



# CIO.FOCUS

---

**Données et documents : un patrimoine  
à valoriser pour le business**

# EN BREF

Qu'il s'agisse de construire les transports en commun de demain pour la région parisienne, de valoriser des programmes audiovisuels ou, plus simplement, d'exploiter la data à des fins commerciales ou encore de traiter des notes de frais, le terme "système d'information" prend tout son sens. Ces données, ces documents, sont autant d'informations qu'il faut collecter, assembler, traiter. De Monoprix au Grand Paris Express, d'Agapes Restauration à France Télévisions, voici quelques exemples de valorisation business des données et documents.

# SOMMAIRE

## / INTERVIEWS

Mohamed Karouia (DSI du Grand Paris) :  
« la collaboration est clé pour la conduite  
d'un tel projet »

3

## / PROJETS

Comment Monoprix met la donnée  
au service des processus métier

7

## / PROJETS

France Télévisions mise sur la donnée  
et l'IA pour valoriser ses programmes

9

## / PROJETS

Samuel Descroix (Manager Data SNCF  
Réseau) : « la représentation géographique est  
une évidence pour tous les métiers  
à SNCF Réseau »

13

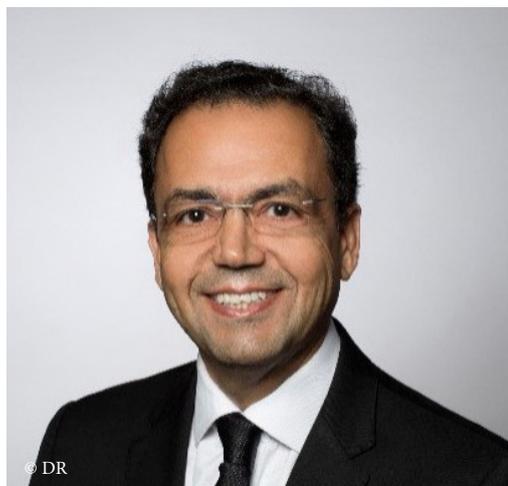


/ INTERVIEWS

## Mohamed Karouia (DSI du Grand Paris) : « la collaboration est clé pour la conduite d'un tel projet »

Les deux cents kilomètres de lignes nouvelles de métro du Grand Paris Express sont réalisées sous l'égide de la Société du Grand Paris (SGP). Maître d'ouvrage et propriétaire des infrastructures, cet EPIC doit disposer d'un système d'information adapté autant au présent, le chantier, qu'au futur, la période d'exploitation. Mohamed Karouia, DSI de la SGP, nous en explique les principes.

© anouchka-istock



© DR

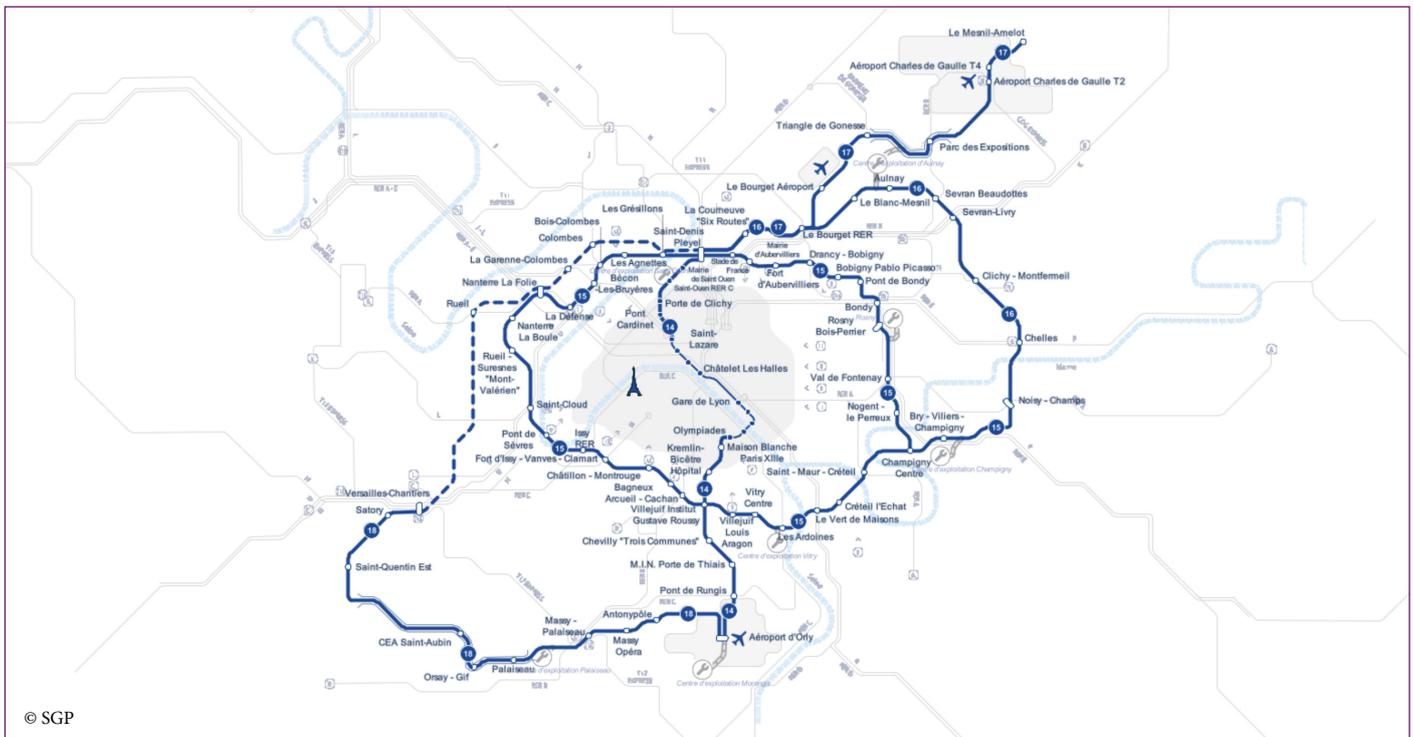
*Pour Mohamed Karouia, le SI doit être compris, accepté et utilisé pour pouvoir parler de transformation numérique*

### **CIO.** Le nom « Grand Paris » est connu mais que signifie-t-il ?

**Mohamed Karouia.** La Société du Grand Paris (SGP) est un EPIC [Établissement public à caractère industriel et commercial, NDLR] d'État dont la mission est de réaliser le Grand Paris Express (GPE). Le GPE, ce sont 200 kilomètres de lignes nouvelles de métro, la prolongation sud de la ligne 14 et la création des lignes 15, 16, 17 et 18. Le planning des inaugurations s'étale de 2024 à 2030.

L'EPIC est propriétaire des terrains où passera le GPE dont il est aussi le maître d'ouvrage et le propriétaire. Notre modèle économique repose donc sur la redevance que l'exploitant du GPE nous versera mais aussi sur la valorisation de nos biens fonciers, y compris via la location des surfaces commerciales. Un point important de la philosophie du GPE est que les gares sont des lieux de vie et qu'elles s'insèrent dans des quartiers où il y a également de la vie. Le GPE et ses gares doivent donc assurer un lien entre ceux qui utilisent le métro et ceux qui résident alentour. Nous voulons donc donner une identité aux gares en continuité culturelle et architecturale avec ce qui les entoure.





## CIO. Le SI de la SGP est donc celui d'un constructeur d'infrastructures ?

**Mohamed Karouia.** Tout à fait. Il y a donc quatre grandes familles structurantes, deux chaînes de valeur et deux socles communs transverses, auxquelles s'ajoutent des systèmes spécifiques.

La première chaîne de valeur est celle de la gestion d'actifs : achat, gestion des marchés, gestion de patrimoine, gestion des contrats sans oublier la gestion financière, la comptabilité, la gestion des immobilisations (au sens comptable mais aussi patrimonial, de real estate), etc.

La deuxième est celle de l'ingénierie. Il s'agit ici de la conception et de la construction du GPE, de ce que l'on fait ou fait faire. On parle ici de génie civil mais aussi de la gestion des livraisons, du contrôle du respect des exigences, de pilotage de projet.

Ensuite, il y a le socle commun des données et des documents, ainsi que du pilotage d'activité. On est là sur une problématique aiguë pour la SGP car nous sommes maître d'ouvrage et nous avons à gérer toutes sortes de ressources (matérielles, financières, humaines...). Et le GPE implique la création d'une grande quantité de données dont la durée de conservation est égale à la durée de vie de l'infrastructure (un métro, rappelons-

le)... Nous avons des milliers de contributeurs pour des millions de documents. La gestion des données et des documents est donc extrêmement critique. Et le pilotage de ces données va bien au-delà d'un simple besoin cadastral. On est dans le pilotage de chantiers. Donc il s'agit de savoir où est tel tunnelier, où sont les déblais extraits de tel endroit, si ces déblais sont ou non pollués... L'enjeu est celui de la maîtrise de l'activité au-delà du seul suivi des travaux. Ces données doivent également permettre de respecter et faire respecter, ainsi que contrôler le respect, des engagements sociétaux et environnementaux.

Le second socle, c'est celui du travail collaboratif de toute la communauté du GPE. La collaboration est clé pour la conduite d'un tel projet. Et j'insiste : ce n'est pas une simple optimisation ou du confort. Il nous faut une capacité à faire collaborer ensemble toutes les personnes actives autour du projet : les 800 collaborateurs de la SGP bien sûr mais aussi les 8000 intervenants sur le GPE.

La DSI doit aussi fournir de nombreux outils métiers tels qu'une salle de marché, un logiciel de gestion des déblais, etc. mais, comme ce ne sont pas des outils transverses, ils ne sont pas dans des socles ou des chaînes de valeur.

## **CIO. La crise sanitaire a-t-elle changé quelque chose en matière de collaboration ?**

**Mohamed Karouia.** Avant la crise sanitaire, tout le monde travaillait sous le même toit. Avec le confinement, il a fallu faire (sans rupture) en nous reposant sur la visioconférence, etc. Une fois le déconfinement arrivé, nous avons eu parfois un fonctionnement hybride avec, par exemple, une partie des participants dans une salle, une autre partie dans une autre et quelques uns à distance. Le traitement de la question a été plus complexe que prévu alors que c'est un enjeu important. D'où le fait que l'on en a fait un socle à part entière.

## **CIO. Quels sont vos grands choix d'architecture ?**

**Mohamed Karouia.** Nous nous orientons autant que possible vers les solutions les plus standards. A la genèse, reconnaissons-le, nous étions plutôt dans l'esprit start-up et on se débrouillait avec les moyens du bord, des outils ad hoc de partenaires, de la bureautique... Mais, avec le temps, nous nous sommes orienté vers un SI basé sur les solutions du marché.

En ce moment, nous consolidons ce SI. La taille de l'entreprise a été multipliée par deux en deux ans. Ce changement d'échelle peut aussi être constaté dans la taille du chantier et les ressources financières mobilisées. Le GPE, c'est 35 à 40 milliards d'euros hors matériel roulant...

Avant, si le SI n'était pas parfaitement intégré, il était possible de compenser de façon humaine : chacun connaissait le collègue du bureau d'à côté et on s'arrangeait avec lui. Aujourd'hui, c'est impossible.

Et le SI doit aussi être assez agile pour s'adapter aux nouvelles activités qui s'ajoutent. Par exemple, avec l'avancement des travaux de génie civil, nous avons à nous préoccuper de la réception comme de la gestion et de la maintenance par le futur exploitant.

Il faut que le SI soit toujours plus consolidé, plus intégré, pour sécuriser le projet et le livrer dans les délais.

## **CIO. Privilégiez-vous le cloud ?**

**Mohamed Karouia.** Nous avons un SI hybride avec, d'un côté, un cloud privé, et de l'autre un recours à un cloud public. Dans le cloud privé, nous avons plutôt le cœur de métier comme le BIM et le SI d'ingénierie et de



© SGP

gestion. Nous utilisons ainsi l'ERP Qualiatic de Cegid et le BIM Kroqi développé dans le cadre du Plan Transition Numérique du Bâtiment sous la double égide de la Fédération Française du Bâtiment et de l'État, utilisé par tous les EPIC et entreprises publiques.

Dans le cloud public, nous retrouvons ce qui relève de la messagerie, de la collaboration, etc. Nous sommes plutôt Microsoft avec Azure et Office 365. Il y a là un vrai vecteur d'agilité et de résilience, notamment en cas de bascule confinement / déconfinement.

## **CIO. D'ici 2030, vous allez basculer progressivement du chantier à la mise en exploitation. Quelles évolutions cela va-t-il impliquer pour l'IT ?**

**Mohamed Karouia.** Une fois le GPE terminé, nous allons administrer un patrimoine.

Mais avoir livré deux cents kilomètres de lignes en une quinzaine d'années fera de nous un maître d'ouvrage de référence. Rien n'est bien sûr décidé à l'heure qu'il est mais je pense que, au-delà de la gestion de notre patrimoine, nous pourrions faire de l'AMOA pour d'autres ouvrages, mettant ainsi notre savoir-faire au service d'autres projets, un peu comme a évolué Aéroports de Paris. Notre système d'information évoluera dans le sens de l'évolution de la société, de l'évolution de son activité et de son modèle économique.

Concernant le GPE lui-même, notre SI sera amené à gérer les redevances, la maintenance, les relations avec les collectivités locales...

## **CIO. Quels sont aujourd'hui vos grands défis ?**

**Mohamed Karouia.** L'entreprise est dans une phase de changement d'échelle : nous passons de la start-up à la taille industrielle. A cela s'ajoute d'abord la nécessité d'un collaboratif de bout en bout pour tous les acteurs du GPE. Concernant les datas et les documents, la masse n'est pas le seul défi : il y a bien sûr un enjeu important de sécurité. Bien entendu, en pleine crise sanitaire, nous avons un enjeu de résilience qui restera au-delà. Et j'ai déjà parlé de notre enjeu d'agilité. Tout ceci étant posé, nous avons deux grands défis à relever.

Le premier est celui que je nomme la question de la posture de la DSI. Nous ne sommes pas un fournisseur des métiers mais un business partner. Nous ne devons pas attendre une commande mais anticiper les besoins qui vont s'exprimer. Cela suppose un changement culturel au sein de la DSI. Nous devons explorer de nouveaux usages, exploiter des outils et garantir la résilience quoiqu'il arrive.

Le deuxième grand défi, c'est la consolidation et l'intégration du SI pour sécuriser et fiabiliser notre information, permettant ainsi de respecter les attentes en matière de délais et de qualité pour le GPE. Le SI peut être un levier seulement s'il est assez consolidé, intégré, connecté, fiable et robuste. Il sera impossible d'atteindre nos objectifs quant à la livraison du GPE si le SI est morcelé, avec des besoins de ressaisies, sans capacité à piloter, à garantir résilience et robustesse.

Mais, de toutes façons, nous ne pourrions pas atteindre nos objectifs et relever ces défis sans intégrer la dimension humaine, la dimension utilisateurs. Notre époque n'est plus celle où l'on se contente de livrer des outils informatiques. Le SI est au service des métiers pour qu'ils atteignent les objectifs business. L'utilisateur est donc l'un des trois axes majeurs, à côté des processus d'entreprise et des outils. C'est cela l'approche digitale.

En effet l'utilisateur doit être au cœur du changement, la technologie doit être au service de l'humain, conçu avec l'humain. Un projet doit toujours être une démarche bicéphale avec un aspect ingénierie / technologie, de l'autre un aspect usages / conduite du changement. Le SI doit être compris, accepté et utilisé pour que l'on puisse parler de transformation numérique.



UN ARTICLE RÉDIGÉ PAR

**Bertrand Lemaire**, Rédacteur en chef de CIO



/ PROJETS

## Comment Monoprix met la donnée au service des processus métier

En s'appuyant notamment sur les solutions d'Abbyy, Monoprix a pu automatiser au maximum la production de rapports pour des décisions business.



© DR

Arnaud Foujols, directeur Performance, Digital Transformation et Data chez Monoprix, a eu recours à l'automatisation pour que la donnée arrive au bon moment à la bonne personne.

Filiale à 100 % du Groupe Casino, le Groupe Monoprix comprend bien sûr 675 magasins en propre (Monoprix, Monop', Naturalia...) et une centaine de franchisés mais aussi Monoprix Online (Monoprix.fr, Sarenza). Comme dans toute la distribution, le traitement des données (essentiellement de ventes) est la base de nombreuses décisions business. Par exemple, déterminer si une promotion a été efficace permettra d'envisager de la refaire (ou une similaire) ou non. A l'invitation de l'éditeur Abbyy, Arnaud Foujols, directeur Performance, Digital Transformation et Data chez Monoprix, a témoigné de la manière dont les données disponibles ont été valorisées au service des métiers chez ce distributeur.

Celui-ci a ainsi souligné la grande complexité des activités économiques dans la distribution : des centaines de magasins, autant de lignes de produits, une cross-canalité... Face à une surabondance de données, les équipes en place avaient de grandes difficultés à établir des indicateurs clés de performance ou à comprendre les réalités révélées par les chiffres. « Il fallait que nous soyons plus structurés » tranche Arnaud Foujols.

### Un problème en premier lieu technique

Le premier problème était technique : l'outil de BI datait de 2012. Son rôle essentiel était d'extraire les données qui étaient retravaillées dans Excel. Bien évidemment, la taille des données

posait des problèmes dans Excel. Et les serveurs étaient également vite saturés. Arnaud Foujols se souvient : « nous avons de grandes difficultés, le lundi, à sortir les chiffres relatifs à l'activité du week-end ». Or les décisions business devaient être prises en fonction des résultats et cela sans attendre. Par ailleurs, il y avait très peu de tâches automatisées, presque tout étant réalisé manuellement, avec par conséquent une forte consommation de ressources humaines et des délais importants.

Pour structurer la démarche analytique, des évolutions techniques ont donc été menées. Pour commencer, une solution Alteryx a été utilisée pour extraire les données. D'autres outils ont aussi été déployés au fur et à mesure, notamment proposées par Abbyy, un éditeur très généraliste dans le traitement de la donnée et des documents (OCR, IA/Machine Learning notamment pour la reconnaissance et la compréhension du langage naturel (NLP), extraction des informations, etc.). La prise en main des différents outils a nécessité quelques mois.

## Une prise en compte indispensable des besoins métiers

Mais, dans un premier temps, les résultats n'ont pas été à la hauteur des espoirs soulevés. « Par exemple, les dictionnaires de données avaient été conçus par et pour la DSI mais étaient absolument abscons pour les métiers » relève Arnaud Foujols. Au bout de six mois de structuration des données, l'équipe data a commencé à construire des tableaux de bord. Arnaud Foujols indique : « nous avons alors cherché à automatiser leur production pour que les analystes analysent au lieu de passer du temps à des tâches sans valeur ajoutée forte. »

Une fois ces tableaux de bords produits, encore fallait-il qu'ils soient utilisés. Or ce n'était pas le cas, simplement parce qu'ils arrivaient en dehors des processus de décision, là où ils auraient été utiles. Un effort a donc été réalisé pour diffuser la donnée en l'intégrant au plus près des processus métier. Cela a notamment pu se faire grâce à des plates-formes de mise à disposition de la donnée. Arnaud Foujols se réjouit : « désormais, la donnée brute est enrichie et mise à disposition à J+1 afin de pouvoir répondre à des questions essentielles

telles que : combien d'euros additionnels générés par telle action promotionnelle ? »

## De l'analyse du passé à la prédiction de l'avenir

Il restait une étape importante à franchir : l'analyse post-mortem permet bien sûr de tirer des leçons mais c'est, par définition, trop tard pour traiter des promotions réalisées. La DSI a donc créé des outils pour que les créateurs d'une promotion puissent prédire – en fonction des précédentes – les effets de telle ou telle promotion. En plus, un outil de gouvernance a également été mis en oeuvre pour avoir en amont une vision globale de l'ensemble des opérations programmées. Enfin, il restait à traiter un lourd travail administratif : chaque promotion entraîne diverses tâches comme la communication, la commande des marchandises, la saisie dans les systèmes de caisse... Plus de 7 000 contrats sont signés chaque année avec des fournisseurs en lien avec la réalisation de telles promotions dont ils partagent le coût. La GED avec signature électronique de DocuSign a donc été interfacée afin que la génération des contrats puisse être automatisée.

Au final, c'est l'équivalent de quatre équivalents temps plein qui a été économisé tout en permettant un gain de 50 millions d'euros en plus de l'évolution du marché. Arnaud Foujols précise quelques bonnes pratiques : « il faut d'abord faire monter en compétences les équipes d'analystes. Surtout, pour lever tous les freins à l'adoption des outils et des rapports générés, il faut toujours partir des besoins des utilisateurs afin d'amener la bonne information, au bon moment, au bon utilisateur. La data s'associe alors à la transformation numérique. »



UN ARTICLE RÉDIGÉ PAR

**Bertrand Lemaire**, Rédacteur en chef de CIO



/ PROJETS

## France Télévisions mise sur la donnée et l'IA pour valoriser ses programmes

Pour mieux valoriser ses programmes, France Télévisions a mis en place une stratégie centrée sur les données, avec la création d'un département Data et intelligence artificielle (IA). Data hub et IA sont mis à profit pour collecter et enrichir les données sur les programmes, qui viennent ensuite alimenter les différents métiers du groupe.

© istock



© France Télévisions

Matthieu Parmentier, responsable département Data et IA (France Télévisions) :  
« une donnée qui circule, c'est de la valeur ajoutée pour l'entreprise. »

Pour accompagner l'évolution des usages en matière de consommation audiovisuelle, France Télévisions a entrepris de moderniser ses outils de production et de diffusion des programmes TV. Le groupe a notamment décidé de renforcer son expertise autour de la donnée et de l'intelligence artificielle, en particulier sur les phases de fabrication des programmes, durant lesquelles sont produites la plupart des données descriptives utilisées ensuite pour la classification et la valorisation des contenus. La hausse constante de la consommation sur les canaux numériques (+25 % de vidéos vues sur France.tv entre 2019 et 2020) nécessite en effet de mettre en place de nouveaux services, pour lesquels la qualité et la richesse de ces données sont primordiales : navigation, recommandations, recherche parmi les milliers de programmes disponibles, etc.

Dans ce but, le groupe a créé fin 2019 un département Data et intelligence artificielle (IA), rattaché à la direction des technologies. Pour Frédéric Brochard, directeur des technologies de France Télévisions, il s'agissait de « fédérer et dynamiser les initiatives de l'entreprise autour de ces sujets de transformation et de modernisation cruciaux pour nous : meilleure efficacité de nos processus, aide à la décision pour nos programmes et nos antennes, meilleurs outils de recherche pour nos journalistes, meilleure exposition de nos contenus, sont autant de leviers que ce département a en charge d'identifier et de développer. » Ce département, qui compte une vingtaine d'experts



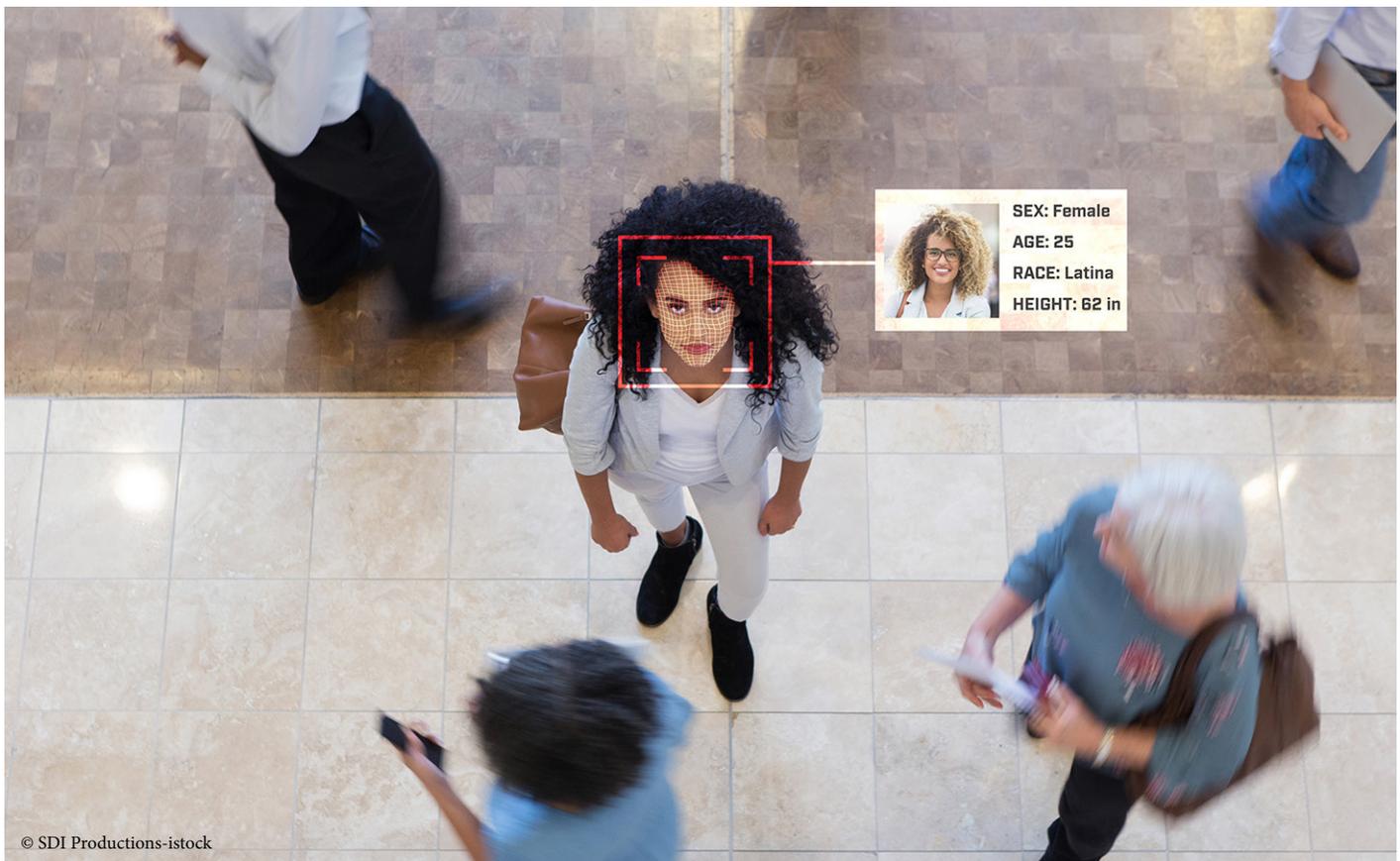
de la donnée, est sous la responsabilité de Matthieu Parmentier. « *Le département fonctionne comme un pôle d'expertise à la disposition des départements d'ingénierie (fabrication, diffusion, distribution, etc.), pour les épauler, voire prendre en charge la dimension data de leurs projets* », explique celui-ci. « *Au démarrage, il a fallu faire comprendre ce qu'était un projet data, faire en sorte que les métiers comprennent la valeur des données qu'ils génèrent ou saisissent. Une donnée qui circule, c'est de la valeur ajoutée pour l'entreprise* », ajoute-t-il. Afin de réfléchir aux nouveaux cas d'usage autour des données et de l'IA dans le secteur audiovisuel, le groupe a choisi de se faire accompagner par Cap Gemini, à la suite d'un appel d'offres européen lancé en avril 2020.

### Plus d'une centaine de données descriptives par programme

Pour favoriser la circulation et la valorisation des données, France Télévisions a également décidé de mettre en place un data hub fédérant l'ensemble des données descriptives sur des programmes. Un appel d'offres public a été lancé, à l'issue duquel le groupe a choisi la solution de DAM (Digital Asset Management) de Perfect Memory. L'outil, hébergé chez OVH, collecte l'ensemble des données de haute valeur disponibles en

interne, notamment celles fournies par les conseillers de programmes, la fabrication et le centre de diffusion, chargé de vérifier la qualité et le respect des obligations légales. « *Ces données décrivent par exemple qui est présent dans quelle scène, les costumes, les époques, les lieux... Certaines d'entre elles ne peuvent être connues à l'avance et ne sont déterminées qu'au montage final* », précise Matthieu Parmentier. À celles-ci s'ajoutent d'autres données ajoutées ultérieurement, comme les données d'audience ou liées aux usages publicitaires. « *Aujourd'hui, il existe au moins une centaine de données de haute valeur par programme, de tout ordre : genres, casting, audience, marqueurs de position des génériques... Chaque plan, chaque scène est associée à une tonalité* », indique le responsable du département Data et IA. Avec plus de 800 programmes quotidiens et 500 000 programmes actifs, ces données représentent une masse d'information conséquente. Pour chaque programme, l'outil de Perfect Memory permet de voir les données disponibles et celles qui manquent.

Quand c'est possible, le département récupère directement les données générées par l'outil de fabrication, comme sur le feuilleton quotidien « Un si grand soleil ». Si ce processus est le plus fiable et le moins coûteux, il n'est pas toujours possible. « *Nous*



© SDI Productions-istock

n'avons pas le droit d'en savoir moins que Google ou Netflix sur nos propres programmes », estime Matthieu Parmentier. Pour enrichir la connaissance des programmes, le département Data et IA utilise donc de nombreux outils d'intelligence artificielle, qui permettent de compléter les données non renseignées : retranscription textuelle du langage (speech-to-text), reconnaissance faciale, d'objets, de mouvements, de lieux... « Pour faire la segmentation automatique du contenu, nous faisons appel à sept réseaux de neurones différents, sur les personnes, les objets, la valeur des plans ou encore le contraste visuel et graphique » illustre le responsable du département Data et IA. Beaucoup de ces outils sont open source, certains provenant du laboratoire de Telecom SudParis, avec qui le groupe travaille sur l'IA appliquée à l'analyse d'image. « Quand ils obtiennent un prototype performant, nous le reprenons et nous l'industrialisons », indique Matthieu Parmentier.

## Optimiser le coût des traitements d'IA

L'industrialisation est indispensable pour maîtriser les coûts. Pour cela, le groupe s'appuie sur une architecture à deux niveaux. L'orchestration des traitements est gérée par une plateforme de microservices développée en interne, le projet open source Media-Cloud-AI.

« Nous avons pour ambition de reverser certains de nos travaux en open source dès que possible, c'est aussi une façon de répondre à notre mission de service public et cela facilite nos collaborations avec d'autres entreprises », explique Matthieu Parmentier. La plateforme est hébergée dans le cloud privé de France Télévisions. En revanche, les traitements sont déployés et exécutés dans des clouds semi-publics d'AWS, Google ou Oracle, hébergés en Europe et intégrés dans l'environnement de cybersécurité de France Télévisions. Appliqués à 800 programmes par jour, les traitements d'IA représentent vite un coût élevé, en particulier quand il faut recourir à des GPU dans le cloud. Grâce à la plateforme d'orchestration, « nous pilotons et nous pouvons rediriger les jobs en fonction des différentes opportunités, là où les GPU sont les moins chers », souligne Matthieu Parmentier. Le département s'organise aussi pour qu'un processus qui a besoin de GPU soit coupé en deux. « La partie GPU est exécutée à part, le plus vite possible, à travers un microservice dédié pour en optimiser le coût ». Enfin, si le traitement GPU nécessite de charger un modèle au préalable, par exemple pour l'analyse de lieu, « nous laissons ce modèle chargé et nous passons le maximum de jobs avant d'en charger un autre », détaille le responsable.



Les données produites par l'IA sont ensuite analysées et mises en relation avec les autres données disponibles, de façon à obtenir un jeu de données le plus fiable possible par émission. « Ces données issues des traitements IA ont du sens quand elles sont combinées. Nous analysons la bande-son, les images et les textes pour comparer les avis et supprimer le bruit : par exemple, si l'outil d'analyse d'image identifie un camion de pompiers et qu'on entend le son, cela conforte le résultat », illustre le responsable du département Data et IA. Inversement, « utiliser l'IA pour savoir où se passe une scène n'a que peu d'intérêt sur un gros plan centré sur un visage, il est utile de discriminer les informations de reconnaissance du lieu en fonction du type de plan. »

## Des données au service des collaborateurs et partenaires

Une fois intégrées dans le data hub, les données des programmes viennent alimenter et enrichir de nombreux services. « L'outil de DAM nous permet d'exposer différemment ces données, notamment dans les moteurs de recherche et d'indexation. Grâce à lui, nous réalimentons la Media Factory du groupe, qui fournit ces données aux différentes offres numériques de France Télévisions (france.tv, Salto, Okoo, Slash...) », indique Matthieu Parmentier. Les données servent également pour l'aide à la décision. Elles seront prochainement utilisées par la régie publicitaire, afin de choisir les bons spots à diffuser sur les supports numériques. « L'objectif est de supprimer le plus possible les publicités avant le début des programmes,

au profit d'une coupure intermédiaire : grâce à l'IA, il est possible de déterminer le moment où cette coupure gêne le moins la narration, mais aussi de choisir des spots en adéquation avec le contenu des programmes, pour éviter par exemple des publicités pour les 4x4 sur un programme nature », illustre le responsable. « Le but est que cela perturbe le moins possible les téléspectateurs, tout en rendant service aux annonceurs, qui sont indispensables à l'équilibre financier de nos offres numériques gratuites. »

Les programmes restent dans le système Perfect Memory jusqu'à neuf mois après leur première diffusion, après quoi la plateforme est purgée et les données sont reversées à l'INA (Institut National de l'Audiovisuel). « Nous travaillons actuellement sur l'accostage de nos deux modèles de données, afin de perdre le moins d'information possible. La finalisation de ce travail est prévue pour cette année », confie le responsable. Une intégration bidirectionnelle est même prévue, de manière à inclure directement le moteur de l'INA dans l'environnement DAM, pour que les collaborateurs de France Télévisions puissent facilement retrouver leurs programmes archivés.

## À PROPOS DU GROUPE FRANCE TÉLÉVISIONS

France Télévisions est un groupe audiovisuel public français. Il propose notamment un bouquet de quatre chaînes nationales (France 2, France 3, France 4, France 5), un réseau régional et outre-mer et une offre d'information sur tous les supports, Franceinfo, ainsi que des plateformes numériques pour tous les âges. En 2020, France Télévisions a atteint 28,8 % de part d'audience et 1,3 milliard de vidéos ont été vues sur sa plateforme France.tv.



UN ARTICLE RÉDIGÉ PAR

**Aurélie Chandeze**, Rédactrice-en-chef adjointe

/ PROJETS

## Samuel Descroix (Manager Data SNCF Réseau) : « la représentation géographique est une évidence pour tous les métiers à SNCF Réseau »

Visualiser le patrimoine de données de SNCF Réseau dans un Système d'Information Géographique géré sous Esri permet la meilleure prise de décision. Les nombreux « gisements de données » sont connectés par API pour faciliter leur exploitation par tous les utilisateurs potentiels, les possibilités techniques étant optimisées avant la définition d'usages.

© DariuszPa iStock

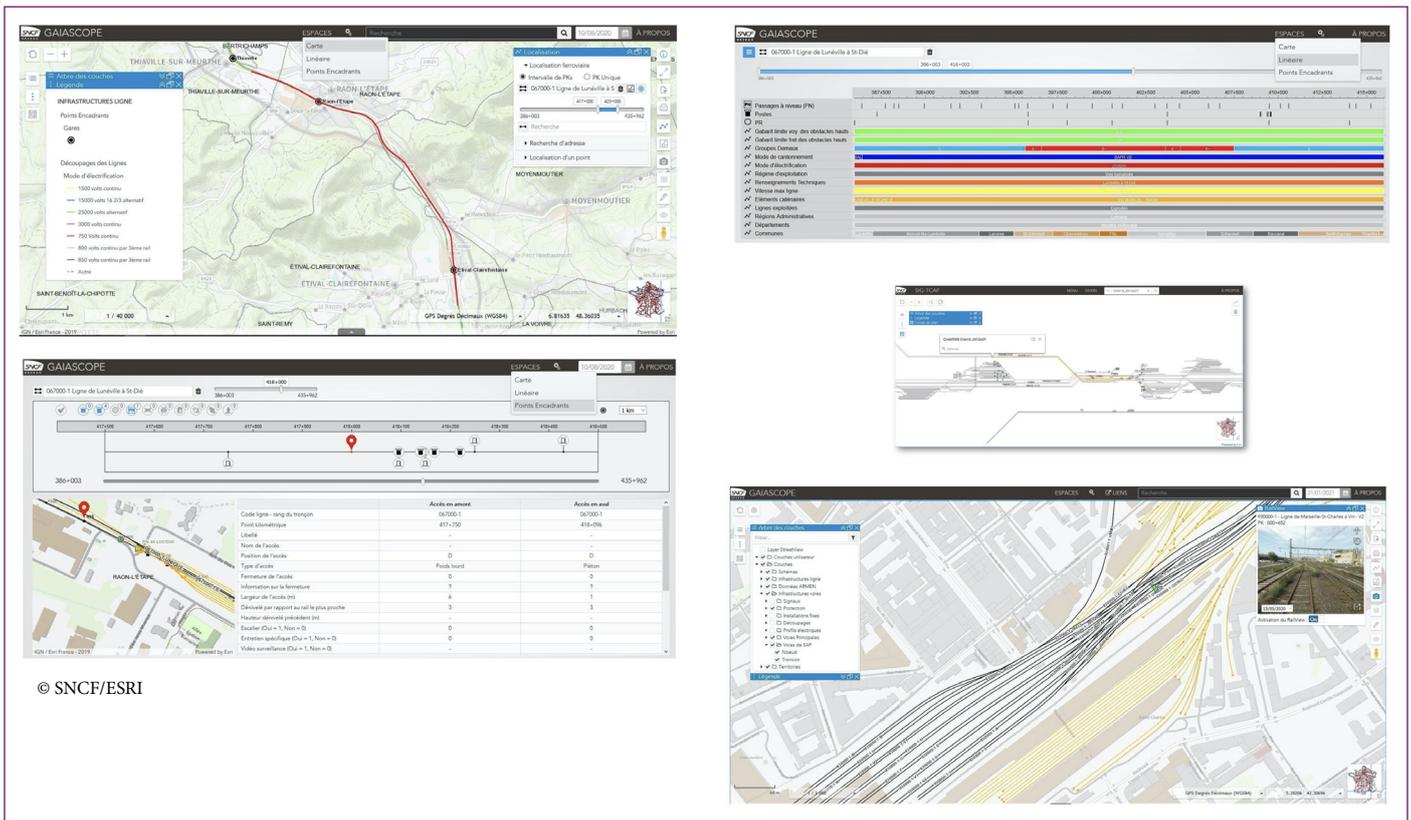


© DR

Samuel Descroix, Manager Data Géographique & Analytique de SNCF Réseau, a comme rôle majeur de valoriser un important patrimoine de données.

35 000 kilomètres de lignes actives sur le territoire national où circulent 16 000 trains par jour, auxquelles s'ajoutent les lignes à intérêt technique (desserte des technicentres, etc.) : voilà le domaine de SNCF Réseau. Cette société issue de la dernière réforme en date est en charge de la gestion de l'infrastructure ferroviaire, de sa construction, de sa maintenance, de son exploitation et de sa mise à disposition équitable de tous les opérateurs de transport ferroviaire. Quelques lignes ferroviaires, de dessertes locales ou au sein d'emprises portuaires notamment, échappent à son périmètre et il faut également ajouter des lignes qui ne sont plus comptabilisées car plus exploitées et plus maintenues. Pour réaliser ses missions, SNCF Réseau a besoin d'une masse considérable de données à valoriser efficacement. C'est le rôle de Samuel Descroix, Manager Data Géographique & Analytique de SNCF Réseau. Notons, avant de poursuivre, que la filiale Gares & Connexions de SNCF Réseau est, d'un point de vue système d'information, bien séparée.

« Les responsables des données, garants des données elles-mêmes et de leur qualité, sont toujours, au sein de SNCF Réseau, les acteurs métiers » précise pour commencer Samuel Descroix. Il ajoute : « le rôle de la direction du numérique, à laquelle j'appartiens, va être de valoriser ce patrimoine de données, d'être garant de sa conservation, d'une bonne représentation et d'une mise à disposition efficace. » Les données sont ainsi stockées dans des « gisements » avant



Diverses visualisation dans GaïaScope, sous formes cartographiques ou schématiques.

d'être utilisées dans et par des outils très variés. En matière de données géographiques, les responsables des données sont d'une part la filière en charge de la filière description du réseau au sein de la direction de la stratégie, d'autre part la direction de l'ingénierie. Le « gisement » de ces données a été appelé Gaïa.

## Gaïa, la mère de tout

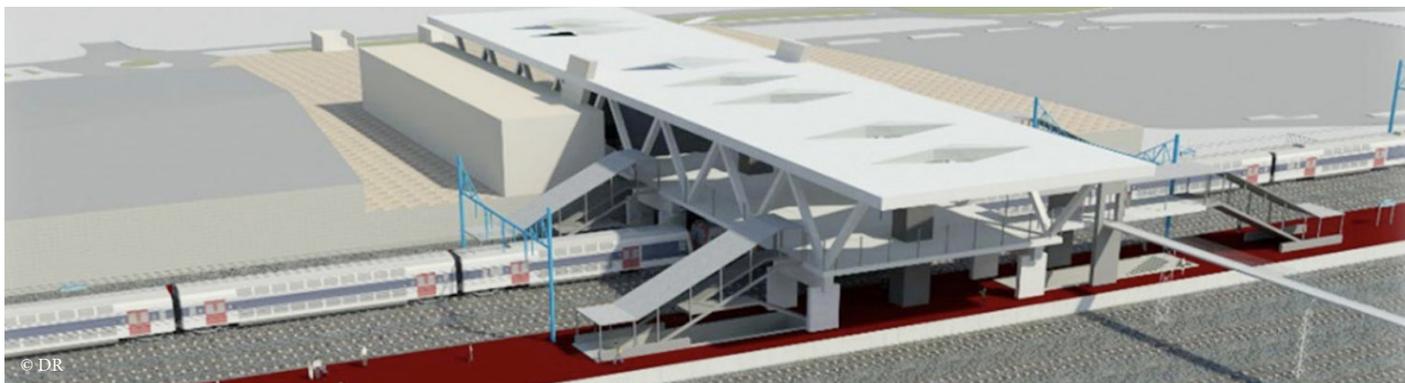
Chaque « objet » du réseau est décrit en fonction du modèle de données Ariane dans chaque « gisement de données ». Le modèle est ainsi utilisé pour décrire autant des éléments d'infrastructures que des circulations, des travaux, etc. Gaïa est le gisement de données qui décrit l'infrastructure ferroviaire et tous ses composants : rails, éléments de signalisation, aiguillages, crocodiles, autres « appareils de voies », etc. En fonction des besoins, selon la nature des données, selon que le gisement va être essentiellement en lecture ou bien en lecture-écriture, etc., SNCF Réseau utilise des technologies variées (PostgreSQL, Oracle DB, MongoDB...). En l'occurrence, Gaïa utilise MongoDB. Gaïa inclut des métadonnées et des liens vers des données externes pour qualifier précisément chaque objet du réseau. Evidemment, parmi ces données complémentaires, on trouve la ferrolocalisation (point

kilométrique du début et de fin sur le schéma de la voie) et la géolocalisation (coordonnées spatiales) des objets mais aussi des descriptions et documentations, des usages (trafic), des états (usure, réparations, opérations de maintenance programmées ou effectuées...).

« Le système d'information géographique est le point central pour gérer une infrastructure ferroviaire » constate Samuel Descroix. De ce fait, « la représentation géographique est une évidence pour tous les métiers à SNCF Réseau ». Il s'agit en effet, pour chaque objet du réseau, de savoir exactement de quoi on parle et de disposer de toutes les informations relatives à l'objet nécessaires pour prendre la moindre décision. En fonction du trafic, de l'usure et d'autres critères, il s'agira ainsi de savoir s'il faut procéder à une opération de maintenance, prévoir un remplacement... Samuel Descroix relève : « il s'agit de représenter géographiquement l'usure d'une voie. »

## Des briques reliées par API

Par dessus le gisement Gaïa, SNCF Réseau utilise une représentation géographique gérée sous Esri baptisée GaïaScope. Cette représentation cartographique peut être sous forme de schéma ou de carte à proprement



Exemple de plan de mise en œuvre avec une approche BIM.

parler : GaïaScope propose de multiples moyens de représentations. Cette plate-forme centralisée est destinée à tous les usages de Gaïa. « *Tout est lié donc tout doit être centralisé : par exemple, la maintenance a un impact sur la disponibilité de l'infrastructure* » note Samuel Descroix. Il insiste : « *le SIG [Système d'Information Géographique, NDLR] n'est pas un bout de chaîne décisionnel. Il y a bien un seul système d'information où tout le monde peut utiliser des briques disponibles avec des IHM diverses selon les métiers et les usages.* » GaïaScope propose ainsi, sur la représentation du réseau, un total de 140 couches de types d'objets différents.

Tous les gisements et tous les systèmes sont donc appelables par API pour être utilisés autant que nécessaire. Une bibliothèque de scripts en javascript permet d'utiliser le patrimoine de données pour des fonctions métier, par exemple rechercher un document dans la GED contenant la description technique détaillée de tel élément d'infrastructure. Le SIG peut être aussi un outil pour la performance de l'entreprise et l'optimisation de l'exploitation. Par exemple, il est possible de projeter le trajet d'un train et les dérives par rapport à sa trame horaire prévue, de rapprocher cette dérive de ce qui a été constaté sur d'autres trains et ainsi d'analyser les origines des problèmes. C'est une approche beaucoup plus opérationnelle que de simples indicateurs clés de performance globaux sur la ponctualité des trains traités dans un décisionnel classique.

## Beaucoup plus qu'un SIG

GaïaScope a donc des usages bien plus vastes qu'un SIG classique où l'on se contente de cartographier un endroit. Mais cela ne signifie pas que l'outil ne continue

pas d'évoluer. Samuel Descroix mentionne ainsi : « *nous avons beaucoup de besoins actuellement sur la représentation schématique du réseau* ». Une telle représentation schématique, à la manière d'un plan de métro, contient tous les éléments nécessaires avec leurs liens mais pas forcément en respectant une échelle : l'important est de disposer d'une présentation adaptée de l'information utile à afficher. Samuel Descroix spécifie : « *notre but est de générer les représentations schématiques directement à partir du BIM [Building Information Modeling, NDLR].* » Si le BIM est surtout connu pour les chantiers de bâtiments, tous les projets ferroviaires actuels sont réalisés avec du BIM. « *A l'heure actuelle, nous avons encore besoin de retravailler manuellement les données du BIM pour les inclure dans le SIG* » regrette Samuel Descroix.

La deuxième famille de progrès attendus relève de l'« hypervision ». Samuel Descroix explique : « *l'idée est de disposer sur un seul SIG de toutes les données de l'entreprise en temps immédiat.* » Par exemple, il s'agirait de disposer en temps quasi-réel des données sur la circulation d'un train sur l'infrastructure avec sa description complète et toutes les alertes contextualisées pour les interpréter. Collecte et intégration des données ont débuté l'été 2020 mais les usages sont encore en phase de découverte.



UN ARTICLE RÉDIGÉ PAR

**Bertrand Lemaire**, Rédacteur en chef de CIO