



T

H

É

M

A

Essentiel

Commissariat général au développement durable

Challenge RST-Analytics

Une expertise ministérielle en science de la donnée au service des politiques publiques de la transition écologique et solidaire

NOVEMBRE 2019

Le Commissariat général au développement durable (CGDD) est chargé, en lien avec le secrétariat général et en appui aux directions générales du ministère, de la supervision générale des données¹. Cette fonction a pour objectif principal de valoriser les données, notamment en aidant le ministère à mieux utiliser ses propres données et en mettant les sciences de la donnée à son service.

Différentes démarches s'inscrivent dans ce cadre : la GreenTech verte², la Fabrique Numérique³, les Entrepreneurs d'Intérêt Général⁴, ou encore le challenge RST-Analytics organisé par le CGDD. Il s'agit d'un hackathon* de deux jours réservé aux organismes du réseau scientifique et technique du ministère (RST)⁵.

PRINCIPES ET OBJECTIFS

Le challenge RST-Analytics consiste à relever sur deux

jours des défis sur les données, proposés par des porteurs de politiques publiques, en réalisant des prototypes rapides, des preuves de concept. Cette démarche d'innovation se veut ouverte et à faible coût : c'est à la fois un appel à défis de *datascience* auprès du pôle ministériel et un dispositif pour les relever rapidement en mobilisant l'expertise de son réseau scientifique et technique.

L'action du challenge RST-Analytics se situe en amont des démarches GreenTech verte, Fabrique Numérique et Entrepreneurs d'Intérêt Général, et peut faciliter ou accélérer leurs projets en apportant l'impulsion initiale et une preuve rapide de faisabilité technique. Trois éditions ont eu lieu et sont présentées dans cette publication.

L'approche est celle du « bac à sable » : le développement se fait en pleine confiance, dans une « zone franche » de propriété intellectuelle et en s'affranchissant de tous les coûts de transaction habituels. En effet, pour les organismes du RST, le coût se limite à la mobilisation de leurs experts pendant deux jours. Pour le pôle ministériel, il s'agit de formuler des défis, de mettre à disposition des données et de proposer un léger accompagnement aux experts.

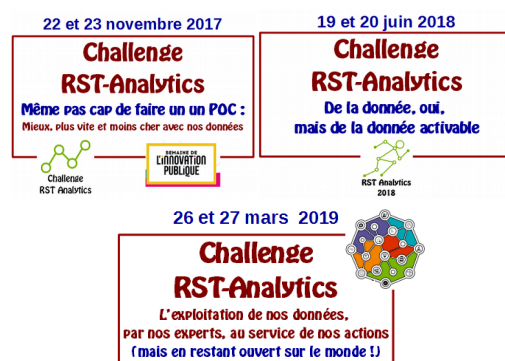


Figure 1: Logos des trois éditions du Challenge RST-Analytics.

1 <https://www.ecologique-solidaire.gouv.fr/missions-du-superviseur-general-des-donnees>
2 <https://www.ecologique-solidaire.gouv.fr/greentech-verte>
3 <https://www.ecologique-solidaire.gouv.fr/fabrique-numerique>
4 <https://entrepreneur-interet-general.etalab.gouv.fr/>
5 <https://www.ecologique-solidaire.gouv.fr/reseau-scientifique-et-technique>

Challenge RST-Analytics - Une expertise ministérielle en science de la donnée au service des politiques publiques de la transition écologique et solidaire

Quatre principaux objectifs sont poursuivis par cette démarche singulière :

1. Le prototype basé sur l'exploitation des données se doit d'être utile, à fort impact, opérationnel et reproductible (possibilité pour le commanditaire d'engager par la suite des développements ultérieurs si cela est pertinent).
2. La démarche a vocation à réaffirmer le potentiel et la réactivité du RST, y compris sur les sujets numériques (notamment l'intelligence artificielle), en allant au-delà de la simple intention ou de l'idée, en réalisant du traitement effectif de données.
3. Il s'agit aussi de démontrer le potentiel de valeur d'une démarche collective au niveau des organismes du RST, en mettant en avant les approches pluridisciplinaires.
4. Le dispositif contribue aux actions du superviseur général des données (SGD), en initiant des réutilisations internes des données ministérielles, en démontrant l'intérêt de traitements sur celles-ci et en participant, en rebond, à leur montée en qualité.

EN TROIS ÉDITIONS, UNE VINGTAINE DE DÉFIS RELEVÉS PAR QUARANTE-DEUX EXPERTS ET DIX ORGANISMES DU RST

De novembre 2017 à mars 2019, trois éditions du challenge RST-Analytics ont eu lieu. La première édition a été labellisée « Semaine de l'innovation publique »*.

À chaque nouvelle édition, le nombre de commanditaires, de défis et de participants (d'horizons toujours plus divers) a augmenté, confirmant l'intérêt de toutes les parties prenantes aux challenges. La profondeur stratégique et la complexité toujours plus importante des défis proposés ont également révélé, de la part des services du pôle ministériel, des attentes toujours plus fortes vis-à-vis d'approches basées sur les sciences de la donnée pour des cas d'usage métiers, à forts enjeux, opérationnels ou exploratoires.

Au total, ce sont vingt-deux défis, listés dans l'encadré ci-contre, qui ont été relevés par les *datascientists* de dix organismes du RST⁶ : AFB, BRGM, Cerema, CSTB, IFPEN, IFSTTAR, IGN, Ineris, IRSN, ONCFS.

Liste des défis relevés par thématique

Aménagement du territoire / Urbanisme :

1. Prédiction de dépôts de permis de construire dans le tertiaire ;
2. Détection automatique de friches industrielles ;
3. Mesure de l'artificialisation des sols en exploitant les documents d'urbanisme ;
4. Test de constitution de bases de données de plan de prévention des risques naturels à partir de sites web de préfectures et identification des différences entre les principaux portails institutionnels nationaux pouvant en contenir (Géoportail de l'urbanisme, Géorisques ...);

Bâtiments :

5. Prédications de diagnostics de performance énergétique (DPE) par machine learning ;
6. Analyse de la fiabilité des DPE ;
7. Calcul de factures énergétiques des ménages à partir de données DPE et INSEE ;

Biodiversité :

8. Détections d'anomalies dans la base de données des ventes des produits phytopharmaceutiques (BNV-D) ;
9. Identification de comportements anormaux dans la BNV-D ;
10. Identification de fronts de propagation d'espèces invasives ;

Énergie :

11. Mesure des économies d'émission de CO₂ et de NOx suite à différentes mesures de politiques publiques ;
12. Détection d'anomalies dans les certificats d'économie d'énergie (CEE) ;
13. Géodatavisualisation de la production d'électricité ;

Transports :

14. Calculs automatisés d'indicateurs de congestion routière ;
15. Ciblage des aires de contrôle des poids lourds ;
16. Complétion de séries temporelles de trafic routier ;
17. Détection des navires non immatriculés ;
18. Détection de bateaux fluviaux sans titre de navigation ;

Divers :

19. Croisement des bases des bailleurs sociaux avec les fichiers « entreprises » de l'INSEE ;
20. Cartographie des graphes de réseaux métiers ;
21. Mise en place d'un moteur de requêtes en langage naturel pour les bases du service statistique ministériel ;
22. Traitement du langage naturel pour exploiter les données du Grand Débat.

Challenge RST-Analytics - Une expertise ministérielle en science de la donnée au service des politiques publiques de la transition écologique et solidaire

Deux exemples de défis relevés sont détaillés ci-après.

Défi – Où sera mon espèce exotique envahissante ?

Sur proposition de la direction de l'eau et de la biodiversité, il s'agissait d'identifier les fronts potentiels de propagation des espèces exotiques envahissantes (ou d'un sous-ensemble de quelques-unes de ces espèces), en croisant les données relatives aux présences d'espèces, aux habitats et à l'occupation du sol. Répondre à un tel défi permettait d'aborder l'un des enjeux du règlement européen sur les espèces exotiques envahissantes entré en application en 2018.

Une équipe AFB / Cerema / ONCFS s'est formée pour répondre au défi. Deux méthodes différentes et complémentaires ont été mises en œuvre pour calculer des probabilités de présence sur le territoire français par maille 10 × 10 km :

- un outil de simulation et d'aide à la prévision de 2018 à 2030 basé sur des calculs d'aires maximales attendues à partir de variables descriptives, permettant de déterminer des caractéristiques de croissance (durée des stades selon le milieu) et de colonisation (« distance » entre deux cellules) ;
- des modèles de *machine learning* en identifiant les variables déterminantes pour prédire la probabilité de présence par analyse de l'auto-corrélation spatiale.

Ce défi a permis de mettre en place une coopération très enrichissante de spécialistes de la donnée et d'experts métiers thématiques. Une valorisation de ce travail est prévue par l'AFB.

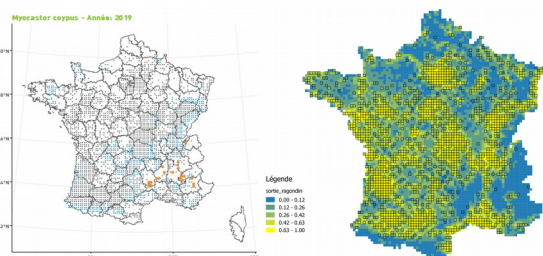


Figure 2. Prédiction de la présence de l'espèce *Moycastot coypus* en 2019 à l'aide de l'outil de simulation et d'aide à la décision à gauche, et d'un modèle de *machine learning* à droite (crédits : AFB / Cerema / ONCFS).

Défi – Détection automatique de friches industrielles

Carte des friches de l'Oise (60)

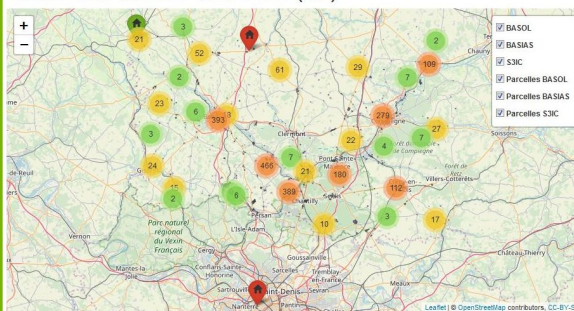


Figure 3. Carte de détection des friches industrielles potentielles sur le département de l'Oise (crédits : Cerema).

Sur proposition de la DREAL des Hauts-de-France, il s'agissait notamment de croiser les différentes données relatives aux friches industrielles : anciens sites industriels, sites pollués, données foncières, répertoires des entreprises...

Le sujet a été traité par une équipe du Cerema qui a établi une carte interactive présentant graphiquement les résultats sur le département de l'Oise. Pour chaque site, un descriptif présente les activités du site à partir des bases de données du ministère.

Pour en savoir plus, consultez l'article publié sur le site du Cerema :

<https://www.cerema.fr/fr/actualites/challenge-rst-analytics-2018-cerema-s-implique-detection>

RST-ANALYTICS : DES PREUVES DE CONCEPT QUI OUVERTENT DE NOUVELLES PERSPECTIVES

Les challenges RST-Analytics sont des moments forts de l'animation du RST : il s'agit de révéler et de réunir des *datascientists*, développeurs et experts de la donnée du RST en leur proposant des défis techniquement motivants et porteurs de sens pour les politiques publiques du pôle ministériel. Il ne s'agit pas seulement de résoudre des problèmes simples, en mode « *quick & dirty* » ou de vérifier des hypothèses formulées par les directions d'administration centrale ; il s'agit aussi et surtout de provoquer des rencontres fructueuses, de mobiliser des compétences rares et de se placer dans un espace « hors du cadre » à la fois intense, productif et tourné résolument vers l'impact.

Lors des trois éditions du challenge RST-Analytics, le réseau scientifique et technique a su convaincre, étonner

et montrer que « *c'est possible* ». Les prototypes réalisés dans un temps très court sont bien évidemment des outils minimalistes à l'impact encore limité. Néanmoins, **ce sont de réels démonstrateurs d'aide à la décision et surtout la preuve que le ministère dispose, en interne, de très fortes compétences sur les données.** Ils préfigurent donc des partenariats fructueux de confiance sur les données entre, d'une part, le réseau scientifique et technique et, d'autre part, les directions d'administration centrale ou les services déconcentrés. Ils démontrent toute la valeur d'une exploitation des données du pôle ministériel dans le cadre de ses missions. En ce sens, la démarche est un succès et apparaît porteuse d'un élan de transformation.

Les challenges RST-Analytics ont été gagnants pour tous les acteurs en présence :

- **Les porteurs de politiques publiques** peuvent exprimer leurs problèmes techniques, stratégiques ou complexes ; formuler des hypothèses ou intuitions autour de l'exploitation des données et de l'intelligence artificielle ; explorer, à faible coût, des pistes de solutions pour leurs projets complexes. Les challenges leur permettent également de côtoyer et d'échanger facilement avec des experts en science des données.
- **Les participants au challenge** sont reconnus en tant que *datascientists* du ministère et rencontrent leurs pairs durant ces deux jours pour, à terme, constituer une communauté d'experts de la donnée du RST. Ils peuvent faire preuve d'agilité, de créativité, de liberté et tester des méthodes et des modes de travail innovants et inédits.
- **Les organismes du RST** expérimentent, dans le cadre de cas d'usage métiers, des sujets émergents (*machine learning*, traitement du langage naturel⁷, appariement complexe de données...). Les challenges sont aussi des espaces de coopération entre

⁷ Ensemble des recherches et développements visant à modéliser et reproduire, à l'aide de machines, la capacité humaine à produire et à comprendre des énoncés linguistiques dans des buts de communication.

organismes.
La poursuite de la dynamique enclenchée avec ces trois premiers challenges contribuera à :

- consolider cette amorce de communauté métier de *datascientists* du ministère ;
- élargir encore le panel des porteurs de politiques publiques impliqués dans les défis et des organismes du RST en matière d'expertise technique mobilisée ;
- tester des dispositifs équivalents, dans des modalités différentes (périodicité, durée, sujets) : traiter un nombre plus limité de défis mais plus stratégiques et sur un temps plus long, lancer des challenges plus thématiques afin de renforcer le dialogue avec les porteurs de politiques publiques concernés...

* GLOSSAIRE

Hackathon : Événement au cours duquel des personnes aux compétences variées (entrepreneurs, graphistes, concepteurs d'interface, *datascientists*, programmeurs informatiques...) se réunissent durant plusieurs jours autour d'un projet collaboratif de programmation informatique ou de création numérique.

Semaine de l'Innovation Publique : rencontre permettant aux citoyens de découvrir et comprendre ce que recouvre l'innovation publique, les nouvelles méthodes de conception des politiques publiques qu'elle propose et les finalités qu'elle poursuit (amélioration de la qualité de service, accompagnement du changement, développement de l'administration numérique...)

REFERENCE

Ministère de la transition écologique et solidaire, La donnée : un nouveau bien essentiel pour la transition écologique et solidaire et la cohésion des territoires – 40 actions de la feuille de route 2018-2019, 2018 (www.ecologique-solidaire.gouv.fr/sites/default/files/SGD%20-%20Feuille%20de%20route.pdf).

Directeur de la publication : Thomas Lesueur, Commissaire général au développement durable
Auteurs : Luc Mathis, Sylvain Pradelle, Béatrice Yoffo
Dépôt légal : novembre 2019
ISSN : 2555-7564

Commissariat général au développement durable

Direction de la recherche et de l'innovation
Sous-direction de l'animation scientifique et technique
Tour Séquoia
92055 La Défense cedex
Courriel : anim-rst.ast3.dri.cgdd@developpement-durable.gouv.fr

www.ecologique-solidaire.gouv.fr

