

# Au-delà du code : accompagner les équipes dans la transition vers l'open source

Raphaël Simon<sup>1</sup>

Vincent Biot<sup>1</sup>

Nathalie Rigollot<sup>1</sup>

Meriem Said<sup>1</sup>

## Résumé

L'ATIH, dont plus d'un tiers des effectifs sont des ingénieurs statisticiens, s'appuie historiquement sur des logiciels propriétaires pour traiter les données massives et complexes du PMSI. Depuis 2023, l'agence a engagé une transition stratégique vers des solutions open source, dans la continuité de son SDSI et à l'image d'autres administrations comme l'INSEE. Les motivations sont classiques mais structurantes : souveraineté numérique, réduction des coûts de licences, diversification des outils, et alignement avec les standards actuels de la data science (R, Python, SQL).

Cette transition ne se limite pas à remplacer un logiciel : elle transforme les pratiques, les identités professionnelles et l'organisation du travail. Le principal enjeu est humain. Les statisticiens de l'ATIH sont des experts métier ; il s'agit donc de valoriser leur expertise plutôt que de la fragiliser. L'agence a adopté une démarche participative, impliquant les profils expérimentés dans la définition des standards R et dans la construction des nouveaux référentiels techniques.

La montée en compétence s'appuie sur un dispositif complet : formations socles, documentation centralisée, communautés internes, séances de partage, ateliers thématiques, audits de code, et trois hackathons réunissant l'ensemble des statisticiens autour de problématiques transversales.

Le changement touche aussi l'infrastructure : introduction systématique du versioning, nouvelles pratiques de collaboration, intégration de forges logicielles et de CI/CD. Ce volet, souvent sous-estimé, nécessite un accompagnement dédié.

Enfin, la migration doit composer avec un patrimoine de programmes statistiques accumulé sur plus de vingt-cinq ans. L'usage du package datadiff et le recours à une méthodologie de traduction assistée par IA, développée avec ThinkR et Théodo, permettent d'avancer sereinement sur le volume de programmes et la stabilité de leurs résultats.

En conclusion, la transition open source repose bien davantage sur l'humain que sur la technique : accompagner, valoriser, structurer et moderniser sont les clés d'une migration réussie dans une administration publique.

**Mots-clefs (3 à 5) :** Statistique - Transformation – Opensource

## Développement

### Contexte général

L'ATIH (Agence Technique de l'Information sur l'Hospitalisation) est historiquement un utilisateur majeur de logiciels d'analyse statistiques : les ingénieurs statisticiens représentent plus d'un tiers de ses effectifs. Leurs rôles sont variés et principalement, mais pas seulement, justifiés par les données du PMSI (Programme de médicalisation des systèmes d'information) : de la collecte au traitement, du nettoyage à la mise à disposition et aux restitutions, les travaux

\* **ATIH**, [raphael.simon@atih.sante.fr](mailto:raphael.simon@atih.sante.fr), [vincent.biot@atih.sante.fr](mailto:vincent.biot@atih.sante.fr), [nathalie.rigollot@atih.sante.fr](mailto:nathalie.rigollot@atih.sante.fr), [meriem.said@atih.sante.fr](mailto:meriem.said@atih.sante.fr)

statistiques à l'ATIH imposent de traiter des données de fort volume, un flux quotidien, et une considérable dose de métier. Ils viennent ensuite servir les tutelles (Ministères de la Santé, Agences régionales de santé, Assurance Maladie ...), les autres agences (Haute autorité de santé, Santé publique France, Agence de la Biomédecine ...), mais aussi les chercheurs, universitaires, hôpitaux et industriels. Depuis janvier 2023, sous l'impulsion de son conseil d'administration, l'agence est engagée dans une transition vers des options logicielles open-source pour ses traitements statistiques.

## **Ce qui motive la transition**

La démarche s'inscrit dans le Schéma Directeur des Systèmes d'Information (SDSI) de l'ATIH. Les raisons sont similaires à celles d'autres administrations françaises, notamment l'INSEE qui a achevé sa propre transition fin 2025 : réduction de la dépendance à un éditeur propriétaire, coûts élevés de licences, et volonté de souveraineté numérique. Les principaux candidats open-source dans ce contexte sont R et Python, ainsi que SQL.

Dans ce contexte, la question ne se résume pas à un simple remplacement d'outil. C'est un **changement de paradigme profond** qui touche les individus, les équipes, les infrastructures et la manière de concevoir la production statistique. Cette présentation propose un retour d'expérience sur les dimensions humaines et organisationnelles de cette transition, à travers trois axes : la conduite du changement, la montée en compétence des utilisateurs, et la gestion du code legacy.

## **Conduire le changement : bien plus qu'un projet technique**

Migrer d'un logiciel propriétaire vers l'open source, c'est d'abord **changer les habitudes de travail** d'équipes qui ont parfois passé dix à vingt ans à maîtriser un outil. À l'ATIH, les statisticiens ne sont pas de simples consommateurs de logiciel : ce sont des experts métier dont la valeur réside dans leur connaissance du PMSI, des nomenclatures, des règles de groupage. Le risque principal n'est pas technique - R et Python peuvent faire tout ce que le cahier des charges exige - c'est que ces experts se sentent **remis en cause ou déstabilisés** dans leur identité professionnelle.

La conduite du changement a donc consisté à bien positionner la question : il ne s'agit pas de « désapprendre » mais de valoriser son expertise sur les données, en mobilisant de nouveaux outils, en **élargissant son champ de compétences**. Les experts métier ont été impliqués tôt dans la définition des standards R de l'agence, transformant leur expertise en actif clé de la transition plutôt qu'en passif à reconvertir. Cette approche participative était la condition d'élaboration d'une trajectoire de transition pertinente et partagée

## **Accompagner la montée en compétence**

La montée en compétence a nécessité une approche **multiniveau** et s'est appuyée sur la solidarité du collectif de travail.

## Des formations socles

- Package, fonctions pour factoriser le métier et le partager entre les services qui partagent des besoins analytiques
- Visualisation de données, applicatif pour les restitutions
- Interaction et analyses avec des bases de données relationnelles

## Des moments de partage entre pairs

- Des séances "Bulles d'R" régulières permettent à un (ou plusieurs) agents de présenter un sujet, un problème, une idée, à la date et l'heure de son choix. L'évènement est d'inscription libre pour les autres, et encourage la discussion et le partage d'expérience.
- Des canaux de discussions transversaux
- De la documentation technique centralisée

## Des évènements dédiés

- Accélération : trois sessions de deux jours chacune, ont permis de réunir l'ensemble des statisticiens autour de thématiques transversales de travail à la recherche d'une élaboration collective de solutions
- Des audits de codes : relecture par des yeux experts d'un projet mature pour un rapport détaillé des pistes d'amélioration pour aller vers la mise en production
- Des ateliers à thème

## Le changement de paradigme infrastructure

Au-delà du langage, c'est **l'environnement de travail** qui change radicalement. Historiquement, la pratique statistique s'effectuait sur des serveurs dédiés et dans des environnements fermés et stables, avec peu de versioning du code. Le passage à R s'accompagne d'une transformation des pratiques :

- **Git et le contrôle de version** deviennent incontournables, alors que beaucoup n'y avaient jamais été exposés.

Ce changement de paradigme infrastructure est souvent sous-estimé. Il demande un **accompagnement spécifique sur les outils** (Forge (Github, Gitlab), CI/CD) et une évolution de la culture d'équipe vers plus de transparence et de partage.

## **La gestion du code historique : ni tout garder, ni tout jeter**

L'ATIH dispose d'un **patrimoine considérable de programmes statistiques** : certains ont plus de vingt-cinq ans, sont peu documentés, mais produisent des chiffres, règles ou indicateurs parfois publiés au Bulletin Officiel, ou dans des arrêtés. La tentation de tout réécrire d'un coup est dangereuse ; celle de tout conserver l'est tout autant.

Techniquement, la stratégie adoptée pour la migration repose en partie sur datadiff, un package développé par ThinkR à cette occasion.

Additionnellement, pour traiter le volume de programmes à migrer dans un calendrier contraint, une méthodologie spécifique de traduction assistée par Intelligence Artificielle a été développée en collaboration avec Théodo et ThinkR.

## **Conclusion**

La recherche d'alternatives open-source à l'ATIH illustre que **la réussite d'une telle transition repose à 20 % sur le choix de l'outil et à 80 % sur l'humain**. S'appuyer sur sens du service public des agents, cohérent avec les valeurs de l'open source, et sur la solidarité du collectif de travail, pour que personne ne se sente seul face à ses difficultés ; accompagner les équipes, valoriser les expertises existantes, différencier les approches selon les profils, moderniser progressivement les infrastructures et traiter le code legacy avec méthode : voilà les vraies clés d'une migration open source réussie dans une administration publique.