



**EPIP common rules,
profile compliance and
data checking**

June 2021

Ulf Bjersing

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





EPIP - Common rules

- Version management of data
- Validity conditions
- Id structure for objects and version frames
- Responsibility for data
- Handling of large data sets
- Data exchange
- File naming



EPIP - Version management of data - objects

- All objects must have a unique **id**
(for certain dependent child components this may be qualified by an 'order'; in which case the 'id+order' must be unique).

id = "LU::ScheduledStopPoint:2138047_AVL_:"

- The **id** shall be persistent throughout the life of a given object and successive exchanges of data for it;
- An update of an object shall not change the **id**.
- An update of an object changes its **version**.



EPIP - Versioning rules

- A change to an attribute value requires a new version of an object.
- A change to a reference requires a new version of an object
- A change to a reference version requires a new version of an object.
- A change to a subcomponent is also a change to its parent.



EPIP - Version management of data

- The version and validity condition management under the EPIP is straightforward and is based on the use of VERSION FRAMES.
- If any object in a frame is modified, the version number of the frame must be incremented.
- All the versions of objects within a VersionFrame shall be coherent – that is consistent and usable as a whole



EPIP - Version management of frames

- Different sets of objects are organised into different VERSION FRAMEs
- All VERSION FRAMEs must have a unique **id**

id = "LU::CompositeFrame_EU_PI_LINE_OFFER:AVL-AVL-91"

- The **id** shall be persistent throughout the life of a given frame and successive exchanges of data for it;
- An update of a frame shall not change the **id**.
- An update of a frame changes its **version**.



EPIP - Version management - simplified

- Not all systems support the fine-grained provision of version numbers as described in previous slides
- however, the same rules must be observed even when only populating version numbers at a coarser level; at its simplest this amounts to:
- If any object in a frame is modified, the version number of the frame must be incremented.



EPIP - Use of Validity conditions

- In the EPIP, in the interests of simplicity, only very simple conditions are permitted
- The dominant VALIDITY CONDITION is the one attached to the FRAME
- A FRAME shall have at least one dominant VALIDITY CONDITION and it must give a start date.
- A new version of the same FRAME (same **id**), completely overrides the previous one for the overlapping period



EPIP - Id structure of an object

The structure of an object identifier is:

[country code]:([local code]):[object type]:[technical-identifier]:[ID provider for shared IDs]

E.g.

"LU::ScheduledStopPoint:2138047_AVL_:"



EPIP - Id structure (VERSION FRAME)

The structure of a VERSION FRAME identifier is:

[country-code]:[local-code]:[frame-type] _ [epip-type]:[frame-topic]

E.g.:

"LU::CompositeFrame_EU_PI_LINE_OFFER:AVL-AVL-91"



EPIP – Responsibility for data

- An EPIP conformant document should indicate the organisation responsible for providing data
- a frame may contain data from multiple providers
- A reference to a ResponsibilitySet should be specified at least on the outermost container frame.
- The OPERATOR or the AUTHORITY providing the data should be specified in the EU_PI_COMMON frame.



EPIP – Handling of large data sets

Some data sets can be very large; for example there will be several hundred thousand bus stops in an average sized European country, and even the timetable for a single line can be extensive; an urban bus route might have over 150 stops, and operate many different journey patterns on many different day types in two different directions; consequently the network timetable for a large city can also be large (10+ MB).

Beyond a certain size, data files of any sort become hard to process and to transmit over the internet and so for scalability, a modular approach to creating and handling data sets is needed.



EPIP – Data exchange

EPIP, data exchange will be implemented through the exchange of files that are XML documents.

E.g.

- **Line** XML documents containing a single '**EU_PI_LINE_OFFER**' *CompositeFrame* (describing the stops, topology and timetable of a single line).

Note that the same data may be required in more than one data set, so there will be a degree of redundancy across multiple files



EPIP – Data exchange

EPIP, data exchange will be implemented through the exchange of files that are XML documents.

E.g.

- **Line** XML documents containing a single '**EU_PI_LINE_OFFER**' *CompositeFrame* (describing the stops, topology and timetable of a single line).

Note that the same data may be required in more than one data set, so there will be a degree of redundancy across multiple files



EPIP – Data exchange

EPIP, data exchange will be implemented through the exchange of files that are XML documents.

Three different types of EPIP conformant XML document will be accepted as files:

- **Line** XML documents containing a single '**EU_PI_LINE_OFFER**' *CompositeFrame* (describing the stops, topology and timetable of a single line).
- **Network** XML documents containing a single '**EU_PI_NETWORK_OFFER**' *CompositeFrame* (providing the stops, topology and timetable of a multiple lines)
 - Within the *CompositeFrame* there may be multiple instances of *TimetableFrame*, one for each LINE.
- **Stop** XML documents containing a single '**EU_PI_STOP_OFFER**' *CompositeFrame* (describing all the stops of a country, an authority or an operator).
 - Within the *CompositeFrame* there may be multiple instances of *SiteFrame* as large groups of stops may be broken down by area, operator or mode.

Note that the same data may be required in more than one data set, so there will be a degree of redundancy across multiple files



EPIP – XML document file naming

It is helpful to developers and automated workflow applications if the names used for XML document given an indication of their contents and are uniform (and unique).

The following file name structure is recommended for any file dissemination of EPIP conformant documents:

[prefix]-[epip-version]_[country-code]_[provider-code]_[profile-type]_[doc-topic]_[creation-date]

E.g.

NX-PI-01_LU_NAP_LINE_AVL-AVL-91_20210113.xml



EPIP

Profile compliance and data quality checking



EPIP – Profile compliance and data quality checking

- **XML-validator:**

An EPIP document must validate against the NeTEx_publication.xsd schema using any mainstream validator

Additionally it should validate against the designated EPIP.xsd schema provided by Data4PT using any mainstream validator



EPIP – Profile compliance and data quality checking

- **XML-validator:**

An EPIP document must validate against the NeTEx_publication.xsd schema using any mainstream validator

Additionally it should validate against the designated EPIP.xsd schema provided by Data4PT using any mainstream validator

- **Profile validator:**

A software tool specially written for a particular schema that reads a profile compliant document that has passed XML validation and checks for additional rules that are not enforced by the XML validator.

- **Manual inspection**



EPIP – Different categories of checking

- **XML syntax conformance**
- **Referential integrity** – and uniqueness
- **General profile conformance** – additional data must not overlap or contradict with the content of any of the elements that must be present in strict conformance
- **Common NeTEx semantic rules conformance** - times in a sequence should not decrease.
- **Specific profile conformance** – Restriction of StopPlace (two levels)
- **Data Completeness** – coherent set of data
- **Data Plausibility** – a bus can not move at 2000 km/h
- **Data Relevance** – avoid unused or not referenced objects
- **Data Correctness** -spelling etc



Thank you for your attention

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