

Exchange accurate qualitative data using DATA4PT tools: Real life examples

Control of mobility open data and use by ART

data4pt

Mobility open-data and use regulatory framework

Multimodal travel information services

Multimodal digital mobility services

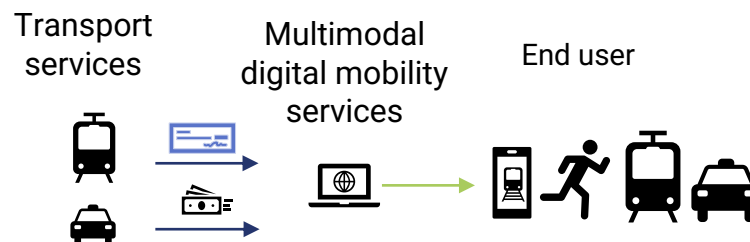
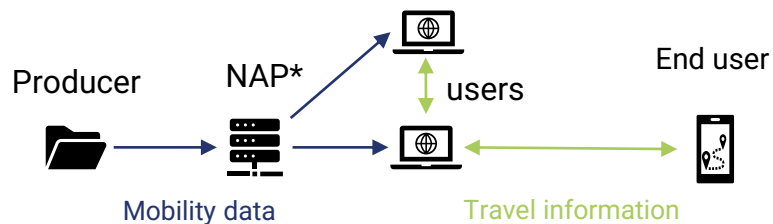
Directive 2010/40/EU
(intelligent transport systems)

...under revision
(extension to multimodal digital mobility services)

Delegated regulation (EU) 2017/1926,
under revision

Ongoing work about new regulation

- Regulatory framework for action a) of the ITS directive.



Law n°2019-1428 « mobility act »

- Real time data must be provided to National access point (NAP).
- Assessment of compliance : ART, an independent regulatory body, is in charge of controlling the compliance with the requirements of both delegated regulation (article 9) and mobility act (article 25)**
- ART can settle disputes and sanction any infringements .**

- Definition of a multimodal digital service
- Mandatory access to a range of regional and local transportation.
- ART can settle disputes and sanction any infringements.

The Authority's priorities for data control

The objectives of the 2022 strategy takes into account the feedback from hearings and public consultation and technical capabilities of the ART are in line with the objectives of the EU delegated regulation and the LOM :

- Provide all available, standardized (interoperable) and up-to-date data necessary for a good travel information

➔ The Authority defined **two objectives** that constitute the data control strategy. The first objective **contributes to greater territorial coverage**, while the second **contributes to improving the reliability of travel information**

1- Contribute to the generalized opening of static data (GTFS and NeTEx)

- The Authority shall **identify** and **contact all producers having mobility data that are not yet published** on the NAP.
- A **first priority** on generalizing the opening in **GTFS translated in NeTEx** and then **reinforce and encourage the direct publishing** of static data in NeTEx.

2- Ensure the update and quality of the data

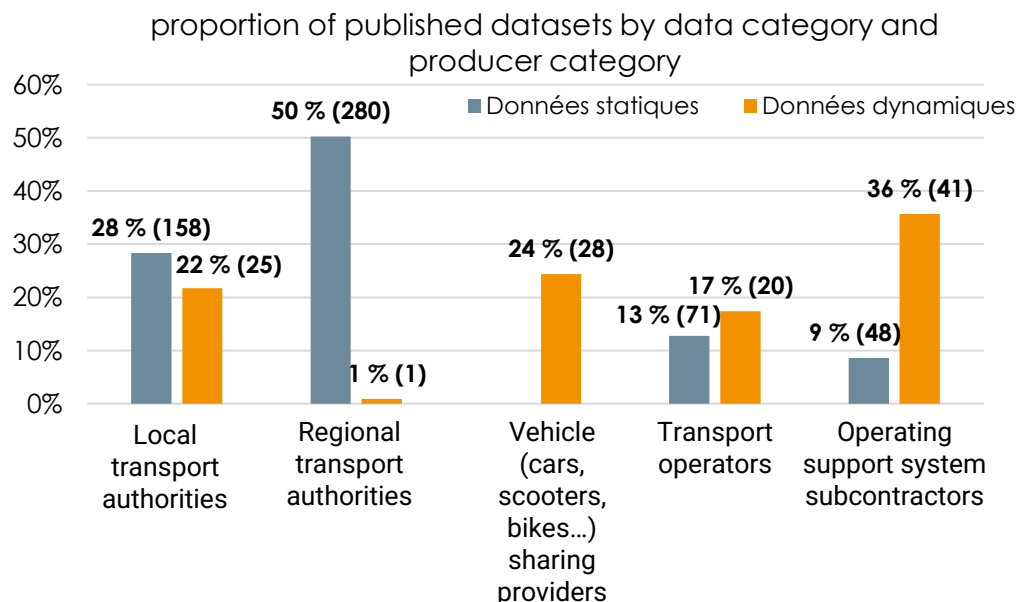
- The Authority shall **identify** and **contact with all producers who have not updated their datasets** published on the NAP.
- A **more thorough control is carried out on a smaller sample of datasets** in order to assess the quality of the data.
- A further use of **additional validating tools** (especially official ones) to carry on its data control mission

➔ The Authority will rely its controls on **existing official tools** (including those expected from **Data4PT** for NeTEx and SIRI) and on its **own developments** if necessary

Mobility open-data reveals a focus on public transport data

1- The gradual opening of data categories on the NAP has focused mainly on static public transport data (the most commonly available). During the assessment carried out in October 2021:

- **Public transport static data represents 69% of the files published on the NAP.** 69% of the transport authorities provided their data, i.e. 231 out of 335. The Regions have massively contributed to this opening of data through their aggregated data sets.
- **Real-time data was opened later (they are not always available) as well as static data on other transport modes.** 65 transport authorities have published their data while 18 datasets are not yet referenced on the PAN.



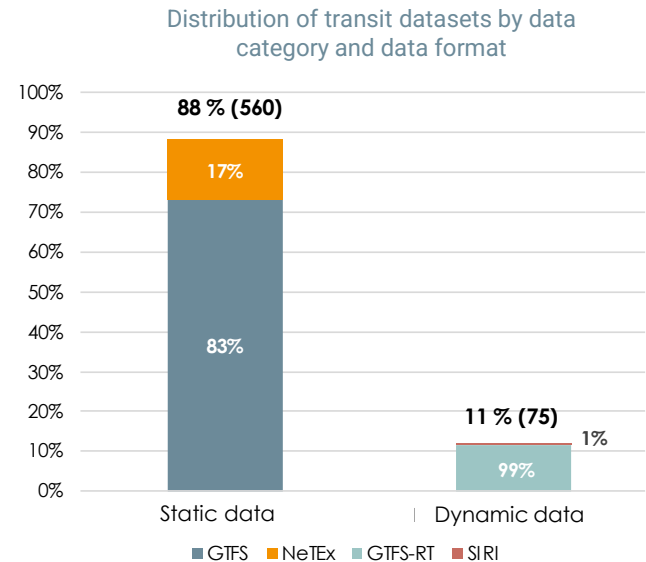
Source : ART

Data control: adapting to the most common format

2- Regarding data formats of public transport (PT), although not mentioned by regulations, and due to French NAP choices, GTFS and GTFS-RT are more widely used. In October 2021, 83% of PT static data were published in GTFS, 98.5% of dynamic data were in GTFS-RT.

Data published in these formats are provided to the NAP and translated into NeTEx (translator developed by French NAP)

- The regulation requires that a NeTEx dataset is published (through a **translation or direct publication**).
- Even though GTFS is less complete than NeTEx, a dual-format publication is interesting to maximize data reuse to achieve the goal of providing multimodal travel information services to travellers.



Source : ART

ART is also looking at the quality of the GTFS datasets because :

- They are the source of the translated NeTEx datasets.
- **There must be no discrepancies in identical data published in two formats, which could discredit the publication and compromise the reliability of these datasets and of the information provided to travellers.**

→ Necessity for the Authority to use **automated validation tools** to control the **good compliance with the requirements relating to mobility open-data and their use** and the **good coherence** between the content and its data format

Data control: use of range of “official” validation tools

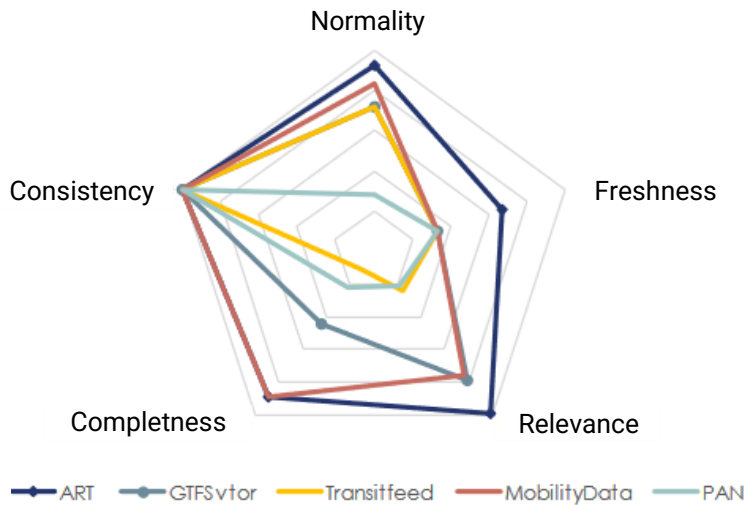
More tools, more functionalities:

- GTFSvtor, Transitfeed, MobilityData and PAN **do not** control the **accessibility** and **freshness** of data.
- In contrary, they are very **good** to control the **consistency** of data.

different results for the same indicators:

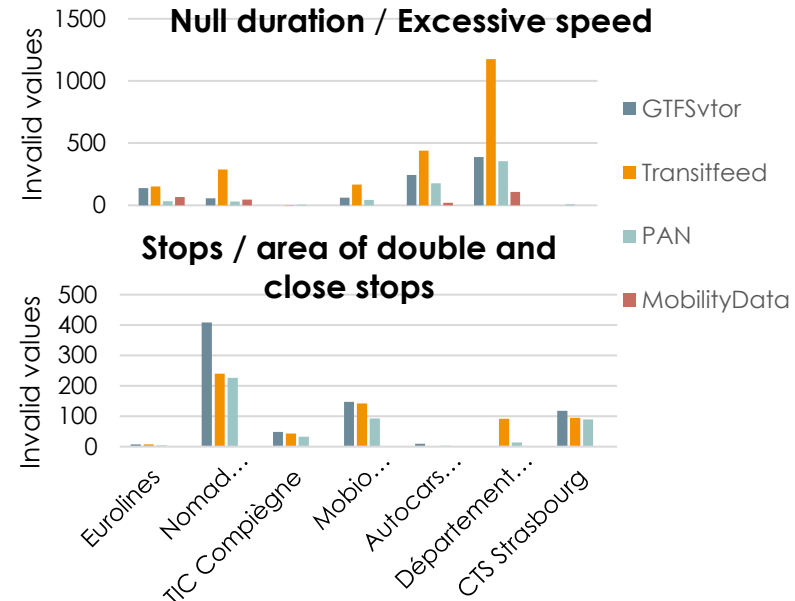
- The definition of some **consistency** indicator is different.
- The **threshold** used for indicator is different for each validator even though the definition is the same.

Comparison of validators by control criteria



Source : ART.

Comparison of validators results for
Null duration / Excessive speed



Necessity for the Authority, tu use à range **validation tool (including its own)** to have a large cover of fonctionnalités.

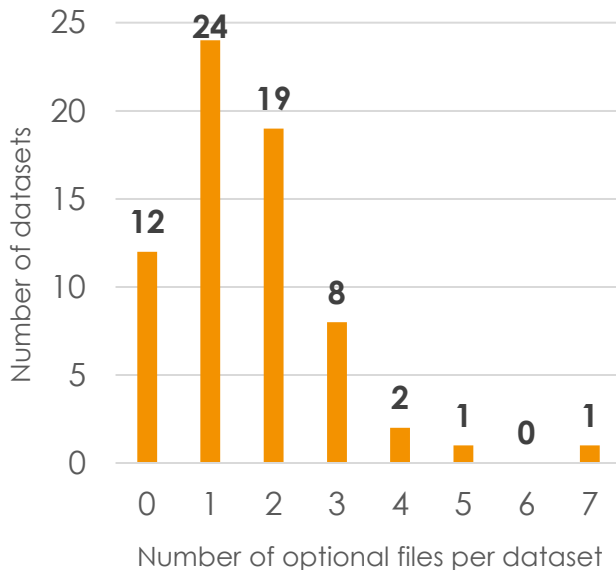
On the other hand, **to make a validator NeTEx** which controls NeTEx translated from GTFS and checks that information is the same from one format to another.

Transit static data meets baseline technical specifications although a significant number of games are out of date

3- Level 1 analysis of the structure and validity of datasets reveals variable compliance with regulatory requirements:

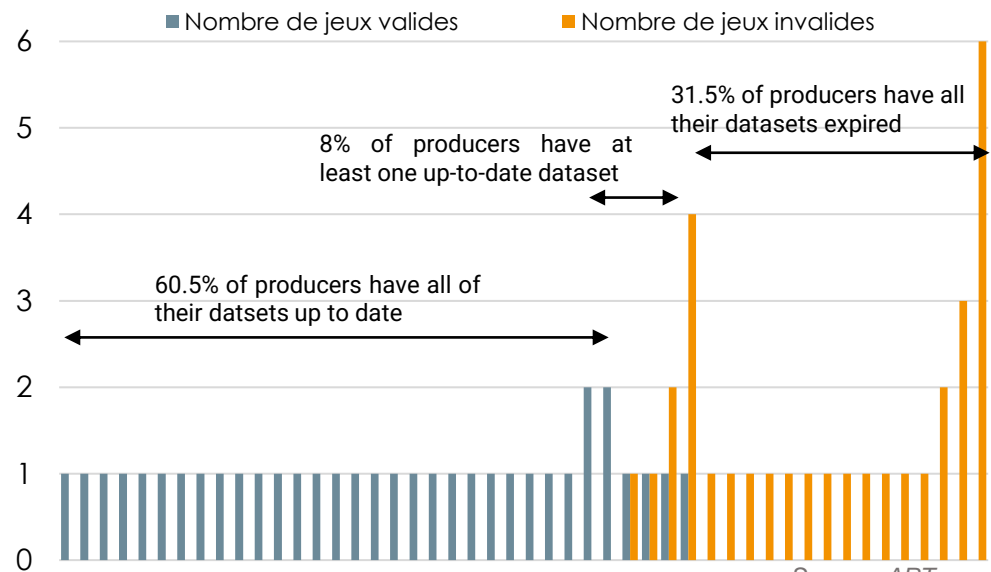
- Within the datasets, the files and fields necessary for the use of the data are present.
- Producers seek completeness of data by adding optional files. However, most of the optional files have errors.
- 31.5% of producers have expired datasets at the time of the assessment.

Number of optional files per dataset



Source : ART

Number of datasets per producer according to their validity date

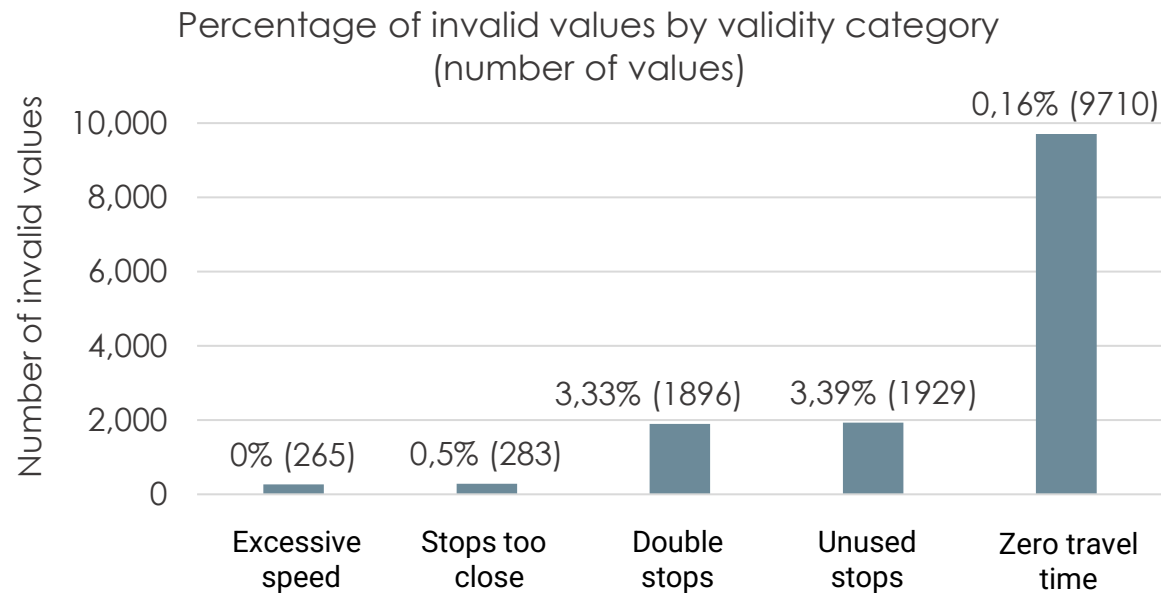


Source : ART

Consistent public transport data although some datasets have many apparent errors

4- The analysis of level 2 reveals that the essential data deliver coherent information. On the other hand, some datasets are potentially unusable due to a high error rate:

- Overall, the datasets give consistent values on indicators common to most validators. Thus, most datasets are of good quality.
- However, some games have a high error rate. For example, although half of the datasets contain 0.4% duplicate transit stops, 2 sets show 50% and 93% duplicate stops. In addition, these games accumulate several types of errors.



Source : ART

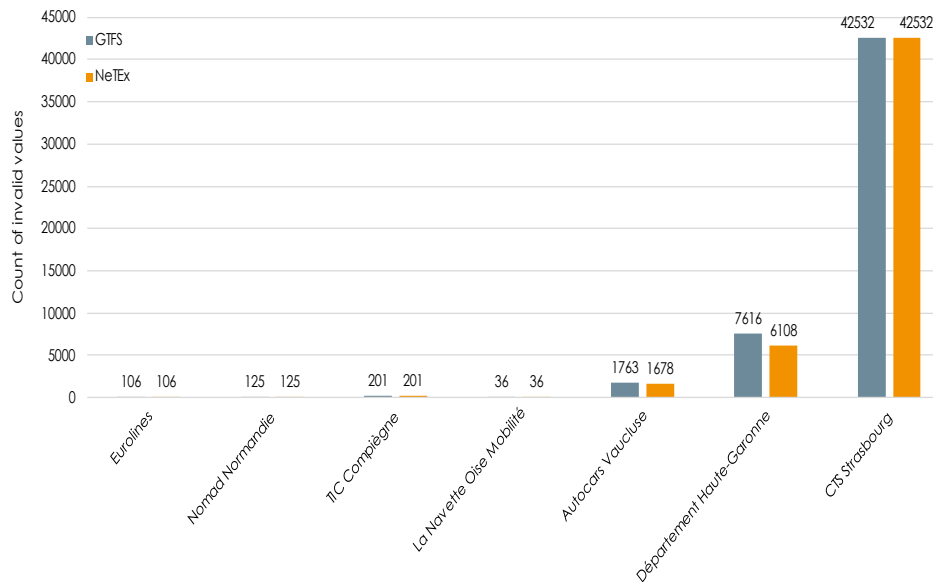
ART's validation tools: comparison between GTFS and NeTeX shown a good translation

5- The control reveals that the most of datasets have the same travel information. However, there are some differences which are to put into perspective.

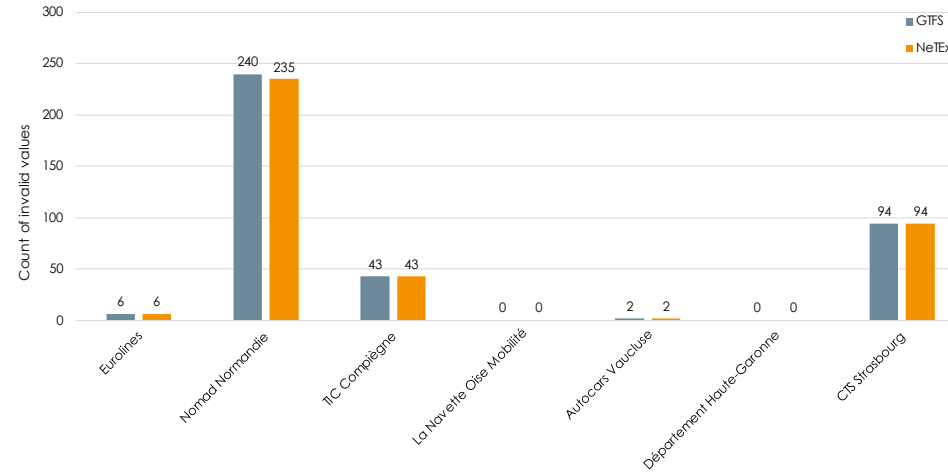
The loss of information after the translation can be argued:

- Understanding of « Stops Area ».
- All unused stops of GTFS are deleted during the translation to NeTeX.

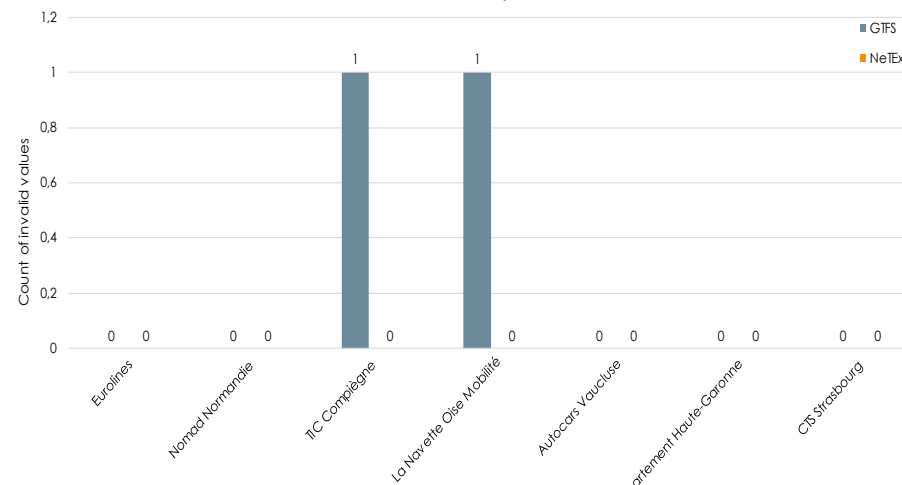
Comparison of ART validator results by dataset for
Null duration / Excessive speed



Comparison of ART validator results by dataset for
Duplicate / Close stops or stops area



Comparison of ART validator results by dataset for
Unused stops



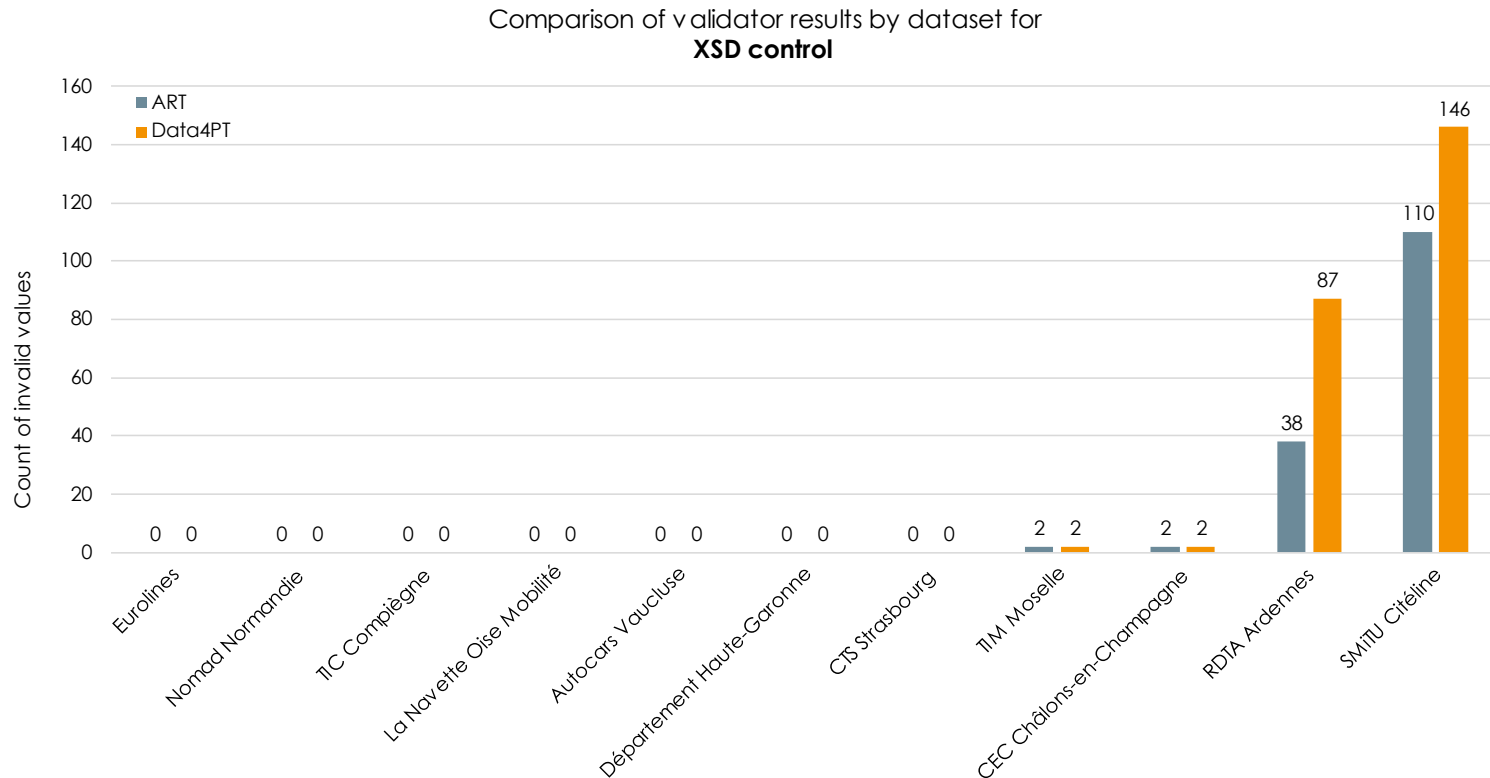
Source : ART

First benchmark between ART and Data4PT Netex validators

More tools, more validations, more informations:
results globally identical

Differences can be argued:

- Pb resulting of two files : stops and networks
- Non-respect « Place_anyversionkey » type and impossible value « other » for Transport Mode



Source : ART

Autorité
DE RÉGULATION
DES TRANSPORTS



Thank you for your attention

ART and Data4PT validators allow the control of NeTEx's structure (about XSD schema) with some difference in results. In addition, ART controls the consistency of values while Data4PT suggests the consistency of documents.

ART

- XSD validation by lxml and xmllint tools;
- Consistency control by some indicators:
 - Null durations (duration between two stops during a route);
 - Excessive speeds (speed between two stops during a route);
 - Duplicate quays or stop places (quays or stop places which have same name and the distance between them is less than 2 meters);
 - Close quays or stop places (quays or stop places which have not same name and the distance between them is less than 2 meters);
 - Unused quays (list of quays which are not used in routes);
 - Invalid URLs (validity of the value of operator URL).

Data4PT

- XSD validation;
- Addition of some rules:
 - Every line is referenced;
 - Every scheduled stop point has a name;
 - Every stop place has a correct stop place type;
 - Every stop place has a name;
 - Every stop place is referenced;
 - Every stop point have an arrival and departure time;
 - Frame defaults have a locale and timezone;
 - Locations are referencing the same point;
 - Passing times have increasing times;
 - Stop place quay distance is reasonable.