Designing WIM data aggregating systems

Vincent Janes, Eric Klein, Sébastien Romon, Franziska Schmidt, Hai Pham Doan

DESIGNING WIM DATA AGGREGATING SYSTEMS

Vincent JANES
CETE de l'Est
Metz, France

Eric KLEIN
CETE de l'Est
Metz, France

Ajouter photo

Sébastien ROMON
ZELT
Toulouse, France

Franziska SCHMIDT
LCPC
Paris, France

Hai PHAM DOAN
Sétra
Paris, France

Abstract

WIM systems are still a fairly young technology. Like many developing technology platforms, these are subject to fragmentation in terms of uses and standards. This paper addresses to efforts aimed at dealing with this fragmentation by observing two on-going projects. First system considered is the FIWI WIM database led by ISWIM, which through a cooperative effort aims at creating a WIM sensors and systems cartography. Second system considered is SITL system, which aims at developing a WIM data processing system. It is based mostly on open-source or free-use software. This system, while having a short term operational goals, is also meant as a demonstrator for promoting a formal WIM data model. Those two projects aim at serving and strengthing WIM users community by mutual understanding and promotion of common tools.

Keywords: Weigh-in-motion (WIM), database, heavy vehicles.

Résumé

Les systèmes de Pesage en marche demeurent une technologie relativement récente. Comme de nombreuses plates-formes technologiques en développement ils sont sujets à la fragmentation en termes de types d'usages et de normes. Cet article s'intéresse aux actions cherchant à contrer ce problème de fragmentation en étudiant deux projets en cours. Le premier est la base WIM FIWI, pilotée par ISWIM, qui par un effort coopératif cherche à établir un panorama des systèmes WIM et de leur déploiement. Le second système considéré est le système SITL qui cherche à développer un système de traitement des données WIM. Ce projet utilise pour l'essentiel sur des logiciels open source ou d'usage libre. Ce système bien qu'étant conçu pour répondre à un besoin opérationnel a aussi vocation a servir de démonstrateur pour un modèle de données WIM. Ces deux projets visent à servir et renforcer la communauté du pesage en marche par une meilleure compréhension mutuel et la promotion d'outils communs.

Mots-clefs: Pesage en marche (WIM), base de données, véhicules lourds

1. Introduction

Compared to other roadside technologies WIM is fairly recent. While a lot of effort has been placed on sensor and measurement technologies, from our experience, data exploitation could be improved. This article deals with two on-going projects aiming at getting a better return on available measurements.

From our experience even inside a single country, WIM data is largely fragmented, both qualitatively and quantitatively.
Qualitatively:

- type of measurable data depends on type of sensor considered,
- type of actually collected data depends on collection architecture,
- measurement specs depend on sponsor's particular needs,
- data format depends on manufacturers' choice.

Quantitatively, even for a single type of measurement:

- data is spread between several different agencies,
- even inside a single agency it is not unusual to have several distinct repositories.

Even when different organisms use the same tools or data format it is still unusual for them to share their data. Also many WIM systems we did encounter had limited exploitation tools implemented.

In addition to those facts, technical progress in recent period, notably in data storage (size and price) and communication network (3G…) has strongly affected the size and type of data storages we are able to consider.

All those elements led us into working on projects aiming at improving data knowledge and processing.

Two on-going projects deal with those observations:

- FIWI database is a website aimed at gathering WIM related information from various sources and thus establishing a description of data panorama. It has an important cooperative aspect.
- SITL is a more technical oriented project as it aims at aggregating data from a two well described fragmented sources into agile entity.

FIWI database project aims at embracing data diversity while BDTL project aims at performing synthesis in a given instance.

2. FIWI-WIM database

FIWI-WIM database is a subset of ISWIM. It aims at providing ISWIM members with information regarding FIWI deployment and activity in other countries and agencies.

This database will be made available on the FIWI webpage www.free.fiwi.fr before Spring 2012.

System relies on a PHP-MySQL architecture web interface for accessing and managing data. There are four parts that can be accessed by tags: a list of all WIM sites that are included in this database, aggregated data (which are the statistics of the WIM data files that can be found on the webpage), the detailed data which are the WIM files that are given by the WIM stations and the membership login and explanation.

2.1 WIM sites database

The WIM sites database webpage lists WIM equipment deployment.
WIM equipments are described through a detailed pre-set template that addresses their main features: location, site characteristics, lanes, WIM type, sensor type, calibration method, data collected…

2.2 Aggregated data

Also in addition to site detailed description is uploaded aggregated data. As opposed to site data that is aimed as describing WIM systems, those statistics provide traffic description elements like traffic volume, weight and vehicle type distributions… These figures give context elements for WIM system implementation and general hints regarding regional traffic conditions and regulation.

2.3 Data samples

This database is not meant as a quantitative exhaustive data storage but rather as a explanatory depiction. Thus the webpage doesn't contain large bulks of data that may be expensive in terms of storage and would probably not be available for confidentiality reasons, but rather small data samples allowing for understanding of data formats and system architecture.
Data samples are documented prior to publication by a standard upload form listing their main features, notably measured parameters (vehicle type, timestamp, lane, speed, length, gross weight, number of axles, axle loads…) and data collection period and duration.

2.4 Principle and membership

This webpage is meant as a cooperation platform:

- Data collection is done by member voluntary contributions. Every member which has WIM data available and wants to share, can login on the webpage and provide some pieces of WIM data (the registration is monitored by the administrator of the webpage).
- Data administration is shared between the system administrator and voluntary members acting as regional data "supervisors" (per country or region).

The webpage is available to guests although most data is only available to registered members. So there are several memberships:

- Simple guest which are just visiting the webpage just see very fiew pieces of WIM data,
- Registered FIWI members (the registration is simple, free and can be done on the FIWI webpage) see more WIM data,
- Data providers are people who own WIM data and upload it on the webpage. This data is not shown directly on the webpage, but must be approved by the supervisors.
- These supervisors control the WIM data uploaded by the providers, check it and if and when they approve it, it will appear for other people.
- The administrator of the system checks the memberships of the supervisors.

Through this cooperative effort, member community is able to develop a shared vision of WIM équipement deployment and ways of use. This effort can prove itself useful in fueling new systems development and joint initiatives.

2.5. Some shotscreens

In this paragraph, we show you 2 shotscreens of this database, giving you a general idea of the formatting and objective of this webpage. The first shotscreen, in Figure 1, shows you how is configured the page giving you access to the aggregated data available. The second one, in Figure 2, is a shotscreen of the page making it possible for the data provider to upload some WIM data.

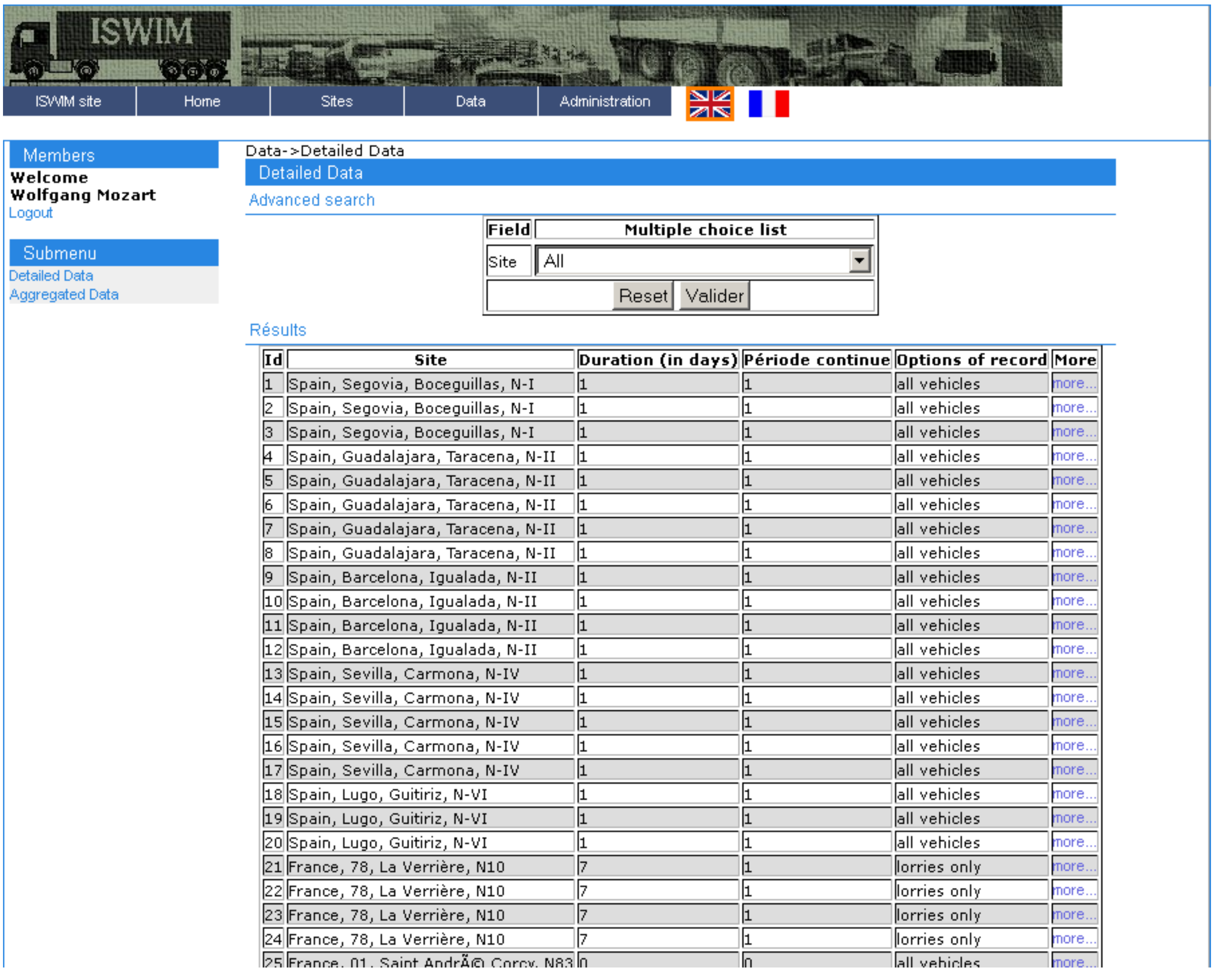

Figure 1: Shotscreen of the webpage of the database FIWI giving the list of WIM stations with aggregated data available.

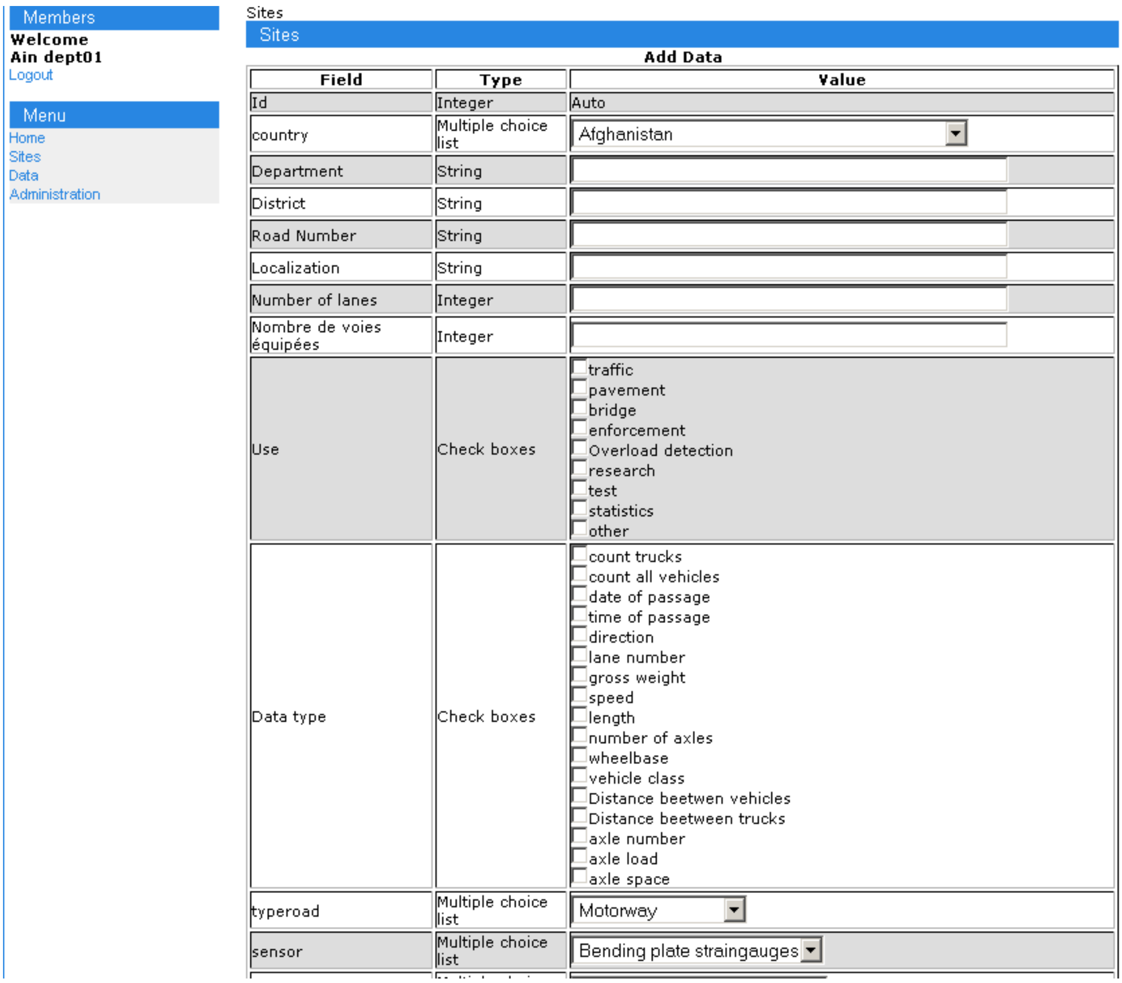

Figure 2: Webpage where the data provider can upload new data. This data is then reviewed by his supervisor, before it is visible on the visible webpage.

3. SITL

3.1 Context

SITL, for "Système d'Information Trafic Lourd" (Heavyweight Traffic Information System), originates mostly from a road structure related need but it also aims at catering several other needs.

SITL idea emerged after an important push in field sensor deployment. HS-WIM sensor network has organically developed. It started with a limited number of sensors that was increased over the years. As for most pioneering projects, most of the support organization was developed "on the spot" as needs made themselves apparent. This evolution notably applied to analysis system that was incrementally up-graded.

Then population grew, and the data stack slowly built itself. The incremental evolution type has several advantages in project early stages but also bears growth problem that tend to sum up over time. While original system catered for most of expressed needs it was becoming increasingly difficult to maintain and new development seemed compromised by early choices.

WIM sensor deployment is an expensive process, both in terms of finance and political will, so it is important it bears fruits. We felt we could improve the return on those equipments by improving our analysis tool, and that for that purpose we needed to go into recreating it from the beginning.

3.2 Objectives

The situation reading was basically that we had amassed a huge amount of data and know-how but we couldn't fully take advantage from it because it was fragmented and lacked a proper architecture.
So our objectives were to:

- capture most of this valuable existing knowledge and to integrate it into a durable agile model,
- make it available and usable to larger audience comprising:
    - specialists such as researchers and R&D users,
    - end-users such as road administrations or law enforcement (regarding max weight enforcement).

3.3 Design principles

Off-the shelf conceptual models
We did not intend to create new concepts. On the contrary, we tried to stick to well-known and proven concepts, as often as possible. In the long run, popular off-the shelf model will prove more durable than obscure in-house particular models. Popular models also mean there is a higher chance people have studied, are familiar or can get information on subject, or that tools exist or are developed for it.

Standard compliance
Prolonging precedent principle, compliance to public standards broadens the audience further and deeper to anyone familiar with them. Anyone skilled in used standards can get quickly started, all existing standard-compatible tools can be quickly applied.

Clarity over optimization
We ruled that clarity had to be considered a prime objective over low-level optimization.
While we aimed at a robust and functional model we ruled we wouldn't embark in fine, product-specific tuning and would rely on general software and hardware improvement for improvement, which good standard compliance would help us integrate.

Open-source approach
Project was completely developed on free open-source software by non-specialist developers.
That approach might be slightly detrimental in terms of performance over custom designed project but difference could be remedied at lower cost over time by hardware investment and flexibility and compatibility with community input could largely make up for this limitation.
It also allows for quick deployment and distribution without any licensing hassle.

3.4 Choices

Deciding to use off-the-shelf models saves time in conception but also requires investing some time in researching and choosing among existing models. After some digging we elected a few concepts that seemed us strong and rich enough to base our architecture on them. Those concept were:

- relational database model,

- business intelligence model.

Relational Model (SQL)
Regarding relational database it was also a very strong pillar, relational database (often known as SQL databases) have been invented in the 70s. They are widespread and a very mature field:

- RDBS offering is very mature. There are number of very powerful systems including open-source. The off-the-shelf associated tools offering is also constantly growing.
- We settled for Postgresql which seemed technically adequate, completely free, as opposed to other systems like MySQL, and backed by open-software foundation rather than a commercial firm.
- Databases and SQL subject have become a mainstream academic subject which means it is easy to find people competent on subject.

Figure 2 Main components used

Business Intelligence (BI)
Historically, business intelligence was invented as a way to analyze business data for large commercial firms, like sales per shop or per city, industrial costs, hence the "business" term, but those techniques are actually applicable to any decision-making activity. SAP-Business Object or SpagoBI are example of well-known BI tools.

In an initial step we chose not to implement any particular BI interface but we retained a typical BI star-schema. While initial interface would be a simple web query interface, this provision would allow to later implement most common BI tools or to connect some third party application.

Hardware improvement
We took into account trends into hardware improvement, notably we thought recent and future improvement in communication network speed and improvements in data-storage. The foresight in hardware improvement led us to consider large exhaustive storage.

Functionally oriented architecture
Our architecture is functionally oriented, which means it is based on WIM principles rather than purely software rules. While this is a downside in the sense it is probably less optimized than a more software-oriented architecture, it is also an upside in the sense that it makes the architecture easier to apprehend for any WIM worker.

4. Implementation

General principles and technical choices were settled we started developing core system tools.

4.1 Source Data

We have been working about two different formats; in the context of this article we'll mostly discuss the per-vehicle micro-data format. Data is produced by "EPM" WIM equipment network. Each equipment is out-fitted with a dedicated ADSL internet connection which allows for high data volumes.
Data for all stations is currently gathered on a centralized data server. Data repository consists in XML format, under a classic "tree-type" model. Measurements are stored per vehicle.

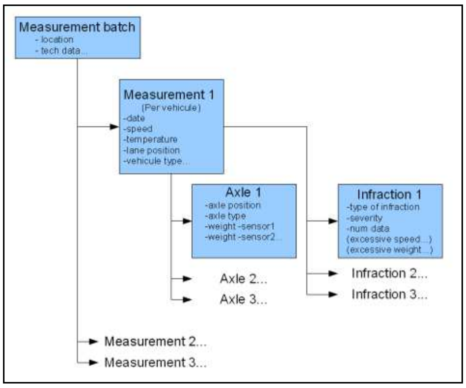

Fig 1 Source Data Model - xml markup

One may note that technically all data are not independent (for example vehicle type is deduced from axle positions). We did keep redundant data when those seemed expensive to recomputed afterward and/or when those redundant data was produced by WIM hardware itself. One can note that this choice doesn't prevent to recomputed redundant information (e.g. if vehicle types are to be redefined and/or refined) and it could be used to benchmark WIM's firmwares.

4.2 Integration tools

One of the main works to do was to transfer database from an XML repository to SQL database. For that purposed we used Talend Open Studio software. This software is free and open source. It is an ETL tool that, through a graphical interface, generates independent Java programs that can be deployed on any architecture featuring a Java Virtual Machine (Windows, Linux, Unix…).
The tool has a large array of extensions, including most formats (cvs, xls, xml…) and DBMS (Oracle, MySQL, SQL-server, Sybase), which allowed us to quickly cope with various data sources.

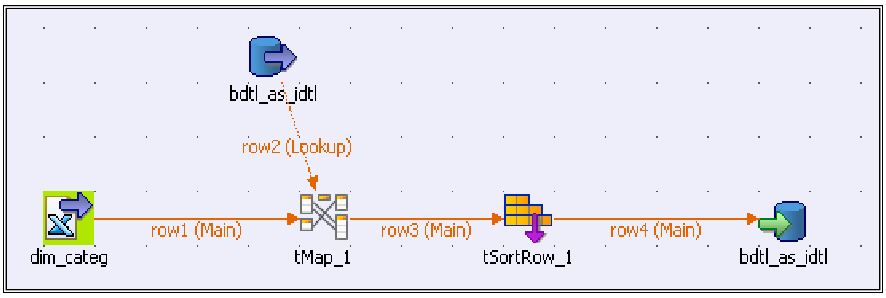

Fig 2 Simple Talend integration project - MS-Excel to PostgreSQL

4.3 Database

Model is mostly based on star schema, which is classic BI database architecture.

Fact table

This table includes one row per heavyweight vehicle. For each vehicle is recorded date, speed, total weight, and weight per axle for up to eight axles. Are also recorded all measurements in case there are several.

This table is meant to harbor the bulk of raw numeric measurement data, at the lowest granularity level available.

This is the main table with, currently, over thirty million lines.

measurement date location...	lane position length speed...				axle types		axle positions		axle weights per sensor		
date timestamp(6)	ecart integer	larg integer	long integer	vit integer	ke_1 integer	ke_2 integer	d1 integer	d2 integer	poids integer	pmoy_1 integer	pmoy_2 integer
2009-07-25 19:54:31	9	154	59	100	1	1	14	266	19	9	10
2011-05-30 11:14:03	75	309	184	87	1	1	108	566	45	14	14
2010-08-15 09:01:35	17	156	122	90	1	1	37	302	41	12	13

Fig 3 Abbreviated Fact table model (axle data, up to eight axles)

Dimension tables

Highly disaggregated data, while retaining most information, is almost impossible to use. In order to give some meaningful information it has to be aggregated.

Aggregation families will depend on purpose and organism, for example, depending on standard or organism considered there are several ways to classify heavy-weight vehicles, based on their axle distribution. Dimension tables' role is to list those aggregation rules inside a table. This allows every user to access information under an aggregation model that suits his needs.

In that regard we consider those tables mostly list rules and conventions (as opposed to fact table raw data).

There is a potential dimension table for each column of fact table. Each table is supposed to list all aggregation rules for the column. Whenever possible, classes are intricate.

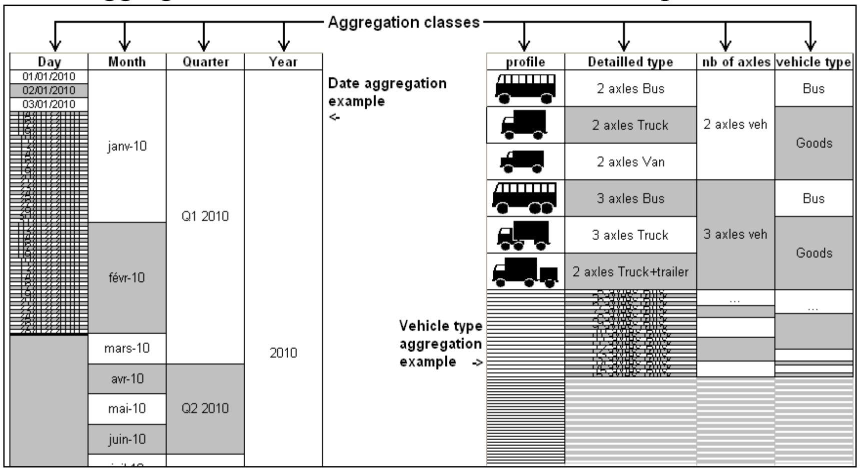

Fig 4 Examples of dimension tables for intricate aggregation classes

Restitutions

The data model allows for several types of analysis among which:

- Aggregations by month, year, vehicle types...
- sophisticated aggressiveness evaluation (per vehicle or statistical)
- cross-dimensional analysis

The fact data is almost unaltered and exhaustive allows for sophisticated experimental studies and statistical treatments.

4.4 Data access/ visualization

Data is accessed through different interface, depending on user. Advanced users access interface through ODBC connection. This is the simplest method, it is easy implemented and powerful but can only cater for a limited part of audience as it requires SQL proficiency and exerts direct pressure on DB server workload. Other method aimed at common user visualizes data through a query/restitution web interface. At this point this GUI is still in the works: based on our experience we already have a pretty good idea about restitutions to implement but technology choice is not settled yet.
Data restitutions are either through table, graphics or GIS systems.

5. Further development

Apart from obvious incremental improvements in engine and treatment there are many foreseeable opportunities regarding the system.

Integrating data inside model
Integration tools allow manipulating and processing complex and diverse data storages. So far we've been working on integrating the two most common formats available inside French WIM architecture but there is a strong potential for addressing more exotic data sources allowing for comparisons or cross studies between different systems at a national or an international data, for a reasonable cost.

Integrating model inside data
In the opposite corner to integrating data is making source data to be compatible with us, which means making data producers to adopt a unified data model. This is done by publicity and making resources available to partners.

Creating WIM data community
Which leads us to our final step: standardization is seldom a one-way road; this type of action is more of a community work:

- On one hand it's resources wasted that work by a given party can't be reused by others.
- On the other hand one given party can't do the work for all other parties, at least without feedback.

There is intrinsically, formal or not, a WIM community, comprising WIM producers, WIM system users, WIM data users… There are many patterns along which this community can organize itself.

- Fragmented community where parties mostly ignore each other.
- A particular party, generally a large commercial firm or public entity is able, by its size or influence, to lead alone making its private standards de-facto standards.
- Balanced community, possibly with one or several core members, able to push a general architecture for common good.
- Of course it is likely, particularly in a fragmented community, to have several clusters of various types existing at the same time.

Balanced community is probably the best and most durable option but it requires enough political will. Question is whether WIM community is big enough, community commons are

important enough, to sustain a vivid community life and make it up for the efforts associated with cooperation.
Creating community is done by publicity, making resources available… and creating debate.

5.6. Examples of applications

Pavements

The data aggregating system can be used to evaluate the impacts of a heavy traffic (overall flow or individual vehicle truck) through the assessment of the aggressiveness in the context of studies on weight and dimensions law changes, monitoring of road asset by administrations, or studies on pavement deterioration mechanisms. Traffic data are converted into aggressiveness by applying a power law on axle loads, whose factors depend on type of structure and axle.
The data system can also help to improve the pavement design method by characterizing accurately traffic aggressiveness.

Bridges

In this case also, the detailed data makes it possible for engineers to assess the damage of one particular traffic on one particular bridge, by using the concept of influence lines. Indeed, with these two elements, it is possible to assess the fatigue and the extreme effects to which the structure is submitted.
For example, we had recently to assess the consequences of changes in the weights and dimensions allowed for the trucks on the residual capacity of the bridges. Then, fatigue has to be assessed for steel and composite bridges, extreme loads have to be studied especially in the case of reinforced or prestressed bridges.

Statistics of traffic: overloading, dimensions, …

These two databases have two advantages for this last application: the first advantage, which is linked with the first database, is that one user may know which information one WIM station for one precise country may provide. So for example, French WIM data provide lateral position of the vehicles in the lane, whereas Dutch provide the type of truck.
The second advantage is that, with the second database which unifies all kind of formatting of WIM data, every data file may be used in the same manner. So statistics may be done easily by transferring the original data in the database and using the same treatment for all data afterwards.

6. Conclusion

In this paper, we presented two projects that have been launched and finished recently in France: the first one is the database FIWI, visible on internet, which makes a general description of the type of WIM stations and WIM data one may find all around the word. In this first project, the data providers are volunteers who want to advertise their data, and so share a small amount via this webpage. For more information, the webpage user may contact them and deal directly with them.
FIWI looks for more providers, so if you want to participate you are welcome!
The second project is a kind of tool which makes it possible to transfer all WIM data files in the same format. It is useful to apply the same procedures to all WIM files.