# Rail transport performance, a data science subject!

Rail infrastructure is widely considered as a critical domain because it has to deliver the highest possible availability levels in order to guarantee continuous public service. This infrastructure – which comprises complex architecture – generates huge volumes of heterogenous data containing essential information to understanding the infrastructure's state of health and therefore optimising its use. But future developments won't be possible if these enormous amounts of data are not structured and analysed in depth – something that can be achieved, in part, by artificial intelligence.

**Robert PLANA**
**Chief Technology Officer (CTO)**

PhD in Information and Communication technologies, Robert has extensive expertise in Industrial Internet of Things and Advanced Technologies.

France's rail transport sector is currently facing several major challenges: the ageing of its infrastructure, a sharp increase in population and therefore in user numbers, and the opening up of the rail passenger market to competition, as from 2020 for main lines and by 2024 at the latest for regional lines. For SNCF Réseau (the entity responsible for upgrading and maintaining the French rail network), the aim is to maintain the sector's existing infrastructure and increase its capacity. And the RATP (Paris's public transport operator) will also be affected as the market for its over-ground network (buses and trams) is also going to be opened up to competition. And last but not least, passenger safety is another constant major preoccupation.

## SNCF and RATP, two data-driven companies

In September 2018, Patrick Jeantet, Chairman and CEO of SNCF Réseau, unveiled a project named "Nouvel'R". This strategic roadmap for a new era of rail transport in France sets out the operational responses to the changes called for by the French State to remedy the SNCF's structural deficit. The fundamental objective is to shape the network of the future in order to regain market share for rail, which is the only low-carbon mode of mass transit. To achieve this objective, the SNCF needs to reduce the structural costs of its infrastructure, increase the capacity and strength of its saturated lines and hubs and step up the pace of modernising its operating and signalling systems, including the transition from ERTMS 1 to ERTMS 2 (the European Rail Traffic Management System).

**"Using a systems-based approach guided by data and models, we can calculate the state of health of a rail network's components and therefore plan how to renovate it robustly."**

*« The SNCF needs to build new lines, taking into account the very French phenomenon of sub-urbanisation", says our expert. "In the future, when a single ERTMS standard has been deployed, the network will be able to run at its optimal level. But today, there are two different systems for lines serving medium-sized towns on the one hand, and high-speed lines on the other. Regional traffic therefore has an impact on long-distance journeys... Also, journeys with connections are not currently optimised. »*

In his roadmap, Patrick Jeantet is counting in particular on new technologies to increase the frequency of trains on the same tracks, avoid accidents and improve passenger services. He is seeking significant improvements in terms of reliability, availability and predictive maintenance for the entire rail network, which will be needed to keep pace with market changes. Security and safety criteria – which are inter-dependent and are absolutely vital for rail networks – also have to be harmonised and streamlined in order to facilitate inter-operability. For example, since 2015, the SNCF has invested €950 million in digitalisation, including €300 million for 2018 alone. Breakdowns have decreased by 30% for "smart" trains thanks to remote diagnostics and predictive maintenance, resulting in savings that are very welcome in the combat to stem losses, at a time when the French State is taking over a large portion of the SNCF's debt in return for an effective performance plan.

Renovating the existing network still needs to be carefully managed: *« The network's renovation has not been optimally planned, as illustrated by recent breakdowns (at Gare de Marseille-Saint-Charles, Gare Montparnasse, Gare de l'Est and so forth). And it all comes down to a question of data! If we could compile all of the available data and calculate the state of health of the network's*

*various components, we would be able to identify the critical areas and plan the network's renovation robustly using a systems-based approach guided by data and models »*, our expert adds.

In a competitive operating environment, in which you have to be an international standard-setter to stand out from the crowd, the RATP Group aims to become the spearhead of sustainable mobility and the smart city to provide best-in-class passenger services. The strategic goals of the RATP Group's Chairwoman, Catherine Guillouard, cover domains such as inter-modality, new digital services and upgrading the bus and tram networks.
One particular project worth highlighting is an R&D project supported by the SystemX Technical Research Institute, aimed at proposing on-demand bus strategies based on a big data analysis of passenger flows.

Among the RATP's major preoccupations, we can cite passenger safety (particularly for women), vandalism and terrorism. The approaches to dealing with these issues necessitate collecting video images and proposing deep-learning algorithms to improve detection performance and reduce the number of false alarms.

## Using data and modelling to identify critical areas in systems and sub-systems

One of the key challenges we come across is being able to identify critical and fragile areas of infrastructure, which is not always easy. In industry, and particularly in regulated sectors (such as nuclear, rail transport and pharmaceuticals), the main problem is the enormous volume of documents and data that needs to be processed (incident reports, design specifications, project deliverables etc.). *« For the SNCF, we worked on railway station audits using the different document formats we were given (Word, Excel, e-mails, Power Point, PDF, plans, images and so on). Our system then automatically indexed the information by category, based on the sector-specific problems concerned. This enabled us to predict critical domains and identify fragile areas. »*

We have already applied this approach in the nuclear sector by examining incident and maintenance reports and then combining our industry expertise with the latest artificial intelligence technologies. This enables us to predict future incidents and therefore to anticipate maintenance periods before events actually occur. We also used this approach for the European Organisation for Nuclear Research (CERN) in order to analyse the criticality of a complex piece of scientific equipment. By combining data with models, we were able to provide the equipment operator with a criticality matrix, which is extremely useful for preventing breakdowns and organising maintenance phases.

## Using data and modelling to optimise network renovation

Once criticality matrices have been identified, the next stage is to intelligently plan the works that need to be performed, taking into account all of the project's constraints, in the same manner as logistics specialists. We do this by using heterogenous and contextual data to obtain robust information on the state of health of the systems and their components. *« For example, electrical equipment ages differently depending on whether it's located by the sea or in mountainous terrain where freezing temperatures are common. This is where the concept of predictive maintenance really comes into its own, as all of the available data is used to create models with a view to optimising capital expenditure, delivery lead times and service quality. »*

All the information and technologies required to carry out predictive analyses currently exist, and yet their full potential is only currently used for 5% of industrial equipment. "We don't

**"It's by modelling the relations between different tasks that we will be able to plan simultaneous operations and optimise maintenance work."**

have optimal models yet because not all interactions have been identified", our expert goes on to say. "It's by modelling the relations between different tasks that we will be able to guarantee a more effective renovation process, which includes planning simultaneous operations as well as optimising the workload based on the available resources". For example, we can't put an electrician, a builder and a mechanic to work on the same electrical equipment at the same time!

*« We're working on these areas in partnership with Cosmo Tech in order to analyse how different tasks are scheduled and carried out over time so as to enhance the planning process and optimise maintenance operations. This means we can anticipate any delays in certain tasks and reduce the related costs. »*

## Using data and modelling to calibrate predictive maintenance

People often say that the Industrial Internet of Things (IIoT) – which involves inundating infrastructure with sensors – is going to revolutionise how maintenance work is planned and performed. *« But we need to be careful about this message because it's questionable whether installing more and more sensors to accumulate data will lead to a better understanding of mechanical fatigue phenomena and breakdowns. »* Collecting and circulating data costs money, and the very sensitivity of sensors can sometimes lead to a lot of false alarms. At the moment, it's better to use hybrid approaches combining models with data with a view to building "digital twins", which are like a "test tube" representation of the health of a machine or system and make it easier to implement predictive maintenance approaches.

This type of analysis can be used as a basis for predicting infrastructure fatigue levels. *« These are tried-and-tested methods that are being used today in the nuclear sector where safety and security are paramount. The same issues apply for train signalling in the rail transport sector, where there's no room at all for mistakes. »*

## "Bespoke" solutions as the way forward

The main characteristics shared by all forms of rail infrastructure are their complexity and numerous interfaces (in terms of technologies, professional trades and processes). This requires cross-cutting approaches to ensure digital continuity and, by extension, sustainable information. But it's not easy to obtain cross-cutting information, and the silo-based structures of today's enterprises don't help. So we need to make sure that engineers work in conjunction with the building, commissioning and operations teams.

*« Today, digital continuity and the related systems engineering processes are carried out based on two different approaches. The first concerns integrated platforms, which are proposed by software publishers (3DS, Siemens, PTC, etc.). While these platforms offer a very broad range of services, they are relatively closed and require major data migration work, which poses data integrity issues. And there's also the problem of training users, who are often used to working with other solutions. The second approach proposes more open architectures, based on inter-operable solutions and the possibility of creating bespoke solutions. This second approach is currently the preferred option for rail infrastructure projects. »*

But we don't know what the future will bring of course, and it may be that when the digital transformation has fully taken hold across organisations then it will be necessary to have integrated platforms. However, today these platforms are not open or flexible enough, which is why some clients are reluctant to use them.

It's important to have an overall methodology to use the data collected and take into account the complexity of different infrastructure.

*« It's not a collection of independent entities", stresses our expert. "It's an inter-dependent ecosystem where systems engineering needs to be used. The problems encountered in the industries we've talked about are all due to the fact that organisations have too much of a top-down structure. So we need to identify the interactions and collate all of the available information in a digital model, which will have to be updated each time maintenance is carried out....»* This is when data management becomes all-important, followed by artificial intelligence to mine the documents.

## Using data and modelling to ensure dynamic and robust project performance

The approach of combining data with models enables engineers to more effectively predict potential problems in a project's operation, and therefore

propose corrective measures with a view to better controlling costs and lead times.
At the same time, we are seeing an increase in the number of transport infrastructure projects, which raises the question of being able to allocate the necessary resources in light of workloads. Here again, approaches based on data and models can be used to make engineers' tasks more efficient and capitalise on project experience.
This whole issue is all the more relevant because we're now really beginning to see a lack of skilled resources. *« The people with industry knowledge have retired or are about to retire, and there's a skills gap in professions outside the core business, such as for digital-based jobs and project management. Today, as well as capitalising, using and leveraging knowledge management approaches, we can add text mining and NLP* methods in order to reconstruct ontologies and extract structured knowledge from documents »,* concludes our expert.

## The use-cases of tomorrow

In terms of operational performance, going forward we expect that transport sector assets will be managed more effectively. This concerns tracks, fixed electric traction equipment, signalling systems and also railway stations. For stations, smart sensors can be installed on the premises to enhance the station's energy performance and more effectively manage passenger flows inside stations, particularly at peak times, thanks to machine learning algorithms. Multiple use cases for this approach can be envisaged.

Another approach involves using a digital twin, which would be a virtual copy of the network to better oversee maintenance work and contribute to the network's "industrial robustness", and switching to a smart signalling system (ERTMS standard) so that more trains can run on the tracks. For example, on the Paris-Lyon high-speed line, which is nearing saturation, 16 trains could run per hour in 2024 compared with 13 today.

From a human point of view, this combination of data and industry expertise is creating and will continue to create opportunities for developing transport inter-modality while retaining the resilience that's essential for the service quality required. These approaches will lead to incredible career paths for young engineers and restore the engineering profession to its former glory. ■

**"The problems encountered are all due to the fact that organisations have too much of a top-down structure. A systems engineering approach will allow to collate all of the information at each maintenance."**

(*Natural Language processing)