

# Jean-Séverin Lair, DSI de l'INSEE : » Les statisticiens utilisent davantage les langages open source R et Python »

**Qu'est-ce qui vous a surpris le plus à votre arrivée à l'INSEE ?**

**Jean-Séverin Lair** – L'INSEE n'a pas de direction numérique comme c'est un peu la mode dans beaucoup d'entreprises aujourd'hui. En fait, c'est elle-même qui est une direction du numérique et le rôle de la DSI est de fournir tous les moyens à nos directions métiers de gérer les grands répertoires, les sources de données, les restitutions et les indicateurs.

En outre, nous disposons de fortes compétences internes en développement alors que l'externalisation des ressources de développement est fréquente dans les services de l'Etat.

**Qu'est-ce qui explique cette particularité ?**



L'INSEE travaille depuis toujours sur la data et l'information : c'est son cœur de métier. Les besoins en numérique remontent à une époque où ce type de ressources ne s'externalisait pas. Nous disposons aujourd'hui de plus de 200 développeurs en interne et nous nous attachons à maintenir cette compétence, car le traitement de l'information statistique, qui plus est avec les gisements de données massives, présente de vraies spécificités.

En outre, et c'est sans doute unique, il y a une grande porosité entre les utilisateurs métiers et l'IT. Nos statisticiens ont eux-mêmes la possibilité de définir des opérations dites en « Self », ce qui signifie que la moitié des effectifs sont amenés à composer des traitements et en assurer eux-mêmes une grande partie de la programmation.

Disposer de développeurs en interne permet d'assurer un dialogue constant entre utilisateurs « métiers » et IT. Beaucoup de collaborateurs ont des carrières mixtes. Ainsi, l'actuelle responsable de production informatique était auparavant en charge du registre des personnes physiques. Mais nous devons aussi développer l'expertise et la spécialisation sur certains métiers ou compétences plus techniques ou pointues. C'est un équilibre sans cesse renouvelé entre polyvalence et expertise.

**En se lançant dans l'agilité dès 2008, l'INSEE s'est montré précurseur en France...**

L'intérêt pour l'Agile a été initié en 2008 avec quelques expérimentations, mais c'est véritablement en 2010 que le mouvement a été lancé avec un projet qui constitue notre cœur de métier, c'est-à-

dire le recensement de la population. Depuis, le passage global à l'Agile a été une vraie réussite même si nous ne sommes pas à proprement parler dans une approche « Agile at Scale » car notre caractéristique est de ne pas avoir des systèmes fortement inter-dépendants.

La fragmentation de nos applications rend donc l'agilité relativement simple à mettre en oeuvre sans nécessiter de mettre en place une grande mécanique centralisée pour synchroniser toutes les équipes entre elles. Nous avons pu nous limiter à une notion de domaines nous permettant de gérer les projets et la maintenance d'un certain nombre d'applications dans une même dynamique. Nous fonctionnons ainsi en réseau. C'est simple et dans l'esprit Agile.

### **Après l'Agile, vous voulez pousser vers DevOps ?**

[Le DevOps](#) est pour moi l'étape suivante après l'agilité. L'agilité casse le mur d'incompréhension entre les métiers et les développeurs mais il reste un deuxième mur à faire tomber, celui qui sépare les équipes agiles métiers/développeurs et les exploitants. C'est ainsi que l'on peut atteindre la meilleure continuité et la meilleure efficacité globale.

Outre la technologie à mettre en place pour permettre cette convergence, le sujet majeur reste celui de la transformation de l'organisation et de la culture des acteurs.

La DSI de l'INSEE compte aujourd'hui 450 personnes, avec environ 150 personnes du côté exploitation, 200 côté développeurs et les autres qui sont positionnées sur des fonctions plus transverses d'innovation, de structuration, de gestion etc.

### **Quelles sont vos attentes en termes de flexibilité et de rapidité de lancement de nouveaux projets grâce à DevOps ?**

L'informatique n'est pas forcément sur le chemin critique dans le lancement de nouvelles enquêtes par les métiers. Il y a de nombreuses phases de préparation et de vérification qui font qu'une étude demande du temps.

Néanmoins, plus qu'un enjeu de réactivité brute, il s'agit pour d'un enjeu de souplesse afin de nous adapter en permanence, notamment pour réutiliser des outils d'enquêtes, les modifier, etc.

La crise sanitaire a bien montré qu'il était important de pouvoir monter des enquêtes rapidement. Une approche modulaire permet de lancer une enquête en quelques mois, en nous appuyant à la fois sur des outils et sur [une approche Agile](#). Ceci sera encore simplifié par l'approche DevOps.

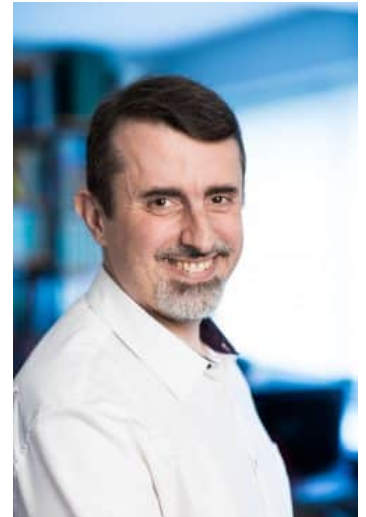
En parallèle à cet objectif, nous devons mener le déménagement de notre datacenter de Metz qui est maintenant obsolète vers deux centres interministériels des douanes dans la région parisienne et de l'agriculture à Toulouse. Nous n'allons pas déménager physiquement nos matériels, mais reconstruire nos infrastructures avec des équipements neufs dans ces deux centres. Nos équipes vont rester à Metz et gérer à distance ces infrastructures.

Les calendriers de ces deux projets – l'essor de pratiques DevOps et l'installation du nouveau datacenter- se télescopent un peu et nous allons devoir les séquencer. En 2021, nous sommes en phase de réflexion et de préparation pour que tout le monde « pense DevOps », pour une réelle mise en oeuvre en 2022.

### **Beaucoup d'entreprises privées déploient leurs infrastructures « Big Data » dans le Cloud**

## **public. Cela n'a pas été envisagé pour l'INSEE ?**

Il y a une vraie problématique d'indépendance et de souveraineté de la donnée statistique de l'État qui rend complexe son externalisation hors de la sphère publique. Notre usage sera donc essentiellement de type Cloud privé sur les ressources techniques dont nous disposons déjà, mais aussi sur des moyens interministériels.



Avoir accès à ces moyens interministériels complémentaires va s'avérer très pertinents pour compléter notre PCA (Plan de Continuité d'Activité). [L'incendie du datacenter d'OVH](#) à Strasbourg a marqué les esprits et les gens ont compris que ce type d'événement pouvait survenir partout. La continuité de service et la sécurisation des données sont essentielles.

En outre, nous avons mené une réflexion puis une expérimentation d'une infrastructure Kubernetes interne en mode DevOps. Nous allons mener quelques pilotes de déploiement sur un Cloud Kubernetes dans le courant de cette année puis, à partir de 2022, développer cette approche beaucoup plus largement. Cela ne concernera pas toutes nos applications, car ce ne sera pas faisable sur certaines applications « legacy », mais j'ai bon espoir que beaucoup d'applications pourront rejoindre ce modèle de déploiement à terme.

Outre le gain en agilité et en rapidité de mise en place d'environnement pour que les métiers puissent lancer de nouvelles enquêtes, l'autre intérêt de cette approche est de faire bouger la frontière entre les Dev et les Ops. Un cluster Kubernetes permet de laisser beaucoup de liberté aux équipes DevOps tout en donnant aux équipes d'exploitation la mission de maintenir une infrastructure solide. La frontière entre Dev et Ops se situe différemment selon que l'on travaille avec des conteneurs ou des VM.

## **Avec 130 projets publiés sur Github, l'INSEE est très active dans le domaine de l'Open Source...**

Cet engagement a été initié depuis une dizaine d'années, et pas seulement comme utilisateur. Nous faisons un fort usage du libre dans les infrastructures avec l'OS Debian, Tomcat et PostgreSQL pour sortir d'Oracle.

Les statisticiens utilisent de plus en plus les langages open source R et Python en lieu et place du logiciel statistique SAS – un changement porté par l'ensemble de l'Institut, dans une démarche volontaire.

Nos réalisations en matière d'Open Source sont nombreuses et [publiées sur GitHub](#) (InseeFr et InseeFrLab). Elles vont de la publication de packages R dans des domaines statistiques tels que le traitement de séries temporelles, jusqu'à la contribution à des briques techniques sans lien direct

avec notre cœur de métier – par exemple nous contribuons à Keycloak pour [France Connect](#).

Parmi nos réalisations, je peux évoquer notamment Métallica (MÉTadonnées Actives, Logiciels Libres et Infrastructure pour une Collecte Assistée). Ce logiciel construit et met en cohérence tous les outils d'une approche modulaire, s'appuyant sur les métadonnées actives, facilitant les développements de la collecte des enquêtes en multimode séquentiel ou concurrentiel (face-à-face, téléphone, internet, papier).

Autre exemple de réalisation mêlant données et technologies Cloud, la solution SSP Cloud permettant de regrouper dans un datalab tous les services de traitement de données à l'état de l'art. Cette solution s'appuie notamment sur le logiciel Onyxia, développé par l'Insee dans une optique d'UX Design afin d'accroître l'expérience utilisateur pour offrir une interface facilitant l'accès aux ressources du datalab.

C'est un projet pensé pour sa portabilité : il est déployable sur le Cloud comme en mode On-Premise et il est vraiment pensé pour être « instancié » afin de mettre en place rapidement un DataLab sur une infrastructure à base de conteneurs. Nous avons implémenté la première version en interne, puis une seconde a été mise à disposition aux acteurs du service public statistique sous le nom de SSP Cloud, pour faciliter la conduite d'expérimentations associant plusieurs ministères.

***Propos recueillis par Alain Clapaud***