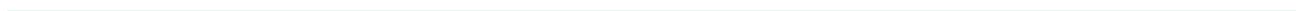




Deliverable D2.4
Testing Procedure Report



Deliverable 2.4 – Testing procedure

Due date of deliverable: 02/07/2020

Actual submission date: 28/02/2022

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

Start date of project: 02/01/2020

Duration: 48 months

Document Control Sheet

Deliverable number:	D2.4
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	30/03/2021	Introduction and Section 1	Anastasia Founta
1.1	10/12/2021	Section 2 and Section 3	Petter Kvarnорfs
1.2	21/1/2022	Section 4	Olle Larson
2.0	15/2/2022	Review and Updates	Anastasia Founta

Legal Disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced consortium members shall have no liability to third parties for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. © 2020 by Data4PT Consortium.

EXECUTIVE SUMMARY

Key activity of DATA4PT project is the development of validation tools for NeTEx and SIRI datasets. As NeTEx and SIRI are the EU standardised formats for public transport data in National Access Points (NAPs), the purpose of validation is to ensure a certain level of quality of the published data. The quality dimension is aligned with the overall objective of the project to enable the implementation of ITS Directive Delegated Regulation EU 2017/1926 and therefore the interoperable exchange of travel and traffic data across Europe.

The first step for the development of validation tools was described in the deliverable D.2.1. and includes the definition of functional requirements and the benchmarking survey of the currently available tools.

This report, deliverable D.2.4, presents the second step for the tools development and consists of the definition of the testing procedure. The testing procedure encompasses the high-level architecture of the test platform, the type of tests and the validation and verification rules. Moreover, in the report it is described the adopted methodology for defining the different aspects of the testing process, taking into consideration the high-level requirements defined in D.2.1 and the overall objectives of the project.

The proposed architecture will be the basis for the development of the test platform. However, it will be furthermore modified and improved based on user's feedback during the pilot implementation phase (Work package 4).

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	MMPİ	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 Infrastrutture e dei Trasporti	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

Abbreviations and Acronyms

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

Table of Contents

Deliverable 2.4 – Testing procedure.....	2
Executive Summary.....	4
Introduction.....	10
1 Testing (validation) procedure definition.....	12
1.1 Validation and overall data flow.....	12
1.2 Tools Benchmarking results.....	15
1.3 Functional requirements assessment.....	15
1.3.1 Requirements on usage of tools.....	20
1.3.2 Integration requirements.....	21
1.3.3 Hosting of tools Requirements.....	21
1.3.4 Nonfunctional requirements.....	21
1.3.5 Validation related requirements.....	22
1.3.6 Secondary functionalities.....	23
1.4 Summary of requirements assessment.....	23
2 Test platform Architecture.....	25
2.1 Introduction and input from the assessment.....	25
2.2 Overview of use cases and workflow.....	25
2.3 Recommended architecture.....	26
2.3.1 Description of the most important blocks.....	27
2.3.2 Execution environment.....	28
2.4 Details of the architecture.....	29
3 Validation Implementation.....	30
3.1 XML Validation.....	30
3.2 Syntax and constraint validation.....	30
3.3 Performance considerations.....	30
3.4 Reporting.....	30
3.5 Rules implemented to verify the architecture.....	31
4 Planning implementation.....	32
4.1 Methodology, execution and review process.....	32
4.2 Intermediate deliverables.....	32
4.3 Roadmap.....	34
4.4 Review process.....	35
Conclusions.....	36
References.....	37

List of Figures

Figure 1. Examples of data flows from Austria and Denmark	13
Figure 2. Examples of data flows from Italy.	13
Figure 3. Examples of data flows from Portugal and Sweden.	14
Figure 4. Chouette tool example.	14

List of Tables

Table 1. High level requirements as defined in D2.1	15
Table 2. Analysis of collected requirements and high-level assessment	16
Table 3. Selected criteria for requirements assessment	18
Table 4. Summary of the assessment of the requirements regarding the usage of the tools	20
Table 5. Summary of the assessment of the integration requirements	21
Table 6. Summary of the assessment of hosting of tools	21
Table 7. Summary of the assessment of hosting of tools	21
Table 8. Summary of the assessment of hosting of tools	22

INTRODUCTION

Based on the findings of the deliverable D.2.1 [1], data validation process for NeTEx and SIRI profiles seems to be an unexplored field while the need of testing public transport (PT) data against semantic, validity and accuracy rules is increased. The provision of validation tools from DATA4PT will encourage PTAs/PTOs, especially the ones with limited resources and expertise, to provide accurate and complete data [1], complied with Delegated Regulation EU 2017/1926.

Nowadays, the validation for public transport data published in NAPs is not a common procedure or concept among the different countries. Depending on the degree of EU standards adoption in a Member State (MS), validation focuses on the compliance either with the NeTEx/SIRI XML schema or with the corresponding national regulations. The MS, where implementation of EU standards is not yet a reality, seek for tools that serve data translation from the current national format to NeTEx/ SIRI format. The more advanced MS on this domain, employ also tools to enable the creation of standardised data and validation is in a way part of it. However, the available opensource tools offer the possibility of data translation and creation, but not their validation against rules beyond XSD schema¹.

In fact, data validation is a dynamic and complex procedure [2]. One of the reasons is that any adaptation of validation rules may influence more than one domain of a data set, or the process itself. Moreover, the testing procedure should be seen as an integral part of the whole data production process. It should also provide clear and coherent allocation of actions and responsibilities for the users, to ensure the highest performance and reduce the probability of mistakes.

DATA4PT adopts the methodology of the life cycle for data validation process, based on European guidelines [2]. This process is described by four (4) phases: design, implementation, execution, and review.

The design phase for the testing procedure is the key topic of this report while implementation, execution and review will be part of next tasks of the project, within WP2 (Development) and WP4 (Operational Use).

In particular, the design of the testing procedure requires setting up the validation rules, defining subprocesses and interfaces, and planning implementation, testing, review, and documentation of the progress. Therefore, several activities take place in this phase such as the assessment of high-level requirements collected during Task 2.1., the study of the data sets to be tested, the assessment of responsibilities and roles in the validation process, the estimation of resources and the definition of a methodology for the feedback collection. The results of those activities will lead to the:

- Definition of the functional architecture of the test platform and the embedded tools. In this context, they are specified the subprocesses, the interfaces, the different actors and their responsibilities in the validation process (who is using the tools, who is providing data, who receives the reporting, who runs the testing and maintains the test platform etc.).
- Determination of satisfactory set of validation rules. This is a key output of the design phase that requires subject matter specialists. The purpose of validation is to ensure a certain, acceptable level of quality of the final data. However, quality has several dimensions: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, coherence, completeness [2]. This report includes the first defined set of validation rules for NeTEx European Passengers

¹ List of available tools supporting implementation is available for SIRI at <https://data4pt.org/wiki/SIRI> and for NeTEx at <https://data4pt.org/wiki/NeTEx>.

Information Profile (EPIP) based on quality and performance criteria. More rules will be assessed and added in the next versions of the tools, based on the results and feedback of pilot implementation.

- Plan implementation, execution, and review phases. The users' feedback that will be received regarding the testing platform and the embedded tools during the pilot, will be leveraged in order to improve the process. It is also important to define a review and implementation process after the end of the project. In this context, some instructions on how to proceed with the next steps for testing procedure improvement are also embraced in this report.

Considering all these aspects, the report is structured as follows:

- *Chapter 1: Testing procedure definition.* In this chapter it is analysed, assessed and consolidated the collected input from the previous tasks of the project (Task 2.1 and T.2.2). It helps the next step which is the test platform architecture.
- *Chapter 2: Test platform architecture.* The chapter includes the presentation of the proposed architecture.
- *Chapter 3: Validation rules.* The validation rules presented at this stage concern NeTEx EPIP. The validation rules it is expected to be enriched after the preparation of EU minimum profile on SIRI, NeTEx accessibility and fares.
- *Chapter 4: Planning implementation.* In this framework, it is described the pilot implementation methodology, the review process and the future potential of the platform.

1 TESTING (VALIDATION) PROCEDURE DEFINITION

Testing procedure, in the case of DATA4PT, refers to the methods and tools used to check public transport datasets that aim to conform with EU standards such as NeTEx and SIRI.

To define the methods and tools for this specific purpose, it was necessary to consolidate and analyse the different inputs collected during the previous phase of the project while introducing assessment criteria.

Specifically, it is considered the following material:

1. Stakeholders' requirements, identified by partners and observers MS.
2. Benchmarking results
3. Experts' work based on received requests, trends and general needs
4. Use cases of data flow (by partners, observers...).

As main objective of DATA4PT project is to assist stakeholders in overcoming their challenges, their perspective is considered as the most critical one for the definition of the testing procedure.

On the other hand, benchmarking results and experts' input is leveraged to assess and prioritise stakeholders' requirements.

Furthermore, some of the requirements concern the integration of validation in the data production pipeline. Therefore, the use cases of data flows are also summarised, highlighting the key points which may affect the DATA4PT test platform architecture.

The conclusions of the assessment, taking also into account the current practices, will define the testing procedure.

1.1 VALIDATION AND OVERALL DATA FLOW

As mentioning in D.2.1 and D.4.1, each country adopts a different approach to collect and make available public transport data in NAP. Usually, different kind of entities and systems are involved and/or at different stages of the process.

The figure 1 illustrates the collected data flows in a diagrammatic way. The interfaces, the stakeholders (entities and systems) involved, their responsibilities and the flow itself are elements that help indicating when, in the process, validation may be occurred, and which entity is responsible for.

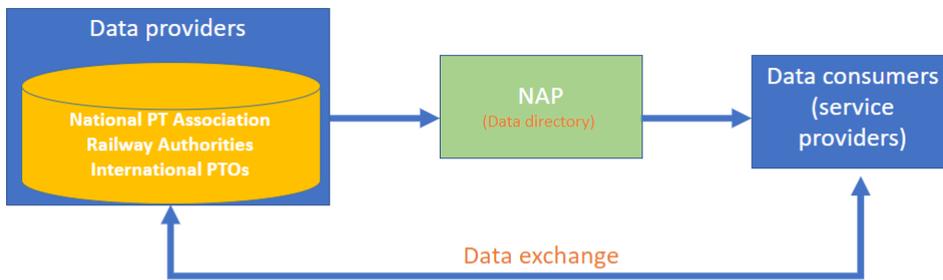
Based on the collected sample of applied or proposed use cases, the data after being shared by the data producers/providers follow different path until being either published in NAP or being used by data consumers (like journey planners). However, data providers (transport authorities, or transport operators, or transport associations) do not necessarily share their data in EU standardised formats (NeTEx/SIRI).

The possibility of validating the datasets in the data providers level, before sharing data to any other entity, is highly relevant to those already using EU standards and it increases the capacity of those not familiar with them. This option might ensure a certain level of quality and interoperability for all data users, regardless the purpose of data sharing (publication in NAP, journey planning, service provision etc.). Enabling the validation by data providers, the implementation of data standards will be further promoted in operational level for both providers and consumers.

The other common brick of data flow is NAP. Validation can be occurred during the collection phase of the datasets. This option will ensure qualitative standardised open data for any potential data requester through NAP. Depending on the current role of NAP in the different countries, validation results can be addressed by NAP entity or by data providers but through NAP.

Finally, validation can be available for all stakeholders and/or systems involved (data providers, NAP, data consumers) supporting all different data flows and data users.

AUSTRIA



DENMARK

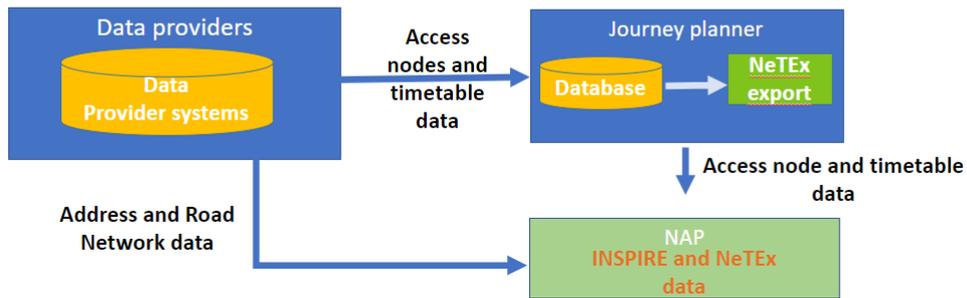
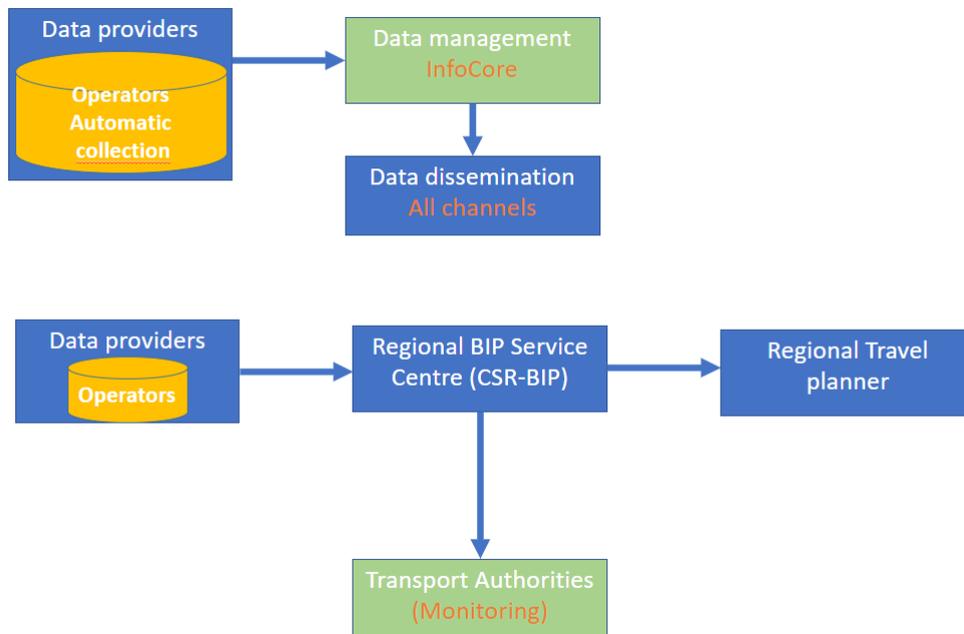


Figure 1. Examples of data flows from Austria and Denmark

ITALY

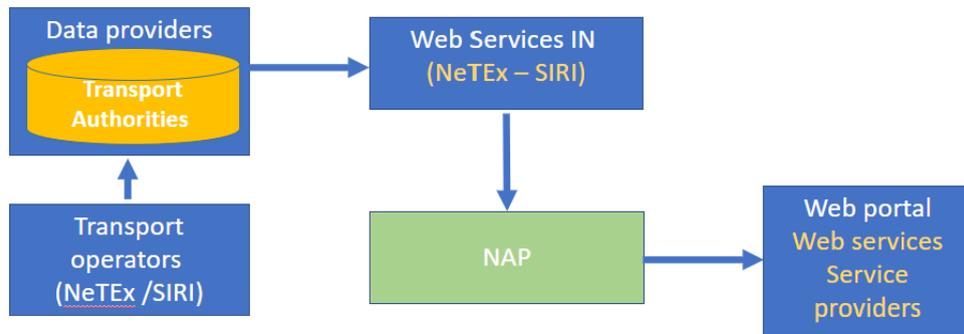
Figure 2.



Examples of data flows

from Italy.

PORTUGAL



SWEDEN



Figure 3. Examples of data flows from Portugal and Sweden.

FRANCE (ENROUTE)

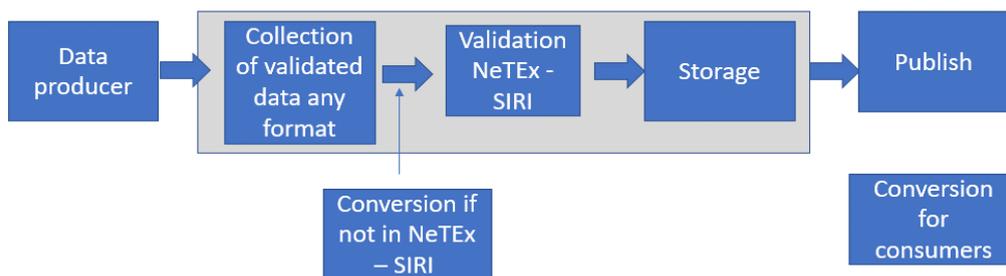


Figure 4. Chouette tool example.

1.2 TOOLS BENCHMARKING RESULTS

Key findings of the benchmarking showed that there are several conversion tools, where most of them implement only a single profile, meaning a Dutch export cannot be used within a German system. In addition, validating conformity as a holistic approach for European wide data exchange has not been addressed by any identified tools. The number of tools implementing NeTEx related programming code are limited and they seem to be built around a single programming language: Java. While this language is popular overall, its equivalent C# lacks any way to automatically load and export files syntactically conforming to the standard. The selected solution needs to pave the road in order to have other programming languages flourish within the ecosystem, allowing developers to create new tools and thereby creating competition in execution on quality, in a similar way that existing standards such as GTFS can be loaded in virtually any programming language and is commonly used for data exchange.

Moreover, processing large amounts of XML can create challenges for validation and parallel execution of code that is able to parse the document. Workarounds have been implemented in Norway, splitting content in several parts. While it becomes easier to consume more information, the inability to use de facto standard XML validation and processing suites creates a new Achilles heel. All validation that have been implemented about 20 years ago has to be done again but is typically tailor made for only this specific use case. Stimulating state-of-the-art operations and having real world datasets may start new research.

1.3 FUNCTIONAL REQUIREMENTS ASSESSMENT

Based on deliverable D.2.1, the stakeholders' requirements concern different parameters of the tools and test platform. Table 1 summarises the identified functional requirements.

Table 1. High level requirements as defined in D2.1²

Parameters	Requirements
Purpose of use/ scope of tools	<ul style="list-style-type: none">• 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.
Users and validation responsibility	<ul style="list-style-type: none">• Data providers and related Stakeholders (e.g. PTOs and PTAs) to use the tools to validate the data they provide.• NAPs nominated bodies to be able to assess provided data.
Administrator/host	Embedded tool on NAPs websites or on neutral body's platform with increased security systems.
Technical characteristics	<ul style="list-style-type: none">• easy to maintain (less legacy)• easy to use• plug in architecture
Periodicity of tests	without limitations

² https://data4pt-project.eu/wp-content/uploads/2021/06/Data4PT_D2.1_Requirements-Report_v2.0_Final.pdf



Parameters	Requirements
Datasets content and format	<p>SIRI and NeTEx included elements</p> <p>Priority to validation of timetables, network, topology elements, fares/pricing</p>
Type of tests	<ul style="list-style-type: none"> • Compliance with NeTEx / SIRI XML-Schema • Other basic XML document compliance checks (non-XSD based) • Internal consistence (ex: "Journey patterns shall be unique") • External consistence (ex: "Reference to Quay %<code>{source_objectid}</code> not found in stop registry") • Plausibility checks (ex: "The speed between 2 stops in a connection shall not be too high") • Temporal validity (In Norway, requirement of time tables and fare data 120 days ahead. • Frequency of data delivery • Compliance check deadlines in the delegated act • Completeness of data • Conformity with the national NeTEx and SIRI profiles. • Checking for trust emblem • Check correspondence between data content and data provider.

After collecting the above requirements, it was obvious the need of prioritisation, in order to create the first version of the tools and test platform, ensuring that corresponds to the users' expectations.

The first step to be able to prioritise was to group the requirements and elaborate high level assessment, considering the results of the tools benchmarking and the collected use cases. Specifically, the requirements were first split into *primary* and *secondary functionalities*. As *secondary functionalities* were considered those that currently address needs of specific countries and do not correspond to most of the use cases described by the MS partners and observers, or because they are not directly linked to data validation scope. Under *primary* requirements, further grouping led into the creation of five (5) sub-categories (Table 2).

Table 2. Analysis of collected requirements and high-level assessment

Category	No	Functionality	Details
Primary functionalities			
Usage	1	Validation for data providers	The tools can be used by data providers to validate the data they provide. Data providers can for example be PTA, PTO, IT Suppliers.
Usage	2	Validation for NAPs	The tools can be used by NAPs nominated bodies to assess provided data.

Category	No	Functionality	Details
Primary functionalities			
Usage	3	Validation for other entities involved in the data flow pipeline	The tools can be used by other entities to access data received or sent between other systems. An example of other entities is national regulation authorities such as Autorité de régulation des transports.
Integration	4	Integration in production pipelines	Easy to integrate with other systems / data flows
Integration	5	Real time tool	That checks every registry when import in NAP for example.
Integration	6	The tool is used only when there is a new implementation	Checks sample of datasets when updates are implemented in data format and content.
Hosting	7	The tool can be hosted by bodies nominated by NAPs	The tools is built so that it is possible to host it in different environments, both in cloud and on premise.
Hosting	8	The tool can be hosted by neutral bodies that is secure	The tool is provided as a secured service (SaaS) by third parties, e.g. ITxPT.
Non functional	9	Easy to maintain	Open source, with a code base that is well structured and documented.
Non functional	10	Easy to use	An interface that is easy to understand and that is well documented.
Non functional	11	Plug in architecture	Modular architecture that can be extended with new validations
Validation	12	Compliance with full NeTEx / SIRI XML-Schema	Validates that the file is compliant with basic XSD-schemas
Validation	13	Validate different NeTEx profiles, e.g. National	Validates that the file is compliant with given national profile XSD-schema.
Validation	14	Validate different SIRI profiles, e.g. National	Validates that the file is compliant with given national profile XSD-schema.
Validation	15	Internal consistency	Validates internal dependencies and rules (ex: "Journey patterns shall be unique").
Validation	16	External consistency	Validates external references (ex: "Reference to Quay %{source_objectid} not found in stop registry")
Validation	17	Plausibility checks	Validates that the information in the file is plausible (ex: "The speed between 2 stops in a connection shall not be too high")
Validation	18	Completeness of data	Validates fields and structures that can not be validated via XSD. E.g.: All fields (required by national profile or by EU minimum profile) have a value
Secondary functionalities			

Category	No	Functionality	Details
Primary functionalities			
Other	1	Temporal validity of the content of the datasets, depending on the national rules that define expiring deadlines (e.g. in Norway datasets for timetables and fares data need to be submitted to NAP 120 days ahead).	
Other	2	Frequency of data delivery by the data provider	
Other	3	Compliance check with deadlines in the delegated act	
Other	4	Checking for trust emblem	
Other	5	Check correspondence between data content and data provider.	
Other	6	Support the transition from other formats to EU ones as an integrated feature	

The next step was the definition of suitable criteria. The criteria are chosen considering different parameters that influence the decision making and concern the current needs of the future users, the available resources, the technical challenges and the long-term impact of the tool. On this basis, four (4) main criteria are selected: the **relevance** to project scope, the **applicability** to national use cases and conditions, the **performance** of the tool in terms of speed of data processing, the **market trigger** as the possibility that the tool may give to the market for further extensions and new features. In addition to those, **feasibility** of the prioritised requirements is also examined by the technical management team of DATA4PT, in order to ensure delivery in the framework of the project.

Considering the four (4) main criteria and a rating scale of three (3) levels: high, medium, low, (Table 3) the assessment of each requirement was based on the declared preferences of both MS representatives of the project and DATA4PT experts. In particular, the *Relevance* of the requirement to the initial validation scope of DATA4PT project and the *Applicability* to national use cases and conditions are criteria that are addressed to both MS partners and DATA4PT experts involved in the project. The *Performance* and *Market trigger* criterion were considered only by the DATA4PT experts as more specific and technical flavoured criteria. To collect the preferences two questionnaires were formulated, one MS specific (Annex I) and one expert specific (Annex II).

This process aims to sort out the requirements that fulfil the criteria and help defining the architecture of the platform and embedded tools. The requirements assessed with the highest scores, will be considered as key features of the testing procedure, while the requirements with lower scores will be evaluated again after the pilot implementation of the test platform.

Table 3. Selected criteria for requirements assessment

Criterion	Scale	Score	Description
Relevance	high	3	The requirement is fully aligned with the validation scope.
	medium	2	The requirement is not directly aligned with validation but addresses the specific objectives of the project (like facilitate the operational use of standards by PTOs/PTAs, enables transition from other standards)

	low	1	The requirement generally supports the overall objective of supporting MS in EU public transport data standards implementation but not the specific objective of the validation
Applicability	high	3	The requirement can be applied to the most use cases, and by the most types of users. For example, to not be dependent on NAP architecture, not require advanced technical skills, not be dependent on the size of the file, etc.
	medium	2	It requires specific skills to use, or it supports specific use cases.
	low	1	Only experts can use and can be applied only to specific Member States.
Performance³	high	3	High performance of the tool is very important to enable the requirement in question.
	medium	2	Medium performance of the tool is enough to reply to the requirement.
	low	1	The requirement does not require any particular performance level.
Market trigger	high	3	The requirement helps in the creation of a platform that inspires, supports, and complements the market products, without competing them.
	Medium	2	The requirement helps in the creation of a platform that inspires and supports market products but overlaps at some point with existing tools.
	Low	1	The requirement leads to the creation of a platform that might compete the market products currently and in the future.
Feasibility	high	3	The requirement can fit to the budget and timeframe of the tools and test platform development activity (T2.2. and T.2.3) and can be supported by a long-term business model .
	medium	2	Most likely it requires additional resources but in the boundaries of the overall budget and timeframe of the project.
	low	1	Most likely it requires additional resources outside the current budget or timeframe.

The results of the assessment are presented in the following paragraphs.

1.3.1 REQUIREMENTS ON USAGE OF TOOLS

Regarding the potential users of the tools, priority is given to data providers (PTA, PTO, IT Suppliers), with the aim to validate their data before providing it to NAP or other systems. Nevertheless, NAPs responsible entities or other entities involved in the data flow pipeline are also highly rated.

Table 4. Summary of the assessment of the requirements regarding the usage of the tools

Category	Requirement	Scores					Total 3&4	Total
		Relevance (Criterion 1)	Applicability (Criterion 2)	Total (1&2)	Performance (Criterion 3)	Market Trigger		

						(Criterion 4)		
Usage	Validation for data providers	22	20	42	5	3	8	50
Usage	Validation for NAPs	20	19	39	5	2	7	46
Usage	Validation for other entities involved in the data flow pipeline	16	16	32	4	1	5	37

1.3.2 INTEGRATION REQUIREMENTS

The possibility of integrating the tool into the data production flow is requested in priority. Less priority is given to create a tool that can process datasets in real time.

Table 5. Summary of the assessment of the integration requirements

Category	Requirement	Scores						
		Relevance (Criterion 1)	Applicability (Criterion 2)	Total (1&2)	Performance (Criterion 3)	Market Trigger (Criterion 4)	Total 3&4	Total
Integration	Integration in production pipelines/ other systems	19	18	37	4	1	5	42
Integration	The tool is used only when there is a new implementation	16	14	30	4	2	6	36
Integration	Real time tool	14	14	28	3	1	4	32

1.3.3 HOSTING OF TOOLS REQUIREMENTS

Regarding the hosting, The possibility of integrating the tool into the data production flow is requested in priority. Less priority is given to create a tool that can process datasets in real time.

Table 6. Summary of the assessment of hosting of tools

Category	Requirement	Scores						
		Relevance (Criterion 1)	Applicability (Criterion 2)	Total (1&2)	Performance (Criterion 3)	Market Trigger (Criterion 4)	Total 3&4	Total
Hosting	The tool can be hosted by bodies nominated by NAPs	17	18	35	4	2	6	41
Hosting	The tool can be hosted by neutral bodies that is secure	17	15	32	4	2	6	38

1.3.4 NONFUNCTIONAL REQUIREMENTS

The possibility of integrating the tool into the data production flow is requested in priority. Less priority is given to create a tool that can process datasets in real time.

Table 7. Summary of the assessment of hosting of tools

Category	Requirement	Scores					Total 3&4	Total
		Relevance (Criterion 1)	Applicability (Criterion 2)	Total (1&2)	Performance (Criterion 3)	Market Trigger (Criterion 4)		
Non-functional	Easy to use (web interface)	21	20	41	2	1	3	44
Non-functional	Plug in architecture (modular architecture)	20	19	39	4	1	5	44
Non-functional	Easy to maintain (code base tool)	17	16	33	2	2	4	37

1.3.5 VALIDATION RELATED REQUIREMENTS

The validation related requirements summarise the main tests that the tool should perform. As expected, the stakeholders prioritise the compliance of the datasets with the full NeTEx and SIRI schema. Second priority with slight difference is granted to the compliance against the EU minimum profiles and/or the national ones. Indeed, considering the technical requests received so far by the experts' team, it is obvious that datasets are usually not checked against the available XSD schemas that would enable a first, basic but important validation step. Even though, commercial tools exist, it is crucial that the checking is done against the correct, valid, and updated XSD. Therefore, those 2 similar functionalities are key also for the open-source validation tool. Highly scored is also the validation for completeness of the datasets (all required elements of a profile to have a value) and the rules that ensure internal consistency (e.g. "Journey patterns shall be unique"). On the other hand, rules that enable checking whether the registered values are plausible and validate external references are less prioritized based on the stakeholders' opinion.

Table 8. Summary of the assessment of hosting of tools

Category	Requirement	Scores					Total 3&4	Total
		Relevance (Criterion 1)	Applicability (Criterion 2)	Total (1&2)	Performance (Criterion 3)	Market Trigger (Criterion 4)		
Validation	Compliance with full NeTEx / SIRI XML-Schema	24	20	44	3	3	6	50
Validation	Validate different NeTEx profiles, e.g. National profile	22	21	43	2	2	4	47
Validation	Completeness of data	19	19	38	2	2	4	42
Validation	Internal consistency	20	17	37	3	3	6	43
Validation	Plausibility checks	16	15	31	3	3	6	37
Validation	External consistency	17	13	30	3	3	6	36

1.3.6 SECONDARY FUNCTIONALITIES

Finally, the stakeholders were asked to prioritise, considering the relevance with the overall objectives of the DATA4PT project, also the requirements that they were categorised as secondary after the first qualitative assessment. As the most important requirement from this category was highlighted the temporal validity of the content of the datasets based on the national rules. The second most relevant requirement was assigned to a functionality that checks the correspondence between data content and data provider.

Category	Requirement	Scores
Other	Temporal validity of the content of the datasets, depending on the national rules that define expiring deadlines (e.g. in Norway datasets for timetables and fares data need to be submitted to NAP 120 days ahead).	22
Other	Check correspondence between data content and data provider.	20
Other	Frequency of data delivery by the data provider	19
Other	Support the transition from other formats to EU ones as an integrated feature	18
Other	Compliance check with deadlines in the delegated act	14
Other	Checking for trust emblem	14

1.4 SUMMARY OF REQUIREMENTS ASSESSMENT

The functional requirements of the test platform and the embedded tools were initially distinguished in primary and secondary requirements, considering the main objective of the project, the results of the existing tools benchmarking and the analysed use cases of data flows. This first categorisation raised the question of further prioritisation of the primary requirements that will help to define the test platform architecture and lead to the first version of the tools. Therefore, a more detailed assessment was conducted, collecting the preferences of both MS and experts of DATA4PT project.

Based on the assessment results, the main scope of the test platform should be to assist the checking of compliance and consistency of the datasets against EU public transport data standards (NeTEx and SIRI) and the defined minimum EU profiles. The stakeholders also expressed their need to be able to check datasets against their (NeTEx and SIRI) national profiles. The tools should be mainly addressed to data providers, but not only. As different countries adopt different approaches of data flows to feed NAP, it is highlighted the need that the tools can be integrated in data production pipeline. In line with this requirement, the tools is preferable to have a modular architecture. However, priority is given to the creation of a web interfaced test platform, that would be easily used by several users, without requiring more sophisticated technical skills. Based on stakeholders' opinion, main hosts of the platform can be considered the NAP responsible entities.

On the side of the secondary requirements, the temporal validity of the content of the datasets based on the national rules and the correspondence between data content and data provider were revealed as more important to be considered as validation rules, at a later release of the tools and the test platform.



2 TEST PLATFORM ARCHITECTURE

2.1 INTRODUCTION AND INPUT FROM THE ASSESSMENT

A well-designed architecture is one of the most critical steps to create a useful tool for validating public transport data. The architecture shall ensure that the tool meets the requirements that have been collected during the assessment of current working methods, and the new requirements that comes with the EU-standards and suggested common processes.

The tool should be easy to use, both as a stand-alone program and integrated into different workflows between systems and organizations. A suggestion is that it should be possible to publish the tool as a cloud-based service in the future, this is something that the architecture must consider from the beginning.

Considering the high-level requirements and the use cases provided by the member states, a two-stage approach for the validation procedure is proposed. Within the first stage, the objective is to ensure the documents meet the normative specifications, while in the second stage the content is validated for consistency. Given the specificity of content constraints the total number of evaluations suggests a method where it is possible to select the type of validations to be used, and incrementally output more results. Errors and warnings should be reported in such a way that it is possible to understand what the errors are and where in the data they occur.

The information to be validated often consists of large amounts of data and it is therefore important that the tool takes performance into account. However, a balance between performance and complexity is needed, otherwise there is a risk that the architecture will be difficult to maintain and develop over time.

2.2 OVERVIEW OF USE CASES AND WORKFLOW

There are several scenarios and use cases for the validation tool. The main purpose is validation of data transferred into the National Access Points (NAP). The validation can also be used in other scenarios; inside an organisation when exchanging information between systems or between organisations (for example timetables from PTA to PTO and planned schedules back to PTA)

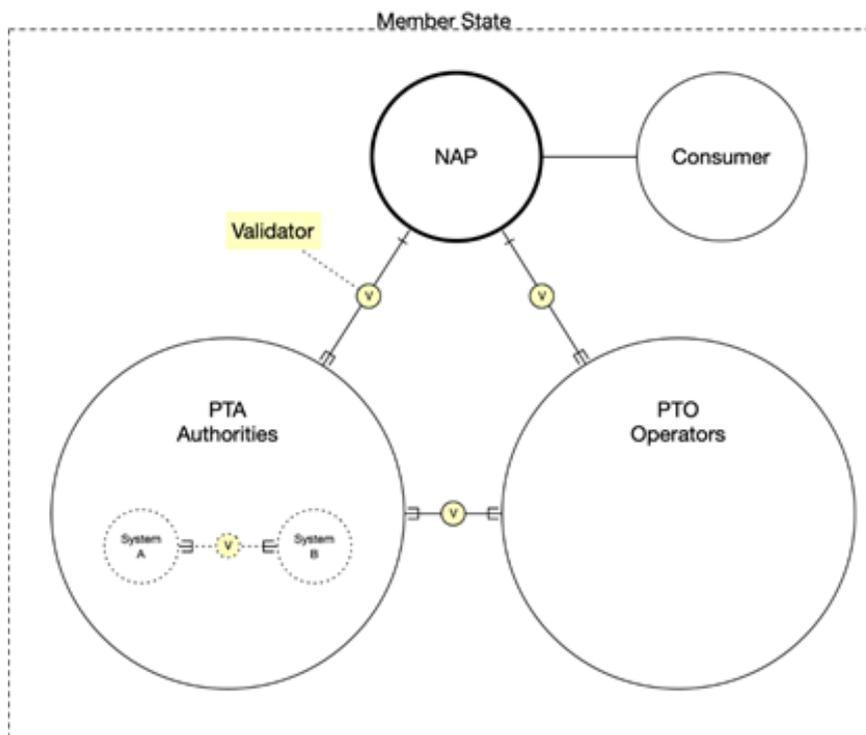


Figure 2: Overview of different use-cases for the validation tool (V).

The workflow used when transferring data contains various aspects of producing, validating and consume information. The data to be validated comes from other systems and tools and is in the form of an XML-file in NeTEx or SIRI format. The details of producing and consuming the data depends on the different systems used and is outside the scope for the validation tool.

Validation is done in several steps which are configurable depending on the data and other circumstances.

1. Validation against standard XSD-schema for NeTEx or SIRI
2. Validation against profile XSD-schema
3. Validation against common validation rules
4. Validation against profile validation rules

When the validation is done the result will contain information about any errors or warnings. To be able to handle different scenarios the result can be configured to be presented in various levels of details. In an automated process the result can be OK or not, and when used manually the result can contain error details and specific row numbers.

2.3 RECOMMENDED ARCHITECTURE

The recommended architecture is built around several functional blocks that together implements the validation workflow. By isolating the functions into blocks the tool will be easy to understand and to maintain over time. Each block can also be tuned for the best performance in relation to their function.

The basic idea is to have blocks that do the processing and validation, these blocks are supported by several configuration blocks that contains XSD-schemas and validation rules.

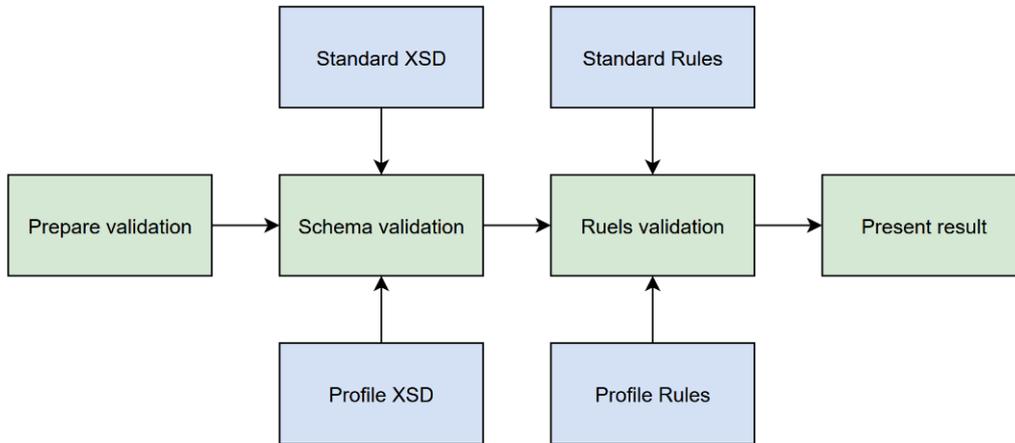


Figure 3. Overview of the blocks in the architecture

2.3.1 DESCRIPTION OF THE MOST IMPORTANT BLOCKS

PREPARE VALIDATION

A configuration is given to the tool as input to start a validation. The configuration contains information about which NeTEx XML-file to validate, which XSD-schema/profile and validation rules to use. The configuration can also contain other information for the tool, e. g. the level of details to report in the result.

SCHEMA VALIDATION

The most important step in the validation is to validate the XML-file against the standard XSD-schema or any specific profiles defined in the configuration, The schema can be the EU minimum profiles such as NeTEx or European Passenger Information Profile (EPIP), NeTEx Accessibility Profile (EPIAP) or any other profile that is available. The XSD-files to be used are provided in configuration blocks which reads the schema to be used. If the data validates OK or if the errors are defined as non-blocking the process continues, otherwise the validation ends, and the result is processed as next step.

RULES VALIDATION

When the validation against the XSD is done and OK the next block to run is validation of specific rules. The rules can be generic rules that is applied to all files, and there are rules specific for different profiles. The rules to be used when validating the file are specified in the configuration given to the validation. The rules themselves are defined as separate scripts. A set of basic script are provided with the tool, more scripts can be written and added if needed.

PRESENT RESULT

The output from the stages is collected during the validation. The result is summarised and depending on the configuration, various levels of details is sent as output from the validation process. The result specifies the used validation rules, if there are any errors, and the exact line numbers in the validated input where the error occurred.

2.3.2 Execution environment

To be able to use the tool under different circumstances the architecture suggests that the main validation is complemented with some more blocks. First a user interface via a command line tool that can be used to start the tool from a terminal. Other types of interfaces can be added later, for example REST via https. The different interfaces are then calling an internal API that provides validation of the input parameters, collection of configuration parameters and then starts the main validation. After the validation of the datafiles the result is sent back via the API and is converted to a format suitable for the calling interface.

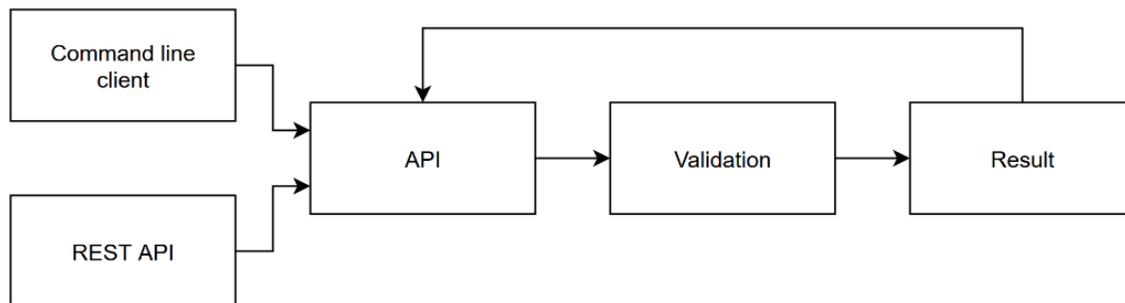


Figure 4. Validation tool with interfaces for Command Line and REST API

2.4 DETAILS OF THE ARCHITECTURE

The suggested implementation of the validation tool is aimed to take both performance, extensibility, and complexity in consideration. To get a good performance a compiled language such as Go is recommended and to enable extensions by end users, validation rules can be written in JavaScript.

Go is a compiled language and is syntactically like C, but with more safety features and structural typing. Go have high performance support for multiprocessing. There are also a well-maintained ecosystem of frameworks and libraries for Go. One important aspect is the support for XML and XSD validation.

JavaScript were originally used only in web browsers but are now a core component in different servers and applications. The validation extensions written in JavaScript are compiled “just-in-time” to give better performance.

To make it easy to set up the runtime environment for the tool, and to support the possibility to later host the tool as a cloud service, a container service such as Docker is suggested. Installation of the tool and all dependencies can be done via provided scripts that downloads the correct base image and loads configuration and datafiles to be validated.

3 VALIDATION IMPLEMENTATION

3.1 XML VALIDATION

For syntactic validation we aim for an industry standard or equivalent open-source XML validator, supporting the validation of XML documents using the NeTeX XML Schema as a base. A change to the XML Schema should not require any additional development effort. The first implementation of the tool uses the open-source library libXML and x-path. One goal is to be able to do as much as possible of the validation without generating a large object-model in the tool, and only do that for smaller object-trees if necessary for checking certain rules.

The tool is independent of the XSD-schema and can be used for different schemas without any changes. That makes it possible to use the same tool to validate Standard NeTeX, EU Minimum NeTeX, EPIP and different national profiles.

3.2 SYNTAX AND CONSTRAINT VALIDATION

The rules written in code will provide a flexible way to add various validations. That can be both simple validations that checks individual data elements exists and that they contain acceptable values. More complex rules can check that information in a list of elements forms have a consistent order. Rules can also check that references between elements in the file, or different files, exists and matches.

Rules are written in JavaScript, other languages are also supported, and the code is compiled before execution to get better performance.

The configuration contains which rules to use in a specific scenario. Some rules will apply to all NeTeX files and others are to be used for specific profiles or situations. The aim is to provide a means to collect rules and collaborate in development and sharing.

3.3 PERFORMANCE CONSIDERATIONS

Performance is a balance between speed and complexity, the validation rules are means to get reliable performance and at the same time have a maintainable solution that is not too demanding on the execution environment. The base of the tool is written in Go and is compiled into an executable file. The rules including the framework for XSD-validation uses compiled JavaScript in combination with different support libraries. This solution gives that the scripts are easy to maintain and develop, at the same time they can run in parallel to use the performance of multi core environments.

The tool is opted to be useful for most use cases and input files, that will give most value for most of the users. If the tool instead is opted to have superior performance for every situation it needs to be more complex, will be less portable and will put higher demands on the execution environment.

3.4 REPORTING

Error reporting will include the original line numbers, element name and failing rule. For single line files, XML formatting may be applied when the user requests so, which should lead to a new output XML document, and an error report matching the new document.

3.5 RULES IMPLEMENTED TO VERIFY THE ARCHITECTURE

Various rules have been implemented to test the architecture both considering implementation, correctness, and performance. The implemented rules exemplify and verify key validation methods to use later when more rules are:

1. Check for correct Locale and Currency properties

Implemented to verify that properties in the file header can be read

2. Check for valid date range of the complete file

Implemented to verify that properties inside another element can be read

3. Check that every Stop has an arrival and departure time

Implemented to verify that a list of elements can be parsed and verified. This rule also verifies the possibility to validate information that is stored in separate files, e.g., lines.xml, stops.xml and shared_data.xml.

4. Check that times do not decrease during a journey

Implemented to verify that elements in a list can be verified as dependent on each other

5. Check that times are handled correct between days

Implemented to verify that rules can span over dated elements

4 PLANNING IMPLEMENTATION

The long-term goal with this project is to create an intuitive tool that will be used in and between EU countries to validate the mobility data required by MMTIS 1927/2017. However, making the tool useful and used by all (or at least most of) the member states require us to plan implementation thoroughly.

Before this tool create a significant added value for the end users, few sub-deliverables need to happen. Some of key parts are:

- Rules engine / management system to help facilitate shared and specific country rules and profiles
- Structured and easy to follow program code
- Script API and instructions on how to create new rules and perhaps use the tool
- A gatekeeper with established governance is important for this to scale well (because of the nature of many contributors that do not know each other)
- Web service and user interface
- Proper repo structure
- Answer the question who is responsible for the tool once this project is considered delivered?

Moreover, main part of the implementation is the pilot use and the review phase that will enable the final deliverable to meet expectations.

Therefore, this section describes how we plan to implement the pilot version, the method used and a potential future for the tool.

4.1 METHODOLOGY, EXECUTION AND REVIEW PROCESS

The methodology we use for this project is agile and in close collaboration with all stakeholders. Work is done in short cycles, so called sprints, and the development team meets almost daily to discuss current tasks and roadblocks.

The Data4PT steering group meets the development team on a weekly basis to discuss progress and prioritization – more often if necessary.

Meetings with member states and other experts are facilitated once a month to get instant feedback and to evaluate the tool, specific code or to discuss something related. The group of experts consists of representatives from different countries and areas.

Review of work and deliverables is an ongoing effort and part of the agile process. The principle is to release updates and deliverables often so that we can get feedback fast. Finding and highlighting potential failures can catapult learning and optimise solutions much quicker than the old waterfall model. The concept of fail fast is strongly connected to the agile methodology we use.

4.2 INTERMEDIATE DELIVERABLES

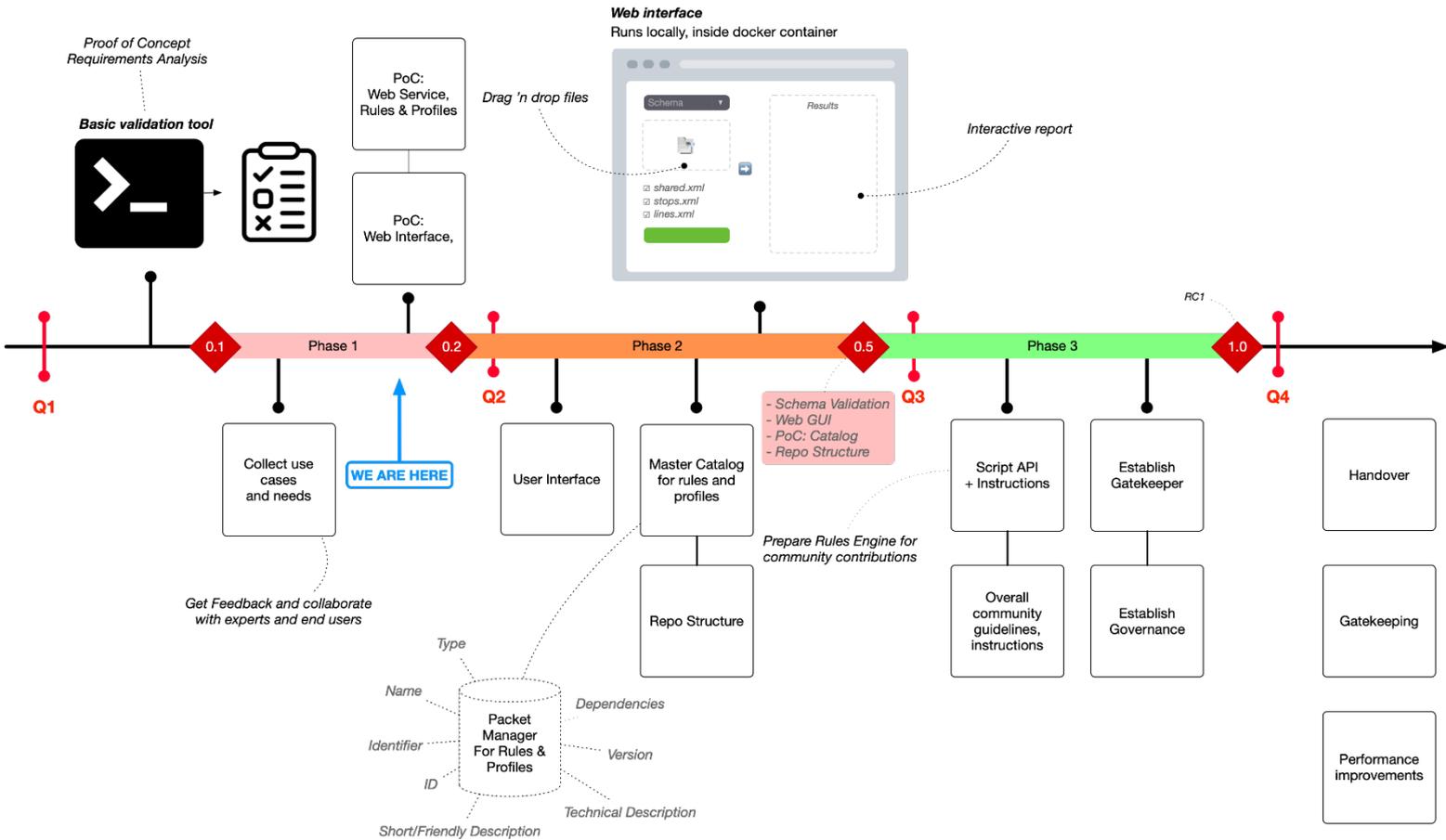
The two important milestones of this work are a) the first release of the tool (Version 1.0.) in June 2022, and the second and final release (Version 2.0) in June 2023. However, intermediate deliverables have been defined to guarantee the efficiency of the tool already from the first release, including the proof of concept on different levels.

- Proof of Concept – Pilot version – Terminal, docker
- Proof of Concept – Web interface, web service

- Proof of Concept – Rules and profiles management
- Script API and instructions
- Version 1.0 – Release Candidate 1
- Plan for handover, governance, and gatekeeping

4.3 ROADMAP

The roadmap shows current progress as of 2022-02-23 and a rough estimate on where we think we are at the end of 2022, assuming that we are allowed to continue making progress.



Proof of Concept – Pilot version – Terminal, docker
 Basic terminal version to make sure that we can verify schemas and a few simple validation rules.

Proof of Concept – Web interface, web service
 Early version of web service and web interface to show experts and end users to make sure we are on the right track.

Proof of Concept – Rules and profiles management
 The ability to scale this tool is crucial. We need to make sure the tool can scale to handle most of not all European countries, profiles and rules (both shared rules and those that are unique to specific countries and profile).

Script API and instructions
 The API will allow for member states and end users to create their own validation rules. The API gives controlled access to certain functions (e.g., xpath validation, xml related functions, javascript) which allows

for both future compatibility and increased security. Instructions and API definitions are an important part of this.

Version 1.0 – Release Candidate 1

Release Candidate 1 is the first attempt of the final public release for review by the experts. Version 1.0 will contain all the features mentioned above and a rules engine framework and management system that will allow for member states to start writing their own validation rules. Decision point for “when to go public”.

Plan for handover, governance, and gatekeeping

A gatekeeper, responsible for governance and the tool development, will be required in order to scale national profiles and rules to the point that member states can learn from and re-use existing validation rules.

4.4 REVIEW PROCESS

The review process, once released to the public, need to adhere to ideas, issues and suggestions from the wider audience. We suggest using specific surveys to certain groups or users (when needed) and Github which is great for managing questions and issues related to the continuous development of the tool.

In particular for the review process, as it is one of the core activities of the project, a [dedicated questionnaire](#) (Annex I) is created to collect the high-level feedback from the users. The questions aim to trigger users' evaluation from different perspectives:

- Installation: Getting the tool up and running was easy
- Configuration: Changing settings and configuring the tool was intuitive
- Overall usability
- Use cases: what use cases the tool was tested for.

The review period is also split in different phases: first round of review encompasses the feedback from the experts' team. Besides the off-line interaction, a dedicated workshop was organized to exchange opinions. In the second round of review, users from the pilot sites of the project are envisaged to test and provide feedback. These two rounds of feedback will result to the first public release of the open-source tool. Finally, the third round of review includes the wide use of the tool by all EU stakeholders and the collection of their feedback. This review phase will result to the final release of the tool, at the second quarter of 2023.

CONCLUSIONS

The long-term goal with this project is to create an intuitive tool that will be used in and between EU countries to validate the standardized mobility data required by MMTIS 1927/2017. Focusing on stakeholders needs, priority has been given to the creation of a validator that can first of all check compliance against EU standard formats such as NeTEx and SIRI and can be configurable enough to meet national requirements in terms of profiles. Nevertheless, the ultimate goal exceeds beyond the XSD validation but targets the integration of rules that will ensure the accuracy and plausibility of the data, ensuring their quality and usability to transport systems. To achieve this goal, an agile architecture and implementation process is applied that eventually will lead to the first release of tools with great potential for further development and configuration by the users to meet their specific needs. This factor is another key element that brings added value to the existing tooling. In addition to that, the choice of a web-based interface, where users without advanced or medium IT skills can use it, is also very critical for the acceleration of EU standards implementation. Finally, as already observed, the key challenge for EU standards to be implemented is not the status of standards per se, as they are already complete and advanced enough to cover most of the current needs of mobility systems, but the lack of easy-to-use tools that can facilitate their use for any type of users. In this respect, the impact of the DATA4PT validator on the acceleration of EU standards adoption is expected to be significant.

REFERENCES

- [1] Founta A., (2020). *D2.1 Requirements Report*, DATA4PT Deliverable.
- [2] Di Zio, M. et al. (2016), *Methodology for data validation 1.0*. Accessed 16 June 2021, https://ec.europa.eu/eurostat/cros/system/files/methodology_for_data_validation_v1.0_rev-2016-06_final.pdf.

ANNEX I: QUESTIONNAIRE FOR MS



DATA4PT: Tools requirements assessment

During the previous phase of the project, the stakeholders' involved identify requirements for different parameters regarding the validation tools and test platform. In order to consolidate these requirements and to prioritise them, we introduce this survey. **The objective of the survey is to assess the collected requirements against some qualitative criteria such as the relevance to the scope of DATA4PT project and the applicability to national contexts.**

How to rate the requirements

Assign "high", "medium" or "low" to **evaluate the relevance of each requirement depending on how much you believe each requirement is fully aligned with the validation scope of DATA4PT and support NAPs in publishing high quality data.**

Assign "high", "medium" or "low" to evaluate the applicability of each requirement based on **how much you believe each requirement can be applied to most of the use cases you need to support in your national implementation of NAP and does not require advanced technical skills.**

1. Usage of the tools

	Relevance to project scope	Applicability to national conditions and capacities
Data providers can use the tools to validate the data they provide. Data providers can for example be PTA, PTO, IT Suppliers.	Choose an item.	Choose an item.
The tools can be used by NAPs nominated bodies to assess the provided data.	Choose an item.	Choose an item.
The tools can be used by other entities like data consumers or entities in charge of controlling the implementation of national regulations based on EU directive. E.g., Autorité de régulation des transports in France)	Choose an item.	Choose an item.

2. Integration of tools

	Relevance to project scope	Applicability to national conditions and capacities
The tool(s) are easy to be integrated into other systems / data flows.	Choose an item.	Choose an item.
The tool(s) support real time checking, for example for every registry imported in NAP.	Choose an item.	Choose an item.
The tool is used only when there is a new implementation. For example, check sample of datasets when updates are implemented in XSDs.	Choose an item.	Choose an item.

DATA4PT: Tools requirements assessment

3. Hosting of tools

Requirements	Relevance to project scope	Applicability to national conditions and capacities
The tool can be hosted by bodies nominated by NAPs. The tools are built so that it is possible to host it in different environments, both in cloud and on premise.	Choose an item.	Choose an item.
The tool is provided as a secured service (SaaS) by third parties, eg. ITxPT.	Choose an item.	Choose an item.

4. Non-functional requirements

Requirements	Relevance to project scope	Applicability to national conditions and capacities
Open-source tools, with a code base that is well structured and documented.	Choose an item.	Choose an item.
An interface that is easy to understand and that is well documented.	Choose an item.	Choose an item.
Modular architecture that can be extended with new validations (plug n play architecture).	Choose an item.	Choose an item.

5. Validation related

Requirements	Relevance to project scope	Applicability to national conditions and capacities
The tool validates that the dataset is compliant with basic XSD-schemas (compliance with full NeTEx/SIRI XSDs).	Choose an item.	Choose an item.
The tool validates that the dataset is compliant with given NeTEx national profile XSD-schema.	Choose an item.	Choose an item.
The tool validates that the dataset is compliant with given SIRI national profile XSD-schema.	Choose an item.	Choose an item.
The tool validates the internal dependencies and rules (ex: "Journey patterns shall be unique").	Choose an item.	Choose an item.
The tool validates external references (ex: "Reference to Quay %{source_objectid} not found in stop registry").	Choose an item.	Choose an item.
The tool validates that the information in the file is plausible (ex: "The speed between 2 stops in a connection shall not be too high").	Choose an item.	Choose an item.
The tool validates fields and structures that cannot be validated via XSD. E.g.: All fields (required by national profile or by EU minimum profile) have a value.	Choose an item.	Choose an item.

DATA4PT: Tools requirements assessment

Secondary functionalities

During the collection of requirements, some functionalities have been identified by some MS that could be **considered at a later stage in the preparation** of the tools because their applicability it is lower (currently they address needs of specific countries) or because it is not linked directly to data validation scope. For these functionalities you are asked to evaluate them only with regard to your perception on how relevant are to the overall scope of the project.

	Relevance to project scope
Tool that checks the temporal validity of the content of the datasets, depending on the national rules that define expiring deadlines (e.g., in Norway datasets for timetables and fares data need to be submitted to NAP 120 days ahead).	Choose an item.
Tools that check the frequency a data delivery by the data providers.	Choose an item.
Tools that check compliance with the deadlines in the delegated act.	Choose an item.
Tools that check for trust emblem.	Choose an item.
Tools that check correspondence between data content and data provider.	Choose an item.
Support the transition from other formats to EU ones as an integrated feature.	Choose an item.

Name:

Email:

Organisation:

Country

ANNEX II : QUESTIONNAIRE FOR EXPERTS

DATA4PT: Tools requirements assessment

*During the previous phase of the project, the stakeholders' involved identify requirements for different parameters regarding the validation tools and test platform. In order to consolidate these requirements and to prioritise them, we introduce this survey. **The objective of the survey is to assess the collected requirements against some qualitative criteria such as the relevance to the scope of DATA4PT project and the applicability to national contexts.***

How to rate the requirements

Relevance to project scope: Assign “high”, “medium” or “low” to **evaluate the relevance of each requirement depending on how much you believe each requirement is fully aligned with the validation scope of DATA4PT and support NAPs in publishing high quality data.**

Applicability to national use cases/conditions: Assign “high”, “medium” or “low” to evaluate the applicability of each requirement based on **how much you believe each requirement can be applied to most of the use cases you need to support in your national implementation of NAP and does not require advanced technical skills.**

Performance of the tool(s): Assign “high”, “medium” or “low” to **evaluate how important is the performance for each requirement based on your perception/experience.** With performance we consider the **speed of processing the data until the final export/report.**

Market trigger: Assign “high”, “medium” or “low” to **evaluate each requirement regarding how much it contributes to the creation of a platform that inspires, supports, and complements the market products, without competing them.**

Name:

Email:

Organisation:

Country:

DATA4PT: Tools requirements assessment

1. Usage of the tools

	Relevance to project scope	Applicability to national conditions and capacities	Performance of the tool	Market trigger
Data providers can use the tools to validate the data they provide. Data providers can for example be PTA, PTO, IT Suppliers.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tools can be used by NAPs nominated bodies to assess the provided data.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tools can be used by other entities like data consumers or entities in charge of controlling the implementation of national regulations based on EU directive. E.g., Autorité de régulation des transports in France)	Choose an item.	Choose an item.	Choose an item.	Choose an item.

2. Integration of tools

	Relevance to project scope	Applicability to national conditions and capacities	Performance of the tool	Market trigger
The tool(s) are easy to be integrated into other systems / data flows.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool(s) support real time checking, for example for every registry imported in NAP.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool is used only when there is a new implementation. For example, check sample of datasets when updates are implemented in XSDs.	Choose an item.	Choose an item.	Choose an item.	Choose an item.

DATA4PT: Tools requirements assessment

3. Hosting of tools

Requirements	Relevance to project scope	Applicability to national conditions and capacities	Performance of the tool	Market trigger
The tool can be hosted by bodies nominated by NAPs. The tools are built so that it is possible to host it in different environments, both in cloud and on premise.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool is provided as a secured service (SaaS) by third parties, eg. ITxPT.	Choose an item.	Choose an item.	Choose an item.	Choose an item.

4. Non-functional requirements

Requirements	Relevance to project scope	Applicability to national conditions and capacities	Performance of the tool	Market trigger
Open-source tools, with a code base that is well structured and documented.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
An interface that is easy to understand and that is well documented.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
Modular architecture that can be extended with new validations (plug n play architecture).	Choose an item.	Choose an item.	Choose an item.	Choose an item.

5. Validation related

DATA4PT: Tools requirements assessment

Requirements	Relevance to project scope	Applicability to national conditions and capacities	Performance of the tool	Market trigger
The tool validates that the dataset is compliant with basic XSD-schemas (compliance with full NeTEx/SIRI XSDs).	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates that the dataset is compliant with given NeTEx national profile XSD-schema.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates that the dataset is compliant with given SIRI national profile XSD-schema.	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates the internal dependencies and rules (ex: "Journey patterns shall be unique").	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates external references (ex: "Reference to Quay %{source_objectid} not found in stop registry").	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates that the information in the file is plausible (ex: "The speed between 2 stops in a connection shall not be too high").	Choose an item.	Choose an item.	Choose an item.	Choose an item.
The tool validates fields and structures that cannot be validated via XSD. E.g.: All fields (required by national profile or by EU minimum profile) have a value.	Choose an item.	Choose an item.	Choose an item.	Choose an item.

DATA4PT: Tools requirements assessment

Secondary functionalities

During the collection of requirements, some functionalities have been identified by some MS that could be **considered at a later stage in the preparation** of the tools because their applicability it is lower (currently they address needs of specific countries) or because it is not linked directly to data validation scope. For these functionalities you are asked to evaluate them only regarding your perception on how relevant are to the overall scope of the project.

Requirements	Relevance to project scope
Tool that checks the temporal validity of the content of the datasets, depending on the national rules that define expiring deadlines (e.g., in Norway datasets for timetables and fares data need to be submitted to NAP 120 days ahead).	Choose an item.
Tools that check the frequency a data delivery by the data providers.	Choose an item.
Tools that check compliance with the deadlines in the delegated act.	Choose an item.
Tools that check for trust emblem.	Choose an item.
Tools that check correspondence between data content and data provider.	Choose an item.
Support the transition from other formats to EU ones as an integrated feature.	Choose an item.