

data4pt

DATA4PT Webinar

NeTEx Validation Tool

6 December 2022
15.30 - 16.30 CET



Facilitators, the DATA4PT experts

Petter Kvarnfors

Nick Knowles

Stefan de Konink

DATA4PT overview and tool introduction

Anastasia Founta, ITxPT

Emmanuel de Verdalle, ITxPT

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



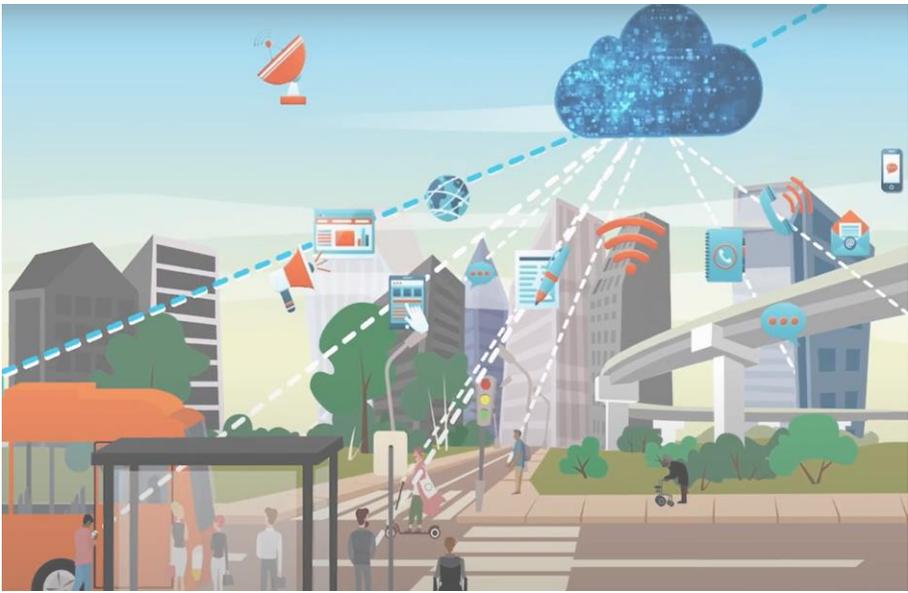


AGENDA

Time (CET)	Topic	Facilitator
15:30 – 15:40	DATA4PT project and introduction of NeTEx validators	Emmanuel de Verdalle
15:40 – 15:50	Demonstration of the Greenlight tool (web interface and terminal version)	Petter Kvarnfors
15:50 - 16:05	Validation rules (current and future developments)	Nick Knowles
16:05 – 16:15	“Real life” examples and relevant validation rules	Stefan de Konink
16:15 – 16:25	Customise Greenlight	Petter Kvarnfors
16:25 – 16:30	Q&A	



A CEF Programme
Support Action
(2020-2023) to
**accompany and
facilitate** Member States
and other relevant
stakeholders
**in the implementation of
MMTIS Delegated
Regulation**





More information on DATA4PT (training material, tools, request of support, technical info...)

<https://data4pt-project.eu/>

The screenshot shows the homepage of data4pt-project.eu. The browser address bar displays the URL. The website header includes the 'data4pt' logo, navigation links for 'ABOUT', 'TECHNICAL SUPPORT', 'KNOWLEDGE BASE', 'LIBRARY', 'NEWS & EVENTS', and 'CONTACT', and a prominent 'DATA VALIDATION TOOL' button. The main content area features a large image of two women on a train platform looking at a smartphone. Below the image, the 'Data4PT' title is followed by a paragraph explaining the project's goal to advance data-sharing practices in the public transport sector. A 'I WANT TO KNOW MORE' button is located at the bottom of the text block.

The screenshot shows the 'NeTEX' page on data4pt.org. The browser address bar displays the URL. The page has a 'Page Discussion' tab. The main heading is 'NeTEX'. Below it is a 'Contents [hide]' section with a numbered list of links: 1 NeTEX overview, 2 NeTEX-Light, 3 NeTEX EPIP, 4 NeTEX national and EU minimum profiles inventory, 5 NeTEX software / tools (with sub-links for conversion, validation, language bindings, editing, data enabled, and planning), 6 NeTEX Part 5 for New Modes (with sub-links for technical artefacts and canonical mapping), and 7 References. A left sidebar contains a 'Main page' section with various navigation links like 'Public Transport models', 'National Access Points', 'Transmodel', 'SIRI', 'NeTEX', 'Support', 'FAQ', and 'Support request'. Below this is a 'Tools' section with links for 'What links here', 'Related changes', 'Special pages', 'Printable version', 'Permanent link', 'Page information', and 'Cite this page'. The 'NeTEX overview' section begins with a paragraph stating that NeTEX was developed under the aegis of CEN and is the most recent system to harmonize European passenger information data.



Introduction to Validators

➤ Why do we need validators?

- Efficient exchange of high-quality data.
- Common tool for exporter & importer to certify conformance.

➤ Nature of Validator Tools

- Static file checkers:
 - ❑ e.g. for NeTEx
- Dynamic API harnesses: request/response
 - ❑ e.g. for SIRI - several national validators
 - ❑ <https://github.com/afimb/siri-validator>

➤ DATA4PT Validator Tool

- Harmonised
- Universal
- Reliable

Screenshot from the wiki page
with the SIRI validators



ABOUT ▾

- Project
- Consortium
- Data Models

TECHNICAL SUPPORT

KNOWLEDGE BASE

LIBRARY

NEWS & EVENTS

CONTACT

DATA VALIDATION TOOL
VERIFY YOUR NETEX DATA



github.com/ITxPT/DATA4PTTools

Greenlight – The DATA4PT validation tool

GO VERSION >=1.17 DOCKER PULLS 420 DOCKER STARS 0

Web · CLI · Source · Configuration

The minimal, customizable, NeTex validation tool

- **Customizable:** configure what you see and how you see it.
- **Scripting** write your own validation rules using javascript
- **Fancy** shows relevant information at a glance.
- **Easy:** quick to install – start using it in minutes.

Validation result

Are you interested in diving deeper? Consider testing it locally with Docker

sigl/Net_NF_2021-04-25_2021-12-11.index.xml

Status	Validation	Errors	Warnings
✓	everytochubbedding@stiffnessAName	0	0
✓	everytochubbedding@stiffnessAName	0	0
✓	everytochubbedding@stiffnessAName	0	0
✗	xsd	1	0

Errors (1/1)

Error (xs) (line 2)

Schema: http://www.w3.org/2001/XMLSchema#dateTime: No matching global declarations available for the validation tool.

Download report

APP STATUS NEXT STATUS

Introduction

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





Greenlight - The Data4PT NeTEx Validator

➤ About the Data4PT NeTEx Validator -

- Open source
 - Go, JavaScript and XPath
 - MIT open license
- Extensible,
 - Common Rules
 - Additional Rules added by users
- Configurable
 - Rules to include, severities,
 - Rule parameters,
- Modular, Configurable Architecture
 - As web service or on local server.
 - With front end GUI calling back-end engine.
 - With batch pipelining of XML documents direct to engine.





Greenlight - ROAD MAP

➤ Beta 1.0 - August 2022

➤ New Revised Beta 2.0 Release - December 2022



- New features e.g. multiple files together
- Test against large NAP data sets
- Usability feedback

➤ Further development?

- Add further common validation rules.
- Support third party customisation of rules.
- Multilingual GUI?

• Strategic issues

- Hosting of high-volume production throughput – who pays?



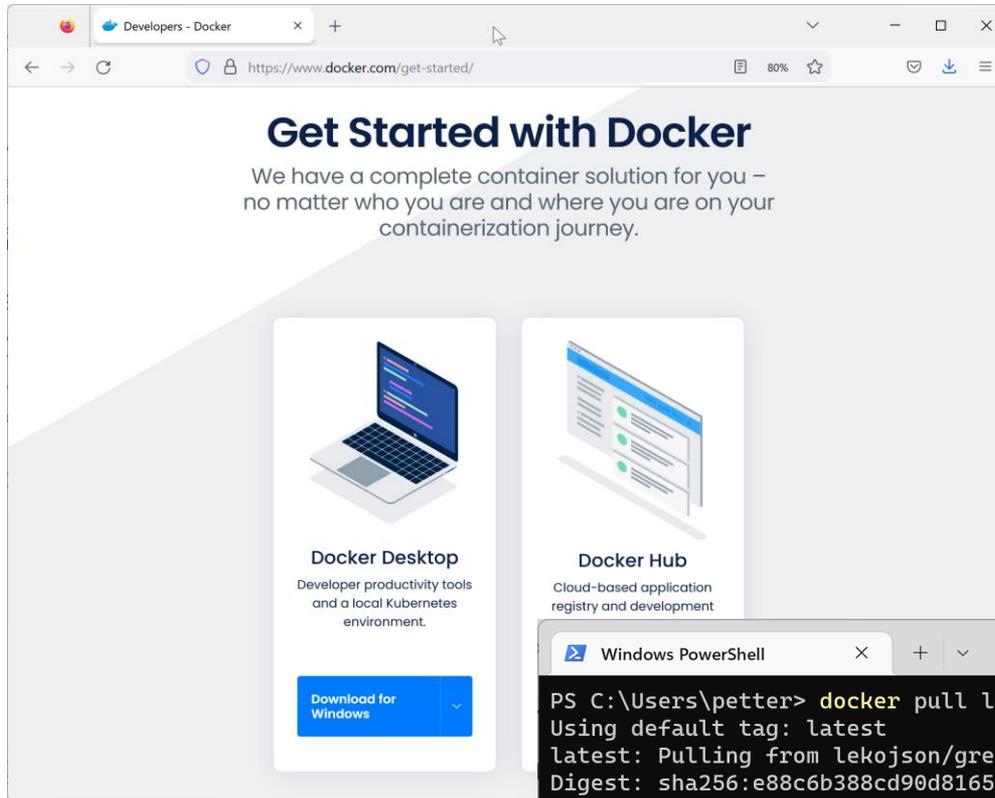
Demonstration of Greenlight

Web interface and Command line





Installation of Greenlight



Download Docker from [docker.com](https://www.docker.com)
Works on Linux, Windows and Mac

Pull the latest image of Greenlight
from DockerHub

```
Windows PowerShell
PS C:\Users\petter> docker pull lekojson/greenlight
Using default tag: latest
latest: Pulling from lekojson/greenlight
Digest: sha256:e88c6b388cd90d8165648577e33adf965cfd4d293ebafdc4cc9715766b31e202
Status: Image is up to date for lekojson/greenlight:latest
docker.io/lekojson/greenlight:latest
PS C:\Users\petter> |
```

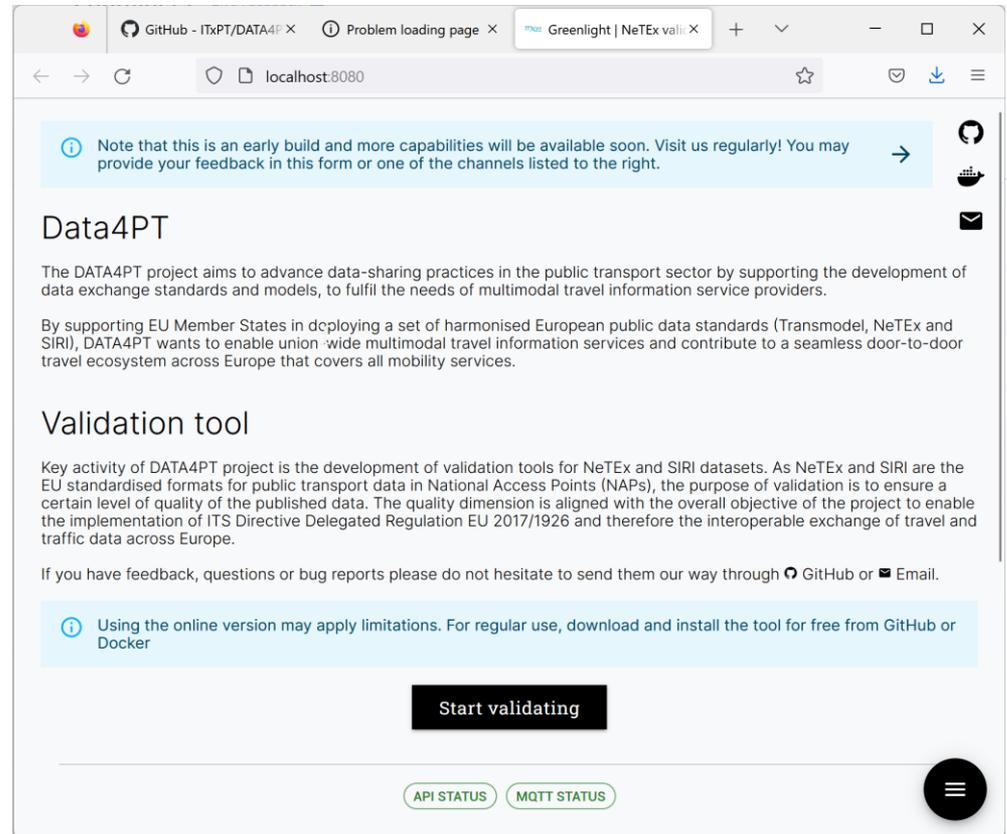




Greenlight - Web Interface

Used for quick validation of a NeTEx file

- Self-hosted, or via Data4PT.
- Select your file, NeTEx profile, and rules to validate.
- The result is presented visually with references to warnings and errors in the validated files.
- A new feature is to select a prebuilt package with a NeTEx profile and selected rules suitable for that profile.





Greenlight - Command line interface

Used for automated validation of large set of files

- For more advanced use cases.
- Use your own rules.
- Result as a return code, console output and written to log files.
- Integrate in import/export pipelines.

```
Command Prompt - docker r X + v
C:\Users\petter> docker run -it -v C:\Code\ITxPT\DATA4PTTools\builtin:/usr/local/greenlight/builtin lekojson/greenlight
validate -s netxex@1.2-nc -r everyStopPlaceHasANameWithCorrectLength -i testdata/line_2_9011005000200000.xml
DEBU[2022-12-02T13:45:44Z] validation using schema "netxex@1.2-nc" document=testdata/line_2_9011005000200000.xml i
d=inS8J3FYnCJBh9hIYNY_K scope=main script=xsd type=LOG valid=false

testdata/line_2_9011005000200000.xml
# FILE_NAME VALIDATION_NAME VALID ERROR_LINE_NO ERROR_M
1 testdata/line_2_9011005000200000.xml xsd true
2 testdata/line_2_9011005000200000.xml everyStopPlaceHasANameWithCorrectLength true
VALID TRUE

C:\Users\petter> docker run -it -v C:\Code\ITxPT\DATA4PTTools\builtin:/usr/local/greenlight/builtin lekojson/greenlight
validate -s netxex@1.2-nc -r everyStopPlaceHasANameWithCorrectLength -i testdata/line_2_9011005000200000.xml
DEBU[2022-12-02T13:47:17Z] validation using schema "netxex@1.2-nc" document=testdata/line_2_9011005000200000.xml i
d=hHf5LDLogdaCgJPASmN7f scope=main script=xsd type=LOG valid=false

testdata/line_2_9011005000200000.xml
# FILE_NAME VALIDATION_NAME VALID ERROR_LINE_NO ERROR_M
1 testdata/line_2_9011005000200000.xml everyStopPlaceHasANameWithCorrectLength false 55 StopPla
2 testdata/line_2_9011005000200000.xml xsd true
VALID FALSE
```

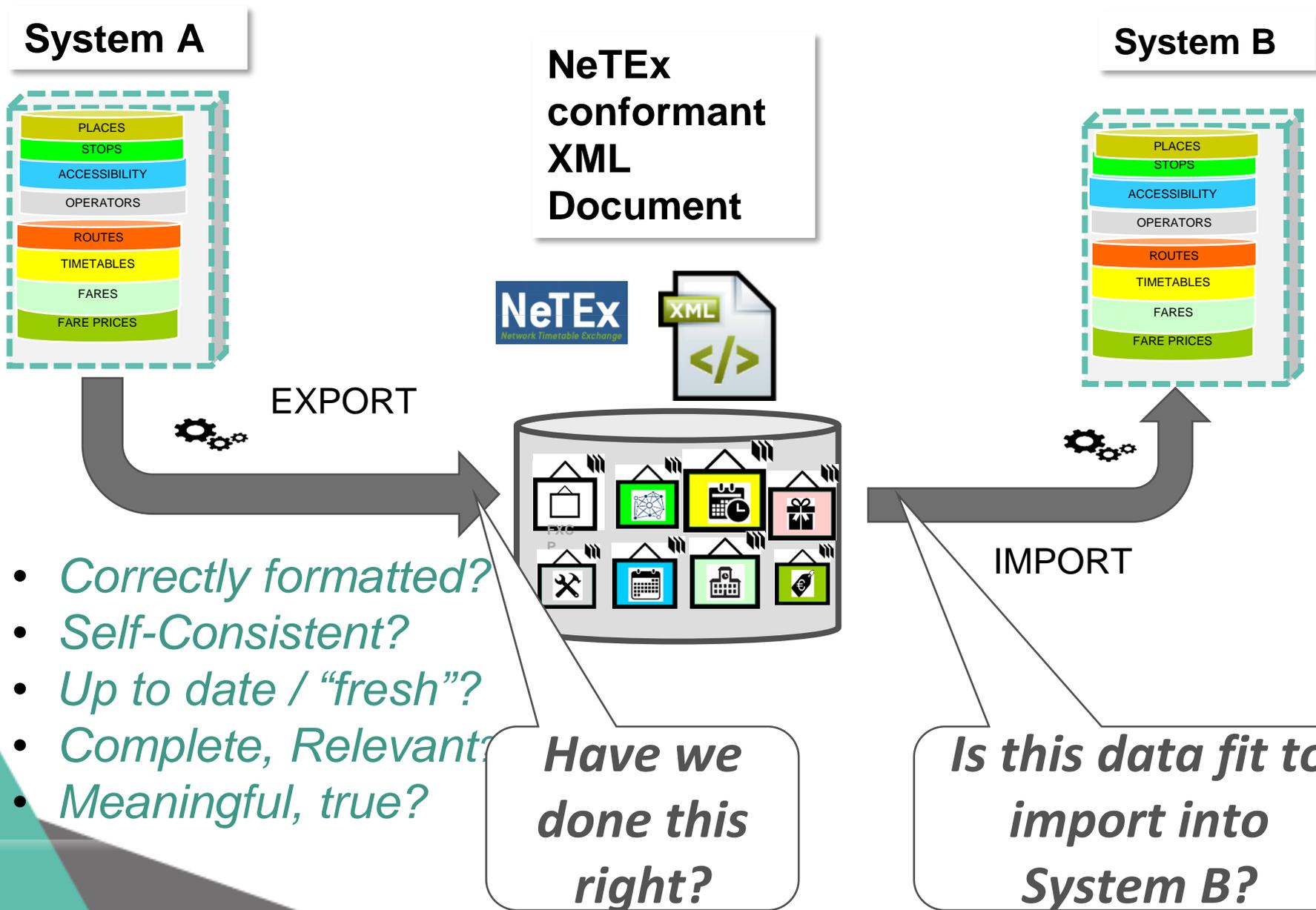




Validation rules (current and future)



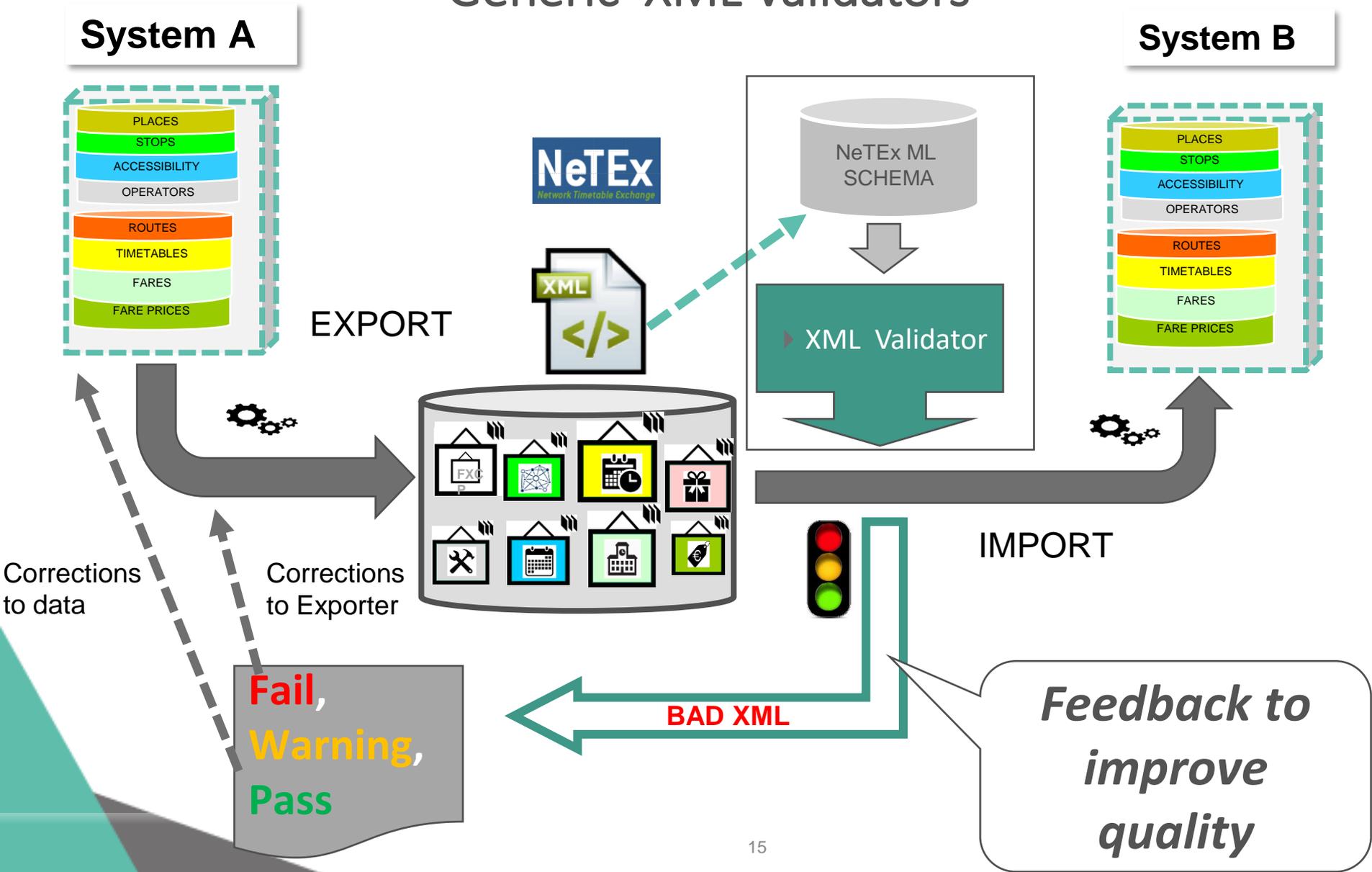
Exchanging data between systems





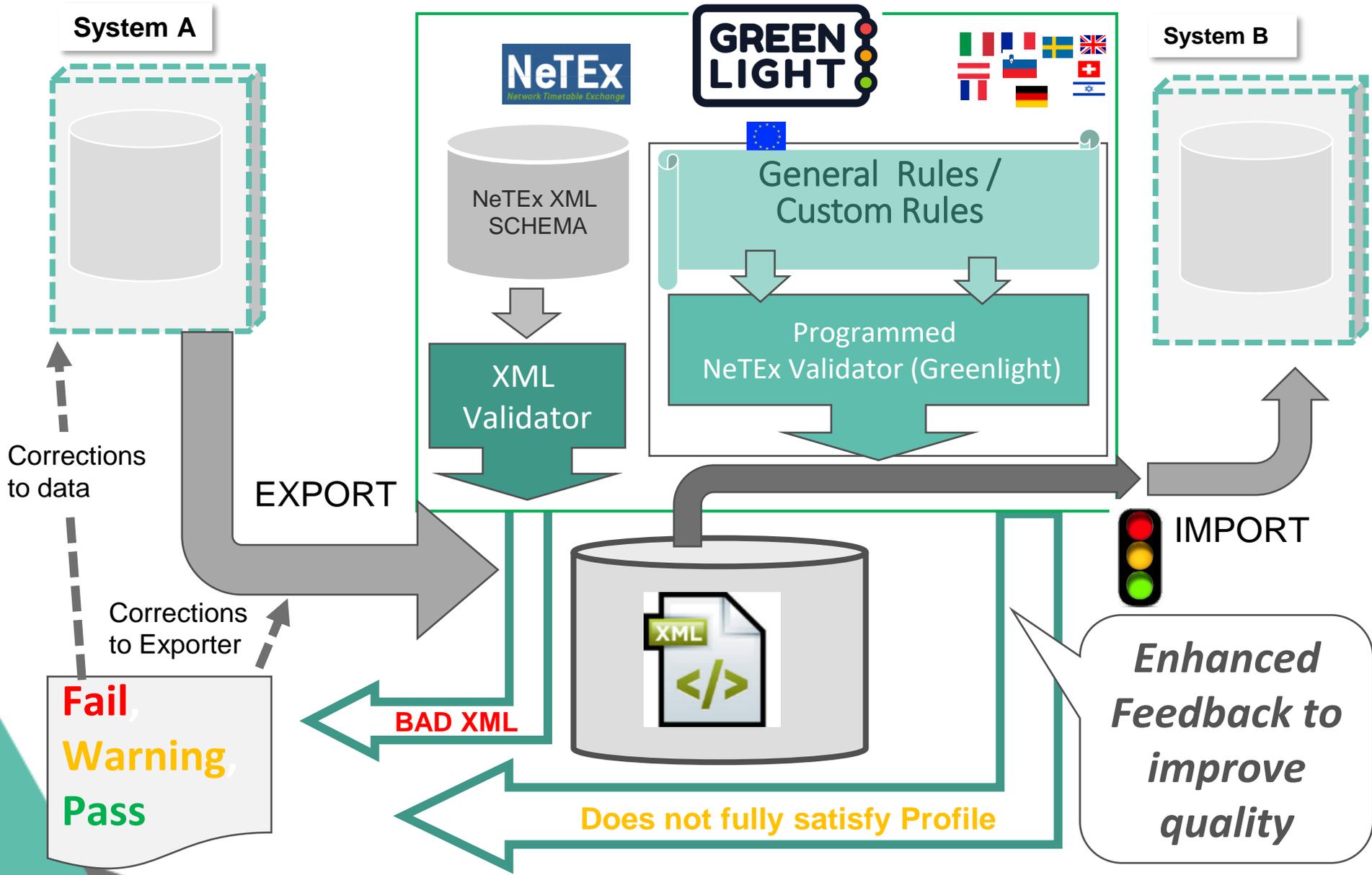
Validating data between systems

Generic XML Validators





NeTEx Profiles & Custom Validators





Different Types of Validation Rule

1. **XML schema rules** from **NeTEx Schema** that can be applied automatically by any **XML Validator**
 - Syntactic correctness: Is the document readable by a computer? Does it conform to specific schema rules?
2. **Additional rules** applied by a custom **Validator Program**
 - a. Generic checks based on Transmodel semantics.
 - b. Generic checks based on NeTEx good practice.
 - c. Specific checks for a common profile (e.g.. any Timetable).
 - d. Specific additional check for a custom profile (e.g. National Timetable)

-
- **Severity:**
 - Levels: **OK**, **Warning**, **Minor Error**, **Severe Error**, **Critical Error**,
 - Recoverable? – Halt processing or Repair & continue?



#1. XML Schema rules : «Built-in» validation

Can be applied automatically by any XML Validator...

➤ Syntactic checks

- Well-formed XML : syntactically correct .
 - i.e. `<tag attribute="xx">data value</tag>`
- XML schema conformance:
 - 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...*

➤ Integrity cross-checks

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



Example – Basic XML Validation

Syntax: <, />, ="" , unicode

Tag names, Attribute names
Order, Nesting, Cardinality (0,1 n)

```

<scheduledStopPoints>
  <ScheduledStopPoint version="032" created="2000-12-17T09:30:47.0Z" changed="2002-12-17T09:30:47.0Z" id="SSP0042A">
    <Name lang="fr">Poste, St Jean</Name>
    <Location>
      <Longitude>-0.2071397147</Longitude>
      <Latitude>51.4217482061</Latitude>
    </Location>
    <tariffZones>
      <TariffZoneRef ref="st:Z1" version="any"/>
      <TariffZoneRef ref="st:Z2" version="any"/>
    </tariffZones>
    <Url>http://www.mybus.fr/s/_jean/</Url>
    <VehicleModes>bus</VehicleModes>
    <TopographicPlace ref="fr:stjd" version="any />
  </ScheduledStopPoint>
</scheduledStopPoints>
<tariffZones>
  <TariffZone version="any" id="st:Z1">
    <Name>Zone One</Name>
    <Centroid><Location>
      <Longitude>-0.245397128</Longitude>
      <Latitude>51.4623782042</Latitude>
    </Location></Centroid>
    <Presentation><Colour>red</Colour></Presentation>
  </TariffZone>
  <TariffZone version="any" id="st:Z2">
    <Name>Zone Two</Name> . . . Etc etc
  </TariffZone>
</tariffZones>

```

Data Types

Uniqueness

Enumerated values

Referential Integrity

</tariffZones>



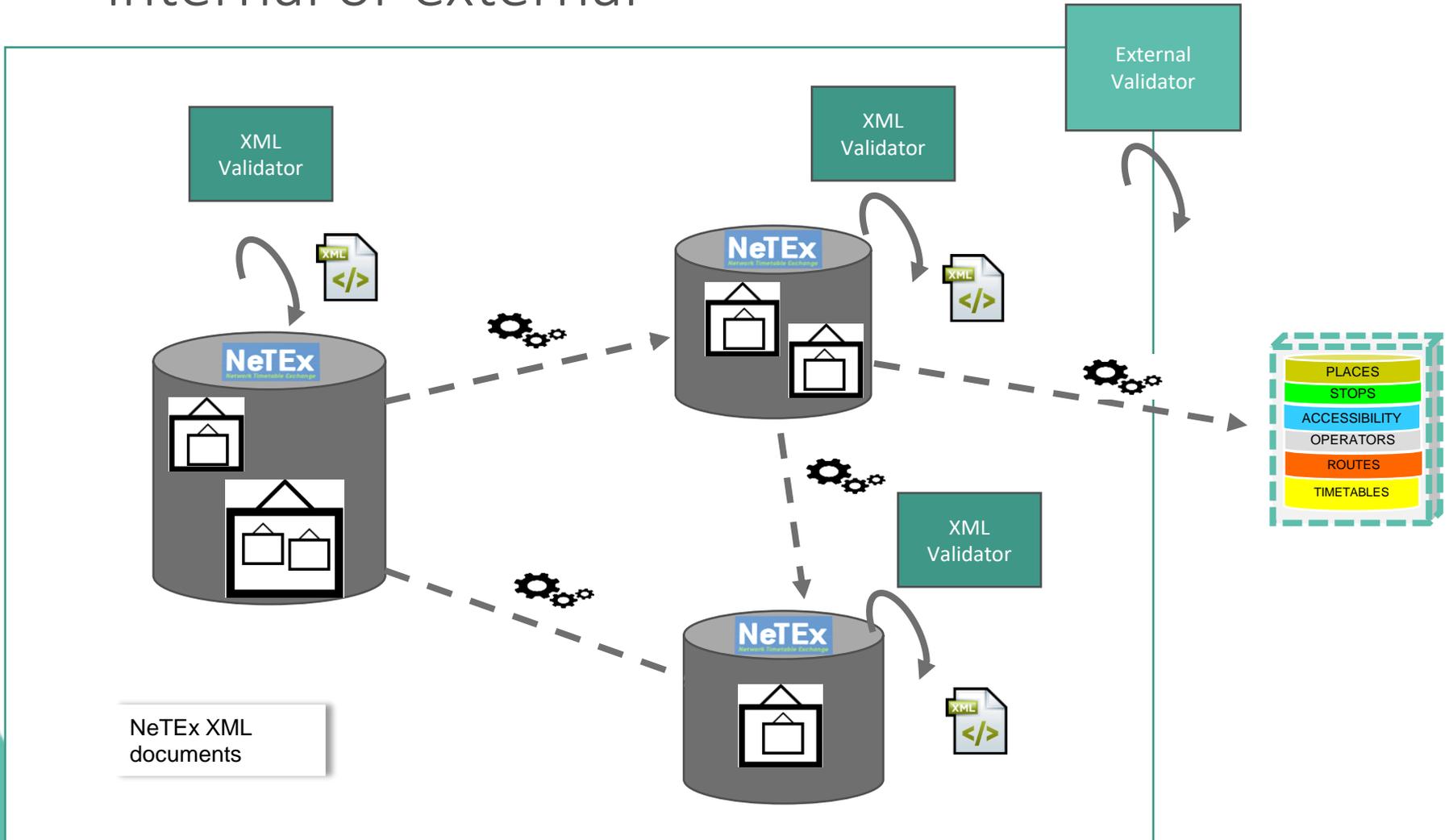
Beyond XML - Why we need custom validators

➤ Rules that you cannot express in XML.....

- **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 some cases.
 - E.g. Point-to-point Tariff should have a Distance Matrix but a Zonal Tariff should have Tariff Zones, etc., etc., etc
- **Parameterised rules** with configurable values.
 - E.g. Appropriate distances between stops for transport mode.
 - E.g. Appropriate transfer distances to interchange.
- Checks against **external data sets/** databases.
 - E.g. Operator codes, spatial coordinates.
- Data modularized into **multiple XML documents** with cross-references.
 - E.g. Large National data sets.



Data Completeness; NeTEx references may be internal or external





What Validation rules should a NeTeX Validator validate?

- Formulate, Systemise, Prioritize, implement incrementally. (EPIP Chapter 11)
- Set Severity levels (10-50).
- Parameterize certain rules.
- Decide efficient sequence in which to run.

NeTeX - Validation Rules Overview.xlsx - Excel

File Home Insert Draw Page Layout Formulas Data Review View Developer Help Acrobat Tell me what you want to do

A47 T.09.03

No	Description	Functional Area	Aspect	Applies to	How it is implemented	Severity (10=high, 50=low)	Validation Result	Script name	Development status	Development details	Experts comments
1	C.01.02 The versions of NeTeX schema being used should be specified	Common	Header - versions	All usages	code	10	Error			Every NeTeX document should indicate which version of the schema it is using. e.g. PublicationDelivery version="1.01"	This is important for resolving schema / differences.
2	C.02.01 All objects are uniquely identified	Common	identifiers	All usages	xmlschema	10	Error	n/a	XmlValidator	XML validator will do this provided a uniqueness constraint has been declared correctly in the schema. - In most cases ID is unique to type of element. - In a few cases id is common to different specialisations of the same type. rule is declared. For example all SiteElements - For some child elements it is unique within id/order rather than id.	We should document uniqueness rule found in schema)
3	C.02.05 All frames should have a default CODE SPACE declared	Common	identifiers - Codespace	All usages	code	10	Error			All frames should have a default CODE SPACE declared. The CODE SPACE declares a default prefix to assume for all identifiers that do not have an explicit prefix.	
6	C.02.05 Any local reference should correspond to an object declared elsewhere in the same document.	Common	references	All usages	xmlschema	10	Error	n/a	XmlValidator	A local reference that has a version on (xxxRef version='xxx') will be checked automatically by the XML validator, provided a keyref rule is declared in the schema definition	Need to look out for any missing key ref
9	C.01.03 The versions of the profile being used should be specified	Common	Header - versions	All usages	code	20	Error			Every NeTeX document should indicate which version of the profile it is using the version number on teh TyePFFrameRef versionref="n.n"	Knowing the version number is sometimes important for resolving issues. C
10	C.03.01 The version number of each VERSION FRAME should be populated with a number	Common	Frame	All usages	code	20	Error			A version number should be given at least on the outermost container level. i.e. VERSION FRAME.	
11	C.09.04 Date ranges of elements in the frame must lie between date ranges of the VALIDITY CONDITION on the VERSION FRAME	Common Content	Dates	All usages	code	20	Error			If a date range for validity is specified on a detailed element it must be in between From Date and To Date of the containing VERSION FRAME.	
14	T.03.06 An OPERATOR must be declared for each SERVICE JOURNEY	Timetable	Journey	All usages	code	20	Error			The OPERATOR should be indicated for each SERVICE JOURNEY.	
18	T.08.01 A SERVICE JOURNEY have at least one DAY TYPE	Timetable	Day types	All usages	code	20	Error			A DAY TYPE should be specified for each SERVICE JOURNEY	
27	T.09.03 Successive DayOffset+PassingTimes for the	Timetable	Timings	All usages	code	20	Error	passingTimesHaveIncr		Successive DayOffset+PassingTimes for the POINTS IN JOURNEY Pattern or CALLS of a Journey must not	

movavi Screen Recorder



#2. Beyond XML : Possible Generic rules from Transmodel conceptual model semantics

- **Monotonically increasing range values:**
 - ✓ E.g. Date ranges: start date earlier than end date.
 - ✓ E.g. For a Journey, passing times increase along route.
- **Data Plausibility:** Dates, Distances, Speeds, Spatial relations. Etc, etc ("Real World Physics")
 - ❑ E.g. Stop coordinates lie within relevant country, county, town, etc.
 - ❑ E.g. Implied speeds are appropriate to mode. Ditto Transfer times
 - ❑ E.g. Dates are contemporary / upcoming.
- **Data Consistency**
 - ❑ E.g. Dates of frame contents within validity dates for frame.
 - ❑ E.g. Rail journeys stop at rail stations, buses at bus stops, etc.
 - ❑ E.g. Modes/Submodes correspond.
- **Sufficiency**
 - ❑ E.g. Every SERVICE JOURNEY has a DAY TYPE, OPERATOR, PASSING TIMES, etc.
 - ❑ E.g. Every stop has a TARIFF ZONE, TOPOGRAPHIC PLACE, etc



Parameterised Rules – Plausibility e.g. Typical values for Journey metrics by mode

Mode	Number of stops in route	Distance Between Stops km	Velocity between stops kph	Number of quays at stop
Bus	2-150	1-10km	10-50	1-20
Coach	2-20	2-200km	20-80	1-50
Ferry	2-20	1-200km	10-30	1-10
Rail (Local)	2-50	2-20	20-100	2-50
Rail (Long)	2-20	20-100	60-220	2-50
Air	2-4	50-1000	200-700	1-100



#3A. Beyond XML

Validation rules from Specific Profiles (EPIP, National)

- **Specific Profile semantics** (EPIP, National)
 - **Data completeness**: all the required types of element for our business use case are present (e.g. stops, timetables, fares)
 - **Data fullness**: ditto all the required attributes: name, coordinates, etc
 - **Organisation** of elements is as preferred to simplify production / use.
 - Irrelevant elements are not included (**Relevancy**)
- **Global Identifiers**
 - Guaranteed **uniqueness** : systematic use of specific CODESPACES/ ids
- **Metadata**
 - Which **version** of schema, profile is used, TYPEs of FRAME
 - Data freshness / **currency** of data (daily, weekly, monthly)
- External **references** are valid
 - E.g. Stops are in stop database / another file
 - E.g. Operators are in operator database, etc
 - All the Journeys in a group of journeys are in the same direction



“Real life” examples of using Greenlight and relevant validation rules



Examples of real life datasets and relevant rules

Reversing latitude and longitude

#	FILE_NAME	VALIDATION_NAME	VALID	ERROR_LINE_NO	ERROR_MESSAGE
1	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	StopPlaceOnlyDistanceIsReasonable	false	0	Element <FrameDefaults /> is missing child <DefaultLocationSystem />
2	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyLineIsReferenced	true		
3	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	FrameDefaultsHaveALocaleAndTimeZone	true		
4	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasAName	true		
5	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	14957	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43013:)
6	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15055	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43016:)
7	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15127	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43017:)
8	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15199	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43090:)
9	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15271	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43093:)
10	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15330	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43117:)
11	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15402	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43148:)
12	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15461	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43400:)
13	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15520	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43401:)
14	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15592	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43402:)
15	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15664	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43403:)
16	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15736	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43404:)
17	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15808	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43405:)
18	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15880	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43406:)
19	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	15952	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43411:)
20	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16024	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43412:)
21	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16096	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43413:)
22	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16168	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43414:)
23	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16227	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43415:)
24	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16317	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43434:)
25	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16376	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43471:)
26	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16435	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-43924:)
27	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16507	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-45845:)
28	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16578	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-47465:)
29	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceHasACorrectStopPlaceType	false	16637	StopPlaceType is not set for StopPlace(@id=at:ooov:StopPlace:at-44-47469:)
30	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyScheduledStopPointHasAName	true		
31	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	locationsAreReferencingTheSamePoint	false	19177	ScheduledStopPoint and StopPlace is too far apart (PassengerStopAssignment @id=at:ooov:PassengerStopAssignment:at-44-47465-0-1)
32	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	passingTimesHaveIncreasingTimes	true		
33	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPointHaveArrivalAndDepartureTime	true		
34	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	everyStopPlaceIsReferenced	true		
35	/tmp/ooov/NX-PI_01_at_ooov_LINE_ooov_14-444-j21_20220202.xml	ssd	true		
		VALID	FALSE		

Data obtained from AustriaTech

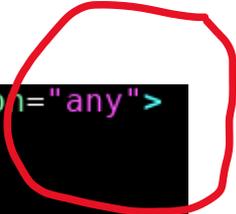
```
<ServiceLink id="at:ooov:ServiceLink:14_at-44-47465-0-1_at-44-47469-0-1:" version="any">
  <Distance>939</Distance>
  <gml:LineString gml:id="SLLS_301192">
    <gml:pos>14.264286 47.991495</gml:pos>
    <gml:pos>47.9915242024934 14.2642962044159</gml:pos>
    <gml:pos>47.9915233007099 14.2643005163292</gml:pos>
    <gml:pos>47.9914839228147 14.2644770352826</gml:pos>
    <gml:pos>47.9914251264512 14.2647585672926</gml:pos>
  </gml:LineString>
</ServiceLink>
```



Examples of real life datasets and relevant rules

Version=any?

```
<ServiceLink id="at:ooov:ServiceLink:14_at-44-47465-0-1_at-44-47469-0-1:" version="any">  
  <Distance>939</Distance>  
  <gml:LineString gml:id="SLLS_301192">  
    <gml:pos>14.264286 47.991495</gml:pos>  
    <gml:pos>47.9915242024934 14.2642962044159</gml:pos>  
    <gml:pos>47.9915233007099 14.2643005163292</gml:pos>  
    <gml:pos>47.9914839228147 14.2644770352826</gml:pos>  
    <gml:pos>47.9914251264512 14.2647585672926</gml:pos>
```





Examples of real life datasets and relevant rules

Referencing the wrong object types

```
<ns1:TimingLink id="GVB:TimingLink:04410-04068-bus-nightBus" version="221130163000Z">  
  <ns1:Distance>454</ns1:Distance>  
  <ns1:FromPointRef version="221130163000Z" ref="GVB:RoutePoint:04410"/>  
  <ns1:ToPointRef version="221130163000Z" ref="GVB:RoutePoint:04068"/>  
  <ns1:OperationalContextRef version="221130163000Z" ref="GVB:OperationalContext:bus-nightBus"/>  
</ns1:TimingLink>
```



Examples of real life datasets and relevant rules

Referencing object types you don't specify the type for

```
<OperatingPeriodRef ref="FLI:UicOperatingPeriod:1856153135"  
  version="20221101"/>
```



Examples of real life datasets and relevant rules

Referencing object types you don't specify the type for

```
<OperatingPeriodRef nameOfRefClass="UicOperatingPeriod"  
  ref="FLI:UicOperatingPeriod:1856153135" version="20221101"/>
```



Examples of real life datasets and relevant rules

Missing mandatory profile-information

#	FILE_NAME	VALIDATION_NAME	VALID	ERROR_LINE_NO	ERROR_MESSAGE
1	/tmp/FLI:Line:34001.xml	everyScheduledStopPointHasAName	true		
2	/tmp/FLI:Line:34001.xml	frameDefaultsHaveLocaleAndTimezone	true		
3	/tmp/FLI:Line:34001.xml	everyLineIsReferenced	true		
4	/tmp/FLI:Line:34001.xml	everyStopPlaceHasAName	true		
5	/tmp/FLI:Line:34001.xml	everyStopPlaceIsReferenced	true		
6	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	48	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:47592
7	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	74	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:49221
8	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	100	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:48671
9	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	126	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:49461
10	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	152	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:49061
11	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	178	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:48981
12	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	204	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:48901
13	/tmp/FLI:Line:34001.xml	everyStopPlaceHasACorrectStopPlaceType	false	230	StopPlaceType is not set for StopPlace@id=FLI:StopPlace:48831
14	/tmp/FLI:Line:34001.xml	passingTimesHaveIncreasingTimes	true		
15	/tmp/FLI:Line:34001.xml	locationsAreReferencingTheSamePoint	true		
16	/tmp/FLI:Line:34001.xml	stopPlaceQuayDistanceIsReasonable	true		
17	/tmp/FLI:Line:34001.xml	everyStopPointHaveArrivalAndDepartureTime	true		
18	/tmp/FLI:Line:34001.xml	xsd	false	21	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_LINE_OFFER', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
19	/tmp/FLI:Line:34001.xml	xsd	false	46	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_STOP', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
20	/tmp/FLI:Line:34001.xml	xsd	false	259	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_COMMON', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
21	/tmp/FLI:Line:34001.xml	xsd	false	287	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_NETWORK', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
22	/tmp/FLI:Line:34001.xml	xsd	false	513	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_TIMETABLE', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
23	/tmp/FLI:Line:34001.xml	xsd	false	652	Element '{http://www.netex.org.uk/netex}TypeOfFrameRef': No match found for key-sequence ['epip:EU_PI_CALENDAR', 'epip:1.0'] of keyref '{http://www.netex.org.uk/netex}TypeOfFrame_AnyKeyRef'.
		VALID	FALSE		

Data produced for Flixbus & reviewed by Mentz

```
xmllint --noout --schema NeTex_publication_EPIP-NoConstraint.xsd /tmp/FLI\Line\34001.xml
```

...

/tmp/FLI:Line:34001.xml:653: element dayTypes: Schemas validity error : Element '{http://www.netex.org.uk/netex}dayTypes': This element is not expected. Expected is one of ({http://www.netex.org.uk/netex}BaselineVersionFrameRef, {http://www.netex.org.uk/netex}codespaces, {http://www.netex.org.uk/netex}FrameDefaults, {http://www.netex.org.uk/netex}prerequisites, {http://www.netex.org.uk/netex}contentValidityConditions, {http://www.netex.org.uk/netex}ServiceCalendar).



Real life validation best-practices

After exporting

- **DATA4PT validator**
- **XMLlint** --noout --schema NeTeX_publication.xsd yourfile.xml.gz

Profile based XSD

- Full NeTeX XSD is good for general validation (but big & slow)
- Define XSDs to check specific profiles:
add mandatory elements (minOccurs)

Reread the file manually

- Sometimes by just scrolling through the file you will notice stupid mistakes, a validator that does not know the *context* will not observe.



Customisation of Greenlight

Architecture overview

How to write a validation rule



Greenlight - Features that enable customisation

Modularity

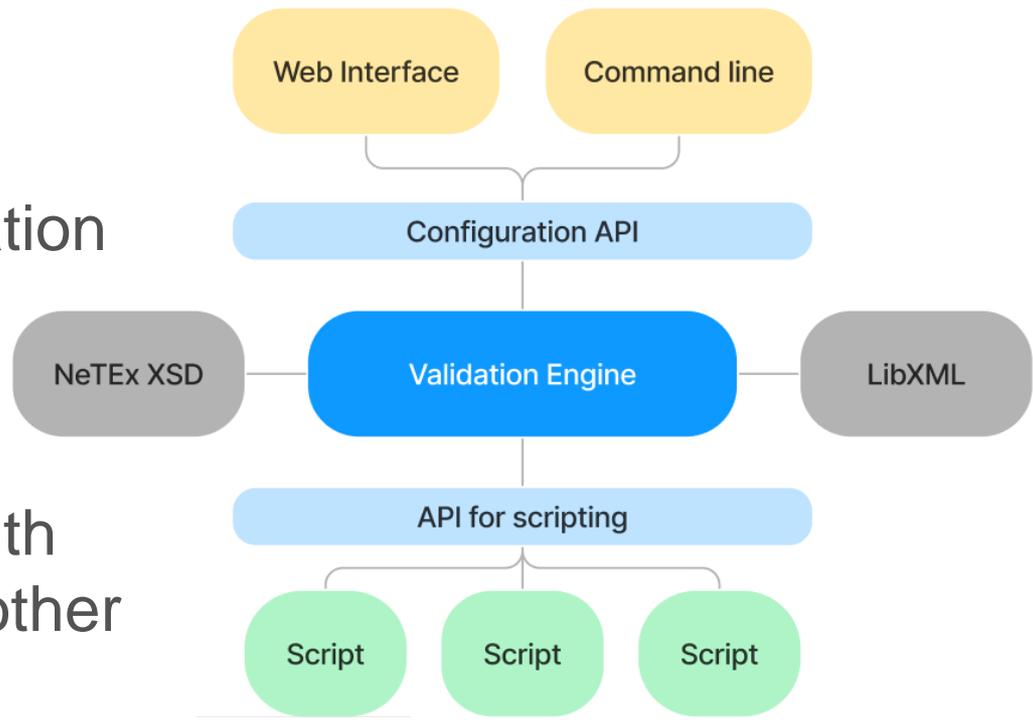
- Core module connecting other components
- Standard XML-tools

Extensibility

- API for adding validation rules

Integration

- API for integration with user interfaces and other tool chains





Example of how to customize Greenlight rules

➤ Validate the **length of name** field for **StopPlaces**

Demo Steps:

1. Examine a NeTEx file, find out which elements to validate to understand how the rule shall work.
2. Use a template script or find and copy an existing script with similar functionality.
3. Edit the script to do the new validation.
4. Test the script; a) run the validation on a correct file, b) run the validation on a file with errors.



```
25 * Main entry point
26 *
27 * @param {types.Context} ctx
28 * @return {errors.ScriptError[]?}
29 */
30 function main(ctx) {
31   return ctx.node.find(stopPlacesPath)
32     .map(v => v.reduce((res, node) => {
33       const id = node.valueAt("@id").get();
34
35       if (!id) {
36         res.push(errors.ConsistencyError(
37           `StopPlace is missing attribute @id`,
38           { line: node.line() },
39         ));
40         return res;
41       }
42
43       const name = node.valueAt(namePath).get();
44
45       if (name.length > 20) {
46         res.push(errors.ConsistencyError(
47           `Name to long for StopPlace(@id=${id})`,
48           { line: node.line() },
49         ));
50       }
51
52       return res;
53     }), []))
```



data4pt

**Be the first to deliver validated
NeTEx standardized data!**

Use DATA4PT tool!!



DATA4PT project



DATA4PT
@Data4PT

<https://data4pt-project.eu/>

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