

CONCOURS INTERNE ET 3^{ème} CONCOURS D'INGÉNIEUR DES SYSTÈMES D'INFORMATION ET DE COMMUNICATION

- SESSION 2021 -

Mardi 1^{er} juin 2021

EPREUVE ECRITE

CONCOURS INTERNE :

Résolution d'un cas pratique à partir d'un dossier à caractère technique permettant d'apprécier les qualités d'expression, d'analyse et de synthèse du candidat et sa capacité à conduire un projet.

3^{ème} CONCOURS :

Résolution d'un cas pratique à partir d'un dossier à caractère technique faisant appel à des connaissances relatives à l'environnement et à la technique des systèmes d'information et de communication et permettant de vérifier les capacités d'analyse et de synthèse du candidat ainsi que son aptitude à dégager des solutions appropriées.

Durée : 4 heures – Coefficient 1

**Le dossier documentaire comporte 24 pages.
(Hors page d'énoncé du sujet)**

Il vous est rappelé que votre identité ne doit figurer que dans l'en-tête de la copie (ou des copies) mise(s) à votre disposition. Toute mention d'identité ou tout signe distinctif porté sur toute autre partie de la copie ou des copies que vous remettez en fin d'épreuve entraînera l'annulation de votre épreuve.

Si la rédaction de votre devoir impose de mentionner des noms de personnes ou de villes et si ces noms ne sont pas précisés dans le sujet à traiter, vous utiliserez des lettres pour désigner ces personnes ou ces villes (A ..., B..., Y..., Z...).

- 1. LES COPIES SERONT RENDUES EN L'ÉTAT AU SERVICE ORGANISATEUR. À L'ISSUE DE L'ÉPREUVE, CELUI-CI PROCÉDERA À L'ANONYMISATION DE LA COPIE.**
- 2. NE PAS UTILISER DE CORRECTEUR D'ORTHOGRAPHE SUR LES COPIES.**
- 3. ÉCRIRE EN NOIR OU EN BLEU – PAS D'AUTRE COULEUR.**
- 4. IL EST RAPPELÉ AUX CANDIDATS QU'AUCUN SIGNE DISTINCTIF NE DOIT APPARAÎTRE SUR LA COPIE.**

Sujet :

Vous êtes affecté(e) en tant que chef de projet au département PNM2 - Produit Numérique Métiers (logement, urbanisme) du service du Numérique (SNUM) du Ministère de la transition écologique.

La Ministre a chargé la direction générale de l'aménagement, du logement et de la nature (DGALN) de mettre à disposition des citoyens, sur le site internet data.gouv.fr, les données relatives aux loyers qu'elle a déjà en sa possession.

Vous êtes chargé(e) d'accompagner cette direction pour répondre à la demande de la Ministre .

En qualité d'ingénieur SIC, vous rédigez, en vous appuyant sur la documentation jointe, une note de cadrage à l'attention de votre sous-directeur qui présente :

- la problématique et les enjeux liés à l'ouverture des données ;
- la conduite du projet ;
- la gouvernance des données ;
- la solution de réalisation proposée.

DOSSIER DOCUMENTAIRE :

Document 1	Direction générale de l'aménagement, du logement et de la nature (DGALN) Source : Ministère de la transition écologique	Page 1
Document 2	Organigramme du service du numérique Source : Ministère de la transition écologique	Page 2
Document 3	Politique de la donnée: état des lieux – Résumé du rapport Bothorel – 23 décembre 2020 – Open data: l'IGN libère ses données – 13 décembre 2020 https://www.zdnet.fr/blogs/l-esprit-libre/open-data-ign-libere-ses-donnees-39914791.htm – Extrait : définition de l'open data https://www.lebigdata.fr/open-data-definition	Pages 3 à 5
Document 4	Eléments d'architecture - Exemple de données sources – mars 2018 https://www.cohesion-territoires.gouv.fr/sites/default/files/2019-05/Comite_scientifique_de_l'observation_des_loyers_Prescriptions_methodologiques_mars_2018.pdf - Architecture générale	Pages 6 à 7
Document 5	Valoriser ses données - Extrait : la préparation des jeux de données https://guides.etalab.gouv.fr/qualite/preparer-le-jeu-de-donnees/#la-structure-du-jeu-de-donnees - Extrait : l'intégration des données https://fr.wikiversity.org/wiki/Pentaho/Int%C3%A9gration_de_donn%C3%A9es - Extrait : Comment publier sur data.gouv.fr – 11 février 2021 https://guides.etalab.gouv.fr/pdf/guide-data.gouv.fr.pdf	Pages 8 à 11
Document 6	Extrait du sommaire-type d'une note de cadrage Source: Site de la Direction du numérique du Ministère de l'intérieur	Page 12
Document 7	Extrait : Qu'est-ce que le dataops ? Source : https://www.lemagit.fr/conseil/Quest-ce-que-le-DataOps-Data-Operations	Page 13
Document 8	RGPD : se préparer en 6 étapes Source : Site internet de la CNIL	Page 14
Document 9	Les métiers de la data Source : https://corporate.apec.fr/files/live/sites/corporate/files/Nos%20%c3%a9tudes/pdf/Les-metiers-de-la-data.pdf	Pages 15 à 17
Document 10	Les 6 étapes incontournables pour bien gérer un projet Source : https://www.conseilsmarketing.com/promotion-des-ventes/les-5-etapes-incontournables-pour-bien-gerer-un-projet/	Pages 18 à 21
Document 11	L'anonymisation des données personnelles Source : Site internet de la CNIL	Pages 22 à 24

Direction générale de l'aménagement, du logement et de la nature (DGALN)

La Direction générale de l'aménagement, du logement et de la nature (DGALN) élabore, anime et évalue les politiques de l'urbanisme, de la construction, du logement, des paysages, de la biodiversité, de l'eau et des substances minérales non énergétiques. Elle s'organise autour de deux directions opérationnelles : la direction de l'habitat, de l'urbanisme et des paysages (DHUP) et la direction de l'eau et de la biodiversité (DEB).

Missions

Les directions, services déconcentrés et opérateurs de la DGALN œuvrent au quotidien à la définition et à l'application des politiques publiques et législations concourant à la transition écologique et solidaire, à la cohésion des territoires et à l'amélioration du cadre de vie. Leurs actions portent notamment sur 7 enjeux prioritaires :

- **Des logements pour tous** : répondre aux besoins de logements accessibles à tous, partout sur le territoire, en particulier dans les territoires en tension et pour les personnes en difficulté. Garantir l'offre de mixité des logements et encadrer les politiques de logements sociaux ;
- **La rénovation énergétique des bâtiments** : réduire la consommation énergétique des bâtiments et les émissions de gaz à effet de serre de ce secteur et améliorer les performances des logements existants, notamment pour lutter contre la précarité énergétique ;
- **Un urbanisme durable** : planifier et aménager le territoire de façon résiliente et dans le but d'assurer un cadre de vie agréable aux habitants, qui facilite la mobilité, la mixité des fonctions et des populations, tout en réduisant la consommation des espaces naturels ;
- **Une bonne gestion des ressources en eau et minérales** : viser une exploitation des ressources naturelles (grand cycle et petit cycle de l'eau, extractions minérales et carrières...) économe, respectueuse de l'environnement et bien insérée dans le territoire, dans une logique de développement durable ;
- **Des milieux marins et littoraux de qualité** : garantir une mer saine, propre et productive, un bon fonctionnement des écosystèmes marins et littoraux et un usage durable des biens et services associés ;
- **Protection des espaces et des paysages** : protéger les espaces et les paysages en s'appuyant notamment sur le principe d'aires protégées (parcs et réserves naturels), sur le réseau Natura 2000 et sur les plans d'action en faveur des milieux humides. Attachée à la protection des paysages, la politique des sites vise quant à elle à préserver des lieux dont le caractère exceptionnel justifie une protection de niveau national ;
- **La préservation de la biodiversité** : lutter contre l'érosion de la biodiversité en France, préserver les espèces et leurs habitats, les espaces naturels et les continuités écologiques.

Organisation

Œuvrant à favoriser les synergies entre les acteurs du l'aménagement durable, de la biodiversité et de la qualité de vie dans les territoires, la direction générale rassemble plus de 590 agents et s'appuie sur quelque 15 000 agents en services déconcentrés.

La DGALN s'appuie également sur une soixantaine d'opérateurs, //. Au sein de la DGALN, **deux directions opérationnelles** appuyées par un service des affaires générales et de la performance se répartissent les sujets d'expertise et contribuent à la mise en œuvre des politiques de cohésion des territoires et de transition écologique.

La direction de l'habitat, de l'urbanisme et des paysages (DHUP)

Elle doit répondre aux besoins en logement et en hébergement de nos concitoyens et notamment contribuer à programmer la production de logements à la bonne échelle du territoire national. Elle vise aussi à améliorer la gestion de l'offre de logements existante, pour répondre notamment aux besoins des plus défavorisés.

Elle favorise un aménagement durable sur tous les types de territoires en veillant à ce que les documents de planification et les opérations d'aménagement répondent aux besoins des populations et intègrent les enjeux de développement durable et en particulier ceux relatifs au logement. Cette politique doit aussi favoriser et accompagner les projets des territoires là où les enjeux sont complexes (Grand Paris, métropoles, Outre-mer, territoires ruraux, montagne, littoral...). Il s'agit également de mettre en œuvre la politique du droit des sols en lien avec les collectivités compétentes dans le cadre du code de l'urbanisme.

Elle contribue à faire du bâtiment un levier effectif du développement durable et à ce titre œuvre à l'amélioration de la qualité, de l'accessibilité et des performances énergétiques des bâtiments, dans le cadre du code de la construction et de l'habitation.

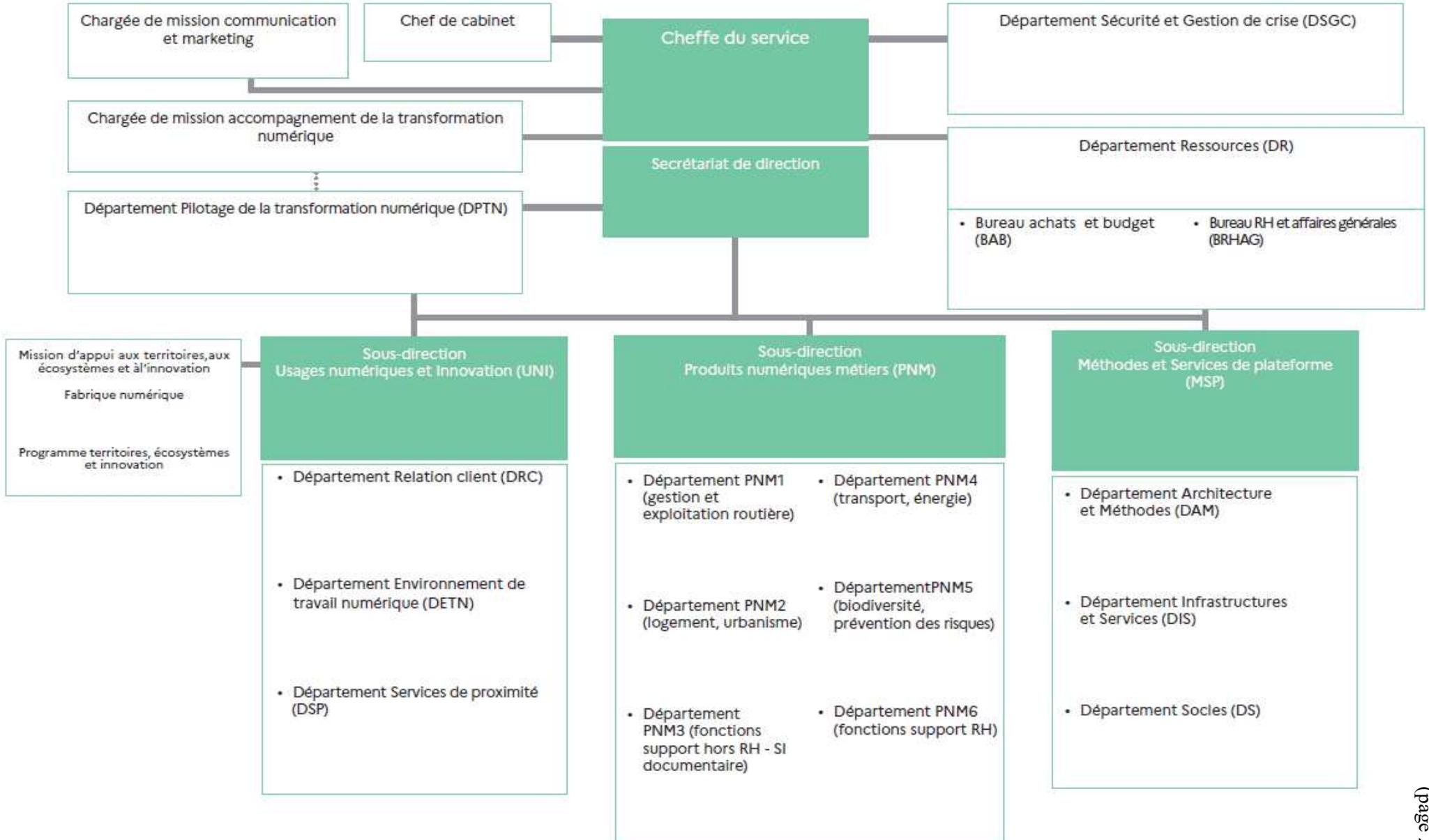
Elle travaille à l'amélioration du cadre de vie de sorte à répondre aux besoins des habitants et aux enjeux de la transition énergétique et écologique, en particulier en matière de paysage.

La direction de l'eau et de la biodiversité (DEB)

La mission première de la direction de l'eau et de la biodiversité est la conception, l'évaluation et la mise en œuvre des politiques de l'eau, des espaces naturels, de la biodiversité terrestre et marine et des ressources minérales non énergétiques en vue de garantir la préservation et un usage équilibré de ces ressources.

Organigramme du service du numérique

Secrétariat général (SG) Service du numérique (SNUM)



2

Politique de la donnée: état des lieux

Rapport Bothorel : pour une nouvelle ère de la politique publique de la donnée

Alors qu'elle était à l'avant-garde des pays européens en matière de politique de la donnée et des codes sources depuis 2013, la France perd peu à peu son avance. Au-delà du constat, le rapport "Pour une politique publique de la donnée" d'Éric Bothorel rappelle les enjeux économiques, scientifiques et démocratiques sous-jacents et prône, à travers une série de recommandations, une politique d'ouverture dans un délai court.

Lever les freins existants à l'ouverture et au partage des données

Après avoir évoqué les freins actuels (procédures de partage trop contraignantes, ouverture des données réduite à l'*open data* par méfiance à l'égard des utilisateurs...), le rapport met en avant une série d'arguments en faveur d'une plus grande ouverture des données. Parmi ceux-ci :

- une plus grande **transparence** vis-à-vis du public ; à l'inverse, une opacité dans la communication des données peut être la source d'une défiance à l'égard des institutions ;
- un **gain en performance et en réactivité** dans des situations d'urgence grâce au recueil d'informations ;
- un levier économique pour la création de nouveaux services ou d'entreprises innovantes, en particulier dans les secteurs de **l'intelligence artificielle** ;
- une **meilleure évaluation** des politiques publiques grâce au croisement de données issues de différentes administrations.

L'IGN ouvre ses données au 1er janvier 2021

Open Source : L'Institut national de l'information géographique et forestière (IGN) diffusera ses données sous licence libre à partir de janvier 2021.

Les utilisateurs de données géographiques vont gagner un an: l'Institut national de l'information géographique et forestière (auparavant Institut géographique national, IGN) annonce l'ouverture de ses données publiques le mois prochain.

Des centaines de téraoctets de données libérées.

«Alors que l'ouverture de l'ensemble des données publiques demandée par le Gouvernement était prévue en 2022, l'IGN anticipe et annonce l'ouverture de ses données publiques à compter du 1er janvier 2021. Cet accès gratuit et libre pour tous, en téléchargement ou par flux, contribue au plan national de relance de l'économie.

L'ouverture des données IGN sous licence ouverte Etalab 2.0 [compatible entre autres avec la licence Creative Commons Attribution, ou CC by], signifie un accès et un usage gratuit pour tous.

Pas de droits de reproduction, ni de diffusion à régler. Seules les données incluant des droits de tiers ou protégées par le droit d'auteur (cartes et SCANS) ne sont pas concernées par cette ouverture. Parmi les données importantes qui seront désormais ouvertes», indique l'IGN, on peut citer la BD TOPO, la BD ORTHO, la BD forêt et le RGE Alté (voir cette page de l'IGN sur ses bases de données sous licence ouverte).

Sylvain Latarget, directeur général par intérim de l'IGN, commente: «Pour tous ceux qui attendaient cette ouverture avec impatience, la mise à disposition de ces données publiques, qui représentent la bagatelle de quelques centaines de téraoctets, se fera progressivement jusqu'à l'ouverture de notre nouvel outil de diffusion massif des données prévue à la fin du 1er trimestre 2021.»

"Ce n'est pas la vente de données qui crée de la valeur, mais sa circulation"

La députée Valéria Faure-Muntian (LREM), qui en juillet 2018 avait présenté un rapport sur les données géographiques souveraines, a salué sur Twitter cette ouverture, «un grand pas pour Internet libre et la démocratisation de données essentielles dans notre quotidien».

Dans son rapport, la députée recommandait notamment (page 11 du PDF) de «diffuser à terme l'ensemble des données géographiques souveraines sous licence ouverte type Etalab, n'autoriser l'IGN à recourir à la licence payante qu'à titre transitoire et au plus tard jusqu'en 2022».

Elle relevait (pages 44-45) que «la gratuité de la diffusion et de la réutilisation des données géographiques souveraines implique de faire financer leur production par la subvention, à défaut de pouvoir le faire par la vente de ces données. Si le modèle d'affaires de l'open data se vérifie empiriquement, un retour vers les caisses publiques s'effectuera par la fiscalisation des richesses supplémentaires créées grâce à la libération des données. (...)

En effet, l'économie de la donnée n'est pas régie pas les mêmes règles que l'économie réelle. On constate que la richesse de la donnée est établie à partir de l'utilisation qui en est faite et donc de l'accès qui lui est donné. Ainsi, 'ce n'est pas la vente de données qui crée de la valeur, mais sa circulation'*. La mise à disposition gratuite de ces données créera beaucoup de valeur, permettant par exemple à de petites entreprises innovantes ou aux citoyens impliqués de participer au débat et de proposer des solutions qui ne pourraient être à l'initiative d'acteurs publics uniquement.»

* La phrase est de Simon Chignard, data editor à Etalab, lors d'une conférence au Syntec numérique du 4 avril 2018, précise le rapport.

Définition de l'open-data

Open data : définition normalisée

Le terme Open Data désigne des données auxquelles n'importe qui peut accéder, que tout le monde peut utiliser ou partager. Les critères essentiels de l'Open Data sont [la disponibilité, la réutilisation et la distribution, et la participation universelle](#). Il s'agit là de la définition donnée par l'Open Knowledge Foundation en 2005.

- **Disponibilité et accès** : Les données doivent être pleinement accessibles, moyennant un coût de reproduction raisonnable. De préférence, elles se téléchargent sur Internet. La forme doit être confortable et modifiable.
- **Réutilisation et redistribution** : Les données doivent être fournies sous des conditions permettant la réutilisation et la redistribution, incluant le mélange avec d'autres ensembles de données.
- **Participation universelle** : Tout le monde doit être en mesure d'utiliser, de réutiliser et de redistribuer les données. Il ne doit y avoir aucune discrimination concernant les fins d'utilisation, ou contre des personnes ou des groupes. Par exemple, des restrictions non commerciales qui empêchent l'utilisation commerciale, ou les restrictions d'usage à certains secteurs, ne sont pas compatibles avec l'Open Data.

Ces trois critères sont l'essence de l'Open Data, car ils autorisent l'interopérabilité. L'interopérabilité désigne la capacité de différentes entreprises ou systèmes à travailler ensemble. En l'occurrence, **l'interopérabilité est la capacité de mélanger différents ensembles de données.**

L'interopérabilité, essence de l'Open Data

Cette interopérabilité est importante, car elle permet à différents composants de fonctionner ensemble. C'est ce qui permet de créer des systèmes larges et complexes. Sans interopérabilité, c'est tout simplement impossible. On peut **prendre pour exemple le mythe de la Tour de Babel**. Dans cette légende, la communication impossible empêche complètement la construction de la tour.

Dans le cas des données, **la mise en commun repose sur la possibilité de mélanger librement ces données**. Cette interopérabilité est essentielle pour tirer des bénéfices de l'ouverture. Il est ensuite possible de développer des produits et des services en plus grande quantité, et d'une meilleure qualité.

Cette interopérabilité ne se repose pas seulement sur le partage de données. Il faut que les jeux de données utilisent un langage de programmation commun ou qu'un élément programmatique fasse l'intermédiaire entre ces informations. **Pour cela, le W3C (World Wide Web Consortium) prône l'application de standards de l'Open Data. Ils permettent de répondre à ce besoin de communication.** Cela passe par l'utilisation du langage Web sémantique, qui facilite le classement et la recherche de données.

Open Data et RGPD : comment concilier les deux

*A priori, le RGPD censé assurer la confidentialité des données semble aller à l'encontre du principe de l'Open Data visant à mettre l'information à disposition de tous. De fait, **les deux concepts pourraient sembler incompatibles.***

*Cependant, **la loi française a anticipé l'entrée en vigueur du RGPD.** Ainsi, la loi Lemaire et l'article L.312-1-2 du code des relations entre le public et l'administration aux collectivités de mettre à disposition l'ensemble des informations concernant le public, mais uniquement en rendant au préalable impossible l'identification des personnes concernées par ces données personnelles.*

*Néanmoins, on peut craindre que le RGPD freine l'ouverture des données, voire même que **l'anonymisation des données publiques les rende tout bonnement inutiles.** On pourrait même considérer que le RGPD va finalement à l'encontre de l'intérêt des citoyens en atténuant la transparence des données publiques...*

*La seule conciliation viable à l'heure actuelle semblerait donc **que les collectivités prennent en compte la notion de vie privée dès la création de données,** afin de pouvoir les diffuser sans avoir à les anonymiser. C'est d'ailleurs l'idée suggérée par le RGPD à travers la notion de « privacy by design » ...*

Éléments d'architecture

Exemple de données sources

1. Identifiant logement de l'OLL -----	2. Source 1. Administrateurs de biens 2. Agences immobilières 3. Personnes morales 4. Autres professionnels (notaires...) 5. Bailleurs en gestion directe 6. Locataires en gestion directe 7. Autres sources	3. Mode de collecte 1. Transfert de fichiers 2. Classique (rendez-vous, téléphone, mail...) 3. Internet 4. Données centralisées				
4. Date d'enquête Enquête précédente	5. Adresse du logement N° ----- B/T/Q _ Type de rue ----- Nom de rue ----- Compl. Adresse ----- CP ----- Commune -----		6. Insee Zone Iris Sec. Cadas. Parcelle -----			
7. Date de référence si autre que le 1 ^{er} janvier -- --	8. Si enquête auprès d'un professionnel N° de carte G (ou T) ----- N° de gestion du logement ----- Raison sociale -----					
<i>21. Remarques pour le centre national de traitement</i> -----						
Description de l'immeuble						
31. Type d'habitat 1. individuel 2. collectif --	32. Epoque de construction 1. avt 1919 à défaut 2. 1919-1945 3. avt 1946 4. 1946-1970 5. 1971-1990 6. 1991-2005 7. après 2005		33. Année de construction -----	34. Type de propriété 1. mono 2. copropriété	35. Equipement 1. oui, 2. non Asc. Interphone Vidéophone Gardien Digicode	
Description du logement						
41. Nombre de pièces --	42. Surface --- m ²	43. Etage --	44. Annexes 1. oui, 2. non Parking, Balcon si inclus dans loyer Jardin	45. Nombre de WC SdB -- --	46. Type de chauffage 1. individuel 2. collectif	47. Type de propriétaire 1. particulier 2. personne morale
Informations relatives au bail						
50. DPE consommation énergétique lettre --	51. Type de location 1. vide 2. meublé	52. Type de gestion 1. déléguée 2. directe	53. Catégorie de location 1. libre 2. conventionnée 3. défiscalisée	54. Date d'entrée du locataire -----	55. Première location si emménagé depuis le 1^{er} janvier N-1 1. oui, 2. non	
56. Loyer mensuel hors charges à la date d'enquête 1 an avant (si évol) -----, ___ € année en cours -----, ___ €		57. Loyer mensuel hors charges à la date de référence (1^{er} janvier en général) 1 an avant (si évol) -----, ___ € année en cours -----, ___ €		58. Provisions mensuelles pour charges à la date d'enquête 1 an avant (si évol) -----, ___ € année en cours -----, ___ €		

(En gras : indispensable, en normal non gras : indispensable à terme, en italique : optionnel)

Statut d'occupation des résidences principales en France

Statut d'occupation des résidences principales en %

Statut d'occupation	2015	2016 (r)	2017 (r)	2018 (p)	2019 (p)	2020 (p)
Propriétaires	57,7	57,7	57,7	57,6	57,6	57,6
Non accédants	37,9	37,8	37,8	37,6	37,6	37,6
Accédants	19,8	19,9	19,9	20,0	20,0	20,0
Locataires	39,7	39,8	39,8	39,9	39,9	40,0
Bailleurs publics (1)	17,2	17,1	17,1	17,1	17,0	17,0
Bailleurs privés (2)	22,5	22,7	22,7	22,8	22,9	23,0
Autres statuts (3)	2,6	2,5	2,5	2,5	2,5	2,4
Total des résidences principales (en milliers)	29 012	29 237	29 480	29 727	29 974	30 231

p : données provisoires.
r : données révisées.

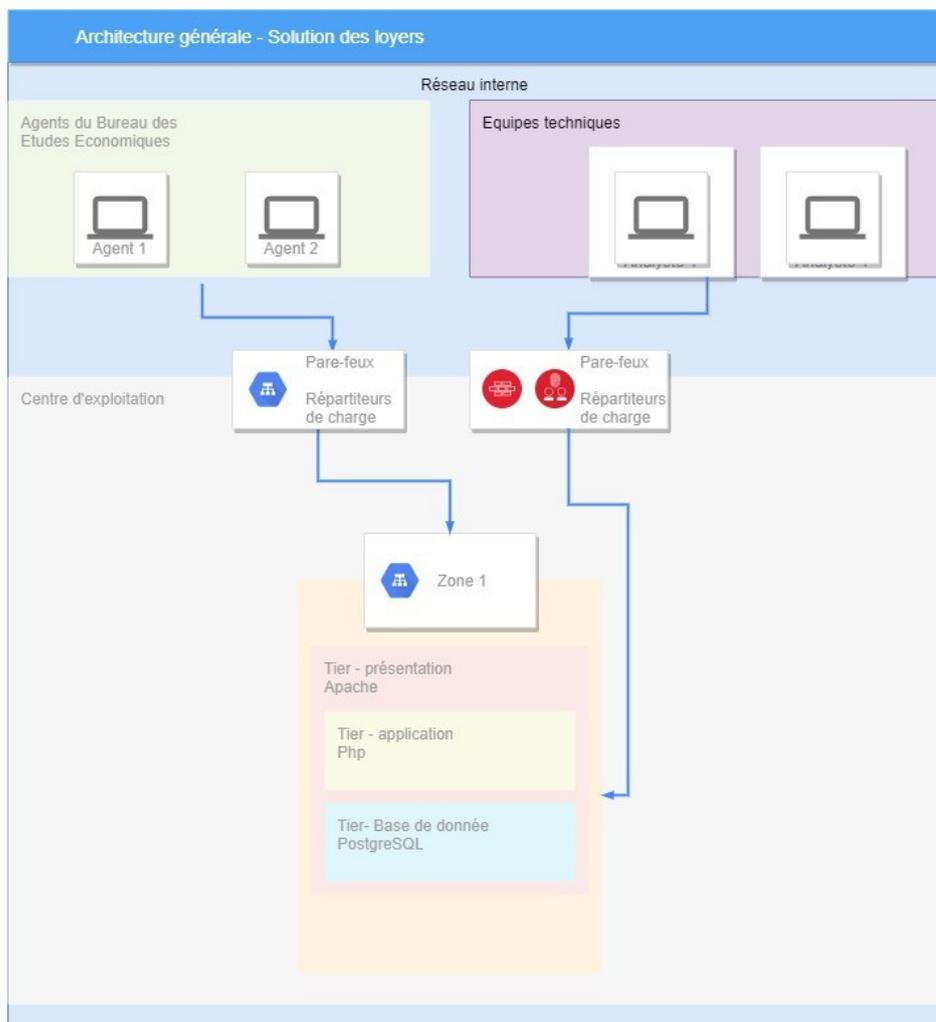
(1) : les bailleurs publics regroupent les organismes d'habitation à loyer modéré (HLM), les sociétés d'économie mixte (SEM), l'État, les collectivités locales et établissements publics ; que le logement soit conventionné ou non et que son loyer relève de la législation HLM ou non.

(2) : les bailleurs privés se composent des ménages propriétaires bailleurs et des autres bailleurs personnes morales.

(3) : regroupe essentiellement les ménages logés gratuitement.

Note : données au 1^{er} janvier.
Champ : France hors Mayotte.
Sources : Insee et SDES, *estimations annuelles du parc de logements*.

Architecture générale



Valoriser ses données

La préparation des jeux de données

Extraire le jeu de données du système d'information de l'organisation

Il est possible que les données que vous souhaitez faire circuler ne soient pas structurées sous la forme d'un jeu de données. Dans cette situation, il est nécessaire de réaliser une extraction des données depuis le système d'information où elles sont stockées. Cette extraction permet d'obtenir un jeu de données structuré, qui ordonne les données en fonction de différentes caractéristiques.

Lorsque vous cherchez à extraire des données d'un système d'information, plusieurs situations peuvent se présenter :

- Le système d'information propose un outil qui permet d'exporter l'ensemble des données depuis le système d'information. Il est nécessaire de sélectionner les données éligibles à la circulation en aval de l'export ;
- Le système d'information propose un outil qui permet d'exporter l'ensemble des données ou de sélectionner un sous ensemble des données à exporter depuis le système d'information ;
- Le système d'information ne prévoit pas d'outil d'exportation des données. Dans ce cas, il est nécessaire de réaliser une opération technique qui permette de réaliser l'export des données. Cette opération est directement liée aux spécificités du système d'information utilisé.

Quel que soit le mode d'export des données, il est recommandé d'automatiser l'opération réalisée afin de faciliter la mise à jour des données publiées. Cette automatisation instaure un processus sur le long terme et fait gagner du temps à l'organisation.

La structure du jeu de données

Les jeux de données qui ont vocation à circuler seront réutilisés par des acteurs tiers qui ne connaissent pas l'environnement de votre administration. Il est nécessaire de proposer une structure de jeu de données compréhensible et appropriable par tous.

Deux approches sont envisageables :

- La structure de votre jeu de données correspond à un schéma de données existant ;
- La structure de votre jeu de données ne correspond pas à aucun schéma de données existant. Un travail de modélisation est nécessaire en amont de la création du jeu de données

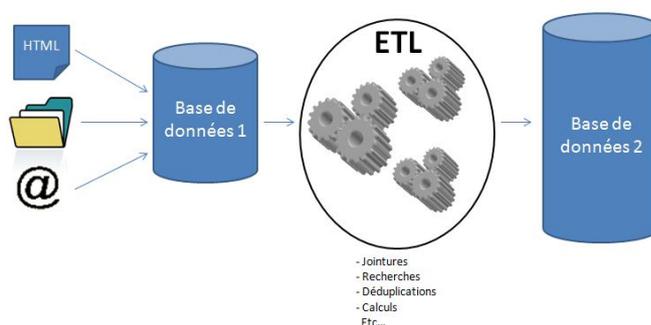
Le choix du format du jeu de données

Afin que le maximum d'utilisateurs, internes ou externes à votre organisation, puisse s'approprier les données, il est conseillé de les faire circuler dans un format :

- **Ouvert** : un format ouvert n'impose pas de spécifications techniques qui entraveraient l'exploitation des données (par exemple l'utilisation d'un logiciel payant) ;
- **Aisément réutilisable** : un format aisément réutilisable sous-entend que toute personne ou machine peut réutiliser facilement le jeu de données ;
- **Exploitable par un système de traitement automatisé** : un système de traitement automatisé permet de réaliser des opérations par des moyens automatiques, relatifs à l'exploitation des données. Par exemple, un fichier CSV est aisément exploitable par un système de traitement automatisé contrairement à un fichier PDF

Intégration des données

L'**intégration de données** appelé ETL (Extraction Transfer Loading) regroupe les processus par lesquels les données provenant de différentes parties du système d'information sont déplacées, combinées et consolidées. Ces processus consistent habituellement à extraire des données de différentes sources (bases de données, fichiers, applications, Services Web, emails, etc.), à leur appliquer des transformations (jointures, lookups, déduplication, calculs, etc.), et à envoyer les données résultantes vers les systèmes cibles.



COMMENT PUBLIER SUR DATA.GOUV.FR ?

Extrait du guide « Publier les données sur data.gouv.fr », Etalab (11/02/2021)

4. Publier un jeu de données

4.1 Définir qui publie le jeu de données

Un jeu de données peut être publié sous le nom de votre compte utilisateur ou sous la bannière d'une organisation.

4.1.1 Quand est-il conseillé de publier sous le nom de votre compte utilisateur ?

Nous vous conseillons de publier un jeu de données sous le nom de votre compte utilisateur s'il n'a pas été produit dans le cadre des activités d'une organisation à laquelle vous êtes rattaché.

4.1.2 Quand est-il conseillé de publier au nom d'une organisation ?

Nous vous conseillons de publier un jeu de données au nom d'une organisation s'il a été produit ou reçu dans le cadre des activités de cette organisation. Tous les jeux de données peuvent être édités par les membres de l'organisation. Les jeux de données seront centralisés et facilement accessibles pour les réutilisateurs et nouveaux membres de l'organisation.

4.2 Décrire le jeu de données

Afin de présenter le jeu de données aux utilisateurs de data.gouv.fr, il est nécessaire de décrire les données que vous publiez. Cette étape est cruciale pour que vos jeux de données soient bien référencés et que les réutilisateurs s'approprient facilement les ressources partagées. Lors de la création de la page de votre jeu de données, vous avez la possibilité de renseigner les informations suivantes :

Information	Description de l'information
Titre*	Le titre de votre jeu de données doit être le plus précis et spécifique possible. Il doit également correspondre au vocabulaire employé par les utilisateurs. Ces derniers recherchent les données le plus souvent dans un moteur de recherche.
Sigle	Vous avez la possibilité d'apposer un sigle à votre jeu de données. Les lettres qui composent ce sigle n'ont pas besoin d'être séparées par des points.
Description*	La description de votre jeu de données permet aux personnes qui le consultent d'obtenir des informations sur le contenu et la structure des ressources publiées, le contexte de production des données, les contacts producteurs, etc. C'est généralement la première chose que les utilisateurs lisent quand ils découvrent votre jeu de données.
Licence	Les licences définissent les règles de réutilisation des jeux de données publiés. En choisissant une licence de réutilisation, vous vous assurez que le jeu de données publié sera réutilisé selon les conditions d'usage que vous avez définies. Afin d'éviter la multiplication des licences, la Loi pour une République numérique a prévu la création d'une liste de licences qui peuvent être utilisées par les administrations. Le site data.gouv.fr a référencé la liste des licences applicables aux informations publiques (données, documents...).
Fréquence de mise à jour*	La fréquence de mise à jour correspond à la fréquence à laquelle vous prévoyez de mettre à jour les données publiées. Cette fréquence de mise à jour reste indicative.
Mots clés	Les mots clés caractérisent votre jeu de données. Ils apparaissent sur la page de présentation et apportent un meilleur référencement du jeu de données lors d'une recherche utilisateur. L'apposition d'un mot clé permet également de classer votre jeu de données dans des thématiques. À partir de chaque mot clé, vous pouvez obtenir la liste des jeux de données pour lesquels le mot clé a également été assigné.
Couverture temporelle	La couverture temporelle indique la portée dans le temps des données publiées.
Granularité spatiale	La granularité spatiale indique le niveau de détail géographique le plus fin que peut couvrir vos données.
Mode privé	L'activation du mode privé permet de ne pas mettre en ligne le jeu de données. Cela laisse la possibilité de l'éditer avant sa publication.

Il est obligatoire de renseigner ces informations.

4.3 Ajouter les ressources

Qu'est-ce qu'une ressource ?

Une ressource est un fichier téléchargeable qui contient des informations à propos du jeu de données. Un jeu de données peut contenir plusieurs ressources (données mises à jour, données historisées, documentation, code source, API, lien, etc.).

Vous avez la possibilité d'importer vos ressources sur data.gouv.fr selon différents modes de mise à disposition.

4.3.1 Mise à disposition directe sur data.gouv.fr

Vous avez la possibilité d'importer vos ressources directement depuis la plateforme data.gouv.fr. Les étapes de cette publication reprend les étapes précédemment citées :



Figure 4 – Création d'un jeu de données

Lors de l'étape "Ajoutez vos ressources", deux options vous sont proposées:

1. Vous pouvez télécharger vos ressources depuis votre ordinateur vers le serveur de data.gouv.fr. Vos ressources seront alors hébergées sur les serveurs de data.gouv.fr.

2. Vous pouvez créer un lien vers une ressource distante existante. Les informations contenues dans le fichier resteront hébergées sur le serveur distant fléché.

(...)

4.3.2 Mise à disposition par API

Qu'est-ce qu'une API ?

Une API est une interface, un contrat passé entre deux systèmes informatiques pour leur permettre de communiquer. Cette solution informatique permet d'automatiser des tâches depuis votre ordinateur ou vos serveurs.

Dans l'administration, les API sont déjà largement utilisées. Par exemple, les entreprises françaises utilisent tous les mois celle de la Déclaration Sociale Nominative. Les entreprises, elles, ne communiquent pas les données de leurs employés manuellement à l'administration : leur logiciel de paie transmet directement ces données grâce à une API.

A partir de l'API de data.gouv.fr, vous pouvez réaliser les mêmes actions que sur la plateforme :

- Vous pouvez créer un jeu de données au nom de votre compte utilisateur ou au nom de votre organisation;
- Vous pouvez décrire votre jeu de données et les ressources associées;
- Vous pouvez ajouter ou supprimer une ressource ou un jeu de données.

L'API de data.gouv.fr propose également des fonctionnalités complémentaires à la publication de jeux de données :

- Vous pouvez récupérer les métadonnées des jeux de données ou ressources publiées sur data.gouv.fr;
- Vous pouvez accéder au contenu des ressources d'un jeu de données.

4.3.3 Quand utiliser l'API de data.gouv.fr ?

À la différence du mode de mise à disposition directe des données sur data.gouv.fr, l'utilisation de l'API permet de réaliser des actions de manière automatisée depuis votre ordinateur ou vos serveurs. Il est par conséquent conseillé d'utiliser une API lorsque la fréquence de publication d'un jeu de données est régulière.

4.3.4 Comment utiliser l'API de data.gouv.fr ?

L'utilisation de l'API de data.gouv.fr se fait par le point d'entrée racine de l'API. Afin de pouvoir exécuter des opérations d'écriture, il est nécessaire d'obtenir une clé API. Cette clé est accessible depuis les paramètres de votre profil administrateur.



Figure 6 – Générer une clé d'API

A partir de la clé API, vous pouvez exécuter différentes opérations. Ces opérations sont référencées dans la documentation disponible sur doc.data.gouv.fr.

Les appels à l'API sont soumis aux mêmes permissions que l'interface web. Par exemple, si vous souhaitez publier ou modifier un jeu de données au nom d'une organisation, vous devez appartenir à cette organisation.

Il ne vous reste plus qu'à publier votre jeu de données !

4.4 Publier un catalogue de données existant par moissonnage

Qu'est-ce que le moissonnage ?

Le moissonnage est un mécanisme permettant de collecter les métadonnées sur un catalogue distant et de les stocker sur une autre plateforme afin de proposer un second point d'accès aux données.

Le service de moissonnage mis à votre disposition permet de référencer sur data.gouv.fr les jeux de données publiés sur d'autres catalogues de données en ligne. De cette manière, vous n'avez pas besoin d'importer à la main sur data.gouv.fr les jeux de données que vous avez déjà importés sur votre propre plateforme.

4.4.1 Quand utiliser le service de moissonnage ?

Si vous mettez en ligne des données publiques sur une plateforme ouverte, dans un format dont les métadonnées correspondent à la syntaxe ODS, CKAN, ou DCAT vous pouvez les référencer automatiquement sur data.gouv.fr en utilisant notre service de moissonnage.

4.4.2 Comment utiliser le service de moissonnage ?

Il est possible de demander au moissonneur d'importer l'ensemble des données ou de ne sélectionner que certains jeux de données au moyen de filtres. Il n'est pas nécessaire de créer un moissonneur par jeu de données à importer, un seul moissonneur par portail suffit.

Le principe du moissonnage sur data.gouv.fr se décompose en plusieurs étapes :

1. Vous créez un moissonneur sur data.gouv.fr afin que la plateforme suive l'activité de votre plateforme ;
2. Vous publiez des données sur votre plateforme d'open data ;
3. Le moissonneur de data.gouv.fr vient automatiquement récupérer les données de votre plateforme ;
4. Les données de votre plateforme sont référencées et visibles sur data.gouv.fr.

4.5 Faire vivre son jeu de données

4.5.1 Obtenir des informations à propos de votre jeu de données

Afin de suivre la vie de votre jeu de données sur data.gouv.fr, vous avez la possibilité de suivre ses statistiques d'utilisation depuis votre compte administrateur. Un tableau de bord centralise les informations relatives au jeu de données :

- Sa couverture temporelle, sa fréquence de mise à jour et ses mots-clés ; sa couverture spatiale ;
- son statut de disponibilité et les téléchargements associés à chaque ressource publiée ;
- ses réutilisations ;
- ses anomalies ;
- ses abonnés ;
- ses ressources communautaires.

Des statistiques d'audience et de téléchargement sont disponibles sur le tableau de bord de votre compte administrateur.

Qualité du jeu de données

La qualité de votre jeu de données est fondamentale pour qu'il soit réutilisé par le plus d'utilisateurs possible. Afin de vous guider, un encart "Qualité" est mis à votre disposition dans le tableau de bord de chaque jeu de données. L'objectif est de vous aider à améliorer la qualité des (méta)données à partir de six critères :

- Le jeu de données possède-t-il une description ?
- Des mots clés sont-ils associés à votre jeu de données ?
- Le format du jeu de données est-il ouvert ?
- Des discussions à propos du jeu ont-elles été ouvertes ?
- Le jeu de donnée est-il à jour ?
- Les ressources du jeu de données sont-elles accessibles ?

(...)

4.5.2 Mettre à jour ou modifier un jeu de données et/ou une ressource

Les données publiées sur data.gouv.fr peuvent être mises à jour après leur publication, que la modification porte sur un jeu de données (sa description, ses tags, etc.) ou sur l'une des ressources qu'il contient.

Un producteur qui s'engage dans une logique de publication de ses données a intérêt à actualiser les ressources publiées le plus souvent possible. Ce critère de fraîcheur est critique pour les réutilisateurs qui fondent leurs services et produits sur les données publiées sur data.gouv.fr.

Rappel juridique

Les producteurs de données dont la publication présente un intérêt économique, social, sanitaire ou environnemental sont tenus de mettre à jour régulièrement les jeux de données publiés (Article L312-1-1 du CRPA).

(...)

Sommaire-type d'une note de cadrage

LA DEMANDE INITIALE DU CLIENT.....	5
OBJET DE LA DEMANDE.....	5
ORIGINE DE LA DEMANDE.....	5
OBJECTIFS DU PROJET.....	5
BESOINS IDENTIFIÉS DU CLIENT	5
ATTENTES DES ACTEURS.....	5
PÉRIMÈTRE FONCTIONNEL DEMANDÉ.....	5
PROCESSUS, FLUX ET APPLICATIONS IMPACTÉS	6
NIVEAU DE SERVICE REQUIS.....	6
DONNÉES A REPRENDRE.....	8
CONTRAINTES.....	8
ETUDES DES SOLUTIONS POSSIBLES.....	8
SYNTHÈSE DE L'ANALYSE DE MARCHÉ.....	8
PRÉSENTATION DES DIFFÉRENTS SCÉNARIOS DE SOLUTION POSSIBLES.....	8
SCÉNARIO PRÉCONISÉ.....	9
SOLUTION RETENUE.....	10
PÉRIMÈTRE DE LA SOLUTION.....	10
EVALUATION DES GAINS ET DES COÛTS.....	10
MISE EN ŒUVRE DE LA SOLUTION.....	10
ORGANISATION DU PROJET.....	12
DÉCISION.....	14

Qu'est ce que le dataops ?

Alors que l'approche [DevOps](#) (la contraction de développement et opérations) commence à se démocratiser en entreprise, de nouveaux termes contenant la terminaison « Ops » font leur apparition. Après le « [NoOps](#) » qui consiste à éliminer le personnel d'exploitation IT (les équipes DevOps, donc) par l'automatisation, place au DataOps. Késako ?

Selon une définition du cabinet, le DataOps est « une pratique collaborative de gestion de données consacrée à l'amélioration de la communication, de l'intégration et l'automatisation des flux de données à travers une entreprise ».

Très inspiré des [pratiques Agile](#) DevOps, le DataOps vise à « créer de la valeur plus rapidement et une meilleure gestion des changements de données, des data model et des artefacts associés », précise le Gartner.

« Le DataOps permet d'aligner les processus d'une entreprise et les techniciens, les opérateurs, afin de facilement délivrer des données sécurisées pour ceux qui en ont besoin, des consommateurs », déclare Eric Bischoff, directeur France de Delphix, un éditeur engagé dans ce mouvement.

« Aujourd'hui, d'un côté nous avons les Ops, ceux qui gèrent les [bases de données](#), la sécurité, la confidentialité, les infrastructures. Toutes ces personnes-là contrôlent la donnée. Ce sont eux qui décident de la mettre à un endroit ou un autre. De l'autre côté, il y a ceux qui ont besoin de la donnée : les développeurs, les data analysts, data scientists, les métiers ».

Entre les deux groupes de personnes, il y a des frictions, selon Eric Bischoff. « À chaque fois qu'un consommateur de données souhaite avoir un rafraîchissement d'environnement de données, cela prend du temps, parfois jusqu'à trois semaines chez certains clients », estime-t-il.

Une chaîne d'outillage complexe

Pour les responsables SI, il s'agit donc de mettre en place une chaîne d'outils complète (dite « de bout en bout ») qui permet de gérer la sauvegarde, la sécurisation, le rafraîchissement (ou versioning), la préparation et l'analyse des données afin d'optimiser leur utilisation en libre-service.

Le mouvement DataOps consiste donc en la combinaison de diverses techniques et technologies dont l'exploration ([métadonnées](#), Data classification, [Data Lineage](#)), l'extraction, la transformation et le chargement de données ([ETL](#)), la [virtualisation](#), la mesure de la qualité des données, leur gouvernance, le contrôle des accès ou encore la planification d'espaces de stockage alloués dans un datacenter. Il n'y aurait pas un outil DataOps à proprement parler, mais un ensemble de logiciels et de systèmes à déployer.

Il faut également suivre l'état de santé des systèmes de fichiers, des [SGBD SQL](#) et [NoSQL](#) (MongoDB, [PostgreSQL](#), MySQL, Cassandra, etc.) des environnements de stockage comme les [data lake](#) et [data warehouse](#) (S3, Hive, Hadoop, Snowflake...), des technologies de messaging ([Kafka](#)) et de calcul distribué ([Spark](#)), des containers ([Kubernetes](#), [Docker](#)), des langages ([Python](#), R, Scala, etc.) ou encore des environnements et des outils de développement (Dataiku, AI Platform, [TensorFlow](#), Jupyter, Visual Studio, Git, MLFlow, etc.).

Pour cela, il conviendrait non seulement d'utiliser des [outils de monitoring](#) comme Kibana ou Datadog, de DevOps comme Jenkins, mais également adopter une approche « Lean et Agile ». Celle-ci est définie spécifiquement par le DataOps Manifesto en 18 points. Elle dépend surtout du contrôle statistique des processus (SPC). Cette discipline consiste à la systématisation des tests pour vérifier la consistance des pipelines analytiques et augmenter continuellement la qualité des données.

DevOps vs DataOps : les différences

Selon Saagie, une startup française qui propose un orchestrateur DataOps, cette pratique a des points communs avec l'approche DevOps. L'automatisation ([CI/CD](#)), les tests unitaires, la gestion d'environnements, de version et le monitoring en font partie. Seulement, l'éditeur identifie des particularités liées à la data science.

Les voici : le maintien du flux (ou pipeline) de données, « la reproductibilité des résultats », le « monitoring des performances » d'un modèle statistique ou de machine learning et la « mise à disposition et l'exposition de modèles » pour les utilisateurs finaux.

Selon Bilel Belaid, consultant Big Data et Dr Ahmed Gater, Practice Manager, chez [Experis IT](#), la filiale de recrutement IT du groupe ManPower, le DataOps vient répondre « aux spécificités des projets Data Analytics », qui sont la gestion du code et des données sur l'ensemble de la chaîne analytique, l'automatisation de la création d'un environnement de travail et le rassemblement de plusieurs techniciens et métiers qui utilisent différentes briques technologiques.

RGPD : Se préparer en 6 étapes

Le 25 mai 2018, le règlement européen est entré en application. De nombreuses formalités auprès de la CNIL disparaissent. En contrepartie, la responsabilité des organismes est renforcée. Ils doivent désormais assurer une protection optimale des données à chaque instant et être en mesure de la démontrer en documentant leur conformité.

1 - DÉSIGNER UN PILOTE

Pour piloter la gouvernance des données personnelles de votre structure, vous aurez besoin d'un véritable chef d'orchestre qui exercera une mission d'information, de conseil et de contrôle en interne : le délégué à la protection des données. En attendant 2018, vous pouvez d'ores et déjà désigner un « correspondant informatique et libertés », qui vous donnera un temps d'avance et vous permettra d'organiser les actions à mener.

2 - CARTOGRAPHIER VOS TRAITEMENTS DE DONNÉES PERSONNELLES

Pour mesurer concrètement l'impact du règlement européen sur la protection des données que vous traitez, commencez par recenser de façon précise vos traitements de données personnelles. L'élaboration d'un registre des traitements vous permet de faire le point.

3 - PRIORISER LES ACTIONS À MENER

Sur la base de votre registre, identifiez les actions à mener pour vous conformer aux obligations actuelles et à venir. Priorisez ces actions au regard des risques que font peser vos traitements sur les droits et les libertés des personnes concernées.

4 - GÉRER LES RISQUES

Si vous avez identifié des traitements de données personnelles susceptibles d'engendrer des risques élevés pour les droits et libertés des personnes concernées, vous devrez mener, pour chacun de ces traitements, une analyse d'impact relative à la protection des données (AIPD).

5 - ORGANISER LES PROCESSUS INTERNES

Pour assurer un haut niveau de protection des données personnelles en permanence, mettez en place des procédures internes qui garantissent la prise en compte de la protection des données à tout moment, en prenant en compte l'ensemble des événements qui peuvent survenir au cours de la vie d'un traitement (ex : faille de sécurité, gestion des demandes de rectification ou d'accès, modification des données collectées, changement de prestataire).

6 - DOCUMENTER LA CONFORMITÉ

Pour prouver votre conformité au règlement, vous devez constituer et regrouper la documentation nécessaire. Les actions et documents réalisés à chaque étape doivent être réexaminés et actualisés régulièrement pour assurer une protection des données en continu.

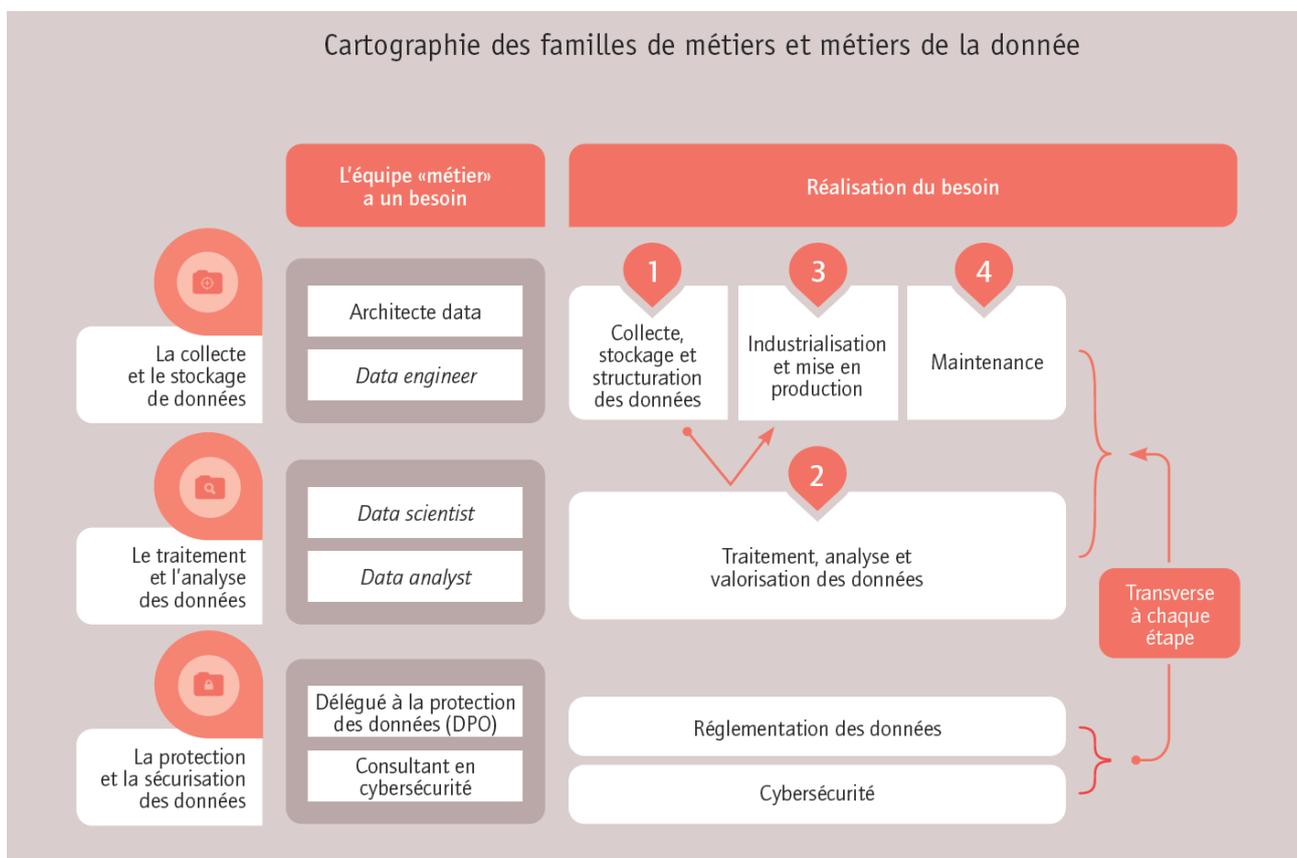
Les métiers de la data

Zoom sur sept métiers qui placent les données au cœur de leur activité

La révolution numérique a fait rentrer nombre d'entreprises dans l'ère du big data

> Le volume croissant des données disponibles, leur variété et la rapidité avec laquelle elles sont créées et échangées, ont fait basculer nombre d'entreprises dans l'ère du big data.

> Le big data joue un rôle de plus en plus stratégique pour les entreprises : l'exploitation des données permettant notamment d'assurer un meilleur suivi de leur performance, d'analyser les comportements des consommateurs, d'identifier des opportunités de marché...



CHIEF DATA OFFICER - DATA MANAGER

INTITULÉS SECONDAIRES : Responsable des données ; Responsable de la stratégie data ; Responsable de la gouvernance des données ; Responsable data management ; Data manager officer ; Data officer

Le-la chief data officer (CDO) ou data manager assure la gouvernance des données et accompagne la transformation digitale de l'entreprise.

ACTIVITÉS PRINCIPALES :

- PILOTER LA STRATÉGIE DE DATA MANAGEMENT
- ASSURER LA QUALITÉ, LA FIABILITÉ ET LA CARTOGRAPHIE DES DONNÉES

ARCHITECTE BIG DATA

INTITULÉS SECONDAIRES : Architecte data ; Architecte de la donnée

L'architecte data préconise les solutions techniques à mettre en œuvre dans les projets nécessitant la collecte, le stockage ou l'utilisation de données en masse (big data). Il·elle intervient ainsi sur de nombreux projets et met en place des solutions génériques pour faciliter le travail des équipes de développeurs. Cela lui permet notamment d'avoir une vision globale des données de l'entreprise. Travaillant dans un écosystème qui évolue rapidement, il·elle assure également la veille technologique autour du big data.

ACTIVITÉS PRINCIPALES :

- COLLECTE, STOCKAGE ET IDENTIFICATION DES DONNÉES
- FACILITATION DE L'USAGE DES DONNÉES DANS L'ENTREPRISE
- CONFORMITÉ, VEILLE ET SENSIBILISATION DES UTILISATEURS AUX PROBLÉMATIQUES DE LA DATA

DATA ENGINEER

INTITULÉS SECONDAIRES : Ingénieur·e data/big ; data Développeur·euse data ; Machine learning engineer Ingénieur·e ; en développement big data

Le·la data engineer est un·e développeur·euse informatique qui a pour mission de mettre en place la collecte et la mise à disposition des données au sein de l'entreprise. Il·elle est également en charge d'industrialiser et mettre en production des traitements sur les données (par exemple : mise à disposition de tableaux de bords, intégration de modèles statistiques) en lien avec les équipes métiers et les équipes qui les analysent.

ACTIVITÉS PRINCIPALES :

- ACHEMINEMENT DE LA DONNÉE
- MISE À DISPOSITION DES DONNÉES AUX ÉQUIPES UTILISATRICES
- MISE EN PRODUCTION DE MODÈLES STATISTIQUES DANS LES APPLICATIONS
- SUIVI DES PROJETS DE DÉVELOPPEMENT

DATA ANALYST

INTITULÉS SECONDAIRES : Analyst dataminer ; Chargé·e d'études en datamining ; Consultant·e dataminer

Le data analyst valorise l'ensemble des données d'une entreprise pour en faire un levier de création de valeur. Il·elle utilise notamment les données recueillies en masse (big data) pour réaliser les nombreux tableaux de bord nécessaires à différents services de l'entreprise (marketing, relations clients, production...). Il·elle est également en charge de construire des modèles statistiques pour éclairer les services opérationnels (segmentations clients ou analyses prédictives).

ACTIVITÉS PRINCIPALES :

- EXTRACTION ET STRUCTURATION DES DONNÉES
- ANALYSE ET EXPLORATION DES DONNÉES
- COMMUNICATION AVEC LES ÉQUIPES CLIENTES
- VEILLE TECHNOLOGIQUE SUR LES OUTILS D'ANALYSE DE DONNÉES

DATA SCIENTIST

INTITULÉS SECONDAIRES : Ingénieur·e data scientist

Le data scientist développe des algorithmes d'apprentissage automatique selon les besoins des équipes métiers. Ses compétences en statistiques lui permettent de construire des modèles de machine learning et ses connaissances en informatique l'aident à anticiper leur mise en production. En amont de ces deux missions, il·elle est également en charge de structurer et analyser les données qu'il utilise.

ACTIVITÉS PRINCIPALES :

- EXTRACTION ET STRUCTURATION DES DONNÉES
- ÉLABORATION DES ALGORITHMES D'INTELLIGENCE ARTIFICIELLE
- INDUSTRIALISATION DES MODÈLES D'INTELLIGENCE ARTIFICIELLE DANS LES APPLICATIONS
- PARTICIPATION ACTIVE AUX PROJETS
- VEILLE TECHNOLOGIQUE SUR LES OUTILS DE DATA SCIENCE

DÉLÉGUÉ·E À LA PROTECTION DES DONNÉES (DPD/DPO)

INTITULÉS SECONDAIRES : Data protection officer (DPO) ; Juriste DPO ; Consultant·e DPO ; RSSI DPO

Entre expertise informatique, juridique, qualité et conformité, le délégué à la protection des données (DPD) ou data protection officer (DPO) accompagne et conseille les structures afin d'assurer leur conformité en matière de traitement de données personnelles. Exerçant ses missions dans tous les types d'organisme (entreprises, associations, administrations et collectivités territoriales), il·elle a également pour rôle d'assurer le contact avec la Cnil et les différentes personnes concernées (salariés, usagers, patients, fournisseurs, etc.).

ACTIVITÉS PRINCIPALES :

- PILOTAGE DE LA CONFORMITÉ DES TRAITEMENTS DE DONNÉES MIS EN OEUVRE PAR L'ORGANISME QUI LE DÉSIGNE (que cet organisme soit responsable de traitement ou sous-traitant)
- MISSION DE VEILLE
- INFORMATION ET CONSEIL
- CONTRÔLE DU RESPECT DE LA RÉGLEMENTATION

CONSULTANT·E EN CYBERSECURITÉ

INTITULÉS SECONDAIRES : Auditeur·rice informatique ; Consultant·e en cyberdéfense ; Consultant·e en sécurité des systèmes d'information ; Consultant·e en sécurité IAM ; Consultant·e e-sécurité ; Consultant·e sécurité en informatique ; Consultant·e SSI

Le·la consultant·e en cybersécurité conseille et accompagne le client sur sa problématique de sécurisation des systèmes d'information. Il·elle analyse les besoins, les risques, participe à la définition de la politique sécurité. Il·elle rédige les propositions techniques et commerciales.

ACTIVITÉS PRINCIPALES :

- AVANT-VENTE ET SUPPORT COMMERCIAL
- PILOTAGE DE PROJETS
- AUDIT SÉCURITÉ ET GESTION DES RISQUES
- FORMATION AUX ENJEUX DE SÉCURITÉ

Les 6 étapes incontournables pour bien gérer un projet !

ÉTAPE 0 – Définir les objectifs du projet

L'étape 0 paraît un "basique", mais elle est trop souvent négligée (c'est pour cela que je l'ai nommée étape 0, tellement elle paraît évidente... et pourtant elle est trop souvent négligée !).

Ne pas définir les objectifs du projet elle peut induire un "effet tunnel" (c'est-à-dire se lancer et avancer jusqu'à ce que l'on se rende compte qu'on a mal compris, mais souvent tardivement), ou encore partir sur de mauvaises bases ce qui obligera plus tard à devoir refaire le travail.

Il est essentiel de clarifier les objectifs du projet (la priorité N°1, et les objectifs secondaires), **ce que l'on attend de vous (livrables, temps, délais, budgets...) et la souplesse que vous pouvez avoir.**

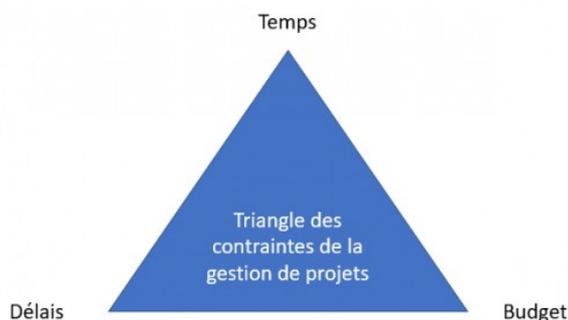
Il est indispensable d'écrire cela dans un document, afin de ne pas laisser de place aux malentendus, et surtout le faire valider par le donneur d'ordres.[...]

Concernant les contraintes classiques d'un projet à prendre en compte lors de l'étape 0, vous avez généralement :

- Le temps pour le réaliser (1 semaine, 1 mois...)
- L'argent pour le faire (budget "no limit" car prioritaire ou au contraire un bricolage vite fait)
- Les fonctions que vous souhaitez y mettre ou les résultats qui sont attendus (quelle est la "killer feature" c'est-à-dire ce qu'il est indispensable de livrer)
- L'importance pour la direction générale ou la stratégie de l'entreprise
- ...

Sachant que ce qui est souvent décisif, c'est les facteurs temps / argent / fonctions.

Ces 3 contraintes forment un triangle, que vous pouvez changer pour faire en sorte de respecter l'une ou l'autre de ces contraintes. Un bon moyen de clarifier le résultat obtenu, et donc de laisser le décideur choisir lui-même afin de ne pas vous retrouver en mauvaise posture...



A noter : si votre commanditaire ne vous propose pas de solutions (ex: plus de moyens, plus de temps...), alors demandez-lui "Mais comment faire alors ?". Cela vous permet de lui renvoyer le problème à la figure au lieu de devoir vous justifier et trouver vous-même des solutions impossibles à tenir.

ÉTAPE 1. Découper le projet et définir la liste des tâches : on ne mange une baguette qu'en la coupant en tranches ! [...]

La première étape à réaliser est celle de dresser la liste des tâches.

C'est cette étape qui est la plus délicate, car elle jette les fondations du projet.

Si l'on oublie une tâche à ce moment-là, le projet sera faussé, et sa gestion le sera donc aussi.

Listez toutes les tâches possibles et imaginables, et soyez très critique vis-à-vis de cette liste : pour chaque tâche, demandez-vous toujours *De quoi ai-je besoin au préalable pour pouvoir faire cela ? Qu'est-ce que cela implique pour le reste du projet ?*

En procédant de cette manière, vous trouverez par déduction et par induction de nouvelles tâches que vous auriez pu oublier, des pré-requis techniques...

[...] vous n'avez pas forcément à lister les tâches par ordre parfaitement chronologique, c'est juste plus pratique pour la suite.

ÉTAPE 2. Définir l'enchaînement logique des tâches [...]

Tout ne doit pas se faire en même temps, et il est essentiel de commencer des chantiers immédiatement.

Par exemple, vous pouvez organiser une réunion de lancement de projets avec les parties prenantes bien avant d'avoir tous les éléments en main.

En anticipant cela, vous allez déjà avancer sur le projet (et ne pas subir des indisponibilités), mais en plus vous allez vous fixer une première date pour avancer sur le projet.[...]

Maintenant que vous avez établi la liste des tâches, il faut les “relier” entre elles, les ordonner.

C'est-à-dire qu'il faut définir, pour chaque tâche, toutes les tâches *prédécesseur*, c'est-à-dire toutes les tâches qu'il faut réaliser en amont, avant de pouvoir réaliser une tâche donnée : ça s'appelle l'ordonnancement. [...]

ÉTAPE 3- Ajouter les durées et des contraintes sur certaines tâches et intégrer des tâches externes



Tout le monde se réserve toujours des marges de sécurité pour réaliser des actions.

Les actions du chef de projet sont donc de :

- Bien formaliser les demandes aux parties prenantes
- Faire une relance avant la date de livraison des livrables (et si nécessaire recalculer une date)

- Trouver des leviers (relationnels, hiérarchique, budgétaires, organisationnels...) pour faire avancer les dossiers
- Caler des dates (réunions, remises de documents...) acceptées par les participants (en leur demandant leurs avis, et si nécessaire de voir ce qui est livré ou compris pour réduire les délais ou les coûts).

[...] Certaines tâches ont des contraintes de dates. C'est le moment de les rentrer. Par exemple [...] :

- début au plus tôt
- début au plus tard
- doit commencer le
- doit finir le
- fin au plus tôt
- fin au plus tard

3.3. Tâches externes

Votre projet dépend également de tâches externes.

Dans ce cas-là, vous pouvez soit choisir de les rentrer directement en tant que tâche standard, soit les rentrer uniquement sous forme de "jalons", c'est-à-dire une tâche de durée zéro, un simple point à valider avant de pouvoir passer à la suite.[...]

ÉTAPE 4- Définir et attribuer les ressources



Sur un projet, vous êtes la ressource principale pour organiser, et souvent faire les choses.

Attention cependant à ne pas faire à la place des autres : non seulement vous risquez de faire moins vite et moins bien, mais en plus en cas de problème vous pourriez être tenu comme responsable...

Faites donc bien attention à apprendre à déléguer, et à être assertif (c'est-à-dire demander sans imposer).

4.1. Dresser la liste des ressources

Ici, le problème est simple : vous devez purement et simplement dresser la liste des ressources. N'oubliez pas le matériel : les ressources n'impliquent pas seulement l'humain.

4.2. Définir les contraintes pour les ressources

Chaque ressource aura des contraintes de disponibilité ou de charge maximale qui lui seront propres.

N'oubliez donc pas de définir leur nombre d'heures disponibles par jour, leurs jours de repos (et à l'inverse leurs jours de travail si elles sont disponibles pendant les week-ends) ainsi que la part de leur temps qu'elles pourront consacrer au maximum sur une tâche.[...]

4.3. Affecter les ressources

Une fois que vous avez votre liste des ressources, vous pouvez les affecter à vos tâches. N'oubliez pas de préciser à ce moment-là le Reste-à-Faire. Le Reste-à-Faire dépend des ressources ET des tâches, ce n'est pas une simple charge de travail par tâche divisible entre les ressources.

ÉTAPE 5- Planifier et assurer le suivi du projet



5.1. Planifier

A présent, il ne vous reste plus qu'à planifier le projet. Cela va vous permettre ensuite de comparer le prévisionnel et le réel, pour voir où vous avez pêché, et ainsi vous améliorer de manière continue.

5.2. Assurer le suivi

[...] Le point principal est de faire vivre votre planning et de le réviser au jour le jour : de cette manière, vous comprendrez en direct ce qu'un changement au niveau d'une tâche implique quant à la date de fin du projet, et vous pourrez prendre les mesures de rattrapage nécessaires.

J'insiste : le planning doit vivre au rythme du projet, sinon il ne sert plus à rien.

Idéalement, vous devez aussi avoir un espace partagé [...] qui centralise l'information pour tous.

Cela vous permet par exemple de dire “Oui, tu as l'information sur l'espace projet...” au lieu de perdre des heures à montrer à tout le monde où se trouve l'information...[...]

Une petite astuce est de passer au bureau des personnes juste pour dire bonjour, faire un coucou... cela permet de physiquement ne pas oublier que les participants doivent vous rendre des éléments !

5.3. Clôturer le projet

C'est une phase souvent négligée, c'est d'indiquer la fin du projet, avec une fête de fin de projet.

Souvent la fin de projet est souvent liée à la mise en place de celui-ci, mais il est conseillé de faire 2 événements séparés :

- Un lancement officiel du projet qui sera le point de départ pour les utilisateurs
- Une clôture de projet qui sera une fête pour les participants au projet

ÉTAPE 6 : Faire un retour d'expérience post projet (REX)

Le Retour d'EXpérience (ou REX) permet de faire une analyse “**post mortem**” de son projet avec ce qui a fonctionné, ce qui n'a pas marché...

Le but est de réunir les principaux acteurs du projet plusieurs semaines après la fin du projet, afin d'avoir un retour “à froid”, c'est-à-dire une fois le projet mis en place.

Le REX permet de ne plus refaire les mêmes erreurs pour les prochains projets, et souvent de récompenser les participants au projet.

L'anonymisation des données personnelles

L'anonymisation rend impossible l'identification d'une personne à partir d'un jeu de données et permet, ainsi, de respecter sa vie privée. La CNIL fait le point sur les techniques utilisables et sur leurs enjeux.

Qu'est-ce que l'anonymisation ?

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible.

L'anonymisation ne doit pas être confondue avec la pseudonymisation.

La pseudonymisation est un traitement de données personnelles réalisé de manière à ce qu'on ne puisse plus attribuer les données relatives à une personne physique sans information supplémentaire.

En pratique, la pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.).

La pseudonymisation permet ainsi de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe. En pratique, il est toutefois bien souvent possible de retrouver l'identité de ceux-ci grâce à des données tierces : les données concernées conservent donc un caractère personnel. L'opération de pseudonymisation est également réversible, contrairement à l'anonymisation.

La pseudonymisation constitue une des mesures recommandées par le RGPD pour limiter les risques liés au traitement de données personnelles.

Pourquoi anonymiser des données personnelles ?

Le règlement général sur la protection des données (RGPD) ne comporte pas d'obligation générale d'anonymisation. Il s'agit d'une solution, parmi d'autres, pour pouvoir exploiter des données personnelles dans le respect des droits et libertés des personnes.

En effet, l'anonymisation ouvre des potentiels de réutilisation des données initialement interdits du fait du caractère personnel des données exploitées, et permet ainsi aux acteurs d'exploiter et de partager leur « gisement » de données sans porter atteinte à la vie privée des personnes. Elle permet également de conserver des données au-delà de leur durée de conservation.

Dans ce cas, la législation relative à la protection des données ne s'applique plus, car la diffusion ou la réutilisation des données anonymisées n'a pas d'impact sur la vie privée des personnes concernées.

Comment anonymiser tout en préservant au maximum l'utilité du jeu de données ?

Puisque le processus d'anonymisation vise à éliminer toute possibilité de ré-identification, l'exploitation future des données est ainsi limitée à certains types d'utilisation. Ces contraintes sont à prendre en compte dès le début du projet.

Pour construire un processus d'anonymisation pertinent, il est ainsi conseillé :

- d'identifier les informations à conserver selon leur pertinence.
- de supprimer les éléments d'identification directe ainsi que les valeurs rares qui pourraient permettre un ré-identification aisée des personnes (par exemple, la présence de l'âge des individus peut permettre de ré-identifier très facilement les personnes centenaires) ;
- de distinguer les informations importantes des informations secondaires ou inutiles (c'est-à-dire supprimables) ;
- de définir la finesse idéale et acceptable pour chaque information conservée.

Ce pré-requis permet de déterminer le procédé d'anonymisation à appliquer, c'est-à-dire l'enchaînement des techniques d'anonymisation à mettre en place. Celles-ci peuvent être regroupées en deux familles : la randomisation et la généralisation.

- La randomisation consiste à modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale. Cette technique permet de protéger le jeu de données du risque d'inférence (cf. ci-dessous).

Exemple : il est possible de permuter les données relatives à la date de naissance des individus de manière à altérer la véracité des informations contenues dans une base de données.

- La généralisation consiste à modifier l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes. Cette technique permet d'éviter l'individualisation d'un jeu de données. Elle limite également les possibles corrélations du jeu de données avec d'autres (cf. ci-dessous).

Exemple : dans un fichier contenant la date de naissance des personnes, il est possible de remplacer cette information par la seule année de naissance.

Comment vérifier l'efficacité de l'anonymisation ?

Les autorités de protection des données européennes définissent trois critères qui permettent de s'assurer qu'un jeu de données est véritablement anonyme :

- 1) l'individualisation : il ne doit pas être possible d'isoler un individu dans le jeu de données ;

Exemple : une base de données de CV où seuls les noms et prénoms d'une personne auront été remplacés par un numéro (qui ne correspond qu'à elle) permet d'individualiser cette personne. Dans ce cas, cette base de données est considérée comme pseudonymisée et non comme anonymisée.

2) la corrélation : il ne doit pas être possible de relier entre eux des ensembles de données distincts concernant un même individu ;

Exemple : une base de données cartographique renseignant les adresses de domiciles de particuliers ne peut être considérée comme anonyme si d'autres bases de données, existantes par ailleurs, contiennent ces mêmes adresses avec d'autres données permettant d'identifier les individus.

3) l'inférence : il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

Exemple : si un jeu de données supposément anonyme contient des informations sur le montant des impôts de personnes ayant répondu à un questionnaire, que tous les hommes ayant entre 20 et 25 ans qui ont répondu sont non imposables, il sera possible de déduire, si on sait que M. X, homme âgé de 24 ans, a répondu au questionnaire, que ce dernier est non imposable.

Comment se prémunir des risques liés à l'anonymisation ?

À défaut de remplir parfaitement ces trois critères, le responsable de traitement qui souhaite anonymiser un jeu de données doit démontrer, via une évaluation approfondie des risques d'identification, que le risque de ré-identification avec des moyens raisonnables est nul.

Les techniques d'anonymisation et de ré-identification étant amenées à évoluer régulièrement, il est indispensable pour tout responsable de traitement concerné, d'effectuer une veille régulière pour préserver, dans le temps, le caractère anonyme des données produites. Cette veille doit prendre en compte les moyens techniques disponibles ainsi que les autres sources de données qui peuvent permettre de lever l'anonymat des informations.

Si un jeu de données publié en ligne comme « anonyme » contient en réalité des données personnelles et qu'aucune des exceptions mentionnées à l'article L.312-1-2 du Code des relations entre le public et l'administration (CRPA) n'est applicable, cela peut être considéré comme une violation de données. Il est alors nécessaire de :

- procéder au retrait du jeu de données en question dans les plus brefs délais ;
- en informer la CNIL si cette violation est susceptible d'engendrer un risque pour les droits et libertés des personnes ;
- d'en informer les personnes concernées si ce risque est élevé.