** in which research information is produced and transformed into knowledge. A "Research Commons" that is collaboratively developed and managed by Europe's research communities. Interconnecting both digital and human resources, EOSC is expected to add new value to existing infrastructures, via **frictionless data flow**, **intelligent discovery** and **retrieval of scientific output**, as well as **homogenized and secure access to services** that transfer, store and analyse data.*

Its envisioned key characteristics are:

- *Openness by default:* A trusted infrastructure with Open Science practices for transparency, reproducibility and accountability embedded at every phase of the research life cycle.
- *A network of digital resources:* A System of Systems, focussing on *business-to-business* sharing and access. A fit-for-purpose, lightweight *regulatory* framework, to ensure that different players have the best possible conditions to develop on their own, with freedom of contract as a cornerstone.
- *Researcher-centric:* An abstraction layer for data manipulation, including reporting, workflows, parallelism, and persistence, enabling a reproducible 'digital laboratory' as a standard way of working, allowing them to communicate with peers and non-peers alike.

[27] Communication from the EC (April 2016): "European Cloud Initiative - Building a competitive data and knowledge economy in Europe" https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555074889405&uri=CELEX:52016DC0178
[28] D2.2: Draft Governance Framework For the European Open Science Cloud, 2017. EOSC pilot project.https://eoscpilot.eu/content/d22-draft-governance-framework-european-open-science-cloud
[29] Manola N., Achieving Open Science in the European Open Science Cloud: Setting out OpenAIRE's vision and contribution to EOSC, 10.5281/zenodo.3610132

- *Open governance:* A shared arena, maintained and preserved by a range of players in collaboration, governed openly at national level (via national EOSC structures), guided by the EC.
- *Training tomorrow's data experts:* A coordinated, professional and certified training infrastructure, embedded in institutions and research groups, as part of the European Digital Skills agenda.

The **EOSC Partnership**[30], one of the 49 proposed European Partnerships under Horizon Europe, is currently being formed, aiming to accelerate the deployment and consolidation of EOSC. It will bring together institutional, national and European initiatives and engage all relevant stakeholders to jointly develop a *science commons* where data are Findable, Accessible, Interoperable, Reusable (FAIR) and where research-enabling and other services are made available throughout Europe. This European contribution to a *"Web of FAIR Data and Related Services for Science"* will enhance the possibilities for researchers to find, share and reuse publications, data, and software leading to new insights and innovations, higher research productivity and improved reproducibility in science. Objectives of the EOSC partnership are:

General objective 1: Open science practices and skills are rewarded and taught, becoming the "new normal"

- *Main milestone: The EOSC ecosystem underpins the reward of open science practices and data stewardship that improve trust, quality and productivity in science.*

General objective 2: Standards, tools and services allow researchers to find, access and reuse results

- *Main milestone: The EOSC provides a trusted platform supporting the development of innovative services and products.*

General objective 3: Sustainable and federated infrastructures enable open sharing of scientific results

- *Main milestone: The EOSC infrastructure is in operation, providing a web of FAIR data and related services underpinning research addressing major societal challenges.*

Crosscutting objective: Boosting the impact of EOSC through collaboration and alliances

5.3 EOSC implementation

Envisioned to be developed in gradual steps, EOSC will be implemented and operated in two conceptual layers[31]: the ***EOSC-Core,*** which encompasses all key elements and components that enable the sharing, accessing, supporting and monitoring functions of EOSC, and the ***EOSC-Exchange***, which allows research communities, industry, public agencies and others to connect via EOSC-Core functions and share their data and services.

Figure 2 illustrates the EOSC layers with research data (FAIR and as open as possible) in the center. It presents the transition to Phase 2, where it is expected that EOSC includes additional functionality and services dedicated to the requirements of end users from the private sector so that they can exploit the open data and associated services for commercial gain without distorting market competition.

[30] EOSC partnership proposal https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf

[31] Solutions for a Sustainable EOSC, A tinman report from the EOSC Sustainability Working Group, Draft 2 December 2019, https://www.eoscsecretariat.eu/system/files/solutions_for_a_sustainable_eosc_-_tinman_draft_02dec19.pdf

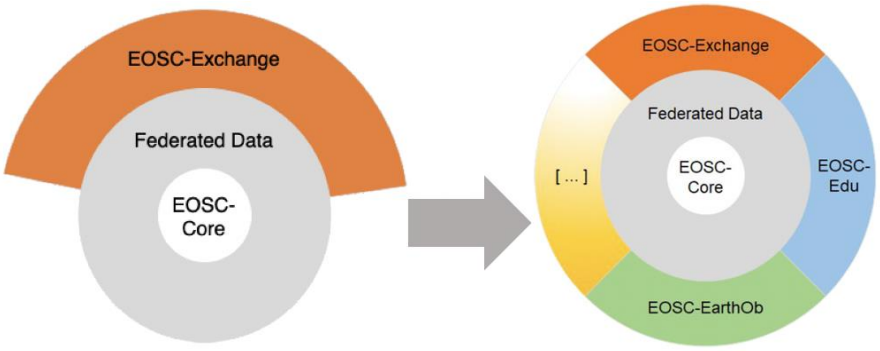

Figure 2. EOSC layers and phases

Box 1. EOSC layers in a nutshell

EOSC-Core components

A shared open science policy framework, which effectively embeds a data compliance framework for open / FAIR data. It dictates and applies the rules of how the data elements are published, shared and re-used.

Authentication and Authorization Interoperability (AAI) framework, a trust and identity service for researchers to seamlessly access all EOSC resources.

Data access framework, whose primary role is to offer data as a service. It enables open interfaces where data consumers are able to discover and use data.

Service management and access framework, whose role is to provide a consistent and agreed upon understanding of e-science services: what they offer, which science problem they address, what is their operational capacity, how they are accessed, who pays for them.

A minimum legal metadata framework, for ensuring openness and interoperability, privacy and security (copyright status, disclosure limitations, patents pending, other IPR on the datasets or workflows, the existence of personal data, designation of data as PSI, etc.)

An open metrics framework, which sets the rules (usage, performance, value for money) for the assessment of EOSC elements, i.e., policies, access framework, services, data, business, funding and usage models. This should include elements to facilitate the incentives and awards mechanism for researchers, as recommended by the EC HLEG on Next Generation Metrics and the EOSC Pilot policy group.

PID services to generate, resolve and validate persistent identifiers (PID).

Helpdesk framework for linking national/thematic/institutional service desks that can provide training/consultancy on EOSC-Core services.

Portal providing one form of accessing and using the EOSC Resources.

EOSC-Exchange

EOSC-Exchange builds on the EOSC-Core to ensure a rich set of services exploiting FAIR data and encouraging its reuse are available to publicly funded researchers. It is expected that rivalrous services, such as those that store, preserve or transport research data as well as those than compute against it, will be made available via EOSC-Exchange. Participation in EOSC-Exchange as a service provider is voluntary and without registration fee. Service providers that participate in EOSC-Exchange will be required to conform to predefined Rules of Participation. While the technical requirements for participation in EOSC-Exchange will be the same for all services, there may be differences in the legal and policy requirements for freely available and payment-based services.

5.3.1 EOSC implementation details

The current governance of EOSC (Figure 3) has put forward actions towards its first implementation phase. The EOSC Executive Board, through the operation of six (6) working groups (EOSC WGs)[32], has gathered a broad community of experts from the national and European infrastructure community to identify, discuss and agree on a series of topics which are key for the *buy-in* from MS/AC and the *uptake* from the research communities. Working in tandem with EU-funded infrastructure projects, these Working Groups are

[32] EOSC Working Groups https://www.eoscsecretariat.eu/eosc-working-groups

developing the foundations for EOSC through a consensus process on technical and non-technical components and propose priorities for their implementation.

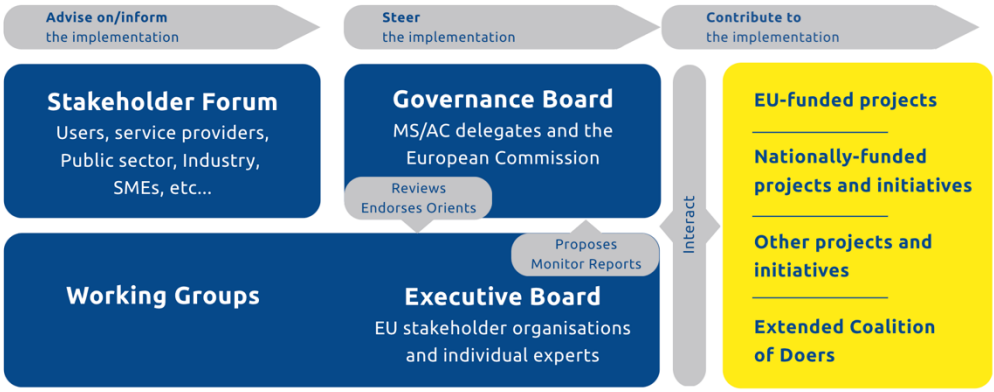

Figure 3. EOSC Governance 2019-2020

EOSC implementation currently revolves around the following themes, corresponding to the EOSC WGs[33]:

1. **Landscape** – Maps the existing research infrastructures which are candidates to be part of the EOSC federation. Studies the readiness levels of national initiatives related to strategies, policies and e-Infrastructures. Develops monitoring specifications and mechanisms to ensure MS/AC are up to speed and synced with EOSC developments.

   Results

   - *Landscape report* of infrastructures, initiatives and policies across Europe: existing policies and investments based on input from the MS and AC.
   - High level specifications for monitoring activities, including KPIs.

2. **Architecture** – Defines the technical framework, i.e., standards, protocols, APIs in EOSC Core to enable the envisioned federation of components and infrastructures.

   Results

   - AAI architecture defines basic architecture elements towards a common global ecosystem for identity and access control infrastructures.
   - PID policy collects expectations about what persistent identifiers will be used to support a functioning environment of FAIR research. Requirements of providers and the basic services they offer are also outlined.
   - Scholarly infrastructures for research software, which aims to place software as a first-class citizen in EOSC.

3. **FAIR** – Implements the FAIR data principles by defining the corresponding requirements for the development of EOSC services, in order to foster cross-disciplinary interoperability.

   Results

[33] Most results are interim pending consultation.

- EOSC Interoperability Framework, defines the general principles that should drive the creation of the EOSC Interoperability Framework, and organises them into the four layers: technical, semantic, organisational and legal interoperability.
- Recommendations for metrics to assess FAIR data and frameworks to certify services that enable FAIR.

4. **Rules of participation** – Defines the rights, obligations governing EOSC transactions between EOSC users, providers and operators: use of the tools, specifications, catalogues and standards and applicable methodologies; principles for regulating transactions in the EOSC (e.g. financial mechanisms and procedures, agreements/bylaws established by the EOSC governance framework); applicable legal frameworks (e.g. GDPR, copyright, data security and cybercrime, dispute resolution and redress mechanisms, e-commerce directive).

   Results

   - Rules of Participation: principles and recommendations for data, services and training.

5. **Skills** – Provides a framework for a sustainable training infrastructure to support EOSC in all its phases and ensure its uptake. Will provide policy and implementation recommendations on the type of digital skills required in EOSC, how to achieve via competence centers and how to embed in national strategies and infrastructures.

   Results

   - Minimum skill set for EOSC.
   - Options for organizational models for competence centres and their coordination.
   - Specifications for training catalogue(s).
   - Recommendations for EOSC skills/training in national digital skills strategies.

6. **Sustainability** – Provides a set of recommendations for the implementation of an operational, scalable and sustainable EOSC federation after 2020.

   Results

   - EOSC governance structure post 2020 (see Figure 4).
   - Establishment of a Legal Entity, the EOSC Association, which will manage the partnership *(to be officially established in July 2020)*.
   - Report on costs and business models.

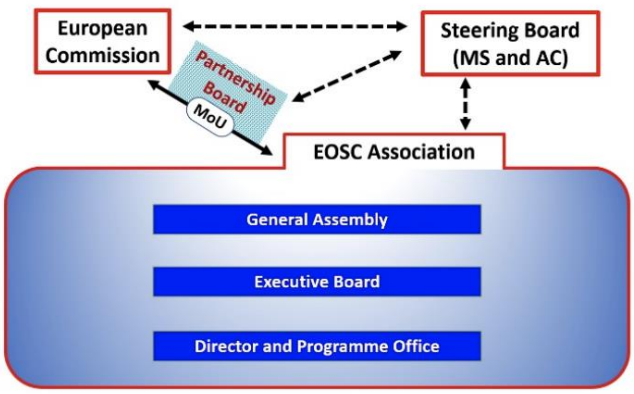

Figure 4. EOSC Governance post 2020

It is important to note that the new EOSC governance structure promotes closer links to MS/AC to improve the alignment of policies, infrastructure and funding. It foresees that each MS/AC may **mandate** a single legal entity established in their territory to represent it in the governance, so that the national EOSC related policies are taken into account and the broader engagement and specificities of each national research system are represented.

5.3.2 EOSC implementation timeline

The EOSC Strategic Implementation Plan for 2020[34] lays out the major milestones and deliverables resulting from the work of the EOSC WGs, illustrated in Figure 5. It is expected that these will produce agreed frameworks on various technical and non-technical issues, which would then need to be concretely implemented via EU infrastructure (both research infrastructures and e-Infrastructures) and national funded infrastructure programmes.

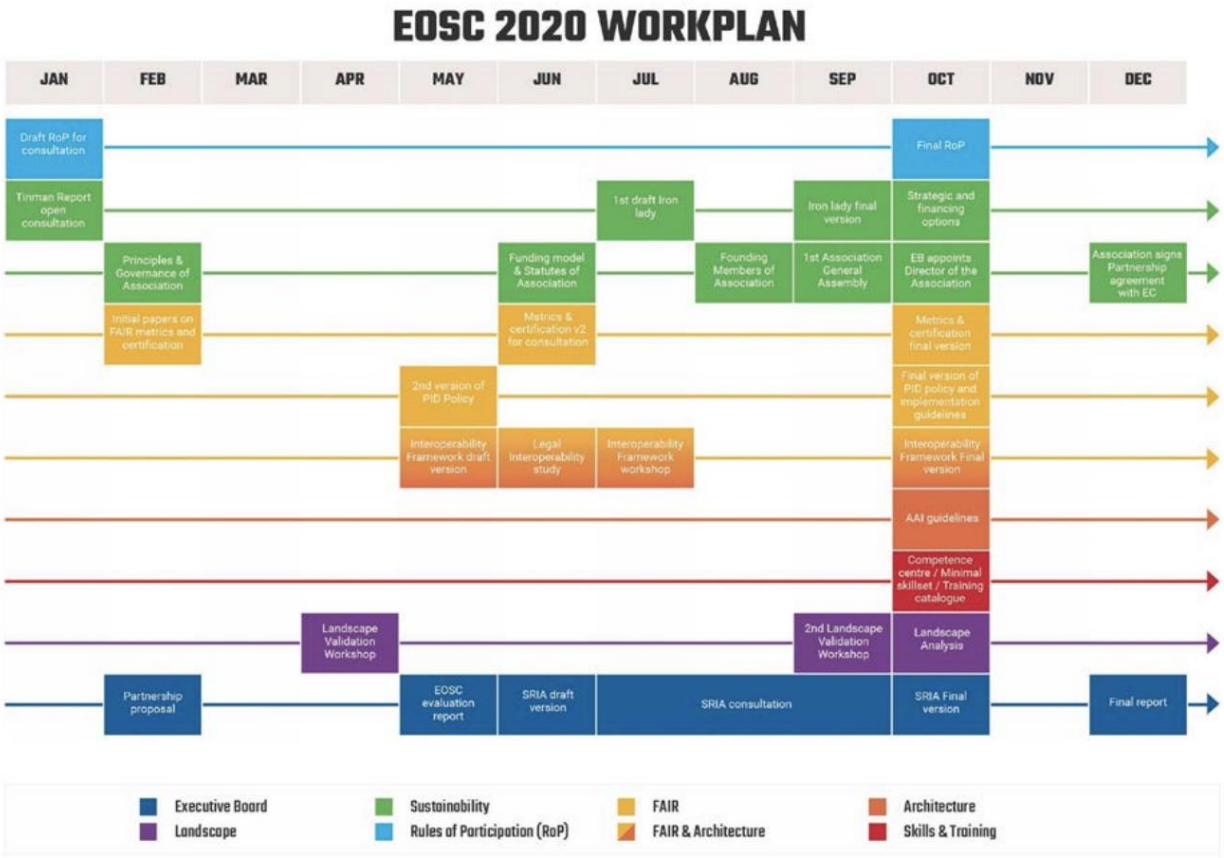

Figure 5. EOSC 2020 roadmap as presented by the Executive Board and EOSC WGs

For this reason, the Executive Board in working to publish the Strategic Research and Innovation Agenda (SRIA) which sets the basic strategic directions for funding and activities for the next 5-10 years, both from the EU and the MS/AC. The current working draft[35] identifies fourteen (14) Action Areas (AA) to help deploy the EOSC ecosystem, with some being more technical in nature, and others relating more to social

[34] EOSC Strategic Implementation Plan https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/ae215698-af7b-11ea-bb7a-01aa75ed71a1

[35] EOSC SRIA, Draft July 2020 https://www.eoscsecretariat.eu/sites/default/files/open_consultation_booklet_sria-eosc_20-july-2020.pdf

dimensions (financial, legal, educational, cultural, policy). These areas, illustrated in Figure 6, are key into identifying overlaps, gaps and points of intervention for the transport research community.

Implementation challenges	Boundary conditions
AA1: Identifiers AA2: Metadata and Ontologies AA3: FAIR Metrics and Certification AA4: Authentication and Authorisation Infrastructure AA5: User Environments AA6: Resource Provider Environments AA7: EOSC Interoperability Framework	AA8: Rules of Participation AA9: Landscape Monitoring AA10: Business Models AA11: Skills and Training AA12: Rewards and Recognition AA13: Communication AA14: Widening to the Public and Private Sectors

Figure 6. EOSC SRIA Action Areas

Furthermore, the EC is expected to actively pursue the implementation of EOSC via the upcoming INFRAEOSC-03 call *"Integration and consolidation of the existing pan-European access mechanism to public research infrastructures and commercial services through the EOSC Portal"*[36] in 2021-2023. This will effectively consolidate existing initiatives, such as EOSC-Hub, OpenAIRE and data infrastructures, essentially advancing and operating *EOSC-Core* and facilitating *EOSC-Exchange*.

[36] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/infraeosc-03-2020

6 Transport research in EOSC

6.1 Transport Research Cloud (TRC)

The DG-RTD 2018 HLEG report on the "Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud"[37] has investigated the potential of a Transport Research Cloud (TRC) as a subset of the European Union's European Open Science Cloud (EOSC) initiative. The experts identified recommendations grouped in five (5) broad topic areas which must be addressed for TRC to become more open to the concept of "open data by default" and have the greatest chance of being a sustainable and high value asset to the transport research community and European Union:

6. **Reusable research data:** Identify what constitutes transport research data and investigate the barriers for the limited use of data collected by others.
7. **Data as a public good:** Properly define what *public data* is, e.g., data collected under contracts paid for by tax payer funds, or data collected under a publically funded initiative/project.
8. **Standards:** Develop and use common standards among members of the transport research community in different sectors, i.e., governmental entities, infrastructure operators, and commercial consumers.
9. **Infrastructure:** Identify operating requirements for a TRC, including user communities, to design and develop infrastructure, ensuring an appropriate level of service provided at a cost that is understood by all stakeholders. Develop a proper governance model for a sustainable TRC.
10. **Incentives, training, and education:** Promote open science rewards and incentives to be adopted in the transport research discipline. Develop training/education programs for existing and future researchers and supporting human capital (e.g., librarians, data stewards, data curators).

As the EOSC is taking shape and the EC European Mobility Data Space will soon be in its design phase the above recommendations are very relevant. The **Transport Research Cloud** is envisioned as a subset of the EOSC, and not as a different entity. It aims to build on top of the *EOSC-Core* and complement the *EOSC-Exchange* with transport specific policies and services, engaging the transport research community in open science. More importantly, TRC plans to build and operate all bridges to the *EC Mobility Data Space*, i.e., infrastructure and policies for sharing data, as well as bring together/combine tools and services (Figure 7).

[37] DG-RTD HLEG 2018 "Analysis of the state of the art, barriers, needs and opportunities for setting up a transport research cloud" https://trimis.ec.europa.eu/content/analysis-state-art-barriers-needs-and-opportunities-setting-transport-research-cloud

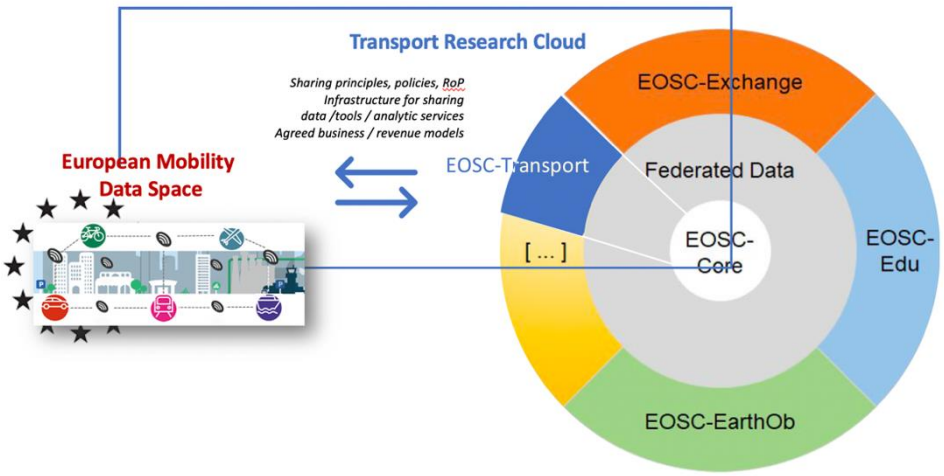

Figure 7. Schematic representation the Transport Research Cloud

6.2 TRC key alignment actions with EOSC

The 2018 EC Staff Working Document on EOSC[38] has defined six action lines for the implementation of EOSC, and to monitor their progress via the contribution of all infrastructures and funded projects (Figure 8).

Architecture	Architecture of the federated infrastructures as the solution to the current fragmentation in research data infrastructures which are insufficiently interoperable.
Data	FAIR data management and tools. A common data language to ensure data stewardship across borders/disciplines based on FAIR principles.
Services	Available services from a user perspective. A rich environment offering a wide range of services covering the needs of the users.
Access & Interface	Mechanisms/interfaces for accessing EOSC. A simple way for dealing with open data obligations or accessing research data across different disciplines.
Rules	Rules of participation for different EOSC actors. An opportunity to comply with existing legal and technical frameworks and increase legal certainty & trust.
Governance	Governance of the EOSC, aiming at ensuring EU leadership in data-driven science but requiring new governance frameworks.

Figure 8. EOSC Model action lines

Along with the 14 SRIA action areas (Figure 6), these action lines constitute the key themes where TRC should build and complement. The following subsections capture infrastructure elements as they currently stand, and describe the overlaps where TRC should participate and shape the discussion to fully follow EOSC implementation, the gaps where TRC should complement with specific requirements the EOSC community, and related points of intervention for the transport research community.

[38] COMMISSION STAFF WORKING DOCUMENT, 14.3.2018, Implementation Roadmap for the European Open Science Cloud, https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

6.2.1 Architecture

EOSC SRIA priority areas: *AA1 – Identifiers, AA4 – Authentication and Authorization Infrastructure, AA6 – Resource Provider Environments, AA7 – EOSC Interoperability Framework, AA14 – Widening to the Public and Private Sectors*

Overlaps: TRC should clearly follow EOSC policies and infrastructure on PIDs and AAI, as these are common to the whole European research community and are provided at institutional/national level by designated service providers and fully covers the federation of relevant research institutions. Due to the nature of transport research, i.e., mix of public-research-industry-user generated data, the PID policy should be extended and sensibly applied when such data is used for research purposes (e.g. maintaining provenance when data is copied, or is transferred from real-time to historical data). AAI should include policies and mechanisms for non-academics, as it is anticipated that public authorities will need to have access to services and data provided and operated by the research community.

The emerging *EOSC Interoperability Framework* not only covers issues related to the exchange of data (across scientific experiments, organisations or even communities), but also of other research artefacts that are commonly used in research (software, workflows, protocols, hardware designs, etc.). It has identified overlaps with the widely adopted *European Interoperability Framework* (ISA^2)[39], which makes it an excellent candidate to adopt in the cross-sector transport research domain.

Gaps: The major gaps originate from the interactions and data exchange beyond the research community, as well as the services to store and manage the big datasets being generated. Privacy, ownership and big/real-time data post challenges which are expected to affect many architectural elements and decisions. The *EC Mobility Data Space* and the industrial cloud platform landscape is likely to be characterized by a plurality of architectural patterns, ranging from approaches characterized by a high level of centralization (e.g. data lakes) to concepts promoting utmost decentralization (e.g. distributed applications using blockchain technology). The proposed architecture of the International Data Spaces Association[40] (Figure 9) indicates roles and processes which will be able to manage trusted exchange of data through brokering services, and such elements should be further considered within the EOSC architecture.

[39] ISA² - Interoperability solutions for public administrations, businesses and citizens, , https://ec.europa.eu/isa2/eif_en

[40] International Data Spaces, Reference Architecture Model, April 2019 https://www.internationaldataspaces.org/the-principles/#architekturmodell

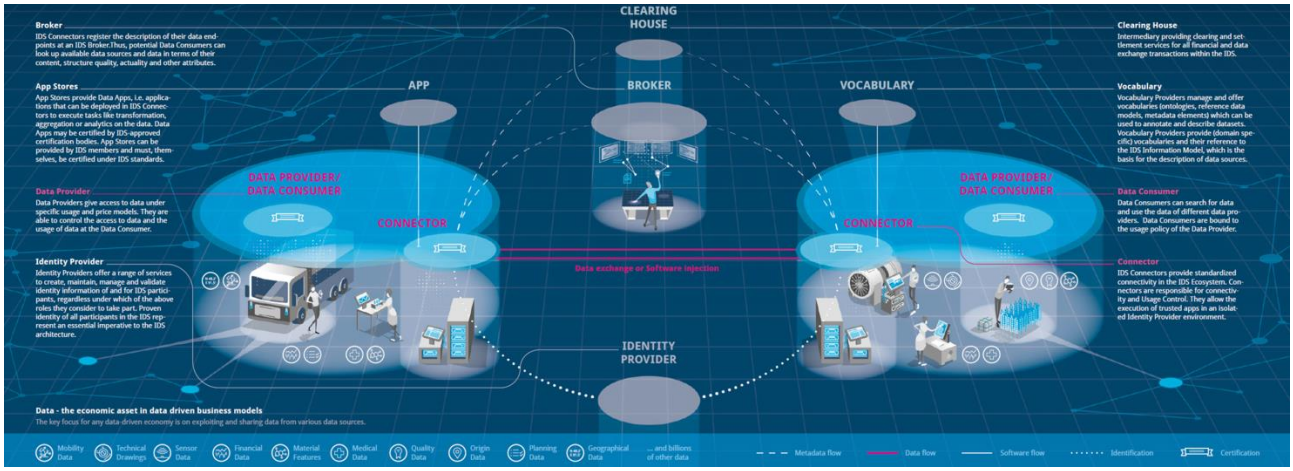

Figure 9. International Data Spaces proposed Architecture[41]

6.2.2 Data

SRIA priority areas: *AA1 – Identifiers, AA2 – Metadata and Ontologies, AA3 – FAIR Metrics and certification, AA6 – Resource Provider Environments, AA7 – EOSC Interoperability Framework, A11 – Skills and Training, A12 – Rewards and Recognition, AA14 – Widening to the Public and Private Sectors.*

Overlaps: EOSC is about making all research digital objects FAIR (data, publications, software, tools, services, etc.) and as open as possible, improving their integration (Linked Open Data). Particularly for data, the EC/HLEG has grouped transport research data in three main categories: **original research data (raw)**, **operational data** directly related to research and **data from published transport research**[37]**.** Reusability relies on the availability of high-quality metadata, and as such, TRC needs to collect, record and maintain various metadata standards, their mappings (if any) and ontologies that describe various assets and are used in the semantic layer. The *EOSC Interoperability Framework* is a solid starting point of reference and schemas like 1) the *DCAT Application Profile* for data portals in Europe (DCAT-AP) can be used as an interoperability bridging mechanism between research and public sector, providing minimum metadata for sharing data[42], 2) the *OpenAIRE Guidelines* for content providers[43] to be used for additional research artefacts for reproducibility, 3) *schema.org* for sharing via web, all to be complemented with additional metadata specific to transport research e.g., mode of transport. Furthermore, certification of content sources (for data, publications and software) is of key importance for sharing and re-usability and TRC should adopt EOSC developed certification mechanisms, potentially mixing them with certification processes from the Data Spaces[40].

TRC should adopt, and adapt where necessary, policies, processes and mechanisms for recognition rewards and incentives as they will be developed in EOSC. These are in the core of open science and one of the primary objectives of the EOSC Partnership.

[41] Conception and design: INFOGRAFIK PRO, https://www.internationaldataspaces.org/the-principles/#overview

[42] DCAT-AP includes: metadata information (date of creation of metadata, metadata language), content information (name of dataset, description of dataset), temporal information (publication date, date of expiry), geographical information, contact information conditions for usage, access information (data format, data structure, access URL), quality information (update frequency, data collection methodology).

[43] OpenAIRE Guidelines for content providers https://guidelines.openaire.eu/

Training researchers and support personnel with digital and information skills for open and data intensive science will mostly take place at institutional level through the development and coordination of Digital Competence Centers (DCC), new education curricula for researchers and life-long training for support personnel. TRC should adopt emerging models, calibrate them to transport research requirements and put them in operation in relevant research institutions (see **Error! Reference source not found.**.

Gaps: As closed as necessary is also a key aspect in sharing data and EOSC needs to develop and implement safeguard measures and mechanisms to protect sensitive data. Furthermore, legal metadata is key into ensuring a trusted space among the different transport sectors, currently missing from EOSC. As proposed in the OpenAIRE vision white paper[29] "*A minimum legal metadata framework, for ensuring openness and interoperability, privacy and security. These will include among others, copyright status, disclosure limitations, patents pending, other IPR on the datasets or workflows, existence of personal data, designation of data as PSI, etc."*

6.2.3 Services

SRIA priority areas: *AA4 – Authentication and Authorization Infrastructure, AA5 – User Environments, AA6 – Resource Provider Environments, AA14 – Widening to the Public and Private Sectors.*

Overlaps: Secure and personalised user environments/spaces are key into bringing trust in TRC, especially as competitive data spaces that involve industries will most likely be part of the interaction with the *EC Mobility Data Space*. TRC should guarantee a controlled execution environment for data related services and support the creation of trusted relationships, with unambiguously published preconditions and post conditions for both resource providers and users (e.g., privacy and security aspects).

TRC should additionally work with other research and e-Infrastructure communities and contribute towards the configuration and deployment of common services used in data processing and analyzing (e.g., computing and storage) and publishing (e.g., DMP, anonymization, linking data and publications) and possibly linked to TRC platforms.

Gaps: For TRC User Environments to foster a trusted open science environment, it is essential for them to develop and deploy specific authorization policies and related services, which will accompany privacy services embedded in data exchange and processing pipelines (e.g., differential privacy, blockchain). It is also important to look into tangible aspects and put in place safety measurements for protecting the "safety data" – sensitive data falling into "wrong hands", agreements for equitable data exchange between industry and research sectors, or to other global infrastructures e.g., in US, Canada or China.

Synergies with the project resulting from the INFRAEOSC-03-2020 – "*Integration and consolidation of the existing pan-European access mechanism to public research infrastructures and commercial services through the EOSC Portal"* call.

6.2.4 Access & Interface

SRIA priority areas: *AA4 – Authentication and Authorization Infrastructure, AA6 – Resource Provider Environments, A11 – Skills and Training, AA14 – Widening to the Public and Private Sectors.*

Overlaps: As already identified in Table 1- *Representative EU Big data transport research projects and their relation to open science*. EU projects are building platforms to facilitate big data/AI driven research, both

using operational or purely research data. In siloed approaches, they create tools and services that do not have clear open science/sharing policies. A dedicated TRC catalogue of services (including repositories as services) would be a main starting point to collect and describe such services, using and extending with TRC specific information (e.g., modes of transport, targeted stakeholders, KPIs) the EOSC Service Description Template and APIs used in the EOSC portal.

In addition, a data catalogue and search mechanism for TRC as a whole, including semantics for intelligent search, must be considered. Such catalogue must build on existing or emerging EOSC guidelines for minimal metadata requirements (e.g., DCAT, OpenAIRE, DataCite, schema.org), extended with TRC specific ontologies (e.g., provenance, legal metadata) and be part of the EOSC portal.

Gaps: A TRC dedicated portal, a subset of EOSC portal[44] but with more rich information and dedicated functionalities, would be a point of reference for TRC. Using the *EOSC-Core* mechanisms the TRC portal should include key functions so as to actively engage the transport research community in the shaping off EOSC and its links to the *EC Mobility Data Space*, while anticipating future developments. A non-inclusive list of such functions is provided below:

1. Service and data catalogues enriched with transport research domain ontologies to allow for more efficient and intelligent search and discovery. Data categorized by type (e.g., datasets, articles, books, etc.) and access rights (e.g., open access vs. restricted) to facilitate data sharing, searching and understanding.
2. Standards used in various aspects in transport research, e.g., metadata for different modes, operational vs. research vs. user generated data, APIs for data exchange, etc., similar to fairsharing.org.
3. Policies, strategies and projects for transport research per country, integrated information from TRIMIS and OpenAIRE as proposed in the implementation of the BE OPEN TOPOS Observatory.
4. Monitoring transport research results categorized via known ontologies for transport or SDGs, e.g., using the OpenAIRE Transport Gateway developed in BE OPEN[45].
5. Communities and user environments classified by mode of transport, target stakeholder group, which include data protection and metadata management services, requirement of DMP.
6. Educational and training material for open science topics: open publications/data/software, FAIR data, legal and ethical issues in using/sharing data, etc.

6.2.5 Rules

SRIA priority areas: *AA3 – FAIR Metrics and certification, AA5 – User Environments, AA6 – Resource Provider Environments, AA8 – Rules of Participation, AA14 – Widening to the Public and Private Sectors.*

Overlaps: Current RoP mostly states the high-level principles of engagement in EOSC, considering services, data and training material, targeting both providers and consumers. TRC must actively participate in the shaping of the next, more refined versions of EOSC RoP to ensure that they capture the specificities of

[44] DG RTD 2018 HLEG report: The TRC must mirror the EOSC as a subsidiary… 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.

[45] Currently in beta, the BE OPEN Gateway https://beta.beopen.openaire.eu provides a place with access to 5K OA publications and 19K other research products.

transport research, especially the interactions with government agencies, industry, mobility providers and citizens. This will give society valuable opportunities such as generating and widening new open data services, products and markets. In the case of *Resource Provider Environments* and *User Environments*, preconditions and post conditions need to be made explicit, and effects on the services/user environment must be outlined (e.g., certification, metrics for monitoring, legal and most often ethical concerns) .

Gaps: For RoP to be operational and therefore effective for transport research, EOSC must consider elements of what constitutes *public data*, especially transport operational data which is directly related to research and is frequently generated by non-research entities, e.g., data collected under contracts paid for by taxpayer funds. In fact, stakeholder groups, like public authorities, businesses and logistics services, if they can relate to a unique operational data sharing ecosystem to efficiently exchange information they need, they will improve their performance and optimize operational processes (e.g. sharing interoperable transport operational data it is of crucial importance since it could facilitate an operational planning synchronization and improve supply chain transparency and clustering capabilities, for a flexible, adaptable and multimodal transport environment). Further work in RoP must consider and embed rules provided by the directive on open data and the re-use of public sector information (PSI)[46], which provides a common legal framework for a European market for government-held data.

As a big volume of transport data is generated and collected by industry and for TRC to be successful, attractive and useful it is essential to consider what would be the possibilities to include this in early stages, requiring clear rules of cooperation. RoP needs to quickly move beyond high-level principles which currently have no reference to any specifics on legal tool-kits necessary to reach the principles (such as licenses, declarations of consent etc.) and address the lack of specific legal guidance.

Finally, the RDRTD 2018 HLEG report[37] has pointed out an existing gap related to developing clear rules on sharing data, to be filled by TRC partly by EOSC and partly by thee envisioned Mobility Data Space: *"Policy must specify the condition in which data is provided, curated, maintained and accessed. Policy must also specify how these data are to be served and how this service function is to be fund".*

6.2.6 Governance

SRIA priority areas: *AA9 – Landscape Monitoring, AA10 – Business Models, AA14 – Widening to the Public and Private Sectors.*

In order to develop a trusted TRC which fully aligns but also complements EOSC, the transport community must be actively engaged and represented in all EOSC governance layers. Bringing solid cross-sector use cases of data and data-intensive services generated and shared from the research-government-industry, with great impact on the emerging Horizon Europe missions[47] (e.g., adaptation to climate change including societal transformation, healthy oceans, seas coastal and inland waters, climate-neutral and smart cities), TRC is envisaged to play an important role in the EOSC Association.

Figure 10 illustrates the complex landscape of actors and interactions, indicating possible, not mutually exclusive routes on how the transport research community can participate in the EOSC governance.

[46] EU Public Sector Information framework https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

[47] EC Horizon Europe Missions https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme/missions-horizon-europe_en

Intermediary organizations will play a significant role in TRC involvement in EOSC as their mission is to coordinate and offer consolidated services: "*mandated*" national organizations and EU e-Infrastructures are singled out, while associations like ECTRI are expected to step up and perform activities which gather its members around EOSC.

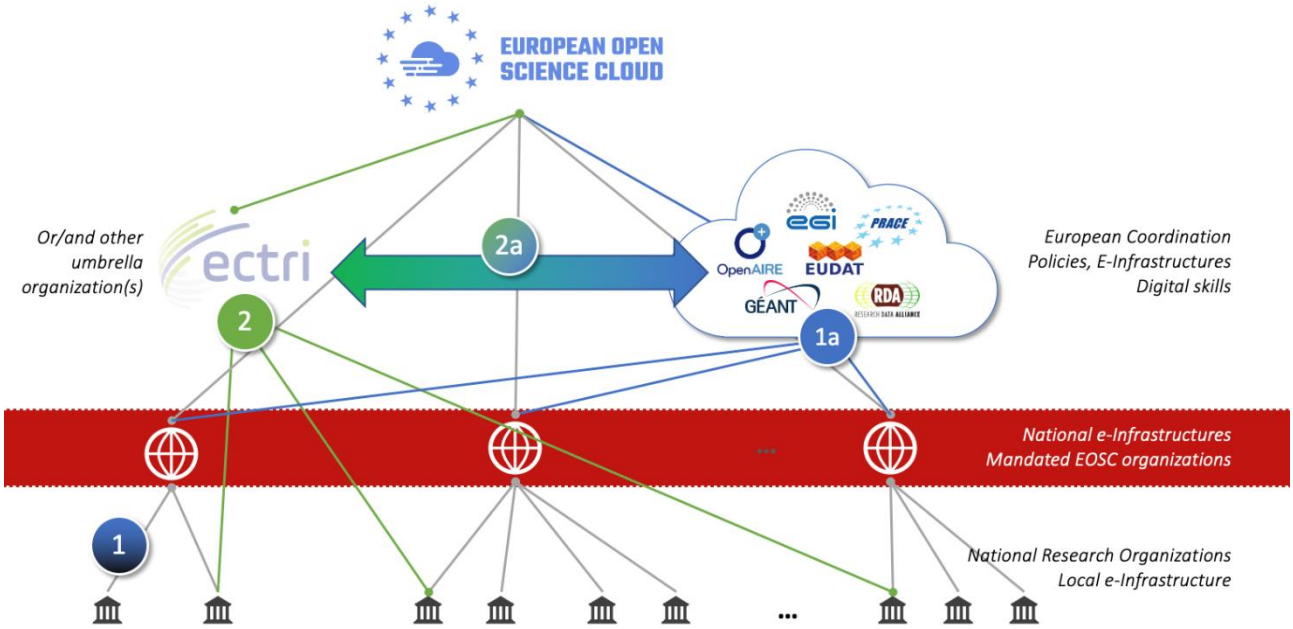

Figure 10. Interrelations and synergies of actors in EOSC

Route 1 – Proxy at the national level: Leading organizations in transport research participate in national EOSC structures, by building synergies with the designated national EOSC "mandated" organizations. This will ensure that existing or emerging national e-Infrastructures for research, which will be linked to EOSC by design, contemplate specific requirements from transport research local needs such as policies, regulation, infrastructure, sectors involved, and business models in a concrete manner and provide the corresponding resources. This is of particular importance if considering, for instance, the advantages that a possible cooperation between researchers and policy makers could bring in performing investigations and analyses to support the decision-making process hence anchoring transport research in policy settings around issues. A key aim would be to build a national transport open science portfolio and infrastructure which includes (non-inclusively):

- Policies and regulations, including data protection and privacy rules.
- Dedicated trusted data sources via EOSC certification mechanisms, including the storing, preserving and publishing operational transport data.
- Registries (e.g., people, initiatives, projects) including semantics.
- Service and data catalogues, ready to be integrated in TRC and EOSC.
- Training competence centers which embrace transport specificities.
- Funding opportunities for implementing local infrastructure compliant to open science and EOSC rules.

The mandated national organizations have close links to EU e-Infrastructures, which are also expected to be members of the EOSC Association, bringing additional elements and components of alignment in practical terms.

Route 2 – Proxy at the EU level: An European umbrella organization, like ECTRI for instance (or FEHRL, EARPA, and other academic networks like EURNEX, HUMANIST, FERSI), represents its member organizations in the EOSC Association and carries out dedicated activities which aim to: bring the requirements of transport research into EOSC and shape emerging developments for standards, roles, interactions, rules, architecture; align its members' policies, practices and infrastructure; develop and introduce an open science portfolio aligned with EOSC to be adopted across its members as part of TRC and solicit funding for its implementation. Such activities include (non-inclusive list):

- Collect and map policies for open science and FAIR data across member organizations, by maintaining and extending the TOPOS Observatory.
- Gather user and service provider needs for big data/AI-driven transport services and infrastructure;
- Promote and align EOSC, open science, RoP to members, including security, privacy and ethical issues of data reuse.
- Align open science practices for rewards and incentives to its members.
- Collect and document different metadata schemata, by building synergies with RDA to avoid duplication.
- Establish mechanisms to monitor open science developments in transport and its sub disciplines, including metrics (and consensus), avoiding duplication by building synergies with OpenAIRE through the TOPOS observatory.
- Study emerging business models and align with industry and the emerging *EC Mobility Data Space.*

Most importantly this umbrella organization must actively liaise with the EU e-Infrastructure community as pan-European service providers addressing different layers of EOSC and especially providers of *EOSC-Core* and enablers of *EOSC-Exchange*.

7 A pilot case study: OpenAIRE services for TRC

The OpenAIRE-BE OPEN liaison aims to promote and support the uptake of Open Science principles among researchers and help building and empowering the transport research community. OpenAIRE with BE OPEN partners is developing an **Open Research Gateway for Transport Research**. The gateway will offer a single entry point to discover research entities (all types of results, projects, organizations, and content providers) relevant to the domain of Transport Research. Within the gateway, researchers may find Open Science publishing tools to:

(v) Deposit research products of any type (i.e. not only scientific literature, but also datasets, software, workflows, methods, etc.) and get a persistent identifier via Zenodo.

(vi) Add details of the research context of a research product by creating links to projects and to other related research products (e.g. a link between a dataset and the software that analysed it).

(vii) Add research products to the community gateway. Via such tools, researchers populate and access an open, participatory scholarly communication graph of interlinked objects through which they can share any kind of products in their community, maximise re-use and reproducibility of science, and outreach the scholarly communication at large.

(viii) View statistics and analysis that support the monitoring of the uptake of Open Science.

The gateway will be further developed and enhanced to become the **TOPOS observatory** including organization related information. A preliminary version of the gateway is available at https://beta.beopen.openaire.eu, with work in progress for cleaning up and validating the underlying data. Once the BE OPEN gateway for Transport Research reaches a good and validated (curated) coverage and upon agreement of the BEOPEN consortium, the gateway will be registered in the EOSC Catalogue & Marketplace.

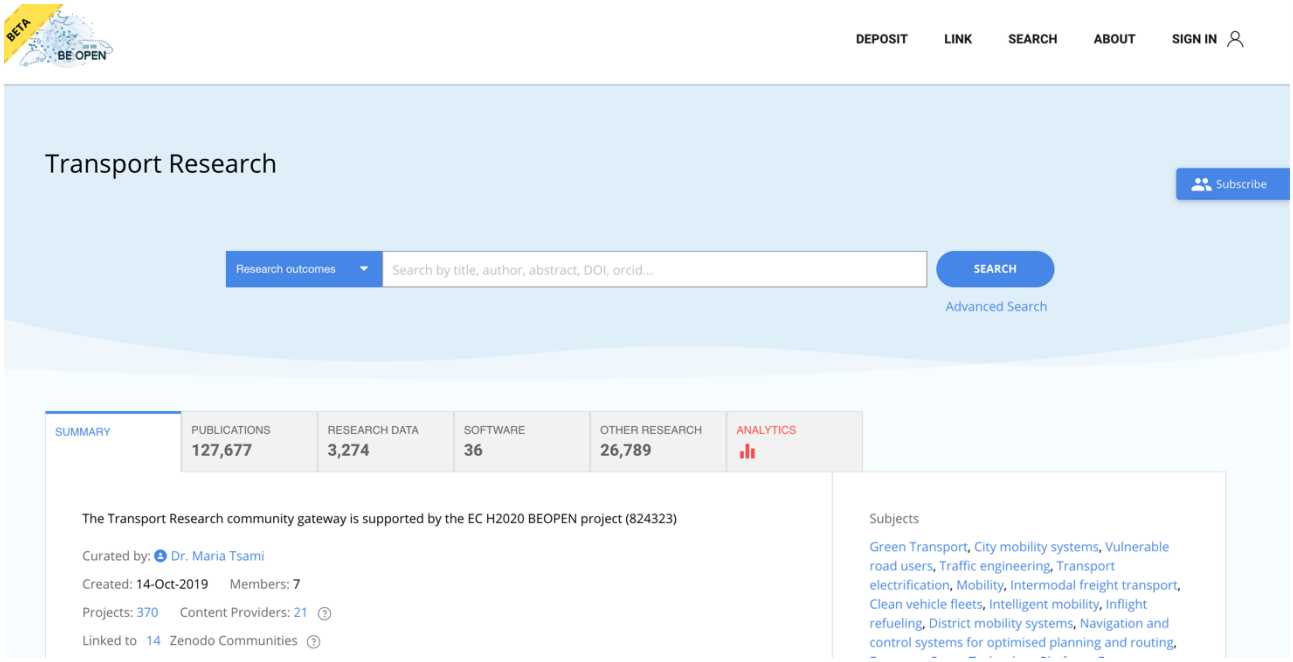

Figure 11. BE OPEN Gateway to transport open science results

Methodology: In order to proceed with the initial set-up of the gateway, the OpenAIRE team analyzed reports and BE OPEN deliverables for input to configure the gateway based on the OpenAIRE Research Graph[48] :

- List of relevant projects – link to Cordis and TRIMIS.
- List of relevant content providers – authoritative list from repository and OA journal registries from re3data.org, OpenDOAR, DOAJ.
- List of relevant research organizations – currently identified by the consortium, with envisioned future registration process.
- Glossary for transport statistics[49] used in text mining for classification purposes.

Detailed information on the methodology and tools is provided in Appendix 2 – Detailed methodology for BE OPEN Gateways.

[48] See overview at https://www.openaire.eu/blogs/the-openaire-research-graph, explore graph at https://explore.openaire.eu. Dedicated site at

[49] 5th edition — co-published by Eurostat, UN and ITF https://ec.europa.eu/eurostat/documents/3859598/10013293/KS-GQ-19-004-EN-N.pdf/b89e58d3-72ca-49e0-a353-b4ea0dc8988f

8 Appendix 1 – Organizations and their role in transport research

Table 4. : Influential research organizations (associations and research centres) in EU

Name	Country	Description
MULTIMODAL		
IFSSTAR *https://www.ifsttar.fr/en/welcome*	France	IFSTTAR is a major player in the European research on the city and the territories, transportation and civil engineering. The French Institute of Science and Technology for Transport, Development and Networks is a Public Institution of a Scientific and Technical Nature, under the joint supervision of the Ministry for an Ecological and Solidary Transition, and the Ministry of Education, Higher Education, Research and Innovation. It's consisted of 6 sites in France.
UITP *https://www.uitp.org*	Belgium	UITP is the International Association of Public Transport, which has more than 1,800 member companies in 100 countries throughout the world and represents the interests of key players in this sector. Its membership includes transport authorities, operators, both private and public, in all modes of collective passenger transport, the industry and research community.
Transportation Research Board (TRB) *http://www.trb.org*	USA	The mission of the Transportation Research Board (TRB) is to promote innovation and progress in transportation through research. In an objective and interdisciplinary setting, TRB facilitates the sharing of information on transportation practice and policy by researchers and practitioners; stimulates research and offers research management services that promote technical excellence; provides expert advice on transportation policy and programs; and disseminates research results broadly and encourages their implementation. TRB is one of seven program units of the National Academies of Sciences, Engineering, and Medicine, which provides independent, objective analysis and advice to the nation and conducts other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine. TRB's varied activities—described below—annually engage more than 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest by participating on TRB committees, panels, and task forces. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.
ITF *https://www.itf-oecd.org/about-itf*	International	The International Transport Forum at the OECD is an intergovernmental organisation with 60 member countries. It acts as a think tank for transport policy and organises the Annual Summit of transport ministers. ITF is the only global body that covers all transport modes. The ITF is administratively integrated with the OECD, yet politically autonomous.

Name	Country	Description
EURO Transportation WG http://www.ewgt.org		EWGT main targets concern providing a forum to share research information and experience, encouraging joint research and the development of both theoretical methods and applications, promoting cooperation among the many institutions and organisations which are leaders at national level in the field of transportation, traffic and logistics.
ROAD		
ACEA https://www.acea.be	Belgium	ACEA is the European Automobile Manufacturers' Association. It is an advocate for the automobile industry in Europe, representing manufacturers of passenger cars, vans, trucks and buses with production sites in the EU.
CEDR https://www.cedr.eu	Belgium	CEDR is the Road Directors' platform for cooperation and promotion of improvements to the road system and its infrastructure, as an integral part of a sustainable transport system in Europe. Its members represent their respective national road authorities or equivalents and provide support and advice on decisions concerning the road transport system that are taken at national or international level. The basic physical and organizational structures and facilities (e.g. buildings, roads, and power supplies) needed for the operation of a society or enterprise (Oxford English Dictionary).
CLEPA https://clepa.eu	Belgium	The European Association of Automotive Suppliers (CLEPA), brings together suppliers for car parts, systems and modules as well as national trade associations and European sector associations. As such it covers all products and services within the automotive supply chain.
ECF https://ecf.com	Belgium	The European Cyclist Federation (ECF) works on the promotion of cycling as a sustainable and healthy means of transportation and recreation, e.g., by encouraging consideration of cyclists' needs in transport planning and management, environment, safety and health, and promote cycle-friendly conditions throughout Europe.
ERF https://erf.be	Belgium	The European Union Road Federation (ERF) is a non-profit association which coordinates the views of Europe's road infrastructure sector and acts as a platform for research and dialogue on mobility issues between stakeholders and institutional players.
ETSC https://etsc.eu	Belgium	The European Transport Research Council (ETSC) seeks to identify and promote effective measures on the basis of international scientific research and best practice for a reduction in transport crashes and casualties. It provides an impartial source of expert advice on transport safety matters to the European Commission, the European Parliament, and Member States.
EUCAR https://www.eucar.be	Belgium	EUCAR is the European Council for Automotive R&D of the major European passenger car and commercial vehicle manufacturers. EUCAR facilitates and coordinates pre-competitive research and development projects and its members participate in a wide range of collaborative European R&D programmes.
IRU https://www.iru.org	Switzerland, Belgium	The International Road Transport Union (IRU) is the global industry association for road transport, supporting trade, economic growth, jobs, safety, the environment and communities. IRU's core constituents are national transport associations and transport operators.
ERTICO https://ertico.com	Belgium	ERTICO – ITS Europe was founded in 1991 at the initiative of 15 industry leaders and the European Commission to fill the gap between research and deployment of mobility services on roads. In the past two decades, ERTICO

Name	Country	Description
		ran dozens of European projects, organised 38 ITS European and World Congresses, grew from 15 to 120 Partners, undertook a key role in public consultations and policy agendas, and established long-term relations with key players in Europe, United States of America, Japan, Russia, and China.
LEVA-EU *https://leva-eu.com/links*	Belgium	LEVA-EU stands for Light Electric Vehicle Association in Europe and is the only trade association that works exclusively for LEVs in Europe. LEVA-EU was founded in 2017 by Annick Roetynck. LEVA-EU has partners and members around the world. Today (March 2018), LEVA-EU has more than 30 members in EU member states, China, Norway, Russia and Switzerland.
RAIL		
UIC *https://uic.org/about/uic*	France	UIC, the worldwide professional association representing the railway sector and promoting rail transport. UIC leads an innovative and dynamic sector, helping Members find continuing success and opportunities. Members are invited to take a proactive role in the UIC working groups and assemblies where the railways' position on regional/worldwide issues is shaped.

9 Appendix 2 – Detailed methodology for BE OPEN Gateways

9.1 Gateway configuration

The following subsections present the current methodological aspects and components of the BE OPEN gateway, which will become a first prototype of using existing services from OpenAIRE. We anticipated that these tools will be enhanced and adapted to TRC requirements until the end of BE OPEN project.

9.1.1 Zenodo communities

Zenodo is a multi-disciplinary repository for sharing, preserving and showcasing research products of any type (e.g. publications, posters, presentations, research data, and software). When a user deposit a product on Zenodo, a DOI (if not already available) is assigned to it and the metadata are pushed in real-time into the OpenAIRE Research graph.

In order to group depositions based on user-defined criteria, Zenodo introduces the concept of "Zenodo communities". They can be seen as "curated containers" of research products. Often, Zenodo communities are used by project managers to group deliverables and outputs of their research products; by institutions that have not an official open access institutional repositories; by research groups working on a specific research topic.

Community gateway managers can search for Zenodo communities whose products are relevant for the domain and add them to the configuration: all research products deposited in the selected Zenodo communities will then be discoverable via the gateway.

Managers can also select or request to create one "main" Zenodo community to be associated to the gateway (e.g. in the Neuroinformatics gateway there is the "OpenAIRE-Connect Neuroinformatics" Zenodo community). Typically, one of the gateway manager is also the curator of the main Zenodo community.

The selection of relevant Zenodo communities for Transport Research has not yet been performed by managers. For testing purposes, the Zenodo community named "Journal of Sustainable Development of Transport and Logistics" has been added to the configuration and led to the inclusion of 41 publications to the gateway.

9.1.2 Content providers

The list of relevant content providers as of November 2019 includes 24 repositories (see the annotated list in https://data.d4science.net/AfQQ).

Most of the repositories are registered on re3data (19). Three of them are already aggregated by OpenAIRE via Datacite (4TU.Centre for Research Data, Centre for International Earth Science Information Network, and Kinder Institute Urban Data Platform) and one is directly aggregated (Copernicus). The four repositories cover a wide range of domains and cannot be considered specific enough to Transport Research and include in the gateway all the products they host. OpenAIRE will be able to identify from those three repositories (and all

repositories it aggregates) the research products relevant for the domain of Transport Research exploiting the information from the subjects.

The 15 repositories that are not available via Datacite do not seem to offer proper API that OpenAIRE can use to collect the metadata about the research products they host.

The initial configuration of the community gateway will therefore not include any content providers.

9.1.3 Organizations

Among the information available in the OpenAIRE Research Graph there are links between organizations and research results (i.e. authors' affiliations), available in some cases in the aggregated metadata of research results and, for Open Access publications, obtained by applying text mining techniques on their full-texts.

Those links can be used to identify research products to include in the gateway, provided the organizations' research focus is the domain of the gateway (Transport Research, in this specific case).

The list of relevant organizations as of November 2019 includes 40 organizations from the public or private sectors, all of them working in the field of Transport Research except for one (Austrian Institute of Technology, excluded from the gateway configuration).

The list available in https://data.d4science.net/AfQQ has been annotated with the identifiers of the organizations assigned by the trusted registry grid.ac and the internal identifiers assigned by OpenAIRE. The availability of an OpenAIRE identifier ensures that it is theoretically possible to have affiliation links to that organization in the OpenAIRE Research Graph. If instead the OpenAIRE id is not available but a grid.ac is, Over a total of 39 organizations, 21 have an OpenAIRE identifier and can be added to the gateway configuration. 5 of the 21 organizations cannot be found on grid.ac, but they are available in OpenAIRE via EC project databases (i.e. those organizations are participants of EC FP7 and/or H2020 projects).

Thanks to the initial configuration for organization, in December 2019, 470 research products were identified as relevant for the community.

9.2 Mining for funders and projects

As described earlier, a number of relevant FP7 and H2020 EC projects were added to the gateway configuration. Whenever a research product is identified as funded by one of those projects, it will also be tagged as relevant for the community and thus included in the gateway. Project mining in OpenAIRE text mines the full-texts of publications in order to extract matches to funding project codes/ID. The mining algorithm works by utilising (i) the grant identifier, and (ii) the project acronym (if available) of each project. The mining algorithm: (1) *Preprocesses/normalizes* the full-texts using several functions, which depend on the characteristics of each funder (i.e., the format of the grant identifiers), such as stopword and/or punctuation removal, tokenization, stemming, converting to lowercase, etc.; then (2) *string matching* of grant identifiers against the normalized text is done using database techniques; and (3) the results *are validated and cleaned results using the context near the match* by looking at the context around the matched ID for relevant metadata and positive or negative words/phrases, in order to calculate a confidence value for each publication-->project link, and a threshold is set to optimise high accuracy while minimising false positives, such as matches with page or report numbers, post/zip codes, parts of telephone numbers, DOI's or URL's, accession numbers, etc. The algorithm also applies rules for disambiguating results, as different

funders can share identical project IDs; for example, grant number 633172 could refer to H2020 project EuroMix but also to Australian-funded NHMRC project "Brain activity (EEG) analysis and brain imaging techniques to measure the neurobiological effects of sleep apnoea".

Project mining works very well and was the first Text & Data Mining (TDM) service of OpenAIRE. Performance results vary from funder to funder but precision is higher than 98% for all funders and 99.5% for EC projects. Recall is higher than 95% (99% for EC projects), when projects are properly acknowledged using project/grant IDs. These rates are based on tests we have performed with manual validation of the results. For BE OPEN, we additionally developed a new mining module to identify sixteen relevant FP5 and FP6 EC projects that are not integrated in OpenAIRE. Its results will appear in the BE OPEN gateway, after the completion of the current data provision run.

9.3 Document classification

OpenAIRE also provides a document classification service that employs analysis of free text stemming from the abstracts of the publications. The purpose of applying a document classification module is to assign a scientific text one or more predefined content classes. In OpenAIRE, the currently used taxonomies are arXiv, MeSH (Medical Subject Headings), ACM and DDC (Dewey Decimal Classification, or Dewey Decimal System). Of these taxonomies, only DDC, which is a hierarchical library classification system[50], partially covers transportation, under the class "300 Social Sciences", and since we have adopted only the 54 most frequent classes, which includes only the higher division "380 Commerce, communications & transportation", this is not adequate for the purposes of BE OPEN. Instead, we decided to work on extending the OpenAIRE classifier by adding a taxonomy for transport research, using the latest (2019) edition of the "Glossary for Transport Statistics"[51], which has been jointly published by the United Nations Economic Commission for Europe (UNECE), Eurostat, and the International Transport Forum, and provides definitions of statistical terms for all modes of transport. Compared to the 4th edition of the Glossary, which is described in Chapter 7 of deliverable D2.1[52], the new 5th (2019) edition comprises 744 definitions, in which many chapters have been substantially revised.

OpenAIRE's classifiers were built using supervised learning machine learning techniques[53]. To add a new classifier for transport research requires a large dataset of relevant publications research, together with annotations/labels for all subcategories of the "Glossary for Transport Statistics". This does not currently exist and therefore a more semi-automatic/manual solution was proposed. The modified and simplified document classification consisted of three parts: *1) Creation of weighted terms from the glossary* by splitting each term in the glossary to individual words or bigrams; *2) Preprocessing of plain texts and vocabulary,* as used with the project mining modules; and *3) Classification* on abstracts of publications: we iterate over each word/bigram of the abstract and JOIN it with the weighted terms using database techniques. When a term is found we increase the confidence of its corresponding category by its weight. Each matched term counts only once even if it appears multiple times in the abstract since according to our experiments this leads to

[50] DCC: https://www.oclc.org/en/dewey.html
[51] UNECE/ITF/EUROSTAT (2009). Glossary for Transport Statistics, 5th Edition. Publications Office of the European Union, 2019. Luxembourg
[52] D2.1: Open access publications and the performance of the European transport research, 2019. BE OPEN. https://beopen-project.eu/storage/files/beopen-d12-open-science-framework-terminology-and-instruments.pdf
[53] D10.2: Clustering Algorithms, 2016. OpenAIRE2020 project. https://doi.org/10.5281/zenodo.1257349

better results. The result of following the above three step process was the creation of just over 820 weighted single-word terms and 174 weighted bigrams, covering all nine of the Glossary's categories, from A (Railway Transport) to I (Passenger Mobility). For each publication that goes through the transport research classifier, the output is one of the categories, the weight, and a set of terms matched (with their respective weights) that can explain the behaviour of the classifier and are helpful for identifying useful or not so useful terms. This allows for improving the classifier. For example, this way we quickly discovered that the term "transport" is actually a bad term to use. If we use it we get hundreds of false positives: papers relating to physics, astronomy, quantum mechanics, etc., such as *"Theory of Fast Electron Transport for Fast Ignition"*. Instead, "transport" is useful when used in bigrams, such as: air transport, freight transport, goods transport, etc.

Depending on the confidence threshold selected, the accuracy of the classifier varied between 55-70%. Higher accuracies substantially reduced recall, and so our overall assessment was that this semi-manual approach was not good enough to be included as part of the OpenAIRE document classification service. As highlighted previously, the proper way to create the classifier would be to first collect a large corpus of publications related to transport, and have them manually labelled by one or more experts in transport research at least at the first level (A-J) of categories, if not at the second level (e.g. I-VI). Then in combination with negative examples (publications unrelated to transport) we could apply machine learning methods and train them appropriately. Since this is not feasible during the duration of the BE OPEN project, another alternative that we are considering is to work with the publications that are currently being collected by the BE OPEN gateway (e.g. from publications linked via the FP7 & H2020 projects tagged as relevant to transport research, or the relevant organisations, institutions, repositories or Zenodo communities provided). This would produce a more targeted dataset for testing and tuning the classifier.

9.4 Context propagation

One of the main principles of Open Science is that all kinds of research products are made available and linked together with semantic relationships. For example, a scientific publication may have a link to the analysed dataset, the software used for processing, and the output data. In many cases, the metadata record describing the publication is richer than the metadata that describes the software and the dataset. Also, metadata of the publication can be enriched with information by applying text mining algorithm on its full-text, as described in section 5.8.

As a consequence, the algorithms described in the previous sections may not identify as relevant for the community the research products whose metadata is not rich enough. However, these products may be linked to others that have been identified thanks to a richer metadata set.

In order to address this case, OpenAIRE is experimenting in its beta infrastructure an approach to "propagate the context" from one research product to the products it links, provided the semantics of the relationship is "strong enough". In particular, information about the community can be propagated to products which are linked via relationships with semantic "is supplement to" and "is supplemented by"[54].

In the first run on December 2019, the algorithm applied to the BEOPEN use case identified additional 17 research products for the Transport Research gateway.

[54] with the semantics as defined by Datacite, see https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf

It is expected that the number will grow over time because of two main factors: (i) the identification of more research products by the other approaches, which would lead to a higher number of products that can be used as starting point for the propagation; (ii) the increasing availability of semantics relationships between the research products: if the uptake of Open Science publishing principles in the community grows, more semantic relationships among the products will be available in the OpenAIRE Research Graph and exploitable by the context propagation.