dataspt

Deliverable D2.3 Validation tools release 2









Version 1.4- 15/01/2024

Deliverable 2.3 – Validation Tools Release 2

Due date of deliverable: 30 June 2023 Actual submission date: 15 January 2024

	Dissemination Level	
PU	Public	х
СО	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: 54 months





Version 1.4- 15/01/2024

Document Control Sheet

Deliverable number:	D2.3
Deliverable responsible:	ITxPT
Work package:	WP2
Main editor:	Anastasia Founta

Editor name	Organisation
Theobald Nutte	ITxPT
Eliot Terrier	ITxPT
Petter Kvarnfors	ITxPT
Jesper Johansson Törnros	ITxPT

		Document Revision History	1
		Modifications Introduced	
Version	Date	Reason	Editor
1.0	20/6/2023	Structure – main input	Anastasia Founta
1.1	7/2023	Technical input	Petter Kvarnfors Jesper Johansson Törnros
1.2	9/2023	Consolidation of input and editing	Theobald Nutte Eliot Terrier
1.3	10/2023	Consolidation of input and editing	Anastasia Founta Theobald Nutte Eliot Terrier
1.4	01/2024	Review	Efe Usanmaz

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

A key activity of DATA4PT project is the development of validation tool 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 the published data is of certain level of quality. Ensuring data quality is important for the overall objective of the project, which is to enable the implementation of ITS Directive Delegated Regulation (DR) 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 <u>D.2.1. Requirements</u> <u>Report</u> and includes the definition of functional requirements and a benchmarking survey of the currently available tools. In the deliverable <u>D.2.4 "Testing Procedure Report"</u>, it was defined the architecture of the tool and the testing procedure. Considering the priority in static data according to MMTIS DR EU 2017/1926, the architecture focused on the validation of NeTEx datasets, as first step. Following this methodology, the development and first release of the validation tool addressing NeTEx datasets was made available in June 2022.

After one year of pilot implementation of the tool, where several users shared their feedback and many updates have been applied, the second, and the final release of the tool in the framework of DATA4PT was delivered in September 2023.

This report constitutes a short summary of the work done during the pilot implementation of the tool but also a manual for the current and future users, accompanying the tool technical artefacts¹.

¹ All technical artefacts and source codes are available in <u>https://github.com/ITxPT/DATA4PTTools</u>. The manual is also incorporated in relevant link.





Version 1.4- 15/01/2024

List of partners

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







Version 1.4- 15/01/2024

Abbreviations and Acronyms

API	Application Programming Interface
AVMS	Automatic Vehicle Monitoring Systems
MMTIS	Multimodal Travel Information Services
EPIP	European Passenger Information Profile
GUI	Graphical user interface
MS	Members States
NAP	National Access Point
NeTEx	Network Timetable Exchange
РТО	Public transport Operators
РТА	Public Transport Authorities
SaaS	Software as a Service
SIRI	Service interface for real-time information
TRANSMODEL	Public Transport Reference Data Model





Version 1.4- 15/01/2024

TABLE OF CONTENTS

Del	livera	ble 2.3 – Validation Tools Release 2	2
Exe	ecutiv	e Summary	4
Tab	ole of	Contents	7
Intr	oduct	tion	11
1	Gre	enlight - The Data4PT Validation tool	12
1	.1	Core part	12
	1.1.1	Web Interface	13
1	.2	Standard XML Library - libXML	13
1	.3	Command Line Interface - CLI	13
1	.4	Individual validation rules - Scripts	13
1	.5	System requirements	13
1	.6	Local installation	13
1	.7	Releases history	15
2	Vali	dation rules	18
2	.1	Rules based on NeTEx XML schema	18
	2.1.1	Completeness checks	19
	2.1.2	2 Improving performance by separating integrity cross-checks from XML schema validation	19
2	.2	Rules beyond NeTEx XML schema	22
2	.3	Build your own rules	28
3	Sou	rce codes inventory	29
4	Mar	nual for Web Interface	32
4	.1	Navigation	32
4	.2	Configuration	33
4	.3	Packages	33
4	.4	Custom configuration	34
4	.5	Select files to validate	34
4	.6	Validation result	36
4	.7	Downloading the result	37
4	.8	Previous validation	38
4	.9	Technical error messages	39
5	Mar	nual for command line interface	40
5	.1	Getting help	40
5	.2	Server command	40
5	.3	Validate command	41
5	.4	NeTEx profile	41
5	.5	Rules	42
5	.6	Providing files	42
C	Data4F	PT has received funding from the European Union's DG for Mobility and Transport	
ι	inder	grant agreement No MOVE/B4/SUB/2019-104/CEF/PSA/SI2.821136 Page 7 c	JT 52





Version 1.4- 15/01/2024

5.7	Output	43
5.8	Completion command	43
6 Buil	ilding from source	44
6.1	Prerequisites	44
6.2	Getting started	44
6.3	Building and running the CLI	44
6.4	Building the Web GUI	45
Conclus	- sions	





List of Figures

Figure 1: Main components of Greenlight Validator.	12
Figure 2: Command of terminal window to call the latest version of the tool.	14
Figure 3 Command to use the web interface to verify that the docker installation works properly.	14
Figure 4 Using the Docker Desktop to run the web interface.	15
Figure 5 Where to upload a custom profile	19
Figure 6 Package selection for fast integrity checks using separated scripts.	20
Figure 7 Custom configuration for <i>fast</i> integrity checks using separated scripts.	21
Figure 8 Custom configuration for fast integrity with customs rules on CLI	22
Figure 9 Data Quality dimensions and correspondance of XML schema based validation rules.	23
Figure 10 Validation tool Greenlight website	32
Figure 11 All steps for a validation	32
Figure 12 Configuration page	33
Figure 13 Premade package available on the website	33
Figure 14 Custom configuration page	34
Figure 15 Setting up parameter of Stop place quay distance rule	34
Figure 16 Upload validation file page	35
Figure 17 Validation of the upload files	35
Figure 18 Running validation page	36
Figure 19 Result page	36
Figure 20 Details of the result on the website	37
Figure 21 Downloading result as json or csv	37
Figure 22 Example of data saved as json	38
Figure 23 Validation history page	38
Figure 24 Example of Error messages	39
Figure 25 Help command on the Command Line Interface	40
Figure 26 Command to build the web interface	40
Figure 27 Example of output from a validation done in the CLI	41
Figure 28 Command line example of rule selection	42
Figure 29 Web interface built from source	46
Figure 30 API status on web interface built from source	47







Version 1.4- 15/01/2024

List of Tables

Table 1 Main features incorporated in the first public release V.0.3.4 in June 2022.	15
Table 2 Main features incorporated in the second public release V.1.0.7 in August 2023.	16
Table 3 Data Quality dimensions and correspondance of XML schema based validation rules.	18
Table 4 Examples of rules based on XML schema.	18
Table 7 Examples of rules that cannot express in XML.	23
Table 8 Summary of all identified rules – scripted or not.	24
Table 9 Currently integrated rules	25
Table 10 Source codes Inventory.	30





INTRODUCTION

Data quality is one of the key factors for the provision of successful services in all domains, including mobility. Therefore, it is also an important parameter in MMTIS Delegated Regulation EU 2017/1926 in order to ensure that the openly available data through the National Access Points are suitable to be re-used, to feed multimodal mobility applications, and to improve mobility services across Europe addressing the changing passenger behaviours after the global pandemic².

In the framework of DATA4PT a validation tool (<u>Greenlight</u>) has been developed, dedicated to check and validate datasets according to the EU CEN technical standards NeTEx CEN/TS 16614.

The tool has been developed following three main steps:

- 1) The definition of requirements regarding functionalities and use cases the tool aims to support. The requirements have been collected by Member States and their NAP operators which participated to the research. The requirements have been also designed after investigating the existing tools, evaluating gaps and important functionalities. Based on the findings of this study, an architecture has been designed, giving priority to static data, aligning also with DR priorities. This step is described in <u>D.2.1. "Requirements"</u> and <u>D.2.4 "Test Platform"</u>.
- The development of version V.0. of the tool. This version has been tested by a sample of users, which consists of the NAP operators of the 9 Member States, partners of the project, leading to a first public release for wider utilisation in June 2022 (<u>V0.3.4</u>).
- 3) The collection of feedback from the users and the final public release (<u>V.1.7</u>) which contains the technical artefacts, the source codes, the web interface and the manual ("<u>readme</u>"). This report summarises the main characteristics of the tool and provide guidelines of how to use the tool. Therefore, it stands also as manual.

² Future of Mobility post-COVID, UITP & Arthur D. Little, July 2020 https://cms.uitp.org/wp/wp-content/uploads/2020/10/2020-ADL-FoM-Lab-UITP_Future-of-Mobility-post-COVID-study.pdf







1 GREENLIGHT - THE DATA4PT VALIDATION TOOL

The tool consists of several components, each with a different responsibility. This will ensure that the tool is modular and that each component is easy to understand and maintain.



Figure 1: Main components of Greenlight Validator.

Its components are:

- The core part
- The standard XML library (libXML)
- The Command Line Interface (CLI)
- The Individual validation rules (scripts)

To facilitate the users, information around system requirements and installation process of the command line Interface is also provided.

The full package of the tool is available in <u>GitHub/ ITxPT/DATA4PTTools</u>. As complementary documentation and material, the users are recommended to consult <u>NeTEx-CEN/NeTEx</u>.

1.1 Core part

The *core* is the main component of the tool, it reads the configuration, handles file imports, calls the validation scripts, and summarizes the result. The *core* provides an API that other components use to control the validation or to get access to shared functions, e.g., in libXML. The API also makes it possible to extend the tool with different front ends, as the CLI and Web Interface.





Version 1.4- 15/01/2024

Deliverable 2.3 – Validation Tools Release 2

1.1.1 WEB INTERFACE

The web interface provides an easy-to-use interface via the web browser. The web interface makes the tool easier to use for the occasional user or for just testing small files. After loading the web page, you can select the NeTEx profile to use, select one or more validation rules and then run the validation. After completion you get the result on the web page but can also download it to a file. More details will be shared in 4 Manual for Web Interface

1.2 STANDARD XML LIBRARY - LIBXML

An open source, standard library integrated into the tool. It is libXML that does all the XSD and XML validation. It is called from the scripts via the API in the Core component.

1.3 COMMAND LINE INTERFACE - CLI

The Command Line Interface is used in a terminal or integrated in an import/export pipeline. Parameters are used to configure the tool and to specify the files to be validated. The result can be read in the terminal or saved as a file.

1.4 INDIVIDUAL VALIDATION RULES - SCRIPTS

Individual validation rules implemented as scripts. The scripts are written in JavaScript that is easy to start with and JavaScript is also well documented. The validation scripts are small programs that each implements one or more validation rules. The scrips provided with the tool implements one rule per script to make it easy to follow and understand how they work. To gain a better performance several rules can be implemented in the same script. Each script uses the API in Core to load the files to validate and to call functions in libXML. XPath provided via libXML is used by most of the scripts to search for and compare different elements in the NeTEx-files.

1.5 SYSTEM REQUIREMENTS

To run the tool locally you must ensure that the machine used has the capability to handle the files to be validated. The validation times can be long, and the tool can stop if the processing power or memory is too low. Below is a recommendation for the configuration of a machine. Be aware that very large or many files affect the performance and can result in longer validation times, even on a machine with the recommended hardware.

Minimum	Recommended	Best performance
4 cores	6 cores	6 cores
8 GB memory	16 GB memory	32 GB memory

1.6 LOCAL INSTALLATION

To use the tool locally, you need to install Docker on the computer that you will use. You can use Windows, Mac or Linux as your base operating system, and you will find Docker and instructions on how to install in the <u>Docker Getting Started</u> guide.





Version 1.4- 15/01/2024

After you have installed Docker, you can get the latest version of the Greenlight image by typing the following command in a terminal window:

docker pull itxpt/greenlight

<pre>// / / / / / / / / / / / / / / / / / /</pre>
[sudo] password for theobald:
Using default tag: latest
latest: Pulling from itxpt/greenlight
31e352740f53: Pull complete
7f9bcf943fa5: Pull complete
f8f5e977ad98: Pull complete
8fa91ef8ef8a: Pull complete
c302c6585799: Pull complete
7767d80d8b5b: Pull complete
f34ca7b837b1: Pull complete
728a9af5d93e: Pull complete
c9c6daea7fc0: Pull complete
7ab33612e61f: Pull complete
7822343a78b7: Pull complete
Digest: sha256:58e4a79e4a701824713ddbb1e34e1e2540fdd1c0fa21f55afb871703602d7747
Status: Downloaded newer image for itxpt/greenlight:latest
docker.io/itxpt/greenlight:latest

Figure 2: Command of terminal window to call the latest version of the tool.

After installing Docker, we suggest that you first start to use the web interface to verify that the installation works and to learn more about the functionality.

To start the web interface, use the following command:

docker run -it -p 8080:8080 itxpt/greenlight server
<pre>//Documents/DATA4PT\$ sudo docker run -it -p 8080:8080 itxpt/greenlight server</pre>
[sudo] password for theobald:
/// _/// //\/ _/// v4.10.2 High performance, minimalist Go web framework https://echo.labstack.com 0/

Figure 3 Command to use the web interface to verify that the docker installation works properly.

If you have used the image before, you can also start the web interface via Docker Desktop. Go to the Containers tab and press Start on the Greenlight container.

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





Version 1.4- 15/01/2024

Docker Desktop Upgrade plan	${\sf Q}$ Search for images, containers, volumes, extensions a	and more Ctrl+K 🕸 💠 Sig	gn in 9 – 🗆 🗙
Containers	Containers Give feedback		
्;) Images			
💼 Volumes	Container CPU usage (i) No containers are running.	Container memory usage (i) No containers are running.	Show charts 🗸
🐑 Dev Environments BETA			
💽 Docker Scout		y snow running containers	
🗢 Learning center	Name Image Status	CPU (%) Port(s) Last started	Actions
Extensions	c2ebe810t	N/A 2 minutes ago	▶ : ∎
Add Extensions			
			Showing 1 item
🐡 Engine running 🕨 📔 🕛 🗄	RAM 1.31 GB CPU 0.12% Disk 62.39 GB avail. of 67.32 GB 🙀 Not	t signed in	v4.25.0 🌘 1

Figure 4 Using the Docker Desktop to run the web interface.

1.7 RELEASES HISTORY

The tool has been developed in three (3) main steps:

- First version (V.0.0) for testing purposes.
- First public release (V.0.3.4) for pilot implementation from different stakeholders and systems.
- Second public release (V.1.0.7)

However, updates have been made throughout the testing and pilot implementation with the aim to support the users, replying to their needs based on the collected feedback. The collected feedback is summarised in D.4.3 Feedback on validation tools 2.

The following table summarises the main features that are incorporated in the first and second public release of Greenlight NeTEx validator, summarising also the main additions, updates and fixes that have been performed. The detailed Release history is available in <u>GitHub/ITxPT/DATA4PTools/releases</u>.

Table 1 Main features incorporated in the first public release V.0.3.4 in June 2022.

Relevant releases	v0.3.4 [2022-05-30], v0.3.3 [2022-05-09], v0.3.1 [2022-04-10], v0.3 [2022-04-10]
Chore (actions no	- Add dependencies
related to the code)	- Add EPIP schema
	- Create README







	- Add NeTEx related XSD links in web interface
	- Improve version reference
	- Add License and texts in web interface
	 Update validation rule texts in README and web interface
	- Add note about limitations on web interface
Refactor (changes	- Remove GFX (graphics) terminal output of reporting
regarding structure)	 Add JSON output of reporting to terminal
	- Add rules configuration
	- Add subsequent validation callback
	- Add docker callback in result
	 Add a text section about rules in configuration window
	- Replace logger, update configuration tab and implement required model
	- Move file upload in preparation of custom XSD
Features	- Basic frontend to recover MQTT
	- Add MQTT broker fork
	- Add MQTT broker fork
	- Publish progress over MQTT
	- Remove old design
	- Remove redundant terminal GUI (web interface)
	 Add cap support and timings to concurrency
	- Proxy MQTT and add report download
	- Re-enable a couple of rules
	 Copy, styling, report download and fixed
	- Add web app build stage
	- Implement new rules
	 Add additional rules to web interface
Bug fixes	 Handle ws ssl (WebSockets Secure Sockets Layer)
	- Handle nextjs (a framework used for web application) parameterized paths
	- Fix readme link and dirname
	 Make sure error count match with returned messages
	- Disable rules until they have been resolved
	- Add a stateless approach to file loading
Notes	- Update readme
Performance	- Only render when needed
a second s	

Table 2 Main features incorporated in the second public release V.1.0.7 in August 2023.

Relevant releases	v1.0.7 [2023-08-07], v1.0.6 [2023-07-28], v1.0.5 [2 v0.5.5 [2023-03-21], v0.5.3 [2023-02-08], v0.5.2 [2 v0.5.0 [2022-12-02], v0.4.3 [2022-11-09], v0.4.2 [2	2023-07-12], v1.0.4 [2023-06-26], 2023-02-06], v0.5.1 [2023-01-25], 2022-09-12]
Chore	 Update dependencies Add diff NeTEx versions Version XSD schemas Clean up dependencies Fix build warnings Update readme Update ignore files 	 Move profiles and scripts to app Stricter linter Simplify building by adding Makefile Add legal documents Add built commands
Refactor	 Move CLI-only relevant code to cmd and cleanup from js API Implement new js API Minor refactor from previous API changes 	 Flatten script compilation Fix configuration after dependencies upd Optimize schema cache Add links to NeTEx, SIRI and TRANSMODEL







	- Add more types and update response	- Link to exact GitHub folder
	structure	- Add navigation buttons and config to steps
	- Update CLI to new API	
	- Update configuration and output	
	- Remove unused code	
	- Gofmt (a tool that automatically formats go	
Fosturos	Romovo tolometry collection	Add profiles to web server
i caluics	- Remove telementy collection	- Add profile and points
		- Add prome endpoints
	- Add Internal API	- Opdated design and app icon
		- Added optional Firebase authentication
	- Add more event types	- Add custom XSD
	- Add cli only docker build	- Update EPIP version 1.1.2
	- Add profiles to validation	- Accent input label
	- Add example profiles	
	-	
Bug Fixes	- Fix memory issue using setcontextnode	 Minor script optimizations and fixes
	 Fix slow queries and type of response 	 Fix unresponsive validation
	objects	 Add better control of tmp files
	 Disable next telemetry collection 	 Modify key ref constraint evaluation
	 Fixed load state on validation 	 Ignore externally referenced keys
	 Fixed js api typing 	 Correctly handle attribute lookups
	 Add missing return result 	 Sort validation results by line number
	 Add missing line no to XSD errors 	- Fix typo
	- Handle nested next js resources	 Backdrop cover entire view
	- Handle configuration state change	- Handle limbo state when moving back
	- Update privacy policy	
	- Fix environment key	
Notes	- Add XSD benchmark	•
	 Add typography style 	
Performance	- Optimize docker build size	
	- Add build arm platform	
Notes	 Modify text sizes and change font 	
	- Rewritten JS API	
	Extended standard library with ability to query mult	iple files, added error types and predefined xpath
	paths for ease of use	
	Extended node API with shorthand methods for pro	operties, attributes and values as well as added
	feature for method chaining	
	 Added support for script configuration (e.g 	. setting max distance between two stops)
	 Added support for different NeTEx schem 	a versions (1.01, 1.02, 1.03 and 1.2)
	 Added support for different NeTEx schem 	a versions (1.01, 1.02, 1.03 and 1.2)
	known issue: Legacy NeTEx versions is	seemingly incompatible with _libxml2_ (or in
	general?)F	
	- Replaced large part of the validation lifecy	cle with a event emitter, giving the user control of
	which information is consumed	
	- Squashed a bunch of bugs related to perf	ormance, validation result, memory security &
	errors	





2 VALIDATION RULES

The Validation Rules relevant to Greenlight NeTEx validator are mainly divided into two major categories:

- The rules based on XML schema
- The rules beyond XML schema

2.1 RULES BASED ON NETEX XML SCHEMA

This kind of rules is relevant to data quality dimensions such as uniqueness, consistency and completeness.



Table 3 Data Quality dimensions and correspondance of XML schema based validation rules.

The NeTEx XML schema rules can be applied automatically by any XML Validator. They concern, *Syntactic checks, XML schema conformance checks, Integrity cross-checks.* Examples of this kind of rules are shown in the following Table.

Table 4 Examples of rules based on	XML schema.
------------------------------------	-------------

Rules category	Examples of rules		
Syntactic checks	Well-formed XML : syntactically correct .		
	i.e. <tag attribute="xx">data value</tag>		
XML schema conformance checks	Valid tags, in valid order. No empty tags		
	Valid cardinality: required, optional, 0,1,n		
	Encoding of Data Types:		
	- Date, Time, text, number, currency value, etc., etc.		
	Enumerated values are valid. E.g. Mode bus, rail, tram		





Version 1.4- 15/01/2024

Integrity cross-checks	 Uniqueness constraints. Identifiers are unique in document Referential integrity constraints. Any referenced entity must also be present in same file.
------------------------	---

2.1.1 COMPLETENESS CHECKS

The existence of the NeTEx profiles allows to check the completeness of a dataset against a particular profile. As full NeTEx schema is quite complete, including all elements that concern public transport, profiles are often used to limit the scope and address national and local specificities and needs. The Greenlight validator offers the possibility to upload your custom profile (e.g. a national or local profile), and choose from the predefined list the European Minimum Profile (EPIP).

NeTEx validation Green x +	~ _ е ×
← → C 🗎 greenlight.itxpt.eu/jobs/pEVLTKS8NLlyauyn0q.lega/custom	< 🖈 🛛 🕫 🗄
G Validations Git	tHub New validation
Using the online version may apply limitations. For regular use, download and install the tool for free from Docker or GitHub. You can read mor here.	re about requirements
Configuration Bes	③ Validate
← Go back	5210/12C-S-9/
Custom configuration	
Profile	
Begin by selecting which profile to use for validation NeTEx - The full NeTEx schema (more info) NeTEx Fast - NeTEx schema without constraint (more info) EPIP - NeTEx European Passenger Information Profile (more info) EPIP Light - NeTEx European Passenger Information Profile (more info)	
Custom	•
Upload custom profile Select which file to use as profile 'Select file(s)'	
Select file(s)	
Supported formats are xml, zip	
Main entry point	• 1/2 ////

Figure 5 Where to upload a custom profile

2.1.2 IMPROVING PERFORMANCE BY SEPARATING INTEGRITY CROSS-CHECKS FROM XML SCHEMA VALIDATION

Integrity cross-checks require a lot of memory which must create issues in performance when checking multiple files simultaneously or big files (of many GB). Therefore, in the tool the option to perform such checks "outside" XML schema validation has been added by using script coded rules.

In particular the relevant rules are:

Validate NeTEx element uniqueness



- Make sure NeTEx references have matching keys

The scripts of these rules are available in <u>builtin</u> folder. Check also *2.3.Build your own* rules You can change or add your own rules by cloning the Greenlight repository from GitHub and modify one of the scripts in the directory builtin. Save it with a new name and then map the *builtin* folder to the docker container with the Docker parameter -v.

-v c:\code\greenlight\builtin:/usr/local/greenlight/builtin

Use the script in the same way as the ones of the standard scripts with the flag -r and name of the script.

Example -r mymodifiedrule

Check also 0







Version 1.4- 15/01/2024

Manual for command line interface.





Version 1.4- 15/01/2024

Source codes inventory.

To apply these rules, choose *NeTEx* schemas without constraints (so called **NeTEx Fast/EPIP light** on the web interface, and **NeTEx@1.2-nc, epip@1.1.2-nc.**

How to perform such checks using the Web Interface

In *Configuration* page

- Select **Packages** → **NeTEx Fast (v.1.2), all rules** package,

elect configu	iration			
netex-light-all-rules v1.0.0				
NeTEx Fast (v1.2), a	ll rules			
NeTEx validation (rule b	based constraints) and all rules included		
INCLUDED RULES				
XSD schema validation	Every line is refere	nced Every scheduled stop has a name	e Every stop place has a stop place type	Every stop place has a name
Every stop place is refere	nced Every stop	point have an arrival and departure time	Frame defaults have a locale and timezone	Locations are referencing the same point

Figure 6 Package selection for *fast* integrity checks using separated scripts.

OR

- Select Custom → Profile → NeTEx Fast OR EPIP light
- Select Rules →
- Validate NeTEx element uniqueness
- Make sure NeTEx references have matching keys





Version 1.4- 15/01/2024

Profile	
. Begin by selecting which profile to u	e for validation
 NeTEx - The full NeTEx schema NeTEx Fast - NeTEx schema with EPIP - NeTEx European Passeng EPIP Light - NeTEx European Passeng 	ore info) ut constraint (more info) Information Profile (more info) enger Information Profile (more info)
Profile	
NeTEx Fast (v1.2)	•
Rules . In addition to the schema validation	ve have also included a few optional rules that validate the consistency of the documents
Rules In addition to the schema validation Every line is referenced Make sure every Line (<line></line>) is reference	we have also included a few optional rules that validate the consistency of the documents
Rules . In addition to the schema validation Every line is referenced Make sure every Line (<line></line>) is reference Every scheduled stop has a nan Make sure every (<scheduledstoppoint></scheduledstoppoint>	we have also included a few optional rules that validate the consistency of the documents Ifrom another element as a (<name></name>) or (<shortname></shortname>)
Rules In addition to the schema validation Image: Strain Str	we have also included a few optional rules that validate the consistency of the documents I from another element as a (<name></name>) or (<shortname></shortname>) Ness

Figure 7 Custom configuration for *fast* integrity checks using separated scripts.

How to perform such checks using the Command Line Interface

When running the docker setup you will have to add the –schema parameter with **NeTEx@1.2-nc or epip@1.1.2-nc,** this will by default use all the rules available.

docker run -it itxpt/greenlight validate -schema epip@1.1.2-nc -i testdata

OR setting only the rules with this parameter :

-r netexUniqueConstraints,netexKeyRefConstraints







Version 1.4- 15/01/2024

.Fl			the	eobald@computer01: ~			Q = - • ×
(base) (/O war Schemas DEBU[20 DEBU[20	theobaldgcomputer01:-5 docker run -it it rning : failed to load external entity "c s parser error : Failed to locate the mai 223-11-04710:19:382] validation using sch 223-11-04710:19:382] validation using sch	xpt/greenlight validate - .hema" .n schema resource at 'che wema "chema" .ema "chema"	schema epip@1.1.2-nc -r ma'. document=testdata/line document=testdata/_sha	netexUniqueConstraints 40_9011005004000000.xr ared_data.xml id=6p3p9r;	s,netexKi nl id=6p: 21Pl1bCTI	eyRefConstraints 3p9r21Pl1bCTNpLf(4pLfOos scope=mat	-i testdata Dos scope=main script=xsd type=LOG valid=false In script=xsd type=LOG valid=false
Schemas Schemas DEBU[20 DEBU[20 I/O war	nting : railed to load external entity of s parser error : Failed to locate the mai 223-11-04T10:19:38Z] validation using sch nning : failed to load external entity "of	nema in schema resource at 'che iema "chema" iema "chema" ihema"	ma'. document=testdata/line document=testdata/line	2_41_9011005004100000.xr 2_39_9011005003900000.xr	nl id=6p: nl id=6p:	3p9r21PlibCTNpLf0 3p9r21PlibCTNpLf0	Dos scope=main script=xsd type=LOG valid=false Dos scope=main script=xsd type=LOG valid=false
Schemas DEBU[20 I/O war	s parser error : Failed to locate the mai 323-11-04T10:19:38Z] validation using sch rning : failed to load external entity "c	.n schema resource at 'che nema "chema" :hema"	ma'. document=testdata/line	2_42_9011005004200000.xr	nl id=6p∶	8p9r21Pl1bCTNpLf(Dos scope=main script=xsd type=LOG valid=false
Chemas 1/0 war Schemas	s parser error : Failed to locate the mai rning : failed to load external entity "c s parser error : Failed to locate the mai	n schema resource at 'che hema" n schema resource at 'che	ema'.				
DEBU[20 DEBU[20 DEBU[20	023-11-04T10:19:38Z] validation using sch 023-11-04T10:19:38Z] validation using sch 023-11-04T10:19:38Z] validation using sch roing : falled to load external entity "c	iema "chema" iema "chema" iema "chema" ihema"	document=testdata/line document=testdata/line document=testdata/line	2_30_9011005003000000.xr 2_3_90110050003000000.xr 2_45_9011005004500000.xr	nl id=6p) L id=6p3p nl id=6p)	3p9r21Pl1bCTNpLf0 99r21Pl1bCTNpLf0 3p9r21Pl1bCTNpLf0	Dos scope≡main script=xsd type=LOG valid=false os scope≡main script=xsd type=LOG valid=false Dos scope≡main script=xsd type=LOG valid=false
chemas /0 war	s parser error : Failed to locate the mai rning : failed to load external entity "c s parser error : Failed to locate the mai	in schema resource at 'che hema" n schema resource at 'che	ema'.				
/O war Schemas	rning : failed to load external entity "c s parser error : Failed to locate the mai	hema" n schema resource at 'che	:ma'.				
Arduin Schemas	sparser error : Falled to locate the mai 223-11-04710:19:412] validation using sch "" 1-04710:19:412] validation using sch ODE : failed to locad external entity " s parser error : Failed to locate the mai	n schema resource at 'che iema "chema" iema "chema" :hema" in schema resource at 'che	ema'. document=testdata/line document=testdata/line ema'.	2_2_9011005000200000.xm 2_38_9011005003800000.xm	l id=6p3p 1l id=6p3	09r21Pl1bCTNpLf00 8p9r21Pl1bCTNpLf0	os scope=main script=xsd type=LOG valid=false Dos scope=main script=xsd type=LOG valid=false
teste	data/line_41_9011005004100000.xml						1
#	FILE_NAME	VALIDATION_NAME	START	STOP	VALID	ERROR_LINE_NO	ERROR_MESSAGE
1	testdata/line_41_9011005004100000.xml	xsd	2023-11-04T10:19:38Z	2023-11-04T10:19:38Z	true		
2 3 4	testdata/line_41_9011005004100000.xml testdata/line_41_9011005004100000.xml testdata/line_41_9011005004100000.xml	netexUniqueConstraints netexKeyRefConstraints	2023-11-04710:19:382 2023-11-04710:19:382 2023-11-04710:19:387	2023-11-04710:19:402 2023-11-04710:19:402 2023-11-04710:19:407	false	32	In violation of key-ref constraint, missing key reference \approx
5	testdata/line_41_96116650641060600.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z 2023-11-04T10:19:40Z	false	38	In violation of key-ref constraint, missing key reference = Th violation of key-ref constraint, missing key reference =
7	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	44	In violation of key-ref constraint, missing key reference a
8	testdata/line_41_90110050041000000.xml	netexKeyRefConstraints	2023-11-04110:19:382	2023-11-04110:19:402	false	47	In violation of key-ref constraint, missing key reference a
10	testdata/line 41 9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:387	2023-11-04T10:19:40Z	false	53	In violation of key-ref constraint, missing key reference a
11	testdata/line 41 9011005004100000.xml	netexKevRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	56	In violation of key-ref constraint, missing key reference #
12	testdata/line 41 9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	59	In violation of key-ref constraint, missing key reference ≈
13	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false		In violation of key-ref constraint, missing key reference ≈
14	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false		In violation of key-ref constraint, missing key reference 🗧
	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	68	In violation of key-ref constraint, missing key reference \approx
16	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false		In violation of key-ref constraint, missing key reference \approx
17	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	81	In violation of key-ref constraint, missing key reference \approx
18	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04T10:19:40Z	false	84	In violation of key-ref constraint, missing key reference \approx
19	testdata/line_41_9011005004100000.xml	netexKeyRefConstraints	2023-11-04T10:19:38Z	2023-11-04710:19:402	false	87	In violation of key-ref constraint, missing key reference ≈
20	Lestuara/line 41 9011005004100000.XML	netexkeykerconstraints	2023-11-04110:19:382	2023-11-04110:19:402	alse	90	in violation of key-ref constraint, missing key reference a

Figure 8 Custom configuration for fast integrity with customs rules on CLI

2.2 RULES BEYOND NETEX XML SCHEMA

The rules that cannot express in XML and have been developed and incorporated in the current tool correspond mostly to validity, consistency, and accuracy.







Version 1.4- 15/01/2024



Figure 9 Data Quality dimensions and correspondance of XML schema based validation rules.

Rules category	Examples of rules
Complex cross-checks	E.g. Validity dates of elements fall within validity dates of frame.
	E.g. Stop spatial coordinates lie within their Tariff Zone spatial coordinates
Conditional rules that only apply in	E.g. Point-to-point Tariff should have a Distance Matrix but a Zonal Tariff should
some cases	have Tariff Zones, etc., etc.
Parameterised rules with configurable	E.g. Appropriate distances between stops for transport mode.
values	
	E.g. Appropriate transfer distances to interchange.
Checks against external data sets/	E.g. Operator codes, spatial coordinates.
databases.	
Data modularized into multiple XML	E.g. Large National data sets.
documents with cross- references.	

Table 5 Examples of rules that cannot express in XML.

In total, we have been identified 138 rules potential rules beyond XML schema basics. These rules have been evaluated in terms of priority and concern both general rules and profile specific rules. It is envisaged that this list will the basis for further developments.





Version 1.4- 15/01/2024

Priority level	Number of rules	Topics	Specific to
High	5	Header – versions; Identifiers – Codespaces; Date ranges – Frame; Journey Parts	General schemasou
Mid- high	31	Journey Parts; Frequencies; Timings	General schema
Medium	52	Stop point; Stop place; Line	EPIP and General schema
Mid – Iow	47	Hierarchy; Topographic; Display	Profile Specific and general schema
Low	3	Unused data; Stop Point; Journey	Profile Specific and EPIP

Table 6 Summary of all identified rules – scripted or not.

We currently handle 15 rules that are listed in the following Table 7 Currently integrated rules. In this table you may find the details of each rule together with the link to the corresponding script.

These rules have been chosen considering several criteria (such as importance, effort to be implemented, relevance to the full NeTEx profile and more).

					Severity	
	Script name	Description	Functional Area	Aspect	(10=high 50=low)	Development details
1	passingTimesIsNotDecreasing	On a VERSION FRAME. ToDate must not be later than FromDate on any date range.	Common Content	Date ranges - Frame	10	Check that <i>from</i> date is before <i>to</i> date on the VERSION FRAME
2	everyStopPointHaveArrivalAndDepart ureTime	Every POINT IN JOURNEY POINT In a JOURNEY PATTERN used by a JOURNEY must have a PASSING TIME with arrival and departure time (except for the first and last stop)	Timetable	Timings	20	Check that an appropriate ArrivalTime and DepartureTime exists in for each PASSING TIME TimetabledPassingTime in a SERVICE JOURNEY.
3	everyStopPlaceHasAName	Every STOP PLACE has a Name or ShortName attribute	Stop	Stop Place	20	Name attribute should be filled in for all STOP PLACEs
4	passingTimesHaveIncreasingTimes	Successive DayOffset+PassingTimes for the POINTs IN JOURNEY Pattern or CALLS of a Journey must not decrease.	Timetable	Timings	20	Successive DayOffset+PassingTimes for the POINTs IN JOURNEY Pattern or CALLS of a Journey must not decrease
5	frameDefaultsHaveALocaleAndTime Zone	The FrameDefaults of a VERSION FRAME should have values appropriate to the content	Common	Frame	30	 Depends on Frame type. For all frames check and DefaultLanguage exists in DefaultLocale in FrameDefaults. For frames that contain spatial coordinates, check that default LOcationSystem is specified (usually WGS84) For Frames that contain elements with Timezones. (e.g. STOP PLACEs etc. in . SITE FRAME. SCHEDULED STOP POINT in SERVICE FRAME, Check that Time Zone is specified. For frames that hold monetary values, e.g. FARE FRAMES or if amount specified. NB can be specified on outmost COMPOSITE FRAME if common to all.

Table 7 Currently integrated rules





Version 1.4- 15/01/2024

					Severity	
	Script name	Description	Functional Area	Aspect	(10=high 50=low)	Development details
6	everyStopPlaceHasACorrectStopPlac eType	Every STOP PLACE has a StopPlaceType attribute with correct value	Stop	Stop Place	30	Each STOP PLACE should have a StopPlaceType attribute. This should match any type on the QUAYs.
7	netexKeyRefConstraints.js	All stop identifiers (QUAY. all STOP PLACEs. GROUPS OF STOP PLACEs and ACCESS) must comply with the profile codification	Stop	Stop Place	30	For Stop specifically [COUNTRY code]: [INSEE common code]: [Type of object]: [Specific stop code]: [Issuer code of the technical code or LOC].
8	locationsAreReferencingTheSamePoi nt	SCHEDULED STOP POINT must have similar spatial coordinates to those of the assigned STOP PLACE	Stop	Stop Point	30	Should be with a certain tolerance of distance, varying by mode Will not necessarily be the same centroid.
9	stopPlaceQuayDistanceIsReasonabl e	Distance Between QUAY and STOP PLACE too long	Stop	Spatial	30	Distance between QUAY and STOP PLACE should not be too far apart. STOP PLACE location is centroid of station. QUAY is centroid of QUAY.
1 0	locationsAreReferencingTheSamePoi nt	The location of QUAY and SCHEDULED STOP POINT should be within reasonable distance of the location or surface of STOP PLACE	Stop	Stop place	30	Take the positions from the QUAY and SCHEDULED STOP POINT and calculate the distance in meters. Hard code 500m in first version. later add parameter to set distance





Version 1.4- 15/01/2024

	Script name	Description	Functional Area	Aspect	Severity (10=high 50=low)	Development details
1	everyScheduledStopPointHasAName	A SCHEDULED STOP POINT must have an instantiated Name field	Stop	Stop Point	30	The name of a stop should be given. Applies s to Both STOP PLACE and SCHEDULE D STOP POINT
1 2	everyLinelsReferenced	A LINE must have one or more ROUTE instances	Timetable	Line	40	Check that a LINE is referenced in at least one ROUTE
1 3	everyStopPointIsReferenced	Any SCHEDULED STOP POINT that is declared should be used. i.e. referenced by an assignment or POINT IN PATTERN etc.	Timetable	Unused data	40	Check that each SCHEDULED STOP POINT is used in one or more JOURNEY PATTERNS.
1 4	everyStopPlaceIsReferenced	Any STOP PLACE that is declared should be referenced by a STOP ASSIGNMENT	Stop	Unused data	50	Every STOP PLACE should be referenced in at least one STOP ASSIGNMENT. Depends on the profile.
1 5	locationsAreReferencingTheSamePoi nt	SCHEDULED STOP POINT must be assigned to a STOP PLACE	Stop	Stop Point	50	Every SCHEDULED STOP POINT should be referenced in at least one STOP ASSIGNMENT. Depends on the profile.

2.3 BUILD YOUR OWN RULES

You can change or add your own rules by cloning the Greenlight repository from GitHub and modify one of the scripts in the directory <u>builtin</u>. Save it with a new name and then map the *builtin* folder to the docker container with the Docker parameter -v.

-v c:\code\greenlight\builtin:/usr/local/greenlight/builtin

Use the script in the same way as the ones of the standard scripts with the flag -r and name of the script.

Example -r mymodifiedrule

Check also 0









Version 1.4- 15/01/2024

Manual for command line interface.





Version 1.4- 15/01/2024

3 SOURCE CODES INVENTORY

This table provides a comprehensive inventory of the assets and components within GitHub repository. This is not a detailed inventory, but it will help you find your way around in the repository.

Table 108 Source codes Inventory.

Folder/File	Definition	Folder/File	Definition	File	Definition
		api	Define how the	e client can communicate with	the application (main script client.ts and type definition in types.ts)
				useApiClient	allow the API
		hooks	handle	useEmail	enable the use of email
	Web	nooks	component	useFirebase	enable the use of email
арр	interface: next.config to			useWebConfig	set up Webconfig using firebase
	use static file	components	all the script th	at are used (example FileUplo	bad handle the uploading, the name are self-explanatory)
		nages	pages	lobs	validation views
		pages	hardcoded	0003	-> custom
		public	icon	1	
		P	css style for th	e pages	
		styles	css style for th	e pages	
builtin	Actual scripts for	or validation rules	coded in javaso	cript (their names are quite self	f-explanatory). Look at the rules part for more details.
		validate.go	Is the main sc	ript	
cmd	Command	profile.go	Handle the pro	ofile type and setup	
Unita	line tool	root.go	Setup the root	github folder needed	
		file.go	Used to handle	e different file and clean tempo	orary one
xml-stream- parser	External library	to find path in xn	nl		
internal	Add intern featu	ure to improve eff	ficiency the tool		
js	Transform java	script feature into	understandable	e feature in GO programming la	anguage
media	Media used in t	he readme file			
testdata	Test xml data s	et			





Version 1.4- 15/01/2024

Folder/File	Definition	Folder/File	Definition	File	Definition
xml	Link library libx	xml2 to .go and ci	eate new bindin	gs	
xsd	NeTEx schema	as from NeTEx G	itHub		
validation.go	Hold the functi	ions to validate a	file		
result.go	Hold the function	ions to generate t	he result		
Makefile	Details of the c	command line			
cliff.toml	Generate char	ngelog file			
CHANGELOG .md	Log with all cha	anges performed	in the GitHub fo	lder	
README.md	The README	file where you ca	In find more info	on the too	l and how to use it

4 MANUAL FOR WEB INTERFACE

After the container has started you can use the web interface by opening a web browser and type the address <u>http://localhost:8080/</u> or you can access the online tool at the address <u>https://greenlight.itxpt.eu/</u>. Then you can click on Start validating to start a new validation session. You can also always use the New validation button in the upper right corner to start over with a new validation.

NeTEx validation Green × +						~	_ @ ×
\leftrightarrow \rightarrow \mathbf{C} \cong greenlight.itxpt.eu						< 😒 😫 🖈	🗆 🗊 :
	G				Validations GitHub New validation	MA)	
	Using the online vers here.	ion may apply limitations. For regular use	e, download and install the tool fo	r free from Docker or GitHub.	You can read more about requirements		24
	Data4PT The DATA4PT project air and models, to fulfil the r By supporting EU Membe enable unlon-wide multim mobility services. Validation tool Key activity of DATA4PT for public transport data quality dimension is align and therefore the interop If you have feedback, que	hs to advance data-sharing practice leeds of multimodal travel informatis ar States in deploying a set of harm modal travel information services an project is the davelopment of valida in National Access Points (NAPs), the erable exchange of travel and traffi estions or bug reports please do no	es in the public transport sect on service providers. onised European public data d contribute to a seamless d nation tools for NeTEx and SIR he purpose of validation is to project to enable the implem c data across Europe. t hesitate to send them our v	cor by supporting the deve standards (Transmodel, N poor-to-door travel ecosyst datasets. As NeTEx and S ensure a certain level of q entation of ITS Directive D vay through G GitHub or =	topment of data exchange standards aTEx and SIRI), DATA4PT wants to em across Europe that covers all IRI are the EU standardised formats uality of the published data. The elegated Regulation EU 2017/1926 Email.		
	LIGHT	Resources	Data4PT	ITxPT	Legal		
		Documentation	Home	Home	License		
		Requirements	About	About	Privacy Policy		
		Changelog	News	News	Terms of use		
		Contact	Wiki				
		GitHub					
Show Applications		Docker					

Figure 10 Validation tool Greenlight website

4.1 NAVIGATION

At the top of the web page is a menu bar and a progress indicator. The logo to the left always take you to the start page. Validations in the menu will show recent done validations, GitHub will take you to our page with documentation and the source code and New validation will start over with a new validation. You can also use Go back to navigate to a previous step.



Figure 11 All steps for a validation







4.2 CONFIGURATION

To start a validation, you first decide if you want to use a premade configuration package or use a custom configuration. The packages are predefined with schemas and rules that are commonly used together. To select your own combination of schemas and rules you can do a custom configuration.

 Note that this is an early build and r 	nore capabilities will be available soc	on. Visit us regularly for updated! Y	ou may provide your feedback in this fo	orm
1 Configuration		Files		3 Valid
Configuration				
Packages				<u></u>
Select from a list of predefined pack	kages of NeTEx profiles and rule	S		3
		OR		
Custom				2
Create your own custom configurat	ion			3

Figure 12 Configuration page

4.3 PACKAGES

If you select to use the premade packages you are presented with a list to select from. Select the one that works best with your validation requirements. When you click on one of the packages you continue to the selection of files.

Configuration	2 Files		– 🗿 Valida
Select configuration			
netex-light-ail-rules v1.0.0			
NeTEx Fast (v1.2), all rules			
NeTEx validation (rule based constraints) and all rules included			
INCLUDED RULES			>
XSD schema validation Every line is referenced Every scheduled stop has a nar	me Every stop place has a stop place type	Every stop place has a name	(in)
	Recent defendes being alle site and alle second	Locations are referencing the same point	
Every stop place is referenced Every stop point have an arrival and departure time	 Frame defaults have a locale and timezone 	Ecoloris are relevencing the same point	
Every stop place is referenced Every stop point have an arrival and departure time Passing times have increasing times Stop place quay distance is reasonable Va	Ilidate NeTEx element uniqueness Make sur	e NeTEx references have matching keys	
Every stop place is referenced Every stop point have an arrival and departure time Passing times have increasing times Stop place quay distance is reasonable Va netex-light v10.0	Frame defaults nave a locale and timezone	 Exclusions are referenced up on the point NeTEx references have matching keys 	

Figure 13 Premade package available on the website





Version 1.4- 15/01/2024

4.4 CUSTOM CONFIGURATION

With the custom configuration you can be more detailed in which NeTEx Profile and combination of rules to use. In the list of rules, you get brief description of each rule. Zero or more rules can be selected by clicking the checkbox for each rule.

Configuration	2 Files	3 Validate
Custom configuration		
Profile		
. Begin by selecting which profile to use for validati	lon	
NeTEx - The full NeTEx schema (more info) NeTEx Light - NeTEx schema without constrair EPIP - NeTEx European Passenger Information EPIP Light - NeTEx European Passenger Inform	nt (more info) Profile (more info) nation Profile	
Schema		
NeTEx Fast (v1.2)		
In addition to the schema validation, we have also Every line is referenced Make sure every Line (<line></line>) is referenced from another e	included a few optional rules that validate the consistency o	f the documents
Every scheduled stop has a name Make sure every (<scheduledstoppoint></scheduledstoppoint>) has a (<name></name>)) or (<shortname></shortname>)	
Every stop place has a stop place type Make sure every (<stopplace></stopplace>) has a (<stopplacetype></stopplacetype>) a	and that it is of correct type	

Figure 14 Custom configuration page

Some rules use parameters as input to the validation. Those rules have a default value that can be changed by clicking on the Configure icon to the right.

Passing times have increasing times	Configuration		
Make sure passing times have increasing times a	nd day offsets Distance		
Stop place quay distance is reasonal Check the distance between a (<stopplace></stopplace>) a	ole 500 state 500		Ξ
Validate NeTEx element uniqueness		Close	



4.5 SELECT FILES TO VALIDATE

The last step is to upload the files to be validated, it can be single files or multiple files compressed in an archive. Click Select file(s) to select which files to upload and then wait until all files has been uploaded, see the Status indicator in the files list.

If you want to get back to the selection of rules you can use the Go back button





Version 1.4- 15/01/2024

(i) Note that this is an early but	uild and more capabilities will be available soon. Visit us regularly for updated! You may provide your feedb	back in this form
Configuration	Biles	3 Valid
← Go back		
Upload files		
Upload files Select which files to validate b	by clicking 'Select file(s)'	
Upload files Select which files to validate b	by clicking 'Select file(s)' Select file(s)	
Upload files Select which files to validate b	by clicking 'Select file(s)' Select file(s) Supported archive/compression formats are zip, gz/p, bz/p and tar	

Figure 16 Upload validation file page

When all files are uploaded you start the validation by clicking on Validate.

 Note that this is an early build and more capability 	lities will be available soon. Visit us regularly for updated! You may provid	de your feedback in this form
Configuration	2 Files	3 v
← Go back		
Upload files		
Select which files to validate by clicking 'Select	file(s)'	
	Select file(s)	
Filename		s
schweiz_20230201.zip		@ uploa

Figure 17 Validation of the upload files

Each file is validated against the selected schema and rules, all validations run in parallel. Depending on the number of files and their sizes the validation can take some time to complete.





	Validations GitHub New valida	ition
Note that this is an early build and more capabilities will be available soon. Visit us regularly for updated! You may available soon.	ay provide your feedback in this form	
Validation result (master-mollusk) FysyligvoXRLF14ZYv-x		
NeTEx Fast (v1.2), no rules XSD schema validation		
(running) zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_COMMON_1_2202302010402.xml (running) zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_RESOURCE_1_1_202302010402.xml	d.	~ ~
(0402.xml	~
(~
(running) zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_SITE_1_1_202302010402.xml		~

Figure 18 Running validation page

4.6 VALIDATION RESULT

When the validation is done you get an overview of the result. You can see the status of the validation for each file. If there are any errors, you can get all the details by clicking on the down arrow to the right of each file.

		10007
Note t	hat this is an early build and more capabilities will be available soon. Visit us regularly for updated! You may provide your feedback in this form	n
Validatio	ON result (master-mollusk) FysylligvoXtRLF14ZYv-x	
NeTEx Fa	st (v1.2), no rules a validation	
Ø valid	zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_COMMON_1_1_202302010402.xml	
(valid	zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_RESOURCE_1_1_202302010402.xml	
(valid	zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_SERVICECALENDAR_1_1_202302010402.xml	
(invalid		
	zip//_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_SERVICE_1_1_202302010402.xml	

Figure 19 Result page

The details display the number of times that specific error occurs in the file, and you can page between them with the arrows to the right. For each error you get information about the type, line number in the file and a more detailed explanation.

Version 1.4- 15/01/2024





Version 1.4- 15/01/2024

Deliverable 2.3 – Validation Tools Release 2

	Validation	Errors	Warning
Ø valid	xsd	0	0
Download rep	rt for zip/./_NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_SERVICECALENDAR_1_1_202302010402 xml	•	
invalid zip//	NETEX_TT_1.10_CHE_SKI_2023_OEV-SCHWEIZ_SERVICE_1_1_202302010402.xml		
Status	Validation	Errors	Warning
∧ (1) inva	d) xsd	46	0
Errors 1/46		<	>
) line: 16421		

Figure 20 Details of the result on the website

4.7 DOWNLOADING THE RESULT

You can download the result for each error or the complete validation to a file in json or csv format to process it further. For example, to give as documentation to someone who can correct the error

Ø valid zip//_N	ETEX_T	T_1.10_CHE_SKI_2023_OEV-SCH	WEIZ_TIMETABLE_99_214_2	02302010402.xml	~
Ø valid zip//_N	ETEX_T	T_1.10_CHE_SKI_2023_OEV-SCH	WEIZ_TIMETABLE_9_214_20	2302010402.xml	*
Download report	-	Ŕ		50/	Validate with this configuration
	json				
	CSV	ources	Data4PT	ITxPT	Legal
	Do	cumentation	Home	Home	License

Figure 21 Downloading result as json or csv





Version 1.4- 15/01/2024

Example of the saved data in json format.





4.8 PREVIOUS VALIDATION

At the bottom of the result page, you have an option Validate with this configuration to go back and start a new validation with the same configuration but with new files. You can also see recent validation by using the menu selection Validations at the to of the page. By clicking on the name of a job you will see the result page for that validation again.

 Note that this 	s is an early build and	more capabilities will be ava	ailable soon. Visit us regularly for updated! Y	ou may provide your feedback in this form	1
Jobs					
Status	Valid	Name	Ref	Created	Duratio
(complete	() invalid	leading-stinkbug	EZTEiz6V356hVFeDJMN5v	2023-04-27, 06:21:51	17 se
@ complete	(invalid	literate-rattler	r-FaNMHOttMGrdyQF3Y5R	2023-04-27, 06:20:20	1 min, 9 se

Figure 23 Validation history page





4.9 TECHNICAL ERROR MESSAGES

Sometimes the web interface will show error messages if the Greenlight tool stops to execute. Often that occurs when the communication to the web server is lost, or the local Docker version has stopped. Check the status of your connection and that the Docker container is running if using it locally.



Figure 24 Example of Error messages







5 MANUAL FOR COMMAND LINE INTERFACE

The CLI (command line interface) is for more advanced use cases where you want more control over the validation or if you want to include the validation in your own pipeline. An example could be to recieve a file via an integration, validate the file with GreenLight and if there are any errors inform via email and otherwise save the file for use in another system.

To use the CLI you must first download the Docker image as described in Getting started

When you use the CLI you first give the command docker and the parameters run -it [docker_image] in this case the docker_image is itxpt/greenlight. After that you give the different commands and flags to greenlight, e.g., help. If you want to use other docker parameters, you have to put them before the name of the image to use. See below for more complex examples of how to invoke the greenlight command.

5.1 **GETTING HELP**

The tool has a built in help system that gives explanations of all commands and parameters in the tool. Use the command below to get an overview of the help you can get.

docker run -it itxpt/greenlight help
NeTEx/Siri validation tool
Usage: greenlight [command]
Available Commands: completion Generate the autocompletion script for the specified shell help Help about any command server Start NeTEx validation server validate Validate NeTEx files
Flags: -h,help help for greenlight
Use "greenlight [command]help" for more information about a command.

Figure 25 Help command on the Command Line Interface

5.2 SERVER COMMAND



Figure 26 Command to build the web interface





docker run -it -p 8080:8080 itxpt/greenlight server

This will start the built-in web interface and it can be accessed via http://localhost:8080/. Se Web Interface for a guide on how to use it.

5.3 VALIDATE COMMAND

A validation is started with the command Validate, it uses the following flags as input to configure the validation.

Flags:	
-h,help	help for validate
-i,input string	XML file, dir or archive to validate
<pre>-l,log-level string</pre>	Set level of log output (one of "trace", "debug", "info",
"warn", "error") (default	"debug")
-o,output string	Set which output format to use (one of "json", "xml",
"csv", "pretty" (default '	'pretty")
-p,profile string	Set path of validation profile (note: flags 'rules' and
'schema' is ignored)	
-r,rules strings	Set which validation rules to run (defaults to all inside
the builtin dir)	
-s,schema string	Which xsd schema to use (supported "NeTEx@1.2",
"NeTEx@1.2-nc", "epip@1.1.	2", "epip@1.1.2-nc") (default "NeTEx@1.2-nc")
silent	Running in silent will only output the result in a boolean
fashion	

To verify that the tool works yo can do a validation with a NeTEx file provided with the tool.

- C	iommand Prompt X + ~					
:\Ca 100: EBU	ode\ITxPT\DATA4PTTools>docker run -it lo 30002000000.xml [2023-04-28T11:15:222] validation using	ekojson/greenlight valida schema "netex@1.2"	te -s netex@1.2 -r ever document=testdata/	yStopPlaceHasAName -i to line_2_9011005000200000	estdata/ .xml id=	line_2_9 doY8i9mx
tes	cuXR_lzH scope=main script=xsd type=LOG stdata/line_2_9011005000200000.xml	valid=false				
#	FILE_NAME	VALIDATION_NAME	START	STOP	VALID	ERROR
1 2	testdata/line_2_9011005000200000.xml testdata/line_2_90110050002000000.xml	everyStopPlaceHasAName xsd	2023-04-28T11:15:22Z 2023-04-28T11:15:22Z	2023-04-28T11:15:22Z 2023-04-28T11:15:25Z	true true	
		VALID	TRUE			



5.4 NETEX PROFILE

To select NeTEx profile use the flag -s or --schema and the name of the profile. Valid names are NeTEx@1.2, NeTEx@1.2-nc, epip@1.1.2, epip@1.1.2-nc. If no schema is selected the NeTEx@1.2-nc is used. -nc at the





end means that the validation is with No Constraints. Which is a faster validation but needs that the noconstraints rule is used instead.

Example of how to use the EPIP schema when validating the built in test file

docker run -it itxpt/greenlight validate -schema epip@1.1.2 -i testdata

5.5 RULES

Select which rules to use with the flag -r or --rules and then give the name of the rules to use. Several rules can be specified by separating them with a comma.

Example



Figure 28 Command line example of rule selection

You can change or add your own rules by cloning the greenlight repo from GitHub and modify one of the scripts in the directory builtin. Save it with a new name and then map the builtin folder to the docker container with the Docker parameter -v.

-v c:\code\greenlight\builtin:/usr/local/greenlight/builtin

Use the script in the same way as one of the standard scripts with the flag -r and name of the script.

Example

-r mymodifiedrule

5.6 PROVIDING FILES

The files to test can be single files, a folder with files or a compressed archive with files. Put the files to be tested in a local folder and use the docker parameter -v to map it with a folder in the greenlight container.

-v C:\code\NeTEx\testdata:/usr/local/greenlight/testdata

Then you can use the greenlight flag -i to include the files in the validation

Command to validate a folder with files





docker run -it -v c:\code\NeTEx\testfiles:/usr/local/greenlight/testdata
itxpt/greenlight validate -s NeTEx@1.2-nc -r everyLineIsReferenced -i testdata

Command to validate an archive with several files

docker run -it -v c:\code\NeTEx\testfiles:/usr/local/greenlight/testdata
itxpt/greenlight validate -s NeTEx@1.2-nc -r everyLineIsReferenced -i
testdata/xt_2023_04_15.zip

Command to validate a single file

docker run -it -v c:\code\NeTEx\testfiles:/usr/local/greenlight/testdata
itxpt/greenlight validate -s NeTEx@1.2-nc -r everyLineIsReferenced -i
testdata/line_2_9011005000200000.xml

5.7 OUTPUT

The result of the validation can be presented in different formats. For example, the pretty will give an output adopted to be read on the screen. The other formats json, xml and csv can be used to pipe the output to a file for further processing.

docker run -it itxpt/greenlight validate -s NeTEx@1.2-nc -r everyLineIsReferenced
-i testdata -o json > greenlight-result.json

5.8 COMPLETION COMMAND

Generate an autocompletion script for Greenlight for different shells. The generated script can be added to your shell profile. Scripts can be generated for bash, fish, zsh and powershell.

Note: This command is for power users who uses the CLI a lot and want to make it easier and faster to type commands and parameters.

As an example, will the command below generate a script for bash

docker run -it itxpt/greenlight completion bash





Version 1.4- 15/01/2024

6 BUILDING FROM SOURCE

6.1 PREREQUISITES

Greenlight is using Go and is powered by libxml2, so make sure those are installed first. If you want to work on the web interface, you will also need Node.js

Go

Download and install the latest version with standard settings

libxml2

Install using

- Mac: brew install libxml2
- Linux: sudo apt install libxml2
- Windows: Build from source or download precompiled binaries

nodejs - only required for the web interface

Download and install the latest version with standard settings

6.2 GETTING STARTED

Open a terminal and navigate to the folder where you want to install the source code.

cd /home/developer/code

Clone repository

git clone https://github.com/ITxPT/DATA4PTTools

Navigate to project

cd DATA4PTTools

Downloading dependencies

go get

6.3 BUILDING AND RUNNING THE CLI

With the source code and all dependencies downloaded you can try the tool with build in test files to verify that all is working

Validate with demo files provided in the source

path definition will differ running on windows

go run cmd/*.go validate -i testdata

You can now start validating your own files by providing the path to your document





Validate using your own files

path definition will differ running on windows

go run cmd/*.go validate -i /path/to/documents

That is all that is needed to start using the tool and to be able to modify the core or work with your own validation scripts (you find the scripts in the folder builtin).

6.4 BUILDING THE WEB GUI

When running the web interface the core tool and the interface are started as two separate servers. The backend server is hosting the core tool and provides the functionallity for the validation itself. The interface is then started in a web server and calls the backend when needed.

1. Open a terminal and navigate to the DATA4PTTools directory

cd /home/developer/code/DATA4PTTools

- 2. Start the backend server
 - go run cmd/*.go server
- 3. Then open a new terminal and navigate to he DATA4PTTools directory again

cd /home/developer/code/DATA4PTTools

4. Set current configuration for backend server

Remember to update with the current path to the backend server.

echo "NEXT_PUBLIC_API_URL=http://localhost:8080" > app/.env.local

5. Navigate to the web app directory

cd app

6. Install dependencies for the web server

npm i

7. Start the web server

npm run dev





Version 1.4- 15/01/2024

8. Open the web interface

Open a browser and navigate to http://localhost:3000 and you will see the web interface of Greenlight.

				71	
\leftrightarrow \rightarrow G	🗘 🗋 localhost 300	0	C ¹	12 E	? P 约
G			Vali	dations GitHub	New validation
•••					
Using the	online version may apply limitations.	. For regular use, download and i	nstall the tool for free from D	ocker or GitHub. You can	read more about
Data4PT					
The DATA4PT pro	ject aims to advance data-shari	ing practices in the public tr	ansport sector by suppor	rting the development	of data exchang
By supporting EU	Member States in deploying a s	set of harmonised European	public data standards (T	ransmodel, NeTEx and	I SIRI), DATA4PT
wants to enable u Europe that cover	nion-wide multimodal travel info rs all mobility services.	ormation services and contr	ibute to a seamless door-	-to-door travel ecosys	tem across
Validation t	tool				
Key activity of DA	TA4PT project is the developme	ent of validation tools for Ne	TEx and SIRI datasets. As	s NeTEx and SIRI are th	ne EU
standardised form quality of the pub Directive Delegat	nats for public transport data in lished data. The quality dimensi ed Regulation EU 2017/1926 and	National Access Points (NA ion is aligned with the overa d therefore the interoperable	Ps), the purpose of valida II objective of the project e exchange of travel and	ation is to ensure a cer to enable the impleme traffic data across Euro	tain level of entation of ITS ope.
If you have feedb	ack, questions or bug reports pl	lease do not hesitate to sen	d them our way through 4	🕽 GitHub or 🗳 Email.	
		Start validat	ting		
		10116258			
GREEN	Pasquireas	Data/DT	ITYPT	Legal	

Figure 29 Web interface built from source

To verify that the web interface has contact with the backend server, you may look at the API Status in the lower right corner of the web page.







Version 1.4- 15/01/2024



Figure 30 API status on web interface built from source





Version 1.4- 15/01/2024

CONCLUSIONS

The Greenlight NeTEx validator aims to support in priority data producers and data consumers that need to check the quality of individual NeTEx files or a set of files to be used as basis for their implementations. It is also addressed to National Access Point (NAP) operators, as an additional asset to ensure quality of the published data. The web user interface does not require technical skills and therefore it is easy to get a quick overview of the status of the datasets. Moreover, the open-source character of the tool offers the possibility to be extended and adapted to specific local and/or national needs. Thus, guidelines on building from source are also available.

Nevertheless, further development is required in the field of additional validation rules to address the specificities of the existing and up-coming European minimum profiles (EPIP, EPIAP, Fares etc.). Defining technical specifications for a collaboratively created library of rules is one of the next steps to increase the added value of the tool. Finally, the improvement of performance is also under consideration aiming at the integration of the tool in platforms where large amount of data is processed (for example in NAP platforms).

