

Les données sont aujourd'hui en majorité associées au logiciel qui les gère, ce qui rend leur extraction délicate. Lorsque la question se pose sur 50, 100, voire 150 applications métiers différentes, les problématiques de réutilisation, d'échange de données ou d'interopérabilité deviennent compliquées voire même impossibles à traiter pour des raisons techniques et budgétaires. Or ce cloisonnement est de plus en plus pénalisant dans les environnements fortement numérisés. L'application "Bordeaux ma ville en poche" qui offre à tout habitant la capacité d'acheter différentes prestations municipales - tickets piscine, accès à la bibliothèque, règlement du stationnement – et de signaler les dégradations constatées sur la voie publique est connectée à six applications métiers différentes ce qui dans une architecture traditionnelle a provoqué une forte dégradation des performances du système de requête. Pour y remédier un serveur relais faisant office de cache a été installé, pour permettre à l'utilisateur de retrouver un confort d'utilisation. A terme, dans une architecture de gouvernance des données, la solution consistera à verser les données actives des applications dans un entrepôt de données. L'application mobile ne communiquera plus directement avec les applications métiers mais uniquement avec l'entrepôt, via des webservices ou des Interfaces de programmation spécifiques. Le système sera plus compact, plus performant et capable d'effectuer lui-même automatiquement les mises à jour sur le compte client de l'utilisateur.

L'exemple bordelais n'est pas fortuit et les besoins de faire autrement se multiplient, ce qui a conduit la plupart des métropoles à s'engager dans des chantiers de transformation. Concrètement cela consiste d'abord à collecter des données existantes sur un territoire en provenance des différentes sources : des systèmes d'information métiers de la collectivité ; des objets connectés (caméras, horodateurs pour les parkings, smartphones et autres mobiles, capteurs environnementaux, systèmes de pilotage des feux tricolores aux intersections...) ; des partenaires, fournisseurs, sociétés titulaires d'une délégation de service public (ERDF, opérateurs de transports...).

Puis cela se poursuit avec la construction brique par brique de plateformes intégrant divers composants : des entrepôts de données plus ou moins spécialisés (données alphanumériques, SIG, big data, Internet des objets, open data) associés à des logiciels ayant pour vocation de filtrer, convertir et extraire certaines données, sans oublier en bout de chaîne les outils de reporting (superviseur et hyperviseur) et d'analyse utilisés pour exploiter les données recueillies (Business intelligence, big data).

La plateforme a ainsi vocation à stocker toutes les données métiers et d'une manière générale les données utiles à de fins de réutilisation mais elle peut aussi se concevoir en réseau, en connexion avec d'autres systèmes. L'entrepôt est plus qu'un réservoir. Pour rendre les données disponibles à l'ensemble des agents de la collectivité et ouvrir certaines vers l'extérieur, celles-ci font l'objet de conversions dans des formats homogènes, elles sont indexées et accessibles via des services de consultation, de visualisation et le cas échéant d'enrichissement avec l'apport d'informations complémentaires.

Le déploiement de telles infrastructures est amorcé, bien qu'il n'existe pas encore à proprement parler de plateforme généraliste traitant des principaux métiers de la ville. La diversité des métiers et des sources d'information à traiter, fait de ce chantier une construction progressive et personnalisée. Chaque projet étant déterminé au démarrage par le contexte local et les priorités fixées par la collectivité.

La stratégie smart data de la métropole de Bordeaux pour valoriser le patrimoine de données cible les domaines métiers, l'open data, les objets connectés et l'échange interactif de données avec les partenaires.

Nice s'appuie sur un entrepôt de données intégrant plusieurs composants dont une plateforme open data et une infrastructure dédiée à l'Internet des objets, et semble progresser sur les données environnementales (pollution et risques naturels) ; ☒ Lyon développe une approche généraliste à partir d'une plateforme ouverte et quelques spécialités axées sur les données de transports et de l'énergie débouchant sur des applications big data ; ☒ Bordeaux ne dispose pas encore de plateforme, mais sa feuille de route prévoit l'ouverture d'un nouveau portail open data en 2016 et d'un entrepôt de données généraliste en 2017 ; ☒ Rennes, collectivité pionnière de l'open data, développe ses projets sur les volets citoyens et décisionnels et axe sa démarche de gouvernance dans la perspective de création d'une régie locale de la donnée ; ☒ Paris, qui dispose déjà d'une plateforme de données ouvertes étendue (notamment sur le partage de toutes les données de mobilité), vient d'achever l'unification de ses systèmes d'information géographiques, et mise maintenant sur déploiement des réseaux intelligents et la création d'un entrepôt basé sur l'Internet des objets et le développement en interne des data sciences tout en se disant en retard sur la constitution d'une plateforme territoriale de données partagée ; ☒ Chartres met en place un centre de supervision urbain (CSU) de nouvelle génération, dans une architecture de gouvernance de la donnée.

➤ Des solutions performantes, ouvertes et qui se diversifient

Les applications de gouvernance des données sont de plus en plus variées et surtout elles couvrent le plus souvent des champs nouveaux, inédits, voire expérimentaux, mais dans des domaines qui restent pour l'instant assez concentrés sur quatre axes :

- Les transports et l'énergie sont des points de focalisation forts ayant déjà fait l'objet d'études collectives²⁴ (Voir encadré ci-dessous). Côté transports et mobilité, plusieurs projets d'optimisation des déplacements en cours (Optimod et Opticities à Lyon ; Gertrude à Bordeaux) reposent sur le développement d'applications d'analyse prédictive du trafic et ne pourraient atteindre le même niveau de performance sans plateforme d'agrégation des données. Côté énergie, les expérimentations de smart grids de quartier ou de gestion intelligente des bâtiments et d'analyse de leurs données de consommation énergétique, via des capteurs, sont notamment en cours dans des villes comme Nice, Rennes et Paris.
- Les réseaux urbains - eau, gaz, électricité, ramassage des ordures, qualité de l'air, circulation, stationnement - deviennent plus intelligents et se placent au premier rang des domaines impactés par le déploiement de l'Internet des objets. Nice expérimente un système d'hypervision multi-métiers permettant le pilotage de grandes fonctions environnementales, dans le cadre d'une expérience de monitoring urbain. De son côté Chartres met en place une centrale de surveillance associant vidéo surveillance, pilotage des carrefours et remontée d'incidents et d'informations sur le trafic.
- Les applications mobiles de e-services sont promises à des développements significatifs dans plusieurs directions : paiement et accès à des services locaux (piscine, stationnement, médiathèque) à Bordeaux ou encore services de vie quotidienne à travers la fourniture d'informations en temps réel sur une gamme étendue de domaines (horaires dynamiques et

personnalisés à l'arrêt de bus, niveau de fréquentation des équipements, événementiels) etc... La faisabilité de ces applications qui mobilisent un nombre croissant de systèmes d'information métiers repose essentiellement sur le développement de plateformes internes regroupant toutes les données y compris les données de flux.

- La dématérialisation administrative avec l'application progressive aux collectivités locales du principe qu'une administration ne peut demander à l'utilisateur une information dont elle dispose déjà (programme "dites-le nous une fois"), va impacter les systèmes d'information locaux. Le lancement de FranceConnect et de quelques applications illustratives comme "marché public simplifié" et "demandes de subventions simplifiées" donne un signal fort sur les évolutions en cours. Autant de services qui commencent à être réutilisés aujourd'hui par les collectivités territoriales (Alpes-Maritimes, Hautes-Alpes...), et qui reposeront de plus en plus sur des logiques de gouvernance de la donnée.

➤ Les datasciences au service de la ville et des territoires

Le développement des data sciences au service de la smart city, de la ville et des territoires en général est aussi expérimental que prometteur. Les premières illustrations relevées confirment le potentiel innovateur de cette nouvelle discipline dans la gestion des collectivités (consommation énergétique des bâtiments), dans la régulation (pilotage de la gestion des feux tricolores à la sortie d'un grand événement populaire) ou encore en matière prédiction ("ce que seront les conditions de circulation dans une heure" afin d'anticiper sur les comportements). Ces fonctions font largement appel au big data et à l'algorithmique et commencent à être déclinées sur certains segments d'activité dans les villes. Les exemples relevés au cours des entretiens apportent quelques illustrations concrètes qui constituent des indicateurs de tendance.

Paris, Lyon, et Nice font figure de postes avancés. Le projet Datacity lancé par Paris il y a quelques mois avec le NUMA, proposait une dizaine de défis urbains, sur la mobilité, l'environnement, la gestion des déchets, l'énergie, les bâtiments et l'espace public afin de déterminer comment résoudre des problèmes complexes par l'usage de la data. Il conduit à la sélection de cinq projets accompagnés et financés par la ville : qui traitent aussi bien du repérage des logements insalubres dans l'habitat diffus, à partir de modèles d'identification, que de la détection des locations temporaires illégales en accompagnement du contrôle que la ville met en place. La ville de Lyon, comme indiqué précédemment, a beaucoup investi sur les déplacements et la prévision de trafic. Avec Optimod'Lyon, la métropole collecte, centralise, et traite l'ensemble des données de la mobilité urbaine sur une plateforme unique dans le but de créer des services innovants qui faciliteront les déplacements et la vie des usagers. Par ailleurs, elle a créé avec les grands opérateurs de services urbains, le Tuba, un laboratoire collaboratif. Ouvert aux porteurs de projets et aux entreprises, il sert de lieu d'expérimentation, d'échange, et de partage de connaissances, notamment sur l'ouverture des données et le big data. Il réunit des startups et des chercheurs du monde académique et peut mobiliser la plateforme métropolitaine pour des expérimentations. La ville est par ailleurs engagée dans plusieurs programmes européens tels que BloTop25 sur les objets connectés ou Smarter Together26 sur la maîtrise de l'énergie. Pour élargir son champ d'action la métropole vient de lancer une nouvelle phase exploratoire en chargeant un stagiaire data scientist d'évaluer l'impact

organisationnel du big data en tant qu'infrastructure. L'enjeu est l'extension du domaine d'expérimentation aux données non structurées et aux données temps réel. De son côté la Métropole Nice Côte d'Azur (MNCA) mène de front plusieurs programmes dont un sur la gestion des risques. Une des applications les plus immédiates est celle de la surveillance de la montée des eaux sur le réseau hydrographique local avec la mise en place de caméras numériques associées à un algorithme permettant d'anticiper les risques éventuels (à deux heures). Par ailleurs, les démonstrateurs de monitoring urbain (déchets, pollution, trafic), ou encore les projets réalisés dans le cadre du programme RéFLeXE27 dédié à la sécurité du réseau électrique, via la gestion de l'effacement et de Cityopt28 programme dédié à l'optimisation énergétique dans les foyers (suivi temps réel des données de consommation via le compteur Linky) constituent d'autres illustrations des chantiers en cours.

On pourrait encore citer bien d'autres manières d'utiliser les data sciences dans le domaine de l'analyse prédictive (risque incendie dans les immeubles, insécurité, vols) ou de l'analyse comportementale. Sur ce point, près d'une cinquantaine de collectivités et d'offices de tourisme ont déjà contractualisé avec la filiale d'Orange qui propose Flux Vision pour produire de l'analyse comportementale des touristes sur un territoire défini. D'autres – et parfois les mêmes – commencent à exploiter leur propres données réseaux (Wifi public et vidéo surveillance) pour effectuer ce type d'analyse, ce qui leur donne plus d'autonomie dans le choix des traitements à effectuer. Les villes de Paris, Strasbourg, Aix en Provence, Chartres, Bordeaux portent actuellement des projets de cette nature.

L'autre versant décisionnel s'appuie sur des technologies plus éprouvées de "Business intelligence" mais relativement peu répandues dans les collectivités territoriales en dehors du champ de la gestion interne. Sur le volet applicatif elles sont en voie de renouvellement en offrant une ergonomie plus intuitive. Quelques collectivités les utilisent notamment pour la réalisation de tableaux de bord de pilotage de l'activité de leurs propres directions ou pour diffuser de l'information à des tiers. Rennes Métropole met par exemple en place un tableau de bord à destination des citoyens présentant des chiffres sur six domaines de l'action publique (mobilité et transports, pilotage de l'action publique, Environnement et énergie, développement économique, aménagement et habitat, culture). Sur ce domaine de l'information le Casa (Centre for advanced spatial analysis de l'Université de Londres) publie un tableau de bord dynamique à l'adresse de chaque grande agglomération britannique. Il est composé d'indicateurs environnementaux et d'indicateurs se rapportant à l'activité urbaine (trafic, incidents éventuels sur les lignes de métro, nombre de bus en service et de rames de métro en circulation...) et pourrait facilement être transposé et enrichi par les villes elles-mêmes.

Big data et aménagement urbain

Voici quelques illustrations complémentaires des domaines dans lesquels l'algorithmique appliquée à l'action publique peut rendre de réels services :

- Constitution d'ilots mixtes (bureaux et logements) susceptibles de mutualiser leur consommation énergétique afin de mieux équilibrer les variations en journée (Paris) ;

- Analyse des flux de circulation (véhicules et piétons) à partir de capteurs vidéo pour monitorer l'usage de l'espace public et réinventer les places en y installant du mobilier urbain déplaçable. L'analyse des flux permet de mieux comprendre les modes d'appropriation des lieux réaménagés dans le cadre d'un urbanisme tactique (Paris) ;
- Analyse de la fréquentation et des mouvements de foule à partir de capteurs de mouvement posés sur les vitrines des magasins (Chartres) ;
- Exploitation des données de Wifi public (réseau de bornes densifié) pour mieux comprendre l'utilisation de l'espace public par les habitants et les touristes à travers la réalisation de cartographies des flux (Strasbourg).

L'autre versant décisionnel s'appuie sur des technologies plus éprouvées de "Business intelligence" mais relativement peu répandues dans les collectivités territoriales en dehors du champ de la gestion interne. Sur le volet applicatif elles sont en voie de renouvellement en offrant une ergonomie plus intuitive. Quelques collectivités les utilisent notamment pour la réalisation de tableaux de bord de pilotage de l'activité de leurs propres directions ou pour diffuser de l'information à des tiers. Rennes Métropole met par exemple en place un tableau de bord à destination des citoyens présentant des chiffres sur six domaines de l'action publique (mobilité et transports, pilotage de l'action publique, Environnement et énergie, développement économique, aménagement et habitat, culture). Sur ce domaine de l'information le Casa (Centre for advanced spatial analysis de l'Université de Londres) publie un tableau de bord dynamique à l'adresse de chaque grande agglomération britannique. Il est composé d'indicateurs environnementaux et d'indicateurs se rapportant à l'activité urbaine (trafic, incidents éventuels sur les lignes de métro, nombre de bus en service et de rames de métro en circulation...) et pourrait facilement être transposé et enrichi par les villes elles-mêmes.

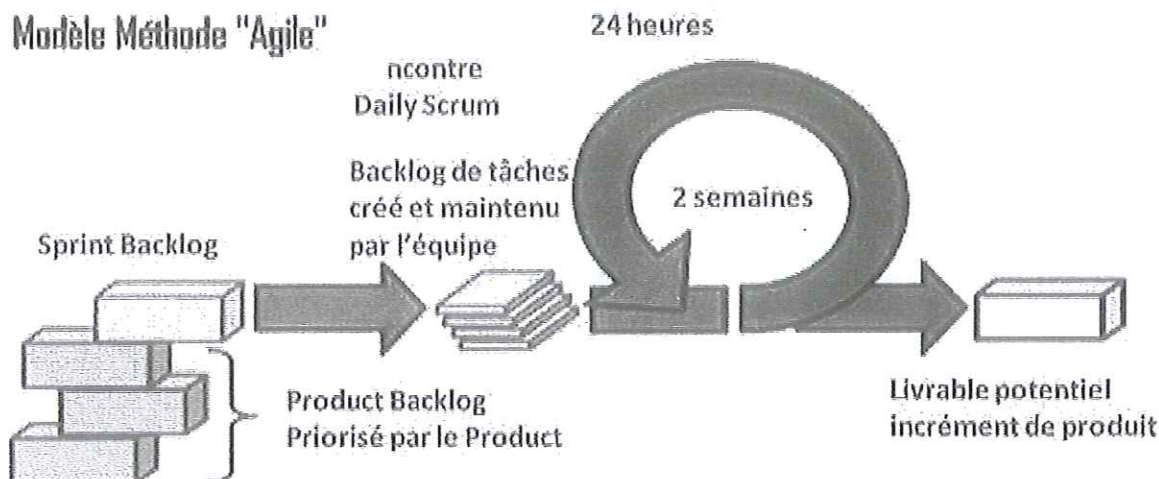
LA METHODE AGILE

La méthode Agile - Optimisation de la relation "client / fournisseur"
JDN - Chronique de Gael Cadoret - le 24/02/14

24

Placer le client au centre des démarches et des personnes. C'est l'objectif des méthodes de développement dites "Agiles". De quoi s'agit-il ? Et surtout quels en sont les avantages et y a-t-il des inconvénients ou des risques d'échecs ?

La méthode Agile est aujourd'hui très répandue dans les sociétés de services ou les agences web. J'ai souvent entendu qu'un des avantages de cette méthode était qu'on pouvait prendre ce qu'on voulait dedans, mais contrairement aux idées reçues, cette méthode ne portera ses fruits que si elle est respectée à la lettre. En effet, chaque étape est importante. Nous verrons ainsi les nombreux avantages ainsi que les inconvénients ou risques éventuels à utiliser cette méthode.



Principes de fonctionnement des méthodes « Agile »

Le principe de base des méthodes « Agile » est qu'il est contre-productif qu'avant de développer un produit, il faille le planifier et en spécifier les moindres détails. En effet, prévoir tous les aspects de la production entraîne dans la plupart des cas frustrations et pertes de temps car les aléas surviennent fréquemment. C'est cette approche prédictive et séquentielle de type waterfall ou cycle en V que les tenants des méthodes « Agile » veulent casser. Ainsi, au lieu de fixer les objectifs lointains, le mieux serait de procéder par étapes c'est-à-dire fixer des objectifs à court terme et commencer le développement sans perdre de temps. Chaque fois que cet objectif est atteint, on passe au prochain et ainsi de suite jusqu'à atteindre le but ultime. Au niveau d'un développement de logiciel, c'est le client qui transmet à l'équipe de développeurs sa vision du produit avec la liste des fonctionnalités qu'aurait ce produit. Il communique ainsi directement avec l'équipe et ensemble ils estiment le coût de chaque fonctionnalité pour aboutir à une idée approximative du budget final. De ce fait, on raisonne plus « produit » que « projet », d'où l'utilisation du terme « gestion de produit » au lieu de « gestion de projet ».

L'Agile manifesto

Ces méthodes « Agile » se sont beaucoup développées et on a pu en recenser une dizaine de variantes jusqu'au début des années 2000. Conscientes qu'une certaine uniformisation était devenue nécessaire, 17 grandes figures du développement logiciel s'étaient réunies aux Etats-Unis en 2001 pour déterminer les critères communs et les principes fondamentaux de ce qui allait devenir un « Manifeste Agile ». D'après ce manifesto, les méthodes « Agile » demandent une plus grande implication du client et permettent une meilleure réactivité des développeurs face à ses demandes. Ce manifesto prône 4 valeurs fondamentales inhérentes aux méthodes « Agile » : l'équipe, l'application, la collaboration et l'acceptation du changement. De ces valeurs découlent 12 principes généraux qui se basent essentiellement sur une relation privilégiée entre le client et les développeurs. La priorité est ainsi donnée à la satisfaction du client et pour y arriver, une implication totale de l'équipe est requise. Cette équipe doit être capable de réagir très vite face aux éventuels changements requis par le client ou aux modifications de la méthode de travail face à des obstacles imprévus. Elle doit également être capable de se remettre en cause et de rechercher perpétuellement à évoluer.

La conception d'un produit « Agile »

La première étape consiste à effectuer une première planification de l'itération ou Sprint dans le jargon des développeurs. Cette réunion fera ressortir les éléments prioritaires de la liste des exigences fonctionnelles du produit. Chaque exigence représente une user storie (US) ou "histoire utilisateur". Une US doit être rédigée dans un bon français et décrit le fonctionnement ainsi que le parcours de l'utilisateur de la fonctionnalité. Elle doit également contenir les cas de tests ainsi que les critères de validation de la fonctionnalité développée.

En accord avec le client, aussi appelé Product Owner, les premières livraisons devraient être effectuées à la fin de cette itération (une itération à une durée d'environ 2 à 3 semaines suivant le nombre d'US présent dans le backlog). Le backlog est l'ensemble des US à développer durant l'itération en cours.

Une autre réunion appelée Revue de Sprint est organisée à la fin de chaque Sprint durant laquelle les développeurs présentent au client les fonctionnalités développées. Ce dernier pourra ainsi tout de suite donner son feedback, ce qui présente l'avantage de gagner beaucoup de temps et d'ajuster les fonctionnalités ou les méthodes de travail le cas échéant. Vient ensuite une rétrospective de Sprint qui permet à tous les acteurs d'améliorer des choses et de s'améliorer également. Une autre particularité de la méthode « Agile » est la réalisation de mêlées quotidiennes qui permettent à l'équipe de développeurs de synchroniser leur travail. Appelée aussi Stand Up meeting, cette réunion qui ne dure pas plus de 15 minutes permet à chacun de déterminer ce qu'ils ont réalisé depuis la dernière mêlée, de ce qu'ils auront à terminer avant la prochaine Stand Up meeting et d'identifier les obstacles qui pourraient les bloquer. En effet, les trois questions "phares" du scrum master sont :

1. Qu'as tu fais hier ?
2. Que vas tu faire aujourd'hui ?
3. As tu des points bloquants ?

Quelques outils d'aide à la gestion d'une équipe « Agile »

De nombreux outils d'aide à la gestion d'une équipe « Agile » ont fait leur apparition ces dernières années. On peut en citer pêle-mêle Agilefant, IceScrum, Agilo, eXPlainPMT ou encore XPlanner. D'autres sont apparus dernièrement comme le JIRA Agile qui a l'avantage d'être facilement utilisable même pour les débutants. Il permet d'élaborer des tableaux de tâches, de créer et d'estimer les user stories (US), d'identifier l'engagement et la vélocité de l'équipe ou encore de créer divers rapports sur l'état d'avancement des projets. En outre, un outil comme PMTool offre un large choix de rapports dans une interface simple mais très complète. Ainsi, les développeurs tout comme les chefs de projet (Scrum master) dispose d'une interface simple pour saisir leur feuille de temps et rendre compte de leurs activités. Un tableau de bord personnalisable est également disponible pour permettre à l'utilisateur d'avoir sous ses yeux les principaux indicateurs pour le suivi de son projet/développement tout en prenant en compte l'itération en cours. Les questions/incidents pourront enfin être gérés de manière très aisée car cette action est centralisée et se fait en un seul clic grâce à un menu très discret qui s'affiche partout (quel que soit la rubrique en cours de consultation).

Conclusion

Avantages

- Méthode fun !
- Excellente réactivité vis-à-vis du client (product owners => PO)
- Qualité accrue du fait de la présence des PO
- Favorise et facilite la communication avec les autres membres de l'équipe
- Développements hors sujet très peu probable

Inconvénients

- Changement radical de gestion de projet/produit : fort impact sur tous les intervenants
- Dans les grosses sociétés : collaboration/synchronisation difficile avec les autres équipes qui n'utilisent pas cette méthode. (délais différents donc réactivité plus longues)
- Contraignante, surtout au début le temps que tout le monde s'y habitue.
- Bonne volonté et bonne entente de tous les participants impératif.
- Gros travail de rédaction pour les PO => création des "users stories" + cas de tests + critères de validation de l'US
- Les développeurs doivent accepter le changement (revenir sur du code pour le modifier voir le supprimer).
- Les développeurs doivent être autonomes (une équipe "Agile" doit être auto gérée)
- La sociabilité doit être une qualité de tous les intervenants (communication)

- Internet : un bienfait ou un danger pour la démocratie ?
- le 11 novembre 2018 De la Direction de l'information légale et Administrative

- Les médias contribuent-ils au débat démocratique ? |
- Internet : un bienfait ou un danger pour la démocratie ? |

Quelques dates clés

1969 : Création par le Département américain de la défense d'un réseau reliant plusieurs gros centres de calcul et des systèmes radio et satellites : Arpanet. Ce serait l'ancêtre d'Internet.

Années 1980 : Arpanet voit son intérêt stratégique faiblir avec la fin de la Guerre froide. Les chercheurs et les universités se dotent de réseaux personnels qui s'interconnectent progressivement (protocole TCP/ IP). Il devient alors possible d'atteindre un ordinateur de n'importe quel réseau depuis un autre ordinateur, à partir du moment où chaque machine est connectée à un réseau.

1989 : Tim Berners Lee, chercheur du CERN (Centre européen pour la recherche nucléaire) à Genève met au point un nouveau mode de communication permettant de faire des liens hypertexte. C'est l'apparition du web qui utilise le langage HTML (*HyperText Markup Language*).

1994-1995 : Sortie de Netscape Navigator et Internet Explorer, les deux premiers navigateurs grand public. Le moteur de recherche Google apparaît en 1998.

Fin novembre 1999 : Manifestation anti-mondialisation à Seattle. Elle a été organisée *via* le Net, permettant ainsi la coordination internationale des participants.

20 novembre 2000 : Condamnation de *Yahoo !* (annuaire de sites, créé en 1994) par le tribunal de grande instance de Paris parce que ce portail permet d'accéder à un site d'enchères où étaient en vente des objets glorifiant le nazisme. Il devra filtrer l'accès des internautes français à ces services.

2004-2005 : Apparition de Facebook et YouTube.

23 avril 2005 : En France, le rapport annexé à la loi d'orientation et de programme pour l'avenir de l'école prévoit que le B2i (Brevet informatique et internet), créé par l'Éducation nationale en novembre 2000, soit intégré au brevet et au baccalauréat.

2007 : Lancement de l'iPhone d'Apple. L'année suivante, premières applications. Le succès des smartphones, lié à la possibilité d'une connexion mobile de plus en plus fiable et rapide, est fulgurant (5 milliards d'unités vendues à l'horizon 2020).

Quelques pistes de réflexion

Internet peut favoriser le développement de la démocratie

- Il est né du souci de communiquer plus facilement, plus rapidement et à moindre coût. Il constitue donc un lieu de débats et permet, par le courrier électronique et la multiplication des forums, les sites web puis les réseaux sociaux, de confronter les points de vue et d'échanger des informations.
- Il contribue à une meilleure information des citoyens et à une plus grande transparence de l'action publique, par exemple par la diffusion d'informations administratives par le Net, de données publiques (Étalab, 2011), des programmes politiques des partis...
- Il favorise une interaction plus grande entre les gouvernants et les gouvernés. Les courriers électroniques permettent de faire connaître aux élus les attentes et les réactions de leurs administrés. On peut imaginer qu'Internet pourrait être un outil pour participer au processus législatif par les forums ouverts au Parlement et l'envoi aux parlementaires de propositions d'amendements. Ainsi, lors de l'élaboration du projet de constitution européenne par la Convention sur l'avenir de l'Europe (février 2002-juillet 2003), un forum ouvert aux organisations représentant la société civile leur permettait de verser leurs contributions aux débats.
- Il favorise par certains aspects les actions collectives permettant à des individus isolés, mais partageant les mêmes idées, d'entrer en contact et de se mobiliser.

Internet peut présenter des dangers pour la démocratie

- La "fracture numérique". Son utilisation est encore réservée à une relative élite, seule une minorité de la population, aux niveaux de revenus et de formation élevés, s'y connectant régulièrement. Des efforts sont fournis par les gouvernements, les fournisseurs d'accès et les fabricants d'ordinateurs pour élargir le public des internautes. Mais, ces initiatives concernent surtout les pays développés. Ainsi, en France, la création du B2i en novembre 2000 a permis de davantage familiariser les enfants aux technologies de l'information et de la communication (TIC). Le B2i de niveau 1 figure dans les programmes de l'école primaire depuis 2002. Avec les niveaux 2 et 3, il est désormais intégré au brevet et au baccalauréat. Mais de nombreux pays non développés sont délaissés par le Net.

- **Internet porte atteinte à la protection de la vie privée.** Ainsi les *cookies*, véritables "mouchards" déposés lors de la fréquentation d'un site, livrent un certain nombre de renseignements, dans un but commercial, concernant l'internaute et ses habitudes. La collecte des données personnelles, d'une façon générale, est une activité à part entière des géants du Net (souvent qualifiés de GAFA : Google, Amazon, Facebook, Apple), comme l'illustre le slogan viral : "si c'est gratuit c'est que vous êtes le produit". Dans l'Union européenne, l'entrée en vigueur du RGPD (Règlement général de protection des données) le 25 mai 2018 renforce et unifie la protection des données pour les individus, à peine de sanction des entreprises concernées (souvent des multinationales) par les autorités de régulation. Internet peut aussi être le support à la diffusion d'informations fausses ou diffamantes à l'égard d'un individu (voir le débat autour de la notion de *fake news* ou "infox" selon le néologisme forgé par la Commission d'enrichissement de la langue française en octobre 2018).
- **Internet peut être le support d'actions ou d'idées en contradiction avec les fondements de la démocratie.** Des sites propageant des idées négationnistes ou racistes y sont accessibles. De même, les *cookies* pourraient être exploités afin de connaître les opinions politiques des citoyens à travers leurs consultations de sites. Du côté des entrepreneurs ou des autorités, le risque d'une société de surveillance n'est pas absent : dans certains pays, comme la Chine ou Singapour, les autorités bloquent l'accès à des sites jugés nuisibles, c'est-à-dire pro-occidentaux, etc.

De la démocratie représentative à la démocratie participative ?

- **Internet pourrait répondre à la crise de la représentation** qui se manifeste notamment par une **abstention** électorale croissante. Il favorise la mobilisation des citoyens sur toutes sortes de questions importantes, sans médiation des institutions et des organisations politiques : pétitions (par exemple contre la peine de mort aux États-Unis), forums de discussion en ligne (*chat*), appel à manifestation. Il convient cependant de s'interroger sur la régularité et la représentativité de ces pratiques érigées en un système, qui serait susceptible de favoriser le populisme et la démagogie.
- **Internet pourrait compléter la démocratie représentative** en corrigeant ses insuffisances par une dimension plus participative, sans en altérer les fondements. On a pu envisager des votes électroniques, déjà testés dans certains pays, mais ceux-ci peinent à s'installer pour des raisons de sécurité (cyberattaques).
- **Attention à l'illusion que pourrait représenter l'e-démocratie.** L'offre d'espaces de liberté supplémentaires aux citoyens n'induirait pas la modification des structures de pouvoir. Dans le cas de la mondialisation, si Internet peut sembler en contester les effets, il en est un des vecteurs principaux.
- **Enfin, il ne faudrait pas qu'Internet soit un média sans contrôle démocratique.** Internet, en s'affranchissant des frontières, pose le problème du contrôle des serveurs et des sites. Les juridictions nationales peinent à imposer leur verdict et les législateurs à suivre le rythme des évolutions technologiques. Quant à la difficulté à distinguer les sources des informations en fonction de leur sérieux, voire de leur neutralité idéologique, ou même à assimiler une profusion toujours plus redondante d'informations sans grande valeur, on a pu parler de la nécessité d'éduquer le citoyen contre l'"infobésité".

Exemple concret

Le rôle de la CNIL

Il existe des institutions nationales qui tentent de réguler le système, notamment pour le prémunir contre les atteintes à la vie privée ou à la liberté individuelle. Ce rôle revient, en France, à la Commission nationale informatique et libertés (la CNIL), créée par la loi du 6 janvier 1978 (modifiée par la loi du 6 août 2004), qui doit ainsi vérifier le respect des droits de chacun : droit à l'information préalable en cas de constitution d'un fichier, droit d'accès aux informations individuelles, droit de rectification des erreurs, etc. Elle dispose de moyens juridiques puissants en cas d'**infraction**, mais qui restent limités aux frontières nationales. La loi de 2004 allège de façon substantielle les obligations déclaratives des détenteurs de fichiers, accroît les pouvoirs de la CNIL en ce qui concerne les contrôles sur place et les sanctions, et renforce les droits des personnes.

L'illectronisme

Le secrétariat d'État chargé du numérique compte 13 millions de Français qui n'utilisent que peu ou pas Internet, dont 6,7 millions qui ne s'y connectent jamais. Une enquête CSA de mars 2018 révélait que 15 % de nos concitoyens trouvaient l'usage des équipements de nouvelles technologies "difficile", et jusqu'à 39 % des personnes âgées de 70 ans et plus. "Si, en 2022, on oblige les gens à accéder aux services publics par le numérique, il y aura de 20 % à 25 % de personnes en difficulté", a alerté le Défenseur des droits, Jacques Toubon, le 18 octobre. Impôt sur le revenu (obligatoirement déclaré en ligne en 2019), et bientôt ordonnances médicales, inscription sur les listes électorales, recours aux prestations sociales, la dématérialisation des démarches administratives est en marche, et risque potentiellement de priver de l'exercice concret de leurs droits nombre de citoyens peu familiers d'Internet. Selon le secrétaire d'État au Numérique Mounir Mahjoubi, l'objectif affiché par le **gouvernement** est de former 1,5 million de personnes confrontées à l'illectronisme par an.

**"Plus que jamais, nous croyons en la Smart City et aux objets connectés",
Pierre Andrade (Suez)**

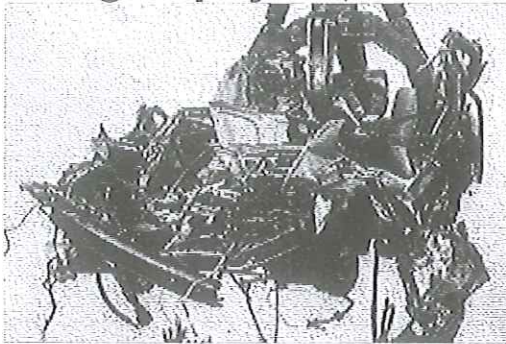
28

ENTRETIEN Avec le lancement de trois applications en quelques mois seulement, Suez muscle son offre en termes d'objets connectés. Pour L'Usine Digitale, Pierre Andrade, son directeur digital pour la France, revient longuement sur le déploiement des compteurs communicants OnConnect ainsi que sur la stratégie du groupe en matière d'IoT.

Pour l'usine Digitale article d'Arthur Le Denn - Publié le 25 février 2019

Pierre Andrade présente les trois nouvelles applications et la stratégie IoT de Suez. © Suez
A LIRE AUSSI

**Pour Suez, l'enlèvement de déchets est un service
en ligne payant comme les autres**



**Smart city : Terradona accélère le déploiement de
ses conteneurs connectés d'incitation au[...]**



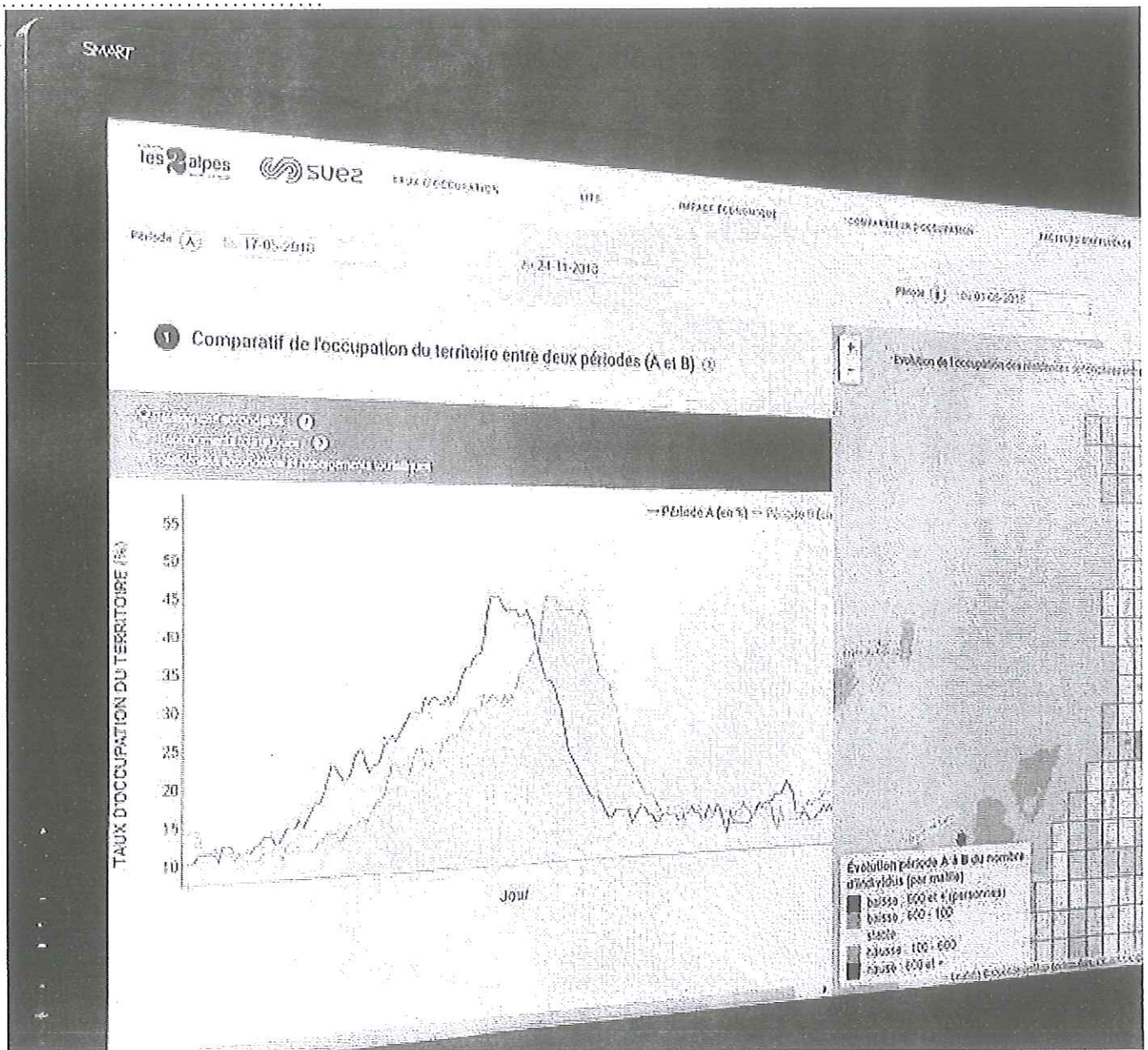
**Les camions poubelles relèvent aussi les compteurs
d'eau**



L'Usine Digitale : Suez a lancé les applications On'Connect Tourism et On'Connect Generation fin 2018. Quel est leur objectif, et quel premier bilan tirez-vous de cette expérience ?

Pierre Andrade : Tout est parti d'une demande de l'un de nos clients. La ville de Biarritz, pour laquelle nous opérons la gestion de l'eau et l'assainissement, voulait notamment optimiser sa gestion des services publics. C'est une ville touristique, qui connaît un pic de fréquentation en été. Elle souhaitait pouvoir optimiser le nombre de collectes d'ordures ménagères et adapter ses campagnes de communication en fonction du taux d'occupation réel.

Or, les compteurs d'eau sont le meilleur traceur d'activité humaine dans le domaine de l'habitat. Les données que nous récoltons sont révélatrices de la présence d'occupants. Les compteurs communicants que nous déployons depuis quelques années permettent de révéler à la rue près le taux d'occupation des logements. C'est la raison pour laquelle nous avons lancé l'application On'Connect Tourism. Il s'agit d'ouvrir des données en notre possession à de nouveaux usages.



En ce qui concerne On'Connect Generation, la démarche est similaire. Le vieillissement de la population amène des problématiques nouvelles, telles que le maintien à domicile. Maintenir les personnes âgées à domicile est synonyme d'économies pour les familles, les collectivités et, plus largement, la société, qui n'aura pas à coliser pour son placement en résidence spécialisée par exemple. Là encore,

les données de consommation d'eau fournissent de précieuses indications. Des algorithmes permettent de surveiller si le "rituel quotidien" a bien lieu, de la toilette à la préparation des repas. Les aidants ou la famille peuvent ainsi détecter la moindre anomalie, via des alertes envoyées en cas d'absence d'activité par exemple. L'application est aujourd'hui testée dans le cadre du plan Alzheimer de la ville de Nice. En plus des collectivités, nous la commercialiserons auprès des particuliers en compléments de médaillons déjà sur le marché.

Vous venez tout juste de lancer On'Connect Coach, application notamment destinée aux particuliers. Qu'entendez-vous proposer avec cette troisième application ?

P. A. : Aujourd'hui, 60 % de l'eau consommée dans nos foyers est chauffée avant utilisation. Cela a pour conséquence de doubler le montant de la facture. On'Connect Coach permettra à chacun d'optimiser sa consommation d'eau en se comparant à d'autres foyers similaires. L'application distillera des conseils selon le modèle suivant : "Une minute de moins sous la douche vous permettrait d'économiser telle somme". Un outil pratique, donc, qui permettra à de nombreux foyers de réaliser des économies tout en diminuant leur empreinte environnementale. On'Connect Coach est entrée en phase de test dans le Valenciennois et le Dunkerquois, où les collectivités ont décidé d'offrir la fonctionnalité à l'ensemble de leurs administrés.

Clé de voûte de la stratégie de Suez, quelque deux millions de compteurs communicants ont été déployés en France à ce jour. Prévoyez-vous d'y apporter des améliorations ?

P. A. : A l'origine, ces compteurs intelligents devaient simplement nous permettre de détecter les fuites, d'améliorer l'expérience client et d'instaurer une facturation en temps réel. Ils nous ont finalement conduits à repenser l'usage que nous faisons des données. Après Tourism, Generation et Coach, d'autres applications devraient sans aucun doute voir le jour. Nous allons en tout cas poursuivre les déploiements de compteurs On'Connect de façon massive. Ils évolueront en même temps que la technologie radio dont ils sont équipés, vers une solution encore moins énergivore et plus économique.

Quels sont vos projets futurs en matière d'IoT ?

P. A. : Avec 20 000 capteurs déployés à ce jour, nous sommes leader en France sur la mesure du remplissage des points d'apport volontaire (PAV) pour le verre et le carton. Cette solution vise à limiter le nombre de vidanges. Nous avons aussi installé de premières bouches d'incendie connectées, qui nous permettent de nous assurer que leur usage est bien réservé aux seuls pompiers.

La gestion des réseaux d'eau potable et d'assainissement constitue un élément clef de nos travaux actuels en IoT. Notre solution Aquadvanced Quality Monitoring permet par exemple de surveiller en temps réel la qualité de l'eau potable dans les réseaux de distribution. Elle mesure divers critères pour surveiller la composition physico-chimique des eaux et adapter le taux de chlore. Pour finir, nous avons déployé des capteurs dans les égouts de la région parisienne dans le but de mesurer le niveau de l'eau. En croisant les données recueillies aux prévisions pluviométriques de Météo-France, nous sommes en mesure de prévenir de potentiels débordements. L'installation d'un tel système a demandé un investissement de 10 millions d'euros au Syndicat interdépartemental pour l'assainissement de l'agglomération parisienne (Siaap). Autant dire une goutte d'eau en comparaison des 200 millions qui auraient dû être dépensés pour la construction de bassins de rétention d'eau de pluie en l'absence de ce dispositif ! Plus que jamais, Suez croit à la Smart City et aux objets connectés.

Machine Learning et Big Data : définition et explications

Bastien L 6 juillet 2018 Data Analytics, Intelligence artificielle

31

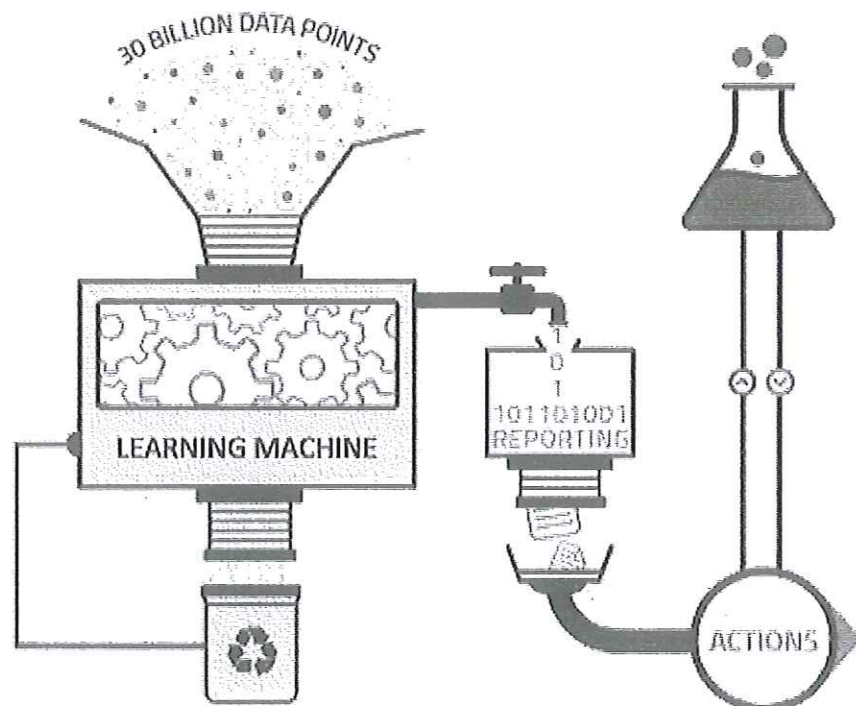
Le Machine Learning est une technologie d'intelligence artificielle permettant aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet. Pour apprendre et se développer, les ordinateurs ont toutefois besoin de données à analyser et sur lesquelles s'entraîner. De fait, le Big Data est l'essence du Machine Learning, et c'est la technologie qui permet d'exploiter pleinement le potentiel du Big Data. Découvrez pourquoi cette technique et le Big Data sont interdépendants.

Apprentissage automatique définition : qu'est ce que le Machine Learning ?

Si le Machine Learning ne date pas d'hier, sa définition précise demeure encore confuse pour de nombreuses personnes. Concrètement, il s'agit d'une science moderne permettant de découvrir des patterns et d'effectuer des prédictions à partir de données en se basant sur des statistiques, sur du forage de données, sur la reconnaissance de patterns et sur les analyses prédictives. Les premiers algorithmes sont créés à la fin des années 1950. Le plus connu d'entre eux n'est autre que le Perceptron.

Le Machine Learning est très efficace dans les situations où les insights doivent être découvertes à partir de larges ensembles de données diverses et changeantes, c'est à dire : le Big Data. Pour l'analyse de telles données, il se révèle nettement plus efficace que les méthodes traditionnelles en termes de précision et de vitesse. Par exemple, pour en se basant sur les informations associées à une transaction comme le montant et la localisation, et sur les données historiques et sociales, le Machine Learning permet de détecter une fraude potentielle en une milliseconde. Ainsi, cette méthode est nettement plus efficace que les méthodes traditionnelles pour l'analyse de données transactionnelles, de données issues des réseaux sociaux ou de plateformes CRM.

Machine Learning et Big Data : pourquoi utiliser le Machine Learning avec le Big Data ?



Les outils analytiques traditionnels ne sont pas suffisamment performants pour exploiter pleinement la valeur du Big Data. Le volume de données est trop large pour des analyses complètes, et les corrélations et relations entre ces données sont trop importantes pour que les analystes puissent tester toutes les hypothèses afin de dégager une valeur de ces données.

Les méthodes analytiques basiques sont utilisées par les outils de business intelligence et de reporting pour le rapport des sommes, pour faire les comptes et pour effectuer des requêtes SQL. Les traitements analytiques en ligne sont une extension systématisée de ces outils analytiques basiques qui nécessitent l'intervention d'un humain pour spécifier ce qui doit être calculé.

Comment ça marche ?

Le Machine Learning est idéal pour exploiter les opportunités cachées du Big Data. Cette technologie permet d'extraire de la valeur en provenance de sources de données massives et variées sans avoir besoin de compter sur un humain. Elle est dirigée par les données, et convient à la complexité des immenses sources de données du Big Data. Contrairement aux outils analytiques traditionnels, il peut également être appliqué aux ensembles de données croissants. Plus les données injectées à un système Machine Learning sont nombreuses, plus ce système peut apprendre et appliquer les résultats à des insights de qualité supérieure. Le Machine Learning permet ainsi de découvrir les patterns enfouis dans les données avec plus d'efficacité que l'intelligence humaine.

Des cours de Machine Learning sont disponibles sur le Web. Ils permettent notamment de débiter l'apprentissage automatique à partir du langage informatique Python. Ce dernier, assez simple à apprendre autorise donc les néophytes à tester des applications utilisant cette technique avec Python. De même, les open classroom permettent de découvrir gratuitement le fonctionnement de cette technique de traitement des données.

Machine Learning et Big Data : pourquoi le Machine Learning n'est rien sans Big Data

Sans le Big Data, le Machine Learning et l'intelligence artificielle ne seraient rien. Les données sont l'instrument qui permet à l'IA de comprendre et d'apprendre à la manière dont les humains pensent. C'est le Big Data qui permet d'accélérer la courbe d'apprentissage et permet l'automatisation des analyses de données. Plus un système Machine Learning reçoit de données, plus il apprend et plus il devient précis.

L'intelligence artificielle est désormais capable d'apprendre sans l'aide d'un humain. Par exemple, l'algorithme Google DeepMind a récemment appris seul à jouer à 49 jeux vidéo Atari. Par le passé, le développement était limité par le manque d'ensembles de données disponibles, et par son incapacité à analyser des quantités massives de données en quelques secondes.

Aujourd'hui, des données sont accessibles en temps réel à tout moment. Ceci permet à l'IA et au Machine Learning de passer à une approche dirigée par les données. La technologie est désormais suffisamment agile pour accéder aux ensembles de données colossaux et pour les analyser. De fait, des entreprises de toutes les industries se joignent désormais à Google et Amazon pour implémenter des solutions IA pour leurs entreprises.

Un exemple de Machine learning appliqué ? MetLife, l'un des principaux assureurs d'entreprise à l'échelle mondiale, utilise cette technique et le Big Data pour optimiser son activité. La reconnaissance de discours lui a permis d'améliorer le tracking d'accidents et de mieux mesurer leurs conséquences. Le traitement de réclamations est désormais mieux pris en charge car les modèles de réclamations ont été enrichis à l'aide de données non structurées qui peuvent être analysées par le biais de cette technologie.

Autre exemple, cette technique permet d'apprendre les habitudes des occupants d'un foyer. Les concepteurs d'objets connectés, notamment de thermostats, peuvent analyser la température du logement afin de comprendre la présence et l'absence des occupants pour couper le chauffage et le rallumer quelques minutes avant leur retour.

Le Deep Learning, un sous-domaine du Machine Learning

L'apprentissage automatique est un sous domaine de l'intelligence artificielle. Le Deep Learning est lui même une sous-catégorie de l'apprentissage automatique. L'exemple d'application le plus commun est la reconnaissance visuelle. Par exemple, un algorithme va être programmé pour détecter certains visages depuis les images en provenance d'une caméra. Suivant la base de données attribuée, il pourra repérer un individu recherché dans une foule, détecter le taux de satisfaction à la sortie d'un magasin en détectant les sourires, etc. Un ensemble d'algorithme pourra également reconnaître la voix, le ton, l'expression d'un questionnement, d'une affirmation et les mots.

Pour ce faire, le Deep Learning repose principalement sur la reproduction d'un réseau neuronal inspiré des systèmes cérébraux présents dans la nature. Les développeurs décident suivant l'application souhaitée quel type d'apprentissage ils vont mettre en place. Dans ce cadre, on parle d'apprentissage supervisé, d'apprentissage non supervisé dans lequel la machine va se nourrir de données non sélectionnées au préalable, semi-supervisé, par renforcement (lié à une observation), ou

par transfert dans laquelle les algorithmes vont appliquer une solution apprise dans une situation jamais vue.

En revanche, cette technique a besoin de beaucoup de données pour s'entraîner et obtenir des taux de réussite suffisant pour être utilisé. Un Lac de données ou Data Lake est essentiel pour parfaire l'apprentissage des algorithmes de Deep Learning. L'apprentissage profond nécessite également une puissance de calcul supérieure pour réaliser son office. Il faut s'équiper

Machine Learning et Big Data : les analyses prédictives donnent du sens au Big Data

Les analyses prédictives consistent à utiliser les données, les algorithmes statistiques et les techniques de Machine Learning pour prédire les probabilités de tendances et de résultats financiers des entreprises, en se basant sur le passé. Elles **rassemblent plusieurs technologies et disciplines comme les analyses statistiques, le data mining, le modélisation prédictif et le Machine Learning pour prédire le futur des entreprises**. Par exemple, il est possible d'anticiper les conséquences d'une décision ou les réactions des consommateurs.

Les **analyses prédictives permettent de produire des insights exploitables à partir de larges ensembles de données, pour permettre aux entreprises de décider quelle direction emprunter par la suite et offrir une meilleure expérience aux clients**. Grâce à l'augmentation du nombre de données, de la puissance informatique, et du développement de logiciels IA et d'outils analytiques plus simples à utiliser, comme Salesforce Einstein, un grand nombre d'entreprises peuvent désormais utiliser les analyses prédictives.

Selon une **étude menée par Bluewolf auprès de 1700 clients de Salesforce, 75% des entreprises qui augmentent leurs investissements dans les technologies analytiques en tirent profit**. 81% de ces utilisateurs des produits Salesforce estiment que l'utilisation des analyses prédictives est l'initiative la plus importante de leur stratégie de ventes. Les analyses prédictives permettent d'automatiser les prises de décision, et donc d'augmenter la rentabilité et la productivité d'une entreprise.

L'intelligence artificielle et le Machine Learning représentent le niveau supérieur des analyses de données. Les **systèmes informatiques cognitifs apprennent constamment sur l'entreprise et prédisent intelligemment les tendances de l'industrie, les besoins des consommateurs et bien plus encore**. Peu d'entreprises ont déjà atteint le niveau des applications cognitives, défini par quatre caractéristiques principales : la compréhension des données non structurées, la possibilité de raisonner et d'extraire des idées, la capacité à affiner l'expertise à chaque interaction, et la capacité à voir, parler et entendre pour interagir avec les humains de façon naturelle. Pour cela, il convient de développer le traitement par algorithme des langages naturels.

Machine Learning et Big Data : l'apprentissage automatique au service du Data Management

Face à l'augmentation massive du volume de données stockées par les entreprises, ces dernières doivent faire face à de nouveaux défis. Parmi les principaux challenges liés au Big Data, on dénombre la compréhension du Dark Data, la rétention de données, l'intégration de données pour de meilleurs résultats analytiques, et l'accessibilité aux données. Le **Machine Learning peut s'avérer très utile pour relever ces différents défis**.

Toutes les entreprises accumulent au fil du temps de grandes quantités de données qui demeurent inutilisées. Il s'agit des dark data. Grâce au Machine Learning et aux différents algorithmes, il est possible de faire le tri parmi ces différents types de données stockées sur les serveurs. Par la suite, un humain qualifié peut passer en revue le schéma de classification suggéré par l'intelligence artificielle, y apporter les changements nécessaires, et le mettre en place.

Pour la rétention de données, cette pratique peut également s'avérer efficace. **L'intelligence artificielle peut identifier les données qui ne sont pas utilisées et suggérer lesquelles peuvent être supprimées**. Même si les algorithmes n'ont pas la même capacité de discernement que les êtres humains, le Machine Learning permet de faire un premier tri dans les données. Ainsi, les employés économisent un temps précieux avant de procéder à la suppression définitive des données obsolètes.

Cette technologie est aussi utile pour l'intégration de données. Pour **tenter de déterminer le type de données qu'ils doivent agréger pour leurs requêtes, les analystes créent généralement un répertoire dans lequel ils placent différents types de données en provenance de sources variées pour créer un bassin de données analytique**. Pour ce faire, il est nécessaire de développer des méthodes d'intégration pour accéder aux différentes sources de données en provenance desquelles ils extraient les données. Cette technique peut faciliter le processus en créant des

mappings entre les sources de données et le répertoire. Ceci permet de réduire le temps d'intégration et d'agrégation.

Enfin, l'apprentissage des données permet d'organiser le stockage de données pour un meilleur accès. **Au cours des cinq dernières années, les vendeurs de solutions de stockage de données ont mis leurs efforts dans l'automatisation de la gestion de stockage.** Grâce à la réduction de prix du SSD, ces avancées technologiques permettent aux départements informatiques d'utiliser des moteurs de stockage intelligents reposant sur le machine Learning pour voir quels types de données sont utilisés le plus souvent et lesquels ne sont pratiquement jamais utilisés. L'automatisation peut être utilisée pour stocker les données en fonction des algorithmes. Ainsi, l'optimisation n'a pas besoin d'être effectuée manuellement.

Une forme pauvre de l'IA ?

Certaines voix s'élèvent au sein des entreprises afin de rappeler que l'humanité est au début du développement de l'intelligence artificielle. Selon Alex Danvy, Evêngéliste technique chez Microsoft France, le machine learning aujourd'hui est une forme simple d'IA. Les algorithmes ne sont pas encore capables d'accomplir les tâches aussi complexes que celles confiées à Skynet, le réseau informatique fictif du film Terminator. Qu'ils traitent des images, des sons, du texte, les algorithmes réalisent des tâches simples. Ce n'est qu'en interconnectant les algorithmes que l'on arrive à créer des systèmes plus intelligents. C'est de cette manière que sont pensées les voitures autonomes. Malheureusement, les acteurs de l'intelligence artificielle créent leurs solutions « dans leur coin », explique Alex Danvy. Selon lui, cela n'empêche pas l'émergence de solutions efficaces basés sur des algorithmes de machine learning « simples ».

<https://www.lebigdata.fr/13-compétences-nécessaires-devenir-data-scientist>

35

Voici les 13 compétences nécessaires pour devenir Data Scientist

Bastien L 20 juillet 2017 Data Analytics, Etudes Commentaires

En janvier dernier, le prestigieux site de recherche d'emploi Glassdoor a élu le travail de Data Scientist en première position de son top 25 des meilleurs métiers du monde. À travers cet article, découvrez les compétences nécessaires pour exercer cette profession au cœur du Big Data.

Chargé de la gestion, de l'analyse et de l'exploitation des données massives au sein d'une entreprise, le Data Scientist est l'évolution du Data Analyst à l'ère du Big Data. Selon l'étude menée par Glassdoor, le salaire annuel moyen d'un Data Scientist s'élève à 116 840 dollars.

Compte tenu de l'extrême spécialisation nécessaire pour exercer cette profession, les opportunités d'embauche sont nombreuses et largement supérieures à la quantité de profils qualifiés. Fin, janvier Glassdoor dénombrait ainsi 1736 offres d'emploi.

À n'en point douter, le métier de Data Scientist est passionnant. Cependant, il s'agit également d'un poste à haute responsabilité, qui nécessite des prédispositions naturelles et une éducation de haut niveau. Voici les compétences indispensables pour espérer faire carrière dans ce domaine.

Comment devenir Data Scientist ? Formations et compétences requises

1 – Une formation d'analyste

À l'heure actuelle, 88% des Data Scientists sont diplômés au minimum d'un master, et 46% d'entre eux sont titulaires d'un doctorat. Cette éducation scolaire semble nécessaire pour développer le niveau de connaissance suffisant à l'exercice de ce métier.

La majeure partie des professionnels (32%) sont issus d'une formation dans le domaine des mathématiques et des statistiques. 19% ont étudié les sciences informatiques et 16% proviennent d'écoles d'ingénieurs.

2 – Le Data Scientist doit avoir des connaissances en statistiques

Il est essentiel pour un Data Scientist d'avoir au minimum des notions de calculs statistiques. Ces connaissances lui permettront de déterminer la bonne technique d'approche et d'analyse pour chaque donnée.

3 – Le Data Scientist doit maîtriser des outils analytiques

Une connaissance approfondie d'au moins un outil analytique tel que SAS ou R est en général exigée. Pour la science des données, la préférence s'accorde principalement vers R, le langage informatique historique et normé pour l'analyse et l'exploration des données.

4 – Les langages de programmation

Les postes de Data Scientist requièrent la maîtrise d'au moins un langage de programmation. Le plus couramment employé est Python, mais il peut être remplacé par Java, Perl ou C/C++.

5 – Des notions de Machine Learning

En complément des outils analytiques, connaître quelques méthodes de Machine Learning peut être un réel atout pour la création d'un produit dirigée par les données. Il peut s'agir des **forêts d'arbres décisionnels, de k plus proches voisins ou encore des méthodes d'ensemble**. Comme ces différentes techniques peuvent être directement implémentées à l'aide des bibliothèques R ou Python, il n'est pas indispensable de savoir comment fonctionnent leurs algorithmes. L'important est de comprendre leur fonctionnement dans les grandes lignes et de savoir **quelle méthode est la plus pertinente selon la situation**.

6 – La compréhension de l'algèbre linéaire et des fonctions de plusieurs variables

L'algèbre linéaire et les fonctions de plusieurs variables constituent la **base de beaucoup de techniques de calculs statistiques et de machine learning**. Même s'ils sont implémentés avec R ou sklearn, certaines entreprises dont le produit est dirigé par les données peuvent décider de **développer leurs propres implémentations** pour améliorer leurs algorithmes ou leurs performances prédictives.

7 – L'utilisation d'Hadoop

Si certaines entreprises ne l'exigent pas, la **maîtrise de la plateforme Hadoop est le plus souvent requise**. De même, une expérience avec les outils de traitement Hive et Pig est un argument supplémentaire en vue d'un recrutement. Les outils de cloud comme Amazon S3 ont également leur importance.

8 – La programmation en SQL

Hadoop et les bases de données NoSQL se sont largement imposés dans le domaine du Big Data. Toutefois, la plupart des recruteurs exigent des candidats la maîtrise de la programmation en SQL pour pouvoir **formuler et exécuter des requêtes**. D'ailleurs, le SQL tend à redevenir le langage prédominant dans le Big Data en 2016.

9 – La gestion de données non structurées

Pour devenir Data Scientist, il est indispensable de savoir gérer des données non structurées **en provenance des réseaux sociaux, ou encore des flux vidéo ou audio**. Ces données sont le principal challenge du Big Data.

Il est également important de savoir traiter les données comportant des imperfections, **telles que des valeurs manquantes ou des chaînes de format incohérentes**. Cette compétence est particulièrement importante au sein des entreprises n'étant pas habituées à l'analyse de données.

10 – Des compétences en ingénierie logicielle

Au sein d'une petite entreprise peu accoutumée à la science des données, un Data Scientist doit avoir des compétences d'ingénieur logiciel. Celles-ci lui permettront notamment de **prendre en charge le développement d'un produit dirigé par les données ou le data logging**.

11 – La curiosité intellectuelle

La curiosité intellectuelle est indispensable pour déceler les données les plus intéressantes et exploitables au sein d'un gigantesque volume de data. Pour mener à bien le travail de Data Scientist, il est nécessaire **d'être créatif et de poser ses propres questions** plutôt que de simplement répondre à celles qui se posent.

12 – L'esprit d'un entrepreneur

Pour parvenir à exploiter le Big Data d'une entreprise, il est nécessaire de comprendre les problèmes à résoudre et les nouvelles possibilités que les données peuvent offrir. C'est pourquoi le Data Scientist doit comprendre **le monde de l'entreprise en général et l'industrie à laquelle il est affilié** plus particulièrement.

13 – Le Data Scientist doit avoir le sens de la communication

Intégré au sein de l'entreprise, le Data Scientist doit impérativement être en mesure de **communiquer ses découvertes techniques aux autres employés**, des pôles marketing ou commerciaux par exemple. Son rôle est d'aider les décideurs à prendre les bonnes décisions, en leur fournissant les informations nécessaires.

Il doit également **comprendre les problèmes des autres équipes** et les aider à relever ces défis grâce à l'analyse des données. Pour ce faire, il est également important de maîtriser les outils de visualisation de données tels que ggplot ou d3.js.

En conclusion, **les compétences requises pour un Data Scientist sont nombreuses et spécifiques**. Avant de décider d'entreprendre une formation ou une carrière dans ce domaine, il est nécessaire de déterminer si vous avez, oui ou non, le profil d'un scientifique des données.

Quelles sont les meilleures formations françaises pour devenir Data Scientist ?

En France, **on compte actuellement une quarantaine de formations au métier de Data Scientists**. Des masters universitaires, des mastères spécialisés, des Master of Science, des spécialisations de 3ème année, et des MBA permettent d'accéder aux compétences nécessaires pour devenir Data Scientist.

On peut diviser ces formations en trois catégories principales. Tout d'abord, **les cursus proposés par des écoles d'ingénieurs ou des universités scientifiques**. L'Ensai, l'Ensa, Polytechnique, Télécom ParisTech, Télécom Nancy, Eisti, Epita, proposent toutes, un programme de Data Science.

En termes de formations universitaires, Reims-Champagne-Ardenne propose un Master Statistique pour l'évaluation et la prospective. Louis-Lumière Lyon-II offre un M2 Data Mining et parcours Business Intelligence et Big data. L'Université Dauphine propose un Executive Master Statistique et Big data. À l'UPMC, les étudiants peuvent obtenir un master et certificat Data science. Un master Informatique-données à Nantes, un master Data Science à Nice-Sophia, et un master Big data et fouille de données à Paris-VIII. À elle seule, l'université Paris-Saclay réunit 45 formations de Data Sciences : 12 masters, 5 certificats, 8 spécialités d'ingénieurs, 4 MBA...

La **deuxième catégorie est celle des écoles de management**. Parmi les écoles proposant des MS, des MSc ou des spécialisations de troisième année, on compte Télécom EM, Neoma, HEC, Audencia, l'Inseec, l'Ieseg, l'ECE, l'ESC Rennes ou encore l'Essca, l'Ecole de management du pôle Léonard-de-Vinci et l'Institut de l'Internet et du multimédia.

La **troisième catégorie est celle des formations conjointes ingénieurs-management**. Parmi les établissements proposant de telles formations, on peut citer l'instar l'Essec et Centrale-Supélec, l'EPSI et l'Esilv.

Quel est le salaire d'un Data Scientist ? Y a-t-il beaucoup d'offres d'emploi ?

En 2017, le **Big Data devrait continuer à dominer le marché de l'emploi américain**. De nouveau, Glassdoor place les Data Scientists au sommet de sa liste des 50 meilleurs métiers. Viennent ensuite les ingénieurs DevOps et les Data Technicians.

Le **métier de Data Scientist est considéré comme le mieux payé, le plus satisfaisant et le plus recherché**. Le salaire moyen d'un Data Scientist américain s'élève à 110 000 dollars. En France, le salaire d'un débutant est en général compris entre 45000 et 50000 euros par an. Par ailleurs, malgré l'apparition de nombreuses formations, les entreprises peinent toujours à trouver des profils suffisamment qualifiés.

Le métier de Data Scientist risque-t-il de disparaître ?

Selon un rapport publié par Gartner Inc, **plus de 40% des tâches effectuées par un Data Scientist seront automatisées d'ici 2020**. Par conséquent, la productivité des Data Scientists va largement augmenter, au même titre que l'usage des données et des outils analytiques par les « citizen data scientists ».

Gartner définit les « citizen data scientists » comme des personnes qui créent ou génèrent des modèles en utilisant des outils de diagnostic avancés ou des outils prédictifs, mais dont la fonction principale n'est pas liée au champ des statistiques et de l'analytique. Ces personnes peuvent réduire l'écart les outils analytiques self-service utilisés par les entreprises, et les techniques analytiques avancées utilisées par les Data Scientists. Il est désormais possible d'effectuer des analyses pointues sans avoir besoin de compétences avancées.

La science des données est désormais un atout convoité par la plupart des entreprises, et c'est la raison pour laquelle les vendeurs de plateformes logicielles de données ou d'analyses cherchent en priorité à simplifier par l'automatisation de tâches variées telles que l'intégration de données et la création de modèles. Malgré tout, il est peu probable que le métier de Data Scientist soit remplacé par l'intelligence artificielle.

Community Manager ou Animateur de communauté ou Gestionnaire de réseaux sociaux

Définition de la fonction	<p>Le Community Manager est un expert des communautés web. Il fédère et anime les échanges entre les usagers internautes, développe et valorise la présence de l'organisation sur les réseaux sociaux et s'en fait le porte-parole.</p> <p>Il définit ou propose des contenus pour chaque cible (images, vidéos, articles, jeux, concours...), planifie la production, publie et évalue l'efficacité des actions des membres en cohérence avec la stratégie déclinée pour la communauté.</p>
Autres appellations	<p>Animateur de communauté, Animateur de réseaux sociaux, Gestionnaire de communautés Internet, Animateur/manager de communautés web, Animateur / Animatrice de communauté virtuelle, Animateur / Animatrice de forum sur internet, Animateur / Animatrice web, Chargé / Chargée des relations avec les internautes, Community manager Modérateur / Modératrice web, Webmaster animateur / animatrice</p>
Facteurs d'évolution	<p>Développement du numérique ou transformation digitale</p> <p>Développement et traitement des données,</p> <p>Application du (RGPD) pour alimenter la communication</p> <p>Développement de la communication multimodale en temps réel</p> <p>Importance accrue de la communication numérique</p> <p>Transparence de la vie publique</p> <p>Ouverture des services et des données aux usagers (Ouverture des données publiques ou open data)</p> <p>Développement des services mutualisés en relation avec l'évolution des organisations</p>
Conditions d'exercice	<p>Cette fonction s'exerce par désignation de l'autorité territoriale, à temps complet ou à temps partiel.</p> <p>Possible partage de la fonction entre plusieurs collectivités et EPCI, centre de gestion.</p> <p>Externalisation possible sur la base d'un contrat de service.</p> <p>Connaissance des règles de collecte et du traitement de l'information</p> <p>Sens du relationnel et de la communication web</p>
Activités techniques	<p>Production de contenus cibles et publication des actions des usagers internautes en cohérence avec la stratégie déclinée pour la communauté</p> <p>Animation et renforcement de la cohésion des membres ou usagers internautes pour échanger et partager des informations utiles et pertinentes</p> <p>Evaluation et suivi des méthodes, des actions de communication pour améliorer et faire évoluer si nécessaire à la stratégie de communication de l'organisation</p> <p>Accompagnement, développement technique et fonctionnel de la plateforme permettant aux membres d'accéder facilement à l'information</p>
Activités spécifiques	<p>Contribution à la politique d'e-réputation de l'organisation</p>

Activités/Compétences "Techniques"

Savoir-faire

Production de contenus cibles et publication des actions des usagers internautes en cohérence avec la stratégie déclinée pour la communauté

Rédiger des contenus spécifiques aux réseaux sociaux

41

Planifier la production des contenus à l'aide d'outils de publication

Publier les actions des membres ou usagers internautes de la communauté dans le respect des règles de fonctionnement de la communauté

Utiliser et communiquer avec les outils multimédia (images, vidéos, articles, jeux, concours...),

Assurer la traçabilité et la visibilité des actions engagées et des besoins complémentaires si nécessaire

Animation et renforcement de la cohésion des membres ou usagers internautes pour échanger et partager des informations utiles et pertinentes

Utiliser et développer les outils et moyens d'animation des réseaux sociaux (jeux, concours, flux d'actualité, compte à rebours, partage d'expériences...)

Contrôler la production de contenu en interne pour assurer une publication régulière des informations sur les forums, les fils d'actualités

Fédérer et donner du sens aux interactions des usagers internautes (internautes, bloggeurs, forums de discussion, les autres communautés...)

Planifier et suivre les actions de recrutement de membres en ligne : partenariats, parrainage, campagne d'e-mailing

Assurer une veille et fidéliser les usagers internautes via des évènements on-line (newsletter, jeux, concours...)

Evaluation et suivi des méthodes, des actions de communications pour améliorer et faire évoluer si nécessaire à la stratégie de communication de l'organisation

Mettre en place une démarche d'évaluation pour la communauté

Utiliser les requêtes et outils d'analyses adéquats pour construire indicateurs et tableaux de bord

Analyser les indicateurs de satisfaction et d'insatisfaction des individus et de l'ensemble de la communauté

Suivre les statistiques de fréquentation et intervenir pour mesurer les motivations et centres d'intérêt des usagers internautes pour mieux répondre à leurs attentes

Proposer des axes d'amélioration en matière de : communication, design, outils, technique d'utilisation...

Accompagnement, développement technique et fonctionnel de la plateforme permettant aux membres d'accéder facilement à l'information

Produire des outils, modes opératoires, guides utilisateurs et outils d'accompagnement au changement pour répondre aux besoins des usagers internautes

Intégrer les suggestions et les propositions des membres ou usagers internautes

Veiller à la disponibilité de la plateforme pour suivre les évolutions techniques et fonctionnelles réseaux sociaux

diagnostiquer les incidents ou dysfonctionnements de la plateforme et travailler avec les équipes de l'organisation pour apporter des solutions,

Savoirs socioprofessionnels

Application du RGPD

Règles particulières de recueil et de traitement des informations de la collectivité ou de l'établissement public

Modes de traitement des informations

Conduite de projets de communication (web)

Gestion de l'information

Politique de confidentialité et de sécurité des informations

Tableaux de bord et indicateurs

Méthodes et techniques de concertation et de communication

Techniques de communication (en particulier de communication numérique)

ÉPREUVES

N° 5 & N° 10