Deliverable D2.1
Requirements Report

T has received funding from the European Union's DG for Mobility and Transport under grant agreement No
34/SUB/2019-104/CEF/PSA/SI2.821136

Deliverable 2.1 - Requirements Report

Due date of deliverable: 02/07/2020

Actual submission date: 14/12/2020

Dissemination Level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Start date of project: 02/01/2020 Duration: 48 months

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Document Control Sheet

Deliverable number:	**D2.1**
Deliverable responsible:	ITxPT
Work package:	WP2
Main editor:	Anastasia Founta

Editor name	Organisation
Emmanuel de Verdalle	ITxPT
Guido Di Pasquale	UITP

Document Revision History			
Modifications Introduced			
Version	Date	Reason	Editor
1.0	21/09/2020	First version	Anastasia Founta
1.1	30/10/2020	Integration of Member States' feedbacks	Anastasia Founta
2.0	15/11/2020	Final release	Anastasia Founta

Legal Disclaimer

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

1 EXECUTIVE SUMMARY

The main objectives of DATA4PT project consist of facilitating the development and deployment of the European public transport data standards Transmodel, NeTEx and SIRI for the provision of Union-wide multimodal travel information services which apply to the TEN-T network including urban nodes. The validation of the collected public transport data is recognised as a key process to enable their wide use. This report focuses on the identification of the appropriate requirements for the development of validations tools which respond to the needs and requests of involved stakeholders, representatives of different EU Members' States (MS).[1]

In particular, DATA4PT aims, among other objectives, to enable the interoperable exchange of travel and traffic data in accordance with the technical requirements outlined in ITS Directive Delegated Regulation EU 2017/1926 across the Union. EU has already set the ground by requiring the development of the National Access Points (NAPs) by all MS. NAPs can mainly used as an open source depository for all collected data. Now the next step is to ensure the accuracy, pertinence and harmonisation with EU standards of the collected data, both in national and regional level.

While the volume of collected data will be increasing day by day, replying to the current need of enabling multimodality across Europe, transport authorities will struggle to control the quality of these data in terms of usability and technical thoroughness. Therefore, the necessity of the development of useful and easy to perform validation tools for public transport data is becoming more and more urgent.

Through this report, DATA4PT aims to the shed light to the needs and expectations of public transport authorities, operators and other involved actors (ministries, transport research institutes, experts etc.) from different European countries (partners[2] and observers[3] countries of the project). To capture their opinions, ideas and experiences at this initial stage, interviews, one survey through questionnaire and one workshop have been carried out among the nine (9) Members States partners of the project and the observers MS. Moreover, the report refers to existing tools and identifies useful and missing features to be considered for tools development and complement MS feedback. The conclusions of this work result in the first set of the requirements for the initiating the validation tools development.

[1] This report summarises the activities that took place under the Task 2.1. (Requirements) of the project. The conclusions of this report will be used as input for the Task 2.2. (Validation tools), focused on the development of the validation tools and their specifications.

[2] Partners are Austria, Croatia, Czech Republic, Denmark, France, Italy, Portugal, Slovenia, Sweden.

[3] Observers are Germany, Norway, Switzerland, UK, Belgium, Netherland, Hungary, Spain, Luxemburg, Ireland, Poland, Finalnd.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

List of partners

Partner's name	Acronym	Country
Union internationale des transports publics	UITP	Belgium
Information technology for Public Transport,	ITXPT	Belgium
Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie	BMK	Austria
Ministry of the sea, transport and Infrastructure	MMPÏ	Croatia
Centrum dopravního výzkumu, v. v. i.,	CDV	Czech Republic
Trafik-, Bygge- og Boligstyrelsen (Danish Transport, Construction and Housing Authority),	TBST	Denmark
Direction générale des infrastructures, des transports et de la mer	DGITM	France
Ministero delle Infrastructture e dei Transporti	MIT	Italy
INSTITUTO DA MOBILIDADE E DOS TRANSPORTES, I.P.,	IMT	Portugal
Republika Slovenija, Ministrstvo za Infrastrukturo	MZI	Slovenia
Trafikverket (Swedish Transport Administration	STA	Sweden

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Abbreviations and Acronyms

AVMS	Automatic Vehicle Monitoring Systems
MMTIS	Multimodal Travel Information Services
MS	Members States
NAP	National Access Point
NeTEx	Network Timetable Exchange
PTA	Public Transport Authorities
PTO	Public Transport Operators
SaaS	Software as a Service
SIRI	Service interface for real-time information
TRANSMODEL	Public Transport Reference Data Model

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Table of Contents

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

List of Figures

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

List of Tables

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

INTRODUCTION

The multimodality and sustainable mobility in general, are concepts based on the public transport data collection, management and exploitation. Beyond the responsibilities of public transport operators, transport authorities and general governments in EU are supported to collect and store travel and traffic data creating and feeding open source databases such as National Access Points (NAPs). While the volume of collected data will be increasing day by day replying to the current need of enabling multimodality across Europe, relevant authorities will struggle to control the quality of these data in terms of usability, accuracy and technical thoroughness.

As public transport is the main pillar of multimodality, the necessity of the development of useful and easy to perform validation tools to ensure the accuracy, pertinence and compliance with the respective European standards (Transmodel, NeTEx, and SIRI), is becoming more and more urgent.

This report aims to shed light on the key elements that need to be considered for the development of such validation tools. Validated data will be used mainly to feed the NAPs.

In this respect, two main axes of investigation are identified:

- What is the **current practice regarding data validation?** Is there any validation tool available? What are their main characteristics (commercial, open source etc.) and main components (features, modules, functions etc.). How do responsible bodies (authorities/operators/suppliers/manufacturers etc.) currently validate the data? Do they use existing tools?
- What are the **main functions a validation tool need to run** based on MS experience, needs and expectations? In order to identify the functions and the components of the tools, other questions can also be raised such as **who** is going to use the tools, **how easy need to be** (specific technical background is required or not), **when** do they need to **be used, who** is going **to use the validated data, who** is going **to host the tools** (online platform, NAPs or a local installation will be needed) etc**.**

The parameters, arising from the above questions, are interrelated. They are furthermore described in Table 1. Several options might be considered for each one of the requirements aspects, such as the examples given in the right-hand column of the table.

The findings from the mapping of current situation, combined with the feedback from MS, will help define/choose the option which will better address stakeholders needs. In Section 2, it is described with details the methodological approach that has been applied to address the aforementioned issues. Then, in Section 3, it is presented current practice with regard to data validation procedures in MS, partners and observers of the project. Additionally, existing tools are listed presenting shortly their main functionalities. Section 4 is dedicated to the presentation of MS needs and expectations for data validation tools. Finally, in Section 5, a set of requirements is defined based on the findings of section 3 and 4.

Table 1. Parameters of requirements to be identified.

Requirements parameter	Description	Examples of options
Purpose of use/ Validation focus	The tools will be mainly used for data validation in respect to EU standards. However, the tools might also serve data preparation and upgrade	Checking - Validation Preparation - Correction All above/Other

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Requirements parameter	Description	Examples of options
	incorporating data quality checking and correction modules. Current capacity gaps and future expectations of the MS will help prioritise different needs and define the different scopes of the tools.	
Users and validation responsibility	The definition of the users in combination with the purpose of use will define tools main functionalities. Different tool features may occur depending on the users. The tools might have one or many different kind of organisations as users, while can also be available to professionals. Current procedures and future expectations of MS will define target group of users and will contribute to the formulation of a business model. Validation responsibility might be assigned to the data providers (responsible to deliver validated data) or to the data recipients (responsible to check and validate the received data). Validation process can also be applied by a neutral nominated body as a intermediate process between data providers and recipients.	Public transport authorities Public transport operators National Research Institutes Public transport international organisations (e.g. ITxPT) Consultants All above
Administrator/host	Administrator refers to the body that hosts the tools and is responsible for maintenance and technical support. This aspect is indirectly link to functional requirements, as security issues (or other issues regarding performance) might taken into account for tools functionalities in case that administrator is not also the user, but just the provider of the tools.	Delegated National research Institutes. Delegated Public Transport Authorities. Public transport international organisations (e.g. ITxPT)
Technical characteristics	Technical capacity and resources of the users entities influences the technical characteristics of the tools. More complex tools or tools based in legacy systems will require more resources for maintenance and adaptation to future improvements or extensions. MS expectations on the technical aspects of the tools need to be examined in combination to the other parameters to enable prioritisation between probable contradictory requirements.	Complex tools to be used by specialised staff. Simple tools easy to use. Tools easy to maintain, improve or extend.
Periodicity of tests	The periodicity of the tests might also influence the development of the tool. It depends on the final user of the validated data and the type of data to test.	To be able to perform unlimited tests per day and per user. Limited number of

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Requirements parameter	Description	Examples of options
		tests/user.
Historical (static) VS **Real time (dynamic) data**	Another parameter to be considered is whether tests need to be performed in real time data or in historical data. SIRI EU standards concern real time data, while NeTEx concerns static data (timetables, fares). Different formats may influence technical parameters of the tools. More than one tools might needed for different cases. Real time validated data might be more considerable in the future to address the needs of advanced multimodal mobility business models.	
Datasets content and format	Based on the type of included information different parameters might be defined particularly regarding to semantic rules (validity of values of the physical quantities etc.). **NeTEx** provides means to exchange data for passenger information such as stops, routes timetables and fares, among different computer systems, together with related operational data[4]. **SIRI** specifies a European interface standard for exchanging information about the planned, current or projected performance of real-time public transport operations between different computer systems[5]. **NeTEx and SIRI** define the type of information to be collected and the way to exchange it. Validation tools might address either to one of these data models of to both types. The format in both cases is XML schema.	Network Topologies (e.g. routes, journey patterns etc.) Timetables (including Journey patterns, vehicle journeys, operating days etc.) Fare data including products, tariff coverage, conditions and prices. Etc.
Type of tests to be performed/functions	The type of tests to be performed is the main output of this activity. Depending on how the validated data will be used and to which technical standards will refer, different kind of testing might be required. MS needs and gaps in their capacity to feed NAPs with standardised data will be taken into account to prioritise the functions and the integration level of the final output (e.g.	Testing technical thoroughness (syntax, coding, etc.) Testing data accuracy/ pertinence Data conversion and testing compliance with EU

[4] http://netex-cen.eu/?page_id=11

[5]

Requirements parameter	Description	Examples of options
	integrated tool, stand-alone tool, toolbox etc.)	standards Etc...

In particular, data content and format is defined by the technical standards NeTEx, and SIRI. Both purpose content and format are highly correlated to the type of testing and tools that it is required. Therefore, before identifying existing tools and future requirements, a short presentation of the main characteristics and information included in NeTEx and SIRI models follows.

1.1 MAIN CHARACTERISTICS OF NETEX TECHNICAL STANDARD [1]

Data in NeTEx format is encoded as XML documents that must conform exactly to the provided XML NeTEx schema. The schema can be used to exchange the following information:

- Public Transport schedules including **stops, routes, departures times / frequencies, operational notes, and map coordinates.**
- Routes with complex topologies such as **circular routes, cloverleaf and lollipops, and complex workings** such as short working and express patterns. **Connections with other services** can also be described
- The **days on which the services run**, including availability on public holidays and other exceptions.
- **Composite journeys** such as train journeys that merge or split trains
- Information about the **Operators providing the service**.
- Additional operational information, including, **positioning runs, garages, layovers, duty crews, useful for AVL and on-board ticketing systems**.
- Data about the **Accessibility of services to passengers with restricted mobility.**
- Data is versioned with **management metadata** allowing updates across distributed systems
- **Fare structures**, (flat fares, point to point fares, zonal fares)
- **Fare products** (Single tickets, return tickets, day, and season passses etc)
- **Fare prices** that apply at specific dates

The schema can also be used to create bindings to different programming languages to assist automating part of the implementation process for creating software that supports NeTEx formats.

Documents in NeTEx format are computer files that can be exchanged by a wide variety of protocols (http, FTP, email, portable media, etc). In addition, a **SIRI based NeTEx protocol** is specified for use by online web services. It defines NeTEx request and response messages that can be used to request and return data in NeTEx format, and also publish/subscribe messages for push distribution.

1.2 MAIN CHARACTERISTICS OF SIRI TECHNICAL STANDARDS [2]

SIRI is a XML protocol for exchange of public transport real-time information. SIRI allows pairs of server computers to exchange structured real-time information about schedules, vehicles, and connections, together with general informational messages related to the operation of the services. SIRI comprises a general communications architecture, and a number of specific services which operate within that architecture. The communications architecture supports two different patterns of interchange.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

- A synchronous request/response protocol: each exchange of data consists of a request message from a client consumer, and a response message from a producer server.
- An asynchronous subscribe/publish protocol: the client subscribes to information by sending a message to the server containing both request information, and sensitivity criteria with which to filter messages. The producer server establishes a subscription for the consumer and will send messages back to the consumer whenever the criteria are satisfied, until the subscription ends. This pattern is 'stateful', that is to say, both parties in the interaction must manage the use of subscriptions that persist over time through successive interactions.

In both cases messages consist of XML documents, whose tags and content are exactly specified by the SIRI XML Schemas.

SIRI comprises eight different concrete services, each consisting of request and delivery message pairs, and all using a common architecture, terminology, reference data.

- Production Timetable Service: Supports the dynamic **exchange of planned schedules, including updates**. These may be used by AVMS systems to predict and monitor vehicle progress.
- Estimated Timetable Service: Supports the **exchange of target schedules in real time, including updates**. These may be used by AVMS systems to predict and monitor vehicle progress.
- Stop Timetable Service: Provides information about **schedules for arrivals and departures at a Stop point.**
- Stop Monitoring Service: Provides information about **arrivals and departures at a Monitoring**, i.e. Stop point.
- Vehicle Monitoring Service: Provides information about the **movement of a vehicle, and its progress against the target schedule**.
- Connection Timetable Service: Provides information about **schedules for interchanges at a connection point**.
- Connection Monitoring Service: Provides information for **interchanges at a connection point** to support guaranteed connection services
- General Message Service: Supports the **exchange of general messages**.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

2 METHODOLOGICAL APPROACH FOR THE IDENTIFICATION OF VALIDATION TOOLS REQUIREMENTS

The identification of validation tools requirements was based on interactive processes involving directly Members States (partners of the project) as well as observers (Members States outside the consortium) and experts.

Three channels of communication were put in place in order to capture both current situation and stakeholders needs: one workshop, one questionnaires survey and one - to - one calls. All three channels focused on questions, to triggers discussions and enable better understanding MS needs. Depending on MS readiness and implementation status of EU standards, some MS provide less details about the different aspects to consider. External factors as pandemic crisis have also affected their preparation. Therefore, this report is considered as a living document that can be updated during the following period if needed. Nevertheless, the different level of progress between MS contributes to a more integrated determine validations tools requirements. In the next sub-sections further details regarding the methodology applied are given.

2.1 WORKSHOP

The first input from MS collected through a dedicated workshop during kick- off meeting of the project. Workshop objective was to collect relevant inputs and engage MS partners at an early stage of the project. Moreover, it was a good opportunity to involve more MS outside of the consortium in the discussion, as “observers”.

Table 2. Observers (users groups) of the project.

Members States	Organisations
Germany	VDV
UK	Department for Transport, Traveline.info
Netherlands	Opengeo, Dova, RWS
Norway	Norwegian Railway Directorate, Entur
Switzerland	Federal Office of Transport, SBB)
Belgium	Belgian ITS Steering Committee: STIB/MIVB, DeLijn)
Ireland	National Transport
Spain	Fomento
Luxemburg	Verkéiersverbond
Poland	Ministry of Transport)
Hungary	(Ministry of Transpport)
Europan Railways Agency	ERA

During the workshop, the participants were asked to present current situation with respect to the data collection, storage and management and link it to the EU standards implementation procedure. This approach contributes to the identification of MS needs in an indirect way. On the one hand, insights of the established procedures, such as who is now responsible for checking and validating data, what is the process between data providers and data recipients, who host the NAPs etc, can help defining the users, administrators and scopes of the tools and addressing organisational challenges. On the other hand, it helps identifying capacity gaps in implementing EU standards and therefore adapt tools development in a way that will have the biggest added value for MS.

Data4PT has received funding from the European Union’s DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

The questions focused on three topics: their vision, the current status and their expectations from the project. For each one of the themes different points were mentioned. In the following table (Table 1), the posed questions are presented.

Table 3. Questions posed to MS with regard to validation tools requirements.

Topic	Questions
Vision	What is your national strategy on Data standards and NAP?
	How does your national law fit into the EU directive?
	What are you key requirements regarding NAP implementation?
	What type of data covered in the NAP (PT, Car-sharing, Parking, etc.)?
	What is your implementation planning?
Current status	What is the state of play in your country against the MMTIS Directive's plan?
	Do you already have the NAP in operation?
	How many PTAs (or PTOs) are already providing data to the NAP?
	What type of data are already collected through the NAP?
	Which standards are used to feed the NAP?
	Are you using some software tool to import/convert data models to feed the NAP?
Expectations	What type of support do you expect from the project?

2.2 SURVEY AND ONE-TO-ONE CALLS

After the first workshop, a questionnaire was distributed between MS and observers to focus on validation tools requirements and pilot sites definition. The questions that address directly validation tools requirements are :

1. What do you expect from a validation tool? (e.g. compliance check, consistency of data etc.). Please detail your expectations.
2. Have you already identified and/or used software tool(s) covering these features? If yes, which one(s) - precise if in-house, open source or commercial.

Based on the answers, type of tests to be included will be determined while the reference to any existing tool will help identify missing features and good aspects, both important for the development of the tools.
Pilot sites definition related questions help reply to parameters such as what type of data need to be considered (type of PT networks, type of PT modes etc.). To complement requested inputs and to better orient MS focus, ITxPT team organised one-to-one calls with MS partners. During the call, insights regarding all the above topics were discussed so as to clarify and specify the different aspects.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

2.3 EXPERTS TEAM INPUT

Experts team in the frame of WP3 (Teaching, Expertise and Support) of the project provided their input regarding existing tools. Further evaluation of the existing tools and feedback on requirements by experts' team is expected in the upcoming period, during validation development.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

3 EXISTING TOOLS AND CURRENT PRACTICE REGARDING DATA VALIDATION PROCESSES

In this section, gathered information from project partners, observers and experts stakeholders regarding existing tools is summarised. Special focus is given to MS current methods and processes for data validation. The aim is to identify gaps in MS capacities to validate data to feed their NAPs and to assess in which extent the available tools can reply to their needs. It is evaluated which of the required features are covered by existing tools and which ones are missing and need to be considered in the future.

Moreover it aims to contribute to the selection of a common open source tools as references. Commercial tools are also taken into consideration in terms of their functions but there will not be used as benchmark.

3.1 EXISTING TOOLS

Existing tools are distinguished in three main categories: the validators, the converters and the editors along with the data management tools. Actually, the validators concern the checking and validation process, while editors and converters support the formulation of the datasets based on the EU standards. Converters, editors and management tools are not directly linked to validation, however they are tools that might incorporate validation modules in order to provide a complete service. The identification of their functions (not in details) aims to better understand current needs and provided services in data creation and management as an integrated process.

Data validators check documents content in terms of coherence, form and completeness. Any XML validator can be used to check the syntax of a NeTEx document. This will check data types, tags and tag order, and referential integrity. Moreover, XML editing tools include a validator to check the content of documents against a schema. Additional validators can apply more specific semantic rules on the use of content. General semantic rules for Transmodel / NeTEx are described in the EPIP profile and refer to reasonable and consistent values for speed, distance, temporal span across related elements. [3]

Furthermore, **data editors and management tools** are available providing means of creating, updating, storing and exchanging transport data. These tools have the capability to produce NeTEx files directly, avoiding the need of format conversion and optimising the whole process of collecting, validating and providing standardised data. For this reason, a reference of their main functionalities is made as follows:

- ***Open source data editors and management tools***

One open source management tool is the **CHOUETTE** software. CHOUETTE is mainly financed by the French Ministry of Ecology, Sustainable Development, Transport and Housing (MEDDTL). Its main purpose is to collect and exchange data, describing the scheduled timetables of public transport networks, in accordance with a standard exchange formats and profiles (NeTEx, NEPTUNE and GTFS). CHOUETTE is complemented by IRYS for SIRI based real-time information. In particular CHOUETTE is able to [4]:

- create PT supply data
- check the compliance of PT supply data with the so-called Neptune standard (Reference: NF P 99506) as well as with the GTFS standard,
- import / export / convert data from one format to another
- visualise transport data on a map (example: positioning of a bus stop, network topology)

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Another open source transit data integration platform for timetables and real time data is **Bliksem Integration,** developed in Netherlands. The toolkit consists of an ETL[6] layer capable of extracting data from national standards into a normalised, versioned database. This database can be transformed in any transmodel based concept such as GTFS or a NeTEx profile. Using the stored schedules and punctuality information it can produce GTFS-Realtime feed.

- ***Commercial data editors and management tools***

Most leading European Software providers of Public Transport data management software (MDV, Trapeze, Hogia, Init, PTV, Itoworld, etc.) have already added or are adding a capability in order to support the EC ITS directive for National Access Points. The commercial tools are integrated platforms where data validation is incorporated to enable accurate data exports and data analysis so to deliver high quality services such as passengers information, journey predictions and planning, considering also any disruptions. Some of the commercial tools are:

- **Hogia provides PT data integration platform**. Hogia PubTrans handles passenger information for 70% of all service journeys in the Nordic countries, including the public transport authorities in Stockholm, Helsinki and Copenhagen. It promises consistent passenger information, regardless of operator, transport mode (bus, metro, train, tram and ferry), or information channel. Some key features of their relevant products (PubTrans Base and DaRT) are the following [6]:
    - Import of data about organisations, contracts, routes, networks, stop points, multi-modal journeys, and vehicle blocks.
    - Export of a comprehensive picture of public transport traffic.
    - Open and standard import and export formats built on Transmodel.
    - A versatile database with a coherent picture of the scheduled services
    - Automated – Data flow requires no manual intervention
    - Real-time information
    - Forecast of arrivals and departures
    - Public Interfaces – external systems can input and retrieve data
    - Integration of data – scheduled traffic, temporary changes, disruptions and real-time
    - Automated – data flow requires no manual intervention
    - Rich data model – many aspects of passenger information and traffic management
- **Ito World PT Data integration platform.** ITO World, a UK based company, provides a platform for delivering real-time transit data feeds for journey planners and a platform for transit authorities & operators. In particular, Ito Transit Hub [7]:
    - ingests transit and related data in any format, for all modes and all operators in a city or region.
    - ingests datasets including transit schedules, real-time updates, disruptions, and fares are constantly ingested and processed
    - creates an always up-to-date, accurate model of the entire transit system providing a single-source of truth to power downstream systems

[6] ETL is the Extract, Transform, and Load of data. ETL collects and processes data from various sources into one data store where it can then be later analyzed (https://www.xplenty.com/blog/what-is-etl/).

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

- provides consistent passenger information through all channels, such as digital street and on-vehicle signage, journey planning apps and websites and via 3rd party applications.
- exports in bespoke or standard feeds including GTFS, GTFS-RT, TxC, SIRI and NeTEx and is also available through developer APIs.
- provides accurate arrival time and journey predictions for all services based on historic and live data
- delivers preferred and alternative routes to journey requests
- manages disruption information, travel news and alerts
- reports and analyses real time and historic data drill down to any level of detail, aggregating data to lines, services, areas, and customisable time periods.

- **IVU.pool,** developed by German company IVU Traffic Technologies, is an integration solution with interfaces to all common formats, serving as a basis for the network-wide timetable information. It can import timetable data from a wide variety of sources and integrate it into a standardised overall network. IVU.pool administers different timetable versions, allowing networks to consult the construction timetable early before operation. In more details, IVU.pool**:**
  - uses standard interfaces for import and export. It integrates the timetables of various companies and generates an overall timetable
  - manages disruption information such as road works, events, holiday timetables. It anticipates future changes automatically
  - performs standard, recurring tasks automatically, which speeds up workflows
  - administers and imports real graphs to generate a clearly arranged visual representation of route layouts on maps
  - adds extra information such as transfer times or walking distances in buildings

- **Mentz Diva,** the Integrated Timetable Planning & Vehicle Scheduling, can plan exact trips, blocks, drivers and duties. This service aims to help finding optimum solution, which is integrated, meaning that timetable planning, duty- and vehicle scheduling are performed in one step and constantly affect one another. It addresses to transport authorities, big and small, urban & regional transport companies as well as data managers & occasional users. Key points of the integration platform are the following:
  - Managing common data in an interactive map, automatically calculating distances and times of route options from GIS data and saved trip time profiles, clustering day types and creating timetables based on wait times or trip patterns that can be edited synchronously for trips of day types.
  - Creating vehicles schedules, checking for errors and conflicts and automatically completing missing block sections or operational trips.
  - Calculating connection relations, optimising and creating interval timetables.
  - Providing statistics for operational cost parameters.

  Additionally, the Dynamic Data Integration Platform (DDIP) from MENTZ collects and distributes real time so that it can be used in passenger notifications (such as electronic schedule information). This platform:
  - collects and disseminates real-time data between computer-aided guidance systems and information systems. A multi-client-capable web interface enables DDIP administrators to configure subscriptions and set up data flows themselves,
  - makes all required subscriptions available to all recipients, and no additional subscriptions are necessary on the supplier's side,
  - converts VDV, SIRI, and GTFS formats and maps data from one service to another wherever it makes sense

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

- manages disruptions by integrating real-time data into routing from the very beginning
- offers 24/7 real time data availability.

Finally, **format converters** are considered as useful tools to help adaptation to EU standards. The converter translates data from one format to another. They support validation process in an indirect way by helping data translation into standardised format. Open source converters are: a) "Hastus to NeTEx" , b) "CHOUETTE" [3]. CHOUETTE tool is able to convert GTFS to NeTEx and vice versa. Data management tools also incorporate converters in their functionalities.

Table 4. Summary of examined existing tools.

Type of tools	Features/functions	Tools name/type of source/key info
XML Validators	XML files syntax	Any typical XML validator
	• data types, XML tags, XML tag order, referential integrity	
	XML content consistency against a schema	Any XML editor include validator to check content against a schema.
	XML content based on semantic rules Transmodel/NeTEx	UK TransXChange Publisher: Legacy National data that applies additional semantic rules.
	• speed values, distance values, temporal span values across related elements	
Converters	GTFS to NeTEX	Chouette (open source) Data management tools (commercial)
	NeTEx to GTFS	Chouette (open source) Data management tools (commercial)
	Hastus to NeTEX	"Hastus to NeTEx" converter (open source)
Editors and management tools	Create PT data in accordance to EU standards (NeTEx, Transmodel)	Chouette (open source) Chouette Core - Enroute -version (SaaS solution). New version of Chouette for creation of NeTEx data with validation capabilities as well. Business model is unclear, however looks like this mainly will be supplied as a SaaS solution[7].

[7] https://github.com/enroute-mobi/chouette-core

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

		Hogia PT integration platform Ito World PT Data integration platform IVU.pool
	Check compliance with national profiles	For French NeTEx profile : Chouette (open source). For Nordic NeTEx profile: Entur[8] has developed the Chouette-AFIMB version of specialised validator for Chouette to handle the validation pipeline for the Nordic NeTEx profile . This is however a lot of "legacy" and we want to replace it with a new tool which are easier to maintain. (https://github.com/entur/chouette)
	Check content consistency and correct automatically	Mentz Diva
	Import / export / convert data from one format to another	Chouette (open source) Bliksem Integration Hogia PubTrans MENTZ DDIP
	Manage disruptions of timetables	Hogia PT integration platform Ito World PT Data integration platform IVU.pool MENTZ DDIP

3.1.1 MISSING FEATURES TO TAKE INTO ACCOUNT

In terms of validation, existing tools enable adequate checking of data integrity and consistency against a XSD schema and/or against defined semantic rules. This kind of tools can actually check compliance with NeTEx and SIRI XSD schema. However, XML validators serve single purposes and address specific needs.

Validation can also be made by data editors and management platforms, as part of data processing in order to be able to provide “sound” information and support adequately the platforms services. Most of this

[8] Entur AS is a government-owned transportation company in Norway. Owned by the Norwegian Ministry of Transport and Communications. It was created to offer sales and ticket solutions for the railways and travel planner for public transport throughout Norway. In the autumn of 2016, the company took over the operation of NSB's ticket systems (https://en.wikipedia.org/wiki/Entur).

Data4PT has received funding from the European Union’s DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

category tools are commercial and support the PT daily operations. They consider EU formats and profiles both for the creation and exchange of data. In this respect they can provide to NAP standardised datasets. Their integrated approach can inspire the formulation of validation tools, underlining the needs of the PT ecosystem.

In conclusion, the availability of tools dedicated to check the content in terms of accuracy and validity is not clear. However, accuracy and validity testing is also important to ensure quality of data and serve the purposes of the NAP. Moreover, there is lack of more integrated validation tools, able to run different kind of tests in respect to standardised data models. Finally, MS needs and expectations will help identify validation tools scope and define functionalities for a short, medium and long term horizon.

3.2 CURRENT PRACTICES

Data validation procedures are different between the EU countries. Depending on the implementation level of EU standards, validation focuses on the data compliance with the EU standardised formats or with the national standards. The consistency of data to the established format is usually checked with a relevant data format validator (e.g. in Czech Republic with JDFkon application). In most cases, tools to check data quality in respect to a national profile are developed in house by the transport authorities or operators (e.g. in Norway, Portugal). For the MS which have already implemented at a certain level the EU standards (e.g. France, Italy, Norway, UK), the developed tools focus on compliance to NeTEx and SIRI technical specifications. Checking for trust emblem is another type of test some of the countries are adopted (e.g. Austria).

Table 5. Validation practices in different EU countries.

EU Members States	Validation practices
Czech Republic	An application called JDFkon is used to check the data in JDF format (national standard) for buses. JDFkon is a console Win32 application that is used to check the accuracy of data in timetables in the JDF format version 1.9. The application searches for input files in the JDF format in the given directory. If an error is found in the input data, a file titled JDFkon.log will be generated in the directory, in which the error will be located and described. The error number, file title, number of the line with the error, and description of the error will be stated in the line.
Norway	Entur[9] developed and uses the AFIMB version of specialised validator for Chouette to handle the validation pipeline for the Nordic NeTEx profile . All PT-data of Norway is collected (both static-timetables- and real-time) and made

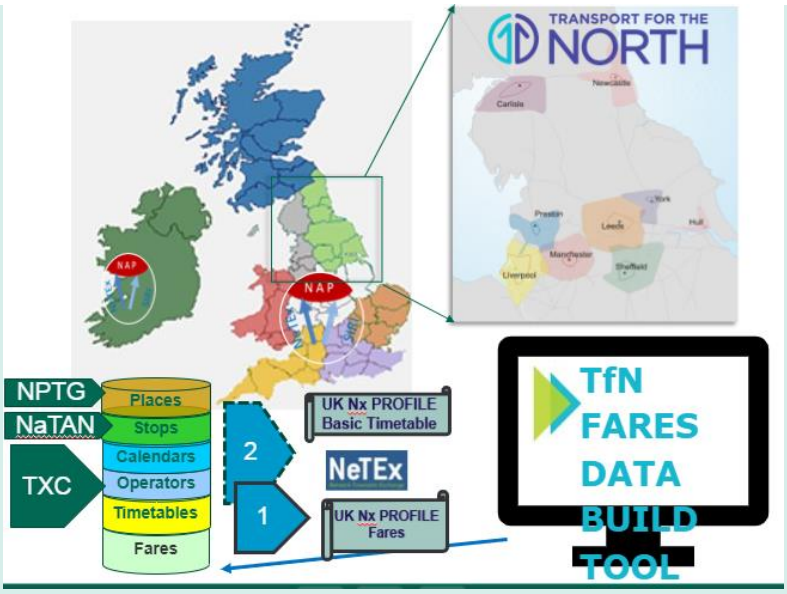
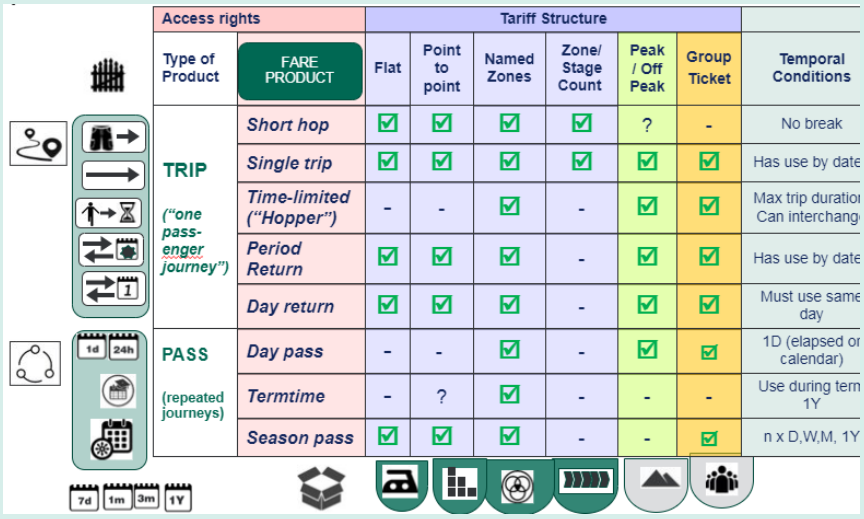
[9] Entur AS is a government-owned transportation company in Norway. Owned by the Norwegian Ministry of Transport and Communications. It was created to offer sales and ticket solutions for the railways and travel planner for public transport throughout Norway. In the autumn of 2016, the company took over the operation of NSB's ticket systems (https://en.wikipedia.org/wiki/Entur).

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

EU Members States	Validation practices
	available through an fully open Journey Planning API with full national coverage, the "Journey Planner Ruter[10]". The platform is already fed by NeTEx/SIRI data for all major PT-traffic in Norway. Validation procedure focuses now on data quality and richness of timetable based on PT-services (both static and dynamic) and is shifted to micro mobility data, demand responsive transport (DRT) services and price/product data.
Portugal	Some Portuguese Transport Authorities have experience with small data validation tools developed in-house, but none with data on European Standards (NETEX, SIRI, etc.). Currently there is no specific tool for the validation on the metropolitan areas of Porto and Lisbon (both are PTA's). The control process is done in ArcGIS and in Excel. The tests are not standardized, and rather follow different procedures according to the type of analysis they want to carry out. Both PTA's, are currently running a tender for public transport services and that following that tender they will receive more information from the PTO's. To handle that data, in parallel, they will also run a tender for implementing a monitoring system.
UK	Transport for the North BODS fares project is using the UK NeTEx fare profile, for the Metropolitan areas Leeds, Manchester, Sheffield, etc. In this respect TfN 'Fares Data Build Tool' will help regional bus operators to comply with upcoming national regulations requiring the 'open' publication of bus journey information, including fares data. The Fares Data Build Tool will be free to use for bus operators and local transport authorities (LTAs) who are responsible for publishing this data.

[10] https://ruter.no/en/journey-planner/

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

EU Members States	Validation practices
	 Figure 1. The transformation process from national UK to EU standards and TfN tool. (Source: Kick-off presentation).  Figure 2. Basic UK bus fares products.
Sweden	Samtrafiken, an organisation owned by all the regional public transport authorities and most of the commercial operators, just started to collect data on NeTEx. Currently using different standards (mostly NOPTIS), they developed and use conversion tools from NOPTIS to GTFS, GTFS-RT and NeTEx. They are also in the process of building (in-house) validation tools for the BoB-standard (National Ticket & Payment Standards for Public Transport). The need for verification tools is highlighted in respect to EU standards.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

4 NEEDS AND EXPECTATIONS OF MEMBERS STATES TOWARDS REQUIREMENTS SET-UP

This section summarises MS needs and expectations as they were expressed during the dedicated workshop, questionnaire and interview in order to define the different requirements parameters. Additionally, current practices and the existing tools assist the understanding of MS opinions and expectations.

4.1 PURPOSE OF USE/VALIDATION FOCUS

Based on the participants' answers, special focus is given on the compliance and consistency to the EU technical standards such as NeTEx and SIRI and to national profiles. However, this presuppose the creation of a common ontology between EU countries, a common minimum EU profile, to ensure a common base for validation. The existence of multiple national profiles in addition to the European one will require configurable tools settings. MS, with less experience in EU standards implementation, mentioned to the versatility to handle data from different modes and from different sources and formats (GTFS standard to NeTEx, GTFS-RT to SIRI) with different characteristics and reliability levels. Moreover, MS expressed their will to have the possibility to integrate the tools in the data production pipeline.

Table 6. Validation focus based on surveys results.

Requirements parameter	Definition
Purpose of use/ Validation focus	• Compliance and consistency to the EU technical standards and to national profiles. • Support the transition from other formats to EU ones - integrated tools • Support the production pipeline - integrated tools.

4.2 USERS AND VALIDATION RESPONSIBILITY

In all MS, the nominated bodies, responsible for NAPs deployment, are also obliged to conduct compliance and quality assessment of the provided data. Nevertheless, PTOs and PTAs are foreseen as the main users of the tools checking their own data before submitted it. Hence, tools should be shared.

Table 7. Suggested users based on surveys results.

Requirements parameter	Definition
Users/ responsible for data validation	• EPIP related Stakeholders (PTOs and PTAs) to use the tools to validate the data they provide. • NAPs nominated bodies to be able to assess provided data.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

4.3 ADMINISTRATOR

Based on received answers, validation tools are expected to be hosted by NAPs websites. In this case, responsibility of tools maintenance and well functioning will lay on NAPs nominated body. However, other neutral bodies may be considered. Security aspects of the systems are highlighted. Tools development need to take into account security issues towards safe testing and secure sharing of data.

Table 8. Suggested administrators based on surveys results.

Requirements parameter	Definition
Administrator/ Host	• Embedded tool on NAPs websites or on neutral body's platform with increased security systems.

4.4 TECHNICAL CHARACTERISTICS

The need of a tool easy to maintain and which is not based on older systems was highlighted. Indeed, maintenance capability differs among the MS. Some of them might struggle to maintain it or extend/adjust it, if it is required in the future. Actually, lack of expertise is one of the problems that more of the MS face. This fact need to be taken into account for the development of tools (tools easy to maintain, use and adjust) or for the assignment of the administrator. Additionally, the "plug in" architecture was underlined by the MS as an important technical characteristic of the tools, for different purposes though (either to allow further national profile rules to be added, or to be integrated in data production pipeline).

Table 9. Technical characteristics/aspects.

Requirements parameter	Definition
Technical characteristics/aspects	• easy to maintain (less legacy) • easy to use • plug in architecture

4.5 DATASETS CONTENT

Data content is defined by the technical standards NeTEx, and SIRI as presented in § 1.1 and 1.2. In spite of the standardised datasets, MS do not collected so far all the information described by the standards. The current availability of data may help prioritising some functionalities with regard to data quality.

In fact, the availability in data is most likely limited to Public Transport schedules, including network and topology parameters (stops, routes, departures times / frequencies, operational notes, and map coordinates etc.) and to fares structure and pricing. In most cases includes all transport modes, but there are also cases which mainly refer to bus and train data. For the MS that have already implemented journey planner platforms, connections with other services can also be described as well as data about the accessibility of services to passengers with restricted mobility.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

Table 10. Datasets content

Requirements parameter	Definition
Datasets content	• SIRI and NeTEx included elements • Priority to validation of timetables, network, topology elements, fares/pricing

4.6 PERIODICITY OF TESTS

Taken into account the users and the scope of the tools, as defined in § 4.1 and 4.2., tests need to be carried out in a daily base. From nominated bodies perspective, responsible for quality and harmonisation of data in NAPs, the usage of the tools might be limited to random tests of the collected data. From PTAs and PTOs point of view, as data providers, the usage of tools should be seamless and adaptable to the demand.

Table 11. Periodicity of tests

Requirements parameter	Definition
Periodicity of tests	• without limitations

4.7 TYPE OF TESTS TO BE PERFORMED/FUNCTIONS

Based on the participants' answers DATA4PT validation tools need to include layered set of validation tests reporting multiple types of error of different levels of severity. All proposals of tests are grouped and presented in Table 12.

Table 12. Type of tests and functions to consider.

	Type of tests	Description/ examples/
1	Compliance of format with NeTEx / SIRI	ex: "Valid .zip Archive and files"
2	Compliance with NeTex/SIRI XML-Schema	ex: "XML Schema validation error" Conformance to XML schema as to data types, tags, tag order, integrity check using a standard XML validator. Correct use of codes.
3	Other basic XML document compliance checks (non-XSD based)	ex: "Missing Name on Line" "Use JourneyPattern instead of ServiceJourneyPattern"
4	Internal consistence	ex: "Journey patterns shall be unique" "ArrivalTime is identical to DepartureTime"

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

	Type of tests	Description/ examples/
		Test of common European profile rules as to codespaces, names, completeness E.g. A Lien timetable document contains lines, timetables, journeys etc.
5	External consistence	ex: "Reference to Quay %{source_objectid} not found in stop registry"
6	Plausibility checks	ex: "The speed between 2 stops in a connection shall not be too high“ “Two successive stops in a journey pattern shall be neither too far, nor too close" General semantic rules for Transmodel / NeTEx as described in the EPIP profile. E.g. reasonable and consistent values for speed, distance, temporal span across related elements.
7	Temporal validity	In Norway, requirement of time tables and fare data 120 days ahead.
8	Frequency of data delivery	
9	Compliance check deadlines in the delegated act	
10	Completeness of data	
11	Conformity with the national NeTEx and SIRI profiles.	
12	Checking for trust emblem	
13	Check correspondence between data content and data provider.	Do data providers use of expected data elements according the transport mode etc.?
14	Warnings	Different levels of warnings. Useful error messages and warnings.

Data4PT has received funding from the European Union’s DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

CONCLUSIONS

As seen by the different available tools and case studies, data validation is not a common procedure or concept. Depending the level of EU standards implementation in a MS, validation focuses on the compliance to the NeTEx/SIRI format and the consistency against their XML schema or to the corresponding national regulations. However, the need of testing is extended in semantic rules and in data validity and accuracy checking. For the MS where adoption of EU standards is not yet a reality, tools for data translation from the current national format to NeTEx/ SIRI format is also a need. More advanced MS employ tools to enable the creation of standardised data rather focusing on validation per se. The need of a common base of profiles for respective validation was underlined by the MS and the experts participating in the survey. Finally, the provision of the validation tools should encourage further PTAs/PTOs and especially the ones with limited resources and expertise, to test their data and eventually provide accurate, complete and harmonised data to the NAPs.

During project upcoming activities, and in particular during operational use, technical aspects will be taken into consideration to ensure feasibility of tools development based on the feedback of Members States, partners and observers of the project.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136

REFERENCES

[1] NeTEX homepage, http://netex-cen.eu.

[2] VDV (Association of German Transport Companies) website, https://www.vdv.de/siri-overview.aspx.

[3] Nicholas Knowles, Welcome to the NeTEx XML Schema wiki!, https://github.com/nick-knowles/NeTEx/wiki/NeTExSoftware.

[4] CHOUETTE homepage, http://www.chouette.mobi/en/.

[5] Plannerstack website, https://plannerstack.wordpress.com/2014/03/18/bliksem-integration/.

[6] Hogia website, https://www.hogia.com/solutions/public-transport.

[7] Ito World website, https://www.itoworld.com/.

[8] IVU Traffic Technologies website, https://www.ivu.com/products-and-solutions/ivusuite/service-planning/ivupool.html.

[9] MENTZ homepage, https://www.mentz.net/en/.

Data4PT has received funding from the European Union's DG for Mobility and Transport under grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136