



CONSEIL GÉNÉRAL DE L'ÉCONOMIE  
DE L'INDUSTRIE, DE L'ÉNERGIE ET DES TECHNOLOGIES

TELEDOC 792  
BATIMENT NECKER  
120, RUE DE BERCY  
75572 PARIS CEDEX 12

Juillet 2015

N° 2014/27/CGE/SG

# Meilleures pratiques pour le « *big data* » et l'analytique dans l'administration : une nouvelle étape

## Rapport à

Monsieur le Ministre de l'Économie, de l'Industrie et du Numérique

Madame la Secrétaire d'Etat chargé de la réforme de l'Etat et de la simplification

Madame la Secrétaire d'Etat chargée du numérique

Établi par

**Jean-Pierre DARDAYROL**  
Ingénieur général des mines

**Marc MEYER**  
Ingénieur général des mines

**Françoise TRASSOUDAINÉ**  
Ingénieur général des mines



## Remerciements

Nous remercions tout particulièrement les 29 responsables de programme LOLF qui se sont impliqués et ont répondu au questionnaire, les 70 personnes rencontrées au cours des entretiens, celles qui ont participé aux ateliers, et notamment lorsqu'elles ont contribué à ce rapport en rédigeant des encarts.

En outre, nous exprimons notre gratitude aux personnes et représentants du SGMAP qui ont apporté à la mission leur aide fructueuse.



## SOMMAIRE

SYNTHESE .....	9
TABLE DES RECOMMANDATIONS .....	11
1 LE RENOUVEAU DES POSTURES VIS-A-VIS DES DONNEES A CONSTITUE LE CŒUR DU TRAVAIL COMMUN A LA MISSION, AUX EXPERTS ET AUX RESPONSABLES .....	13
1.1 <i>Les nouvelles postures vis-à-vis des données sont porteuses d'innovations au-delà de leur seule massification .....</i>	13
1.2 <i>Une méthode de réflexion collaborative avec les responsables de programmes LOLF, les experts et les chefs de projets permet d'appréhender des problématiques multiples et, parfois, contingentes .....</i>	14
2 LES RESPONSABLES LOLF PRIVILEGIENT UNE APPROCHE PAR L'EXPERIMENTATION DU « BIG DATA », JUGE PROMETTEUR, SANS AVOIR ENCORE, EN GENERAL DEFINI DE POLITIQUE OU D'ORGANISATION .....	16
2.1 <i>Le « big data » est généralement projeté comme prometteur dans l'attente des premiers retours d'expérience .....</i>	16
2.2 <i>Les attentes sont variées, elles portent majoritairement sur l'amélioration de la gestion interne des services .....</i>	17
2.3 <i>Les délais d'appropriation seront déterminés par des démarches d'apprentissage choisies progressives et situées dans la continuité .....</i>	17
2.4 <i>Les échanges relatifs au sujet du « big data » avec des tiers sont déjà établis dans près de la moitié des cas avec les partenaires de proximité, l'ensemble des services exprimant un souhait d'échanges, unanimement privilégié entre pairs .....</i>	18
2.5 <i>Le degré d'engagement et de maturité au regard de la réalisation de projets paraît très différencié, dégageant des pionniers et des volontaires .....</i>	19
2.6 <i>Les services privilégient une approche par l'expérimentation, cohérente avec le choix d'un apprentissage progressif .....</i>	19
2.7 <i>La gouvernance et l'organisation autour du « big data » reste à définir dans les trois quarts des cas et à préciser dans les autres .....</i>	20
2.8 <i>La moitié des services ont des attentes vis-à-vis des services du Premier ministre, corrélées à leur degré d'appropriation du « big data » : évolution du cadre juridique et appui politique pour les plus avancés, attentes d'appuis « opérationnels » pour les autres .....</i>	20
2.9 <i>L'analyse sémantique des réponses montre que les services de l'Etat disposent d'une culture commune qui n'est pas exempte d'angles morts mais qui est propice à des innovations et des collaborations .....</i>	21
2.9.1 <i>Les réponses ne font pas apparaître certains aspects du sujet : économies (faire des), efficacité, efficience, emploi, innovation, Europe, international, collectivités territoriales, services territoriaux de l'Etat .....</i>	22
2.9.2 <i>Les réponses font apparaître un vocabulaire commun propre à servir de socle au travail entre pairs : compétences, données personnelles/vie privée et cadre juridique, sécurité. ....</i>	23

<b>3</b>	<b>LES DONNEES DISPONIBLES AU PLAN LOCAL ET DE PROVENANCES TRES VARIEES AINSI QUE LES PROBLEMATIQUES REGIONALES CONSTITUENT UN CREUSET PRECIEUX POUR ALIMENTER LES POLITIQUES PUBLIQUES AU NIVEAU LOCAL .....</b>	<b>23</b>
<b>4</b>	<b>LES REALISATIONS DANS DIVERS DOMAINES DES MINISTERES, ENCORE PEU NOMBREUSES, S'APPUIENT SUR DES RESSOURCES INTERNES ET MONTRENT UN ESPRIT D'INNOVATION REEL.....</b>	<b>25</b>
4.1	<i>L'expérience du recours aux données de caisse pour élaborer l'indice des prix de l'Insee illustre l'intérêt, les conditions et les limites du recours à des tierces données .....</i>	<i>26</i>
4.2	<i>Tracfin et ses équipes refondent l'ensemble de leurs activités sur une stratégie de projet collaborative, itérative et une gestion du changement forte .....</i>	<i>28</i>
4.3	<i>Le projet interministériel de « big data » collaboratif VITAM se distingue par une politique de réutilisation et une gouvernance partagée pluriministérielle.....</i>	<i>30</i>
4.4	<i>La mission pluridisciplinaire de lutte contre la fraude de la DGFIP s'articule étroitement avec les services de contrôle .....</i>	<i>32</i>
4.5	<i>Le succès du « big data » à la direction générale des douanes et des droits indirects repose sur la qualité des relations avec le terrain .....</i>	<i>35</i>
4.6	<i>AF/KLM associe des compétences technologiques et fonctionnelles pour le « big data »... 36</i>	
4.7	<i>Le HUB de BPI France décloisonne l'innovation.....</i>	<i>37</i>
4.8	<i>Les réalisations des ministères couvrent dès à présent une gamme élargie de circonstances et dégagent des lignes de force.....</i>	<i>38</i>
<b>5</b>	<b>LES EXPERTS APPORTENT DES A PRESENT DES VISIONS CONSTRUITES DANS LES DOMAINES MANAGERIAUX, HUMAINS, FINANCIERS, TECHNIQUES ET JURIDIQUES .....</b>	<b>39</b>
5.1	<i>Politique de l'AGD en matière de « big data » et d'analytique.....</i>	<i>39</i>
5.2	<i>La gouvernance des projets « big data » trouvera son utilité à la croisée de celles des systèmes d'information et du décisionnel selon des modes différenciés selon les types de projet et les contingences locales .....</i>	<i>40</i>
5.3	<i>Faire se développer vigoureusement le « big data » dans les administrations de l'État nécessitera des talents variés, animés d'un esprit de communauté, gérés et formés de façon spécifique au sein des métiers mêmes .....</i>	<i>41</i>
5.3.1	<i>L'activité « big data » créera – et nécessitera - directement un nombre significatif d'emplois, exprimé en milliers, notamment de développeurs, avec des salaires supérieurs à la moyenne des emplois homologues .....</i>	<i>42</i>
5.3.2	<i>La constitution et l'animation d'une communauté tonique sont des prérequis au déploiement de formations de qualité.....</i>	<i>43</i>
5.3.3	<i>La gestion des spécialistes du « big data » gagne à être une composante de celle des personnels des métiers.....</i>	<i>44</i>
5.3.4	<i>La gestion des multiples compétences nécessaires à l'Autorité de sûreté nucléaire constitue un exemple de professionnalisation de haut niveau de fonctionnaires issus de corps différents</i>	<i>45</i>
5.3.5	<i>Air France / KLM a construit ses activités et ses compétences de haut niveau dans le domaine du « big data » et des « data sciences » sur la base de ses personnels et de ses équipes existants.....</i>	<i>46</i>

<b>5.4</b>	<b><i>Les questions juridiques multiples et complexes constituent un point d'attention pour les porteurs de projet et justifient la mise en place d'un réseau d'experts.....</i></b>	<b>47</b>
5.4.1	<i>Les « big data » créent de nouvelles responsabilités qu'il convient d'anticiper .....</i>	48
<b>5.5</b>	<b><i>La définition et la mise en œuvre d'un cadre organisationnel des données stimulerait l'usage des données.....</i></b>	<b>50</b>
<b>5.6</b>	<b><i>Des incitations financières françaises et européennes sont mobilisables pour développer des innovations en matière de « big data » administratives.....</i></b>	<b>52</b>
<b>6</b>	<b>LES PERSPECTIVES DU « BIG DATA » CONDUISENT A RECOMMANDER L'ENGAGEMENT D'UNE PHASE COURTE ET ACTIVE DE CONSTRUCTION DES CONDITIONS DE MASSIFICATION DE LEUR USAGE .....</b>	<b>54</b>
<b>ANNEXES</b>	<b>.....</b>	<b>59</b>
	<i>Annexe 1 : Lettre de mission.....</i>	61
	<i>Annexe 2 : Liste des acronymes utilisés.....</i>	63
	<i>Annexe 3 : Liste des personnes ayant participé aux ateliers.....</i>	64
	<i>Annexe 4 : Liste des intervenants aux ateliers.....</i>	65
	<i>Annexe 5 : Programme des ateliers .....</i>	66
	<i>Annexe 6 : Liste des personnes rencontrées ou interrogées .....</i>	68
	<i>Annexe 7 : Liste des programmes interrogés.....</i>	72
	<i>Annexe 8 : Questionnaire aux responsables de programme LOLF Innovations pour tirer parti des données, « big data ». ....</i>	75





## SYNTHESE

Dans un contexte fortement évolutif et innovant, la mission a retenu une définition large du « *big data* » reposant sur une nouvelle exigence de valorisation des données, qu'elles soient internes ou externes.

La mission a procédé à un état des lieux auprès des responsables de programme LOLF et bénéficié de la coopération d'un groupe de responsables de projets venus de différents ministères.

Les nombreuses réponses aux questionnaires complétées par des entretiens montrent un intérêt et une appétence pour des usages innovants des données.

Par ailleurs, des réalisations pionnières permettent d'identifier des bonnes pratiques au sein de l'Etat. Les réalisations existantes sont dans le cœur de métier avec des ressources (modérées) et dédiées.

Les échanges de données entre administrations sont peu développés et il n'y a pas de cadre de référence pour le faire. Il existe au moins une expérience d'utilisation de données externes (privées) en substitut de données internes qui pose des questions spécifiques.

D'autres aspects des « *big data* » devraient être explorés plus précisément, notamment une meilleure représentation visuelle plus compréhensible des résultats pour permettre des décisions ou le partage de la connaissance. Un champ très important (criticité, volume, ...) est en train de s'ouvrir avec les objets connectés.

Il existe un gisement de problématiques important au sein des services déconcentrés de l'Etat qui sont un ferment fort pour des opportunités de mise en œuvre de « *big data*. »

Cependant, la révélation du potentiel du « *big data* » nécessite de réunir des ressources et un contexte fonctionnel adapté, dans les domaines juridique, managérial, technologique et culturel.

Le retour en termes de valeur semble assuré dans le domaine de la lutte contre la fraude. Dans d'autres domaines, le retour en termes de valeur est d'une autre nature : par exemple des externalités positives, un meilleur service au citoyen ou une meilleure efficacité des politiques publiques.

L'utilisation des données tierces dans les domaines fiscaux ou grâce à « dites-le nous une fois » a généré ou a vocation à générer des économies de fonctionnement, tant pour l'administration que pour les administrés. Les processus de l'administration qui reposent sur un ciblage ou un criblage

sont également des candidats potentiels pour démontrer l'utilité économique du « *big data* » tant pour la société que pour l'Etat.

L'Etat a donc intérêt à s'approprier la culture « *big data* » dans des situations variées (production de connaissance, optimisation de processus, services rendus aux usagers, ...) en s'appuyant tant sur ses données par un découplage maîtrisé qu'en ayant recours, dans des cadres sécurisés, à des données de tiers.

L'apprentissage à partir de cette diversité nécessite une gouvernance et de la souplesse.

Enfin, il faut souligner que dans ce document, le terme « données » ou « *data* », conformément à l'usage, désigne tant des données structurées que des informations non structurées.

En conclusion, la mission recommande de s'intéresser en priorité aux modalités d'échanges de données entre ministères par l'établissement d'un référentiel adapté ; elle recommande également de mener une opération de consolidation de l'usage des « *big data* » sur deux ans, puis d'en faire une évaluation avant de l'étendre. Cette étape de deux ans permettrait d'identifier une communauté ouverte de décideurs et professionnels de l'administration du « *big data* ».

\*

\* \*

## TABLE DES RECOMMANDATIONS

<b>Recommandation n° 1.</b>	Engager un programme de deux ans de consolidation en vue de la massification des « <i>big data</i> » ..... 55
	Identifier concrètement cinq programmes LOLF dans lesquels on tirera profit largement du « <i>big data</i> » et de l'analytique..... 55
<b>Recommandation n° 2.</b>	Encourager des POC de découverte de différentes natures (tierces données, objets, visualisation, etc...) notamment dans des domaines pionniers à conduire en mode projet ..... 56
	Evaluer et mobiliser les ressources humaines et financières existantes et nécessaires à la réalisation des chantiers..... 56
	Pour préparer le moyen terme, introduire le « <i>big data</i> » dans la GPEEC et dans les formations initiales et continues, avoir recours de façon préférentielle au personnels présents et parmi eux entreprendre une action de formation en fonction des besoins futurs, en complément s'appuyer sur les entreprises innovantes d'informatique et de data sciences ..... 56
<b>Recommandation n° 3.</b>	Mandater des préfets pour initier une politique régionale de « <i>big data</i> » (par exemple deux régions), en lien en tant que de besoin avec les collectivités territoriales ..... 56
	Leur affecter une équipe pluridisciplinaire et les ressources budgétaires idoines..... 56
<b>Recommandation n° 4.</b>	<b>Constituer une communauté ouverte des décideurs et des professionnels de l'administration du « <i>big data</i> » et l'animer ..... 56</b>
<b>Recommandation n° 5.</b>	57
	<b>Créer l'incubateur bug data de l'Etat en constituant une équipe d'appui et de compétences dont les missions sont :</b> Favoriser l'utilisation des leviers d'incitation français et européens. .... 57
	Créer les conditions à une demande de formations et mettre en valeur des parcours de formation..... 57

Promouvoir des modèles (juridique, économique et technique) d'accès aux données externes (secteur privé, collectivités publiques, délégataires de service public).....	57
Offrir l'accès à une plateforme équipée et armée d'une équipe .....	57
Accompagner les ministères dans la gestion et l'assurance des nouveaux risques .....	57
Animer les relations avec la CNIL, la CADA .....	57
Faire une revue des cadres juridiques pour lever les obstacles au « big data » et proposer les amendements.....	57
Documenter au fil de l'eau les réalisations ainsi que les réalisations homologues en Europe .....	57
<b>Recommandation n° 6.</b>	57
<b>Faciliter les échanges de données entre les services de l'Etat, ses opérateurs .....</b>	<b>57</b>
Disposer d'un cadre contractuel pour des cessions de droit d'usage de données et d'un médiateur pour traiter les conflits entre services d'Etat et leurs opérateurs .....	57
Disposer d'un comité de déontologie pour l'utilisation des données	57
Identifier les charges afférentes à la mise en forme, la complétude, la qualité, à la livraison des données .....	57
Constituer un répertoire des modèles de données .....	57
<b>Recommandation n° 7.</b>	57
<b>Evaluer les réalisations ministérielles, régionales et interministérielles pour proposer une politique de « big data » de l'Etat pour ses politiques publiques .....</b>	<b>57</b>
Retenir une méthode (par exemple MAREVA) et définir des critères sur la base desquels l'évaluation sera faite .....	57

## 1 LE RENOUVEAU DES POSTURES VIS-A-VIS DES DONNEES A CONSTITUE LE CŒUR DU TRAVAIL COMMUN A LA MISSION, AUX EXPERTS ET AUX RESPONSABLES

Par lettre du 24 décembre 2014, le Ministre de l'économie de l'industrie et du numérique, le Secrétaire d'Etat chargé de la réforme de l'Etat et de la simplification, la Secrétaire d'Etat chargée du numérique ont confié au Conseil général de l'économie une mission qui porte sur le « big data ». Jean-Pierre Dardayrol, Marc Meyer et Françoise Trassoudaine ingénieurs généraux des mines, ont été désignés pour effectuer cette mission.

Etablissant le constat de la mise en œuvre de solutions « big data » dans le domaine du marketing ou des services commerciaux en ligne, la lettre de mission précise -au vu d'ouvrages et publications- les secteurs dans lesquels celles-là pourraient apporter des améliorations : « *organisation de la sécurité publique et prédiction criminelle, ciblage et criblage des contrôles, parcours de soins et validation de protocoles<sup>1</sup>, analyse de la réussite scolaire ou universitaire, connaissance des entreprises, statistique publique, gestion de la ville, fraude fiscale et sociale , etc. »*

La mission a donc pour objet d'identifier les prérequis ; les points clés et les bonnes pratiques d'un développement de l'approche « big data » au sein des services de l'Etat.

### 1.1 Les nouvelles postures vis-à-vis des données sont porteuses d'innovations au-delà de leur seule massification

Dans l'ensemble de ce rapport, la notion de « big data » a été prise dans une acception large résultats **incluant toute forme innovante d'usage des données : data mining, acquisition externe de données, etc.**

**Dans les exemples rencontrés, les caractéristiques des données considérées sont généralement « modestes » au regard des critères souvent utilisés pour définir le « big data »**

- leur volume se mesure en téraoctets, ou en dizaines de téraoctets – par opposition au pétaoctet des grandes bases de données (1000 fois plus),

- leur diversité est faible,

- des données structurées et parfois un peu de texte, leur taux de renouvellement se mesure le plus souvent en mois ou semaines et non en minutes ou secondes.

Enfin, si certaines administrations considèrent que le « big data » vise à **la production de savoir et de connaissances, et** ne concernent pas **les cas individuels**, la mission a considéré que le « big data » concernait tant la production de connaissance que l'étude de cas particuliers ou individuels.

Si l'utilisation de données tierces comme la déclaration d'impôt pré-remplie est considérée outre-manche comme un exemple de « big data », la DGFIP ne considère pas cette réalisation comme telle ; la mission n'a pas retenu cet exemple en tant que « big data ». La mise en place de la déclaration fiscale préremplie comme le programme « **dites-le nous une fois** » montrent que des

---

<sup>1</sup> La santé publique est un secteur spécifique qui doit être traité en tant que tel.

gains (des économies) sont potentiels. Si la mission n'a pas pu en mesurer les effets dans les exemples rencontrés, elle estime que des exemples bien choisis pourront apporter des gains, notamment en ce qui concerne le criblage et le ciblage des contrôles.

La mission a constaté que la distinction entre « *big data* » et « *open data* » n'est souvent pas bien intégrée tant au niveau français qu'au niveau européen. Sans en rechercher les causes, la visibilité et le volontarisme d'ouverture des données de l'Etat -portée en France par ETALAB- pourrait éclipser la politique naissante de la gestion des données pour le « *big data* ».

La mission a choisi de travailler à droit constant, malgré les nombreuses évolutions en cours des droits français et européen des données :

- Droit d'auteur ;
- Droit des données personnelles, etc.,
- et jurisprudentielles.

En effet, si l'application du droit en vigueur est impérative, il est apparu que les questions relatives à l'application, et/ou à l'évolution du droit des données ne se révélaient pas déterminantes pour l'appropriation du « *big data* » mais un élément parmi d'autres à prendre en compte.

**La mission définit le « *big data* » comme toute forme innovante d'usage des données. Elle considère cette définition large et holistique comme pertinente, dans la mesure où ces pratiques fondent de nouvelles postures au regard des données de toute nature dans leur diversité et leur hétérogénéité.**

## ***1.2 Une méthode de réflexion collaborative avec les responsables de programmes LOLF, les experts et les chefs de projets permet d'appréhender des problématiques multiples et, parfois, contingentes***

Dans un premier temps, afin de dresser un état des lieux du « *big data* » – pris au sens le plus large-, **la mission a adressé un questionnaire aux responsables de programmes LOLF – RPROG** qui sont les responsables de la conception et de la mise en œuvre des politiques publiques, de la mobilisation des ressources.

Le questionnaire a été adressé à 64 personnes, responsables de 111 programmes – soit la quasi-totalité des programmes- , la mission ayant écarté un tout petit nombre de programmes qui sont des « enveloppes budgétaires » plus que des « programmes ». La liste des programmes interrogés figure en annexe. Toutefois, les secteurs qui sont les moins bien représentés sont la sécurité intérieure et la Défense.

Parallèlement, **la mission a rencontré ou entendu 69 personnes au sein des services centraux de l'Etat et en région, des autorités administratives indépendantes et au sein de la DG Connect de la Commission européenne et a été éclairée par un exemple du secteur privé.**

Enfin, la mission a organisé des ateliers rassemblant des personnes des ministères<sup>2</sup>, issues d'horizons variés, services statistiques ministériels, Insee, direction des systèmes d'information, ou de services « métiers » impliqués dans le traitement des données. La liste des participants à ces réunions figure en annexe, chacun a témoigné d'une assiduité et d'une participation remarquables à ces travaux.

Les trois premiers ateliers ont été consacrés à la présentation :

- de réalisations ;
- des cadres juridiques ;
- d'un exemple de gestion des ressources humaines dans l'administration d'Etat (ASN) ;
- et d'un exemple externe chez Air France/KLM.

Le programme des ateliers figure en annexe. En complément des présentations, les auteurs ont rédigé des contributions qui sont incluses dans ce rapport, en ce qu'elles montrent la diversité et l'originalité de la démarche concernée.

Une séance -en groupe plus restreint- a ensuite été consacrée au travail de mise en commun de l'état des lieux et à préparer les recommandations qui ont ensuite fait l'objet d'un débat en séance plénière.

Le rapport présenté est le fruit de ces démarches croisées : réponses écrites aux questionnaires, ateliers et entretiens, ainsi que les contributions écrites relatives aux expériences et aux expertises.

**Par ailleurs, l'existence même de la mission a induit des réflexions et des premières décisions dans plusieurs administrations, premier résultat positif.**

---

<sup>2</sup> Services du Premier ministre, des ministères de l'éducation nationale de l'enseignement et de la recherche, de la justice, de la culture et communication, des ministères économiques et financiers, des affaires étrangères

## 2 LES RESPONSABLES LOLF PRIVILEGIENT UNE APPROCHE PAR L'EXPERIMENTATION DU « BIG DATA », JUGE PROMETTEUR, SANS AVOIR ENCORE, EN GENERAL DEFINI DE POLITIQUE OU D'ORGANISATION

L'éclairage des entretiens et des échanges au cours des ateliers ont permis notamment d'estimer que **les réponses écrites au questionnaire adressé aux responsables de programme LOLF sont dans leurs lignes de force globalement représentatives de l'état des lieux des administrations de l'État et de ses opérateurs** au regard du « big data » et de l'analytique<sup>3</sup>. Les réponses ont été validées le plus souvent au niveau de chefs de service ou adjoints de directeurs ou directeurs généraux.

Cependant, des expériences intéressantes, importantes ou instructives existent parmi des administrations qui n'ont pas répondu, dans certains cas de façon volontaire et assumée.

Les réponses ont été analysées suivant **leur contenu, en essayant de dégager des typologies ou une mesure du degré de maturité des services**. Dans l'analyse sémantique, ont été identifiés les sujets récurrents ou absents que la mission s'attendait à voir apparaître.

### **2.1 Le « big data » est généralement projeté comme prometteur dans l'attente des premiers retours d'expérience**

*Rappel de la question 1 : Le « big data » vous paraît-il prometteur pour votre programme ? Avez-vous commencé à l'intégrer ?*

Quasi-unaniment, les réponses considèrent que l'exploitation de la posture « big data » est prometteuse. Mais environ la moitié des services n'ont pas du tout commencé à aborder le « big data ».

Les raisons invoquées pour expliquer cette absence d'action sont les suivantes :

- Le service n'a pas trouvé de déclinaison opérationnelle ;
- Ce n'est pas une priorité ;
- Des obstacles l'en empêchent : respect de la vie privée, sécurité.

---

<sup>3</sup> L'analytique est l'application de l'informatique, de la recherche opérationnelle et de la statistique à la résolution de problèmes d'entreprise et industriels. Elle se réalise dans un système d'information. Discipline récente fortement liée à l'expansion de l'informatique, l'analytique s'est au début intéressée à l'usage des ordinateurs puis a évolué vers l'analyse des données. (source wikipédia)



Pour ceux qui ont commencé à l'intégrer dans leur démarche, les situations sont à des stades très différents d'avancement :

- un service a en cours d'établissement une feuille de route ;
- d'autres ont des expériences de *data-mining* ;
- d'autres encore collectent des données externes déjà traitées à partir de données brutes ;
- enfin d'autres ont des projets de collecte de données ou des preuves de concept en cours de conception ou de développement.

## **2.2 Les attentes sont variées, elles portent majoritairement sur l'amélioration de la gestion interne des services**

*Rappel de la question 2 : Quelles sont vos attentes vis-à-vis du « big data » (économie, qualité de service, efficience, pilotage de politiques publiques) ?*

Les attentes des directions de programme LOLF vis-à-vis du « big data » sont très majoritairement relatives à l'amélioration de la gestion interne des services :

- Amélioration du pilotage des politiques publiques ;
- Pilotage des contrôles et lutte contre la fraude, amélioration de l'analyse des risques ;
- Efficience des services ;
- Compréhension de phénomènes et amélioration de la qualité des données et de l'information ;
- Gestion des ressources humaines.

Mais d'autres attentes, minoritaires, sont exprimées portant sur l'identification grâce aux « big data » de meilleurs services et ou comme un moyen pour atteindre une meilleure qualité de service.

Enfin, dans un cas, la mise en place d'un projet pilote « big data » vise à permettre la dissémination de bonnes pratiques et favoriseraient la mobilisation de PME innovantes.

## **2.3 Les délais d'appropriation seront déterminés par des démarches d'apprentissage choisies progressives et situées dans la continuité**

*Rappel de la question 3 : Combien de temps vous donnez vous pour vous approprier le « big data » ? Le porter à maturité ? Avec quelles forces et faiblesses ?*

Les délais d'appropriation évoqués sont liés au choix de l'adoption d'une démarche progressive et continue d'apprentissage. Il n'y a nulle trace de stratégie de rupture, de démarche « séparée ».

Les répondants soulignent les difficultés qu'ils ont à maîtriser certains sujets, notamment :

- Les cadres juridiques ;
- Les délais pour disposer des autorisations ou des conventions d'accès aux données ;
- L'évolution rapide des technologies ;
- La complexité des méthodes d'analyse statistique ;
- L'évolution rapide des formats et des modèles de données.

Ils citent parmi leurs forces :

- les compétences métiers ;
- la culture de sécurité des systèmes d'information ;
- la pratique du croisement des données structurées.

Ils estiment que leurs faiblesses sont relatives :

- aux coûts d'infrastructures et aux limitations budgétaires ;
- à leur faible maturité de maîtrise des technologies exigées par le « big data » ;

à la rareté de leurs compétences sur le sujet. Pour corriger ces faiblesses et atteindre la maîtrise d'opérations industrialisées de « big data », la montée en compétence des équipes et un temps d'appropriation et de montée en compétence seront nécessaires.

En outre, les réponses n'argumentent généralement pas sur la justification d'économies.

#### ***2.4 Les échanges relatifs au sujet du « big data » avec des tiers sont déjà établis dans près de la moitié des cas avec les partenaires de proximité, l'ensemble des services exprimant un souhait d'échanges, unanimement privilégié entre pairs***

*Rappel de la question 4. : Echangez-vous ou aimeriez-vous échanger avec des pairs sur le sujet ?*

Les services qui échangent, partagent, interagissent au sujet du « big data » le font avec des interlocuteurs de proximité : principalement leurs opérateurs de l'Etat, plus rarement avec l'Administrateur général des données, le monde académique ou leurs correspondants étrangers.

Les souhaits d'échanges sont à ce stade assez peu précisés quant à leur champ.

L'échange avec des pairs paraît être la modalité privilégiée.

## **2.5 Le degré d'engagement et de maturité au regard de la réalisation de projets paraît très différencié, dégageant des pionniers et des volontaires**

*Rappel de la question 5. : Etes-vous prêts à vous lancer dans un projet de « big data » ?*

Les réponses font apparaître un degré d'engagement vis-à-vis de la réalisation de projets, expérimentaux ou opérationnels, très variable.

Les réponses ont été réparties sur quatre degrés d'engagement, en fonction de l'existence de projets et des perspectives déclarées de réalisations :

- Premier degré : environ 20 % des répondants n'ont pas de projet et n'ont pas d'appétence pour des projets.
- Deuxième degré : environ 50 % des répondants sont ou seraient prêts à entreprendre un projet, souvent sous diverses conditions : compétences humaines, budget, assouplissement du cadre juridique, etc.
- Troisième degré : 20 % des répondants ont un projet en cours : data mining, GED de masse, ...
- Quatrième degré : 10 % des répondants ont une démarche et des projets ou des réalisations propres ou partagés.

**Dans le cadre de cette situation hétérogène, la situation du groupe des répondants paraît favorable à un engagement collectif d'une partie des services avec :**

- des pionniers qui « avancent » ;
- un groupe significatif de volontaires – dont certains ont besoin d'être confortés – cf. les réponses aux autres questions.

## **2.6 Les services privilégient une approche par l'expérimentation, cohérente avec le choix d'un apprentissage progressif**

*Rappel de la question 6 : Quelles sont les questions que vous souhaitez aborder par le biais du « big data » ? Avec quelles postures d'expérimentation : lab, externalisation ?*

Les services les moins avancés dans la définition de leur approche n'ont pas encore défini de posture. En revanche, ceux qui ont déjà une réflexion souvent liée à la perspective d'un projet, souhaitent majoritairement effectuer une expérimentation dans un premier temps.

Une des réponses cite l'externalisation d'un projet, mais cette externalisation s'appuie sur un service de statistiques de l'Etat. L'idée d'une externalisation du « big data » n'est pas présente dans les réponses.

Une des réponses souligne le souhait de partage des bonnes pratiques – cf. les réponses aux autres questions.

## **2.7 La gouvernance et l'organisation autour du « big data » reste à définir dans les trois quarts des cas et à préciser dans les autres**

*Rappel de la question 7 : Avez-vous nommé un (ou des) responsable(s) de la politique « big data » ? de chacun des projets « big data » ?*

Par ailleurs, dans les autres cas, la gouvernance et l'organisation sont le plus souvent présentées comme provisoires, voire lacunaires, répondant aux besoins « du moment ».

L'aspect gouvernance du « big data » – et des données en général – est un champ d'études et de décisions pour les services, champ jugé prioritaire.

L'idée de doter le « big data » d'une gouvernance et/ou d'une organisation séparée de la data n'est pas présente dans les réponses.

## **2.8 La moitié des services ont des attentes vis-à-vis des services du Premier ministre, corrélées à leur degré d'appropriation du « big data » : évolution du cadre juridique et appui politique pour les plus avancés, attentes d'appuis « opérationnels » pour les autres**

*Rappel de la question 8 : Qu'attendez-vous des services du Premier ministre dans le cadre de la politique « big data » ? De la coopération interministérielle ? Européenne ou internationale ?*

La moitié des services n'ont pas d'attente vis-à-vis des services du Premier ministre, on retrouve parmi eux principalement ceux qui n'ont pas de projets – logiquement, et des services qui se disent prêts à des projets, sans donner plus de précisions – cf. question 5, le premier degré et la majorité du deuxième degré.

**L'autre moitié des services expriment des attentes** ; ils peuvent se diviser en deux groupes dont les attentes ne se superposent pas :

- **ceux qui attendent des « services », des « prestations » des services du Premier ministre.** Ce sont ceux du troisième degré de la question 5 et certains du deuxième degré.

Ces attentes sont variées pour chacun deux, chaque service citant en général deux ou trois attentes.

Sans les hiérarchiser, en raison du petit nombre de réponses pour chaque thème (3 ou 4 au total), on citera les 5 thèmes suivants :

- un appui d'analyse juridique, le cas échéant éthique ;
- un appui méthodologique ;
- la formation, la gestion des ressources humaines et le recrutement ;
- l'échange de bonnes pratiques et leur capitalisation ;
- la technologie : les sujets du stockage, de l'analytique et des moteurs de recherche sont cités chacun une ou deux fois.

- **deux services attendent un appui politique et une évolution du cadre juridique.** Ce sont les deux services qui sont engagés dans une approche globale et multi-projets du « big data » – degré 4 de la question 5.

## ***2.9 L'analyse sémantique des réponses montre que les services de l'Etat disposent d'une culture commune qui n'est pas exempte d'angles morts mais qui est propice à des innovations et des collaborations***

A la lecture globale des réponses, certains mots marquent par leur fréquence, à titre d'exemple : données personnelles, vie privée, compétences alors que d'autres mots apparaissent absents, économie, efficacité et efficience, Europe, collectivités territoriales, services territoriaux de l'Etat notamment.

Il est donc apparu nécessaire de mener une analyse sémantique de l'ensemble des réponses pour rechercher les mots qui n'apparaissent pas, alors que la mission pensait les rencontrer, et les mots qui apparaissent souvent, notamment les mots présents et répétés dans un nombre important de réponses.

Cette approche complète et éclaire l'analyse des réponses du paragraphe précédent.

### 2.9.1 Les réponses ne font pas apparaître certains aspects du sujet : économies (faire des), efficacité, efficience, emploi, innovation, Europe, international, collectivités territoriales, services territoriaux de l'Etat

Les réponses ne font pas apparaître des anticipations ou des interrogations relatives aux économies budgétaires pour les programmes LOLF ou pour les opérateurs de l'Etat. Cette absence peut refléter tant la prudence des répondants que le caractère prématuré de cette question au regard de la maturité réelle du « big data » pour les administrations.

Néanmoins, ce « blanc » surprend au regard d'annonces, il est vrai sans démonstrations appuyées sur des constatations de résultats ex-post, faites par des cabinets d'analystes qui annoncent des bénéfices substantiels pour les économies des Etats de l'Union européenne<sup>4</sup>.

*L'intégration de cette interrogation relative aux économies budgétaires dans la démarche d'appropriation du « big data » et de l'analytique est sans doute une priorité, sans préjuger de ce que sera l'ampleur de l'apport du « big data » aux économies budgétaires.*

Collectivités territoriales, services territoriaux de l'Etat sont des idées absentes alors que le mot Europe est rare, employé par des services dont l'activité a par nature une dimension européenne.

*Compte-tenu de la volonté d'ouverture, d'échange, le cas échéant de mutualisation, des répondants, des projets des institutions européennes et de l'existence de ressources en termes de budgets et d'outils dans des programmes européens, il sera utile d'intégrer un volet Europe dans les feuilles de routes des différents services.*

Le sujet des relations avec les collectivités territoriales sera abordé dans la deuxième phase de la mission.

Le mot emploi n'apparaît pas, au sens de la contribution éventuelle à la création d'emplois hors administrations, emplois induits par les effets des réalisations des administrations.

Enfin, le mot innovation est lui aussi quasi-absent des réponses.

---

<sup>4</sup> <http://ec.europa.eu/digital-agenda/en/what-big-data-can-do-you>

### 2.9.2 Les réponses font apparaître un vocabulaire commun propre à servir de socle au travail entre pairs : compétences, données personnelles/vie privée et cadre juridique, sécurité.

Les mots qui désignent les besoins de compétences humaines et ceux qui désignent le cadre juridique, entendu principalement au sens du respect des règles relatives au respect de la vie privée et des données personnelles se détachent nettement en tête.

Les besoins de compétences humaines sont l'objet d'interrogations quant à leurs caractéristiques et, souvent, quant à la capacité des services à les mobiliser – en interne ou externe.

Le respect de la vie privée et des données personnelles est intégré dans les visions, parfois avec l'ambition de l'adapter. D'autres sujets juridiques importants pour le « big data » – par exemple propriété intellectuelle, responsabilités, ne sont pas abordés dans les réponses écrites.

Les mots relatifs à la sécurité sont présents deux fois moins souvent. Ils sont employés de façon générale bien que les services fassent apparaître leur compétence en matière de sécurité comme l'un de leurs points forts.

*Les compétences humaines, et la gestion des ressources humaines, les cadres juridiques des données, dans leur multiplicité et leur complexité, la sécurité constituent un vocabulaire commun aux services pour poser et construire un travail interministériel.*

## 3 LES DONNEES DISPONIBLES AU PLAN LOCAL ET DE PROVENANCES TRES VARIEES AINSI QUE LES PROBLEMATIQUES REGIONALES CONSTITUENT UN CREUSET PRECIEUX POUR ALIMENTER LES POLITIQUES PUBLIQUES AU NIVEAU LOCAL

Au contact direct des réalités du terrain et de la diversité des contextes locaux, les services territoriaux de l'Etat baignent dans une profusion de données de toutes natures, provenant d'acteurs très variés, public ou privés. Ce contexte est très propice au « big data » comme vecteur d'amélioration et de réalisation des politiques publiques, tant en matière économique, environnemental, santé et éducation, que sécurité intérieure ou lutte contre la fraude. Une attention particulière devrait ainsi être portée au développement d'initiatives locales en la matière.

L'exemple de l'Alsace illustre bien ces enjeux potentiels, en attendant leur réalisation plus complète et prometteuse.

Le Préfet de la région Alsace, préfet du Bas-Rhin : la vision territoriale

Définir précisément la notion de « big data », à l'heure actuelle, demeure complexe, et doit conduire à préciser ou expliciter la perception des acteurs territoriaux, dans un contexte évolutif où les objets connectés, le « quantified self » et autres évolutions numériques impactent les structures et les individus.

Les coûts encore mal connus, la faible maturité des technologies, des gouvernances variables et pas toujours stabilisées en matière de collecte et de gestion des données constituent les premiers points de

diagnostic, face à des attentes portant en priorité sur l'amélioration des connaissances et des services rendus. L'utilisation libre des données et donc leur mise à disposition est autant un sujet éthique et juridique que technique.

Pour ce qui concerne les données de l'État, un sentiment dominant au plan local est que la décentralisation en a amoindri le potentiel. Ces données peuvent ainsi sembler aujourd'hui moins disponibles ou plus dispersées que par le passé (par exemple, visibilité globale sur les investissements en infrastructures).

Des bases de données existent mais ne sont pas nécessairement complètement exploitées ou coordonnées. De multiples croisements engagés ou potentiels sont identifiés : données sur les déchets (ADEME, DREAL) pour l'économie circulaire, lien entre problématiques d'endettement et expulsions, entre autorisations d'urbanisme et risques naturels, croisement des implantations urbaines et des données santé environnement, gestion des ressources à l'échelle territoriale (exemple de la filière bois à l'échelle de la région ACAL).

De nombreuses données individuelles et nominatives (données du Rectorat, VAE, apprentissage) présentent un intérêt affirmé en matière d'emploi, des précautions s'imposant pour garantir leur anonymisation.

L'importance de capitaliser sur les données des plates-formes d'information géographique existantes est soulignée.

Ces constats et analyses amènent à envisager la re-construction d'un ensemble global de données accessibles, couvrant l'ensemble des acteurs Etat, incluant notamment le Rectorat, l'ARS, et débouchent sur le besoin identifié d'un « *cloud Etat* »).

L'accès aux données des collectivités et autres acteurs locaux, leur partage et leur mutualisation, constituent l'autre enjeu majeur au niveau territorial pour le développement du « *big data* ».

La réforme territoriale introduit une dimension supplémentaire, avec la nécessité de mutualiser à court terme les données des régions fusionnées.

En Alsace, le projet Alsacalcul porté par l'Université a suscité des discussions sur les possibilités de mutualisation dans le data center (mésocentre). D'autres projets de mutualisation de données au niveau régional (par exemple le projet du SERTIT de production de données satellite, en lien avec l'entreprise de matériel agricole Kuhn) pourraient constituer des voies d'approfondissement.

Enfin, la dimension transfrontalière constitue un axe de réflexion et de développement. De premiers exemples d'exploitation de données transfrontalières ont été mis en place, ainsi le système d'information géographique du Rhin supérieur (SIGRS) qui associe déjà l'Alsace, les cantons de la Suisse alémanique et une partie des länder allemands voisins, ou le projet Alsacalcul déjà évoqué qui a vocation à intéresser les pays frontaliers.

Face à ces besoins de mutualisation des outils État et collectivités, et d'ouverture vers les entreprises, un cadre bien défini de mise en place et de gestion du « *big data* » apparaît donc souhaitable pour nombre d'acteurs. En particulier, il devra aborder les aspects juridiques de la propriété, de la mobilisation et du partage de ces données, et de leur sécurisation.

Une orientation précise du niveau national pour déclencher des expérimentations locales, en identifiant les marges de manœuvre possibles, pourrait constituer une première étape.

Des sujets prégnants comme la lutte contre la fraude ou la délinquance, l'accélération des projets de construction ou de développement d'entreprises pour leur faire gagner du temps, l'amélioration des échanges avec le secteur de la recherche, et les échanges transfrontaliers de données, pourraient être de premières pistes de travail au niveau régional.



#### 4 LES REALISATIONS DANS DIVERS DOMAINES DES MINISTÈRES, ENCORE PEU NOMBREUSES, S'APPUIENT SUR DES RESSOURCES INTERNES ET MONTRENT UN ESPRIT D'INNOVATION REEL

La mission a identifié, sur la base des réponses à son questionnaire qui couvrent environ le quart des programmes, **des réalisations remarquables qui illustrent les thématiques au fondement du « big data » aujourd'hui dans les ministères**, particulièrement :

- le recours à des tierces données ;
- la refondation ou l'évolution des métiers et des processus ;
- la constitution et le positionnement d'équipes pluridisciplinaires innovantes dans un contexte administratif ;
- l'insertion des activités dans des cadres juridiques multiples, complexes et évolutifs ;
- la mobilisation et la formation des talents et des compétences internes ;
- l'appui sur des start-ups ;
- l'émergence de la gouvernance collaborative.

Il paraît utile d'en présenter cinq plus précisément dans le but de montrer comment :

- l'expérience du recours aux données de caisse pour élaborer l'indice des prix de l'Insee illustre l'intérêt, les conditions et les limites du recours à des tierces données ;
- Tracfin et ses équipes refondent l'ensemble de leurs activités sur une stratégie de projet collaborative, itérative et une gestion du changement forte ;
- le projet interministériel de « big data » collaboratif VITAM se distingue par une politique de réutilisation et une gouvernance partagée ;
- La mission pluridisciplinaire de lutte contre la fraude de la DGFIP s'articule étroitement avec les services de contrôle ;
- Pour la DGDDI, le développement du « datamining » s'appuie sur des relations étroites avec le terrain

Par ailleurs, **l'exposé des politiques d'AirFrance/KLM et de BPIFrance** montre, dans un contexte non-administratif, les enjeux et les solutions attachés à la mobilisation du socle des activités, l'alliage des compétences technologiques et métiers, le travail avec des start-ups.

#### 4.1 *L'expérience du recours aux données de caisse pour élaborer l'indice des prix de l'Insee illustre l'intérêt, les conditions et les limites du recours à des tierces données*



Institut national de la statistique  
et des études économiques  
Mesurer pour comprendre

##### Données de caisse Insee

Afin de produire l'indice des prix à la consommation (IPC), un réseau de 200 enquêteurs, animé par neuf sites prix en régions (environ 75 personnes) et la division des prix à la direction générale (20 personnes), réalise chaque mois environ 180 000 relevés de prix dans 27 000 points de vente situés dans 104 agglomérations de plus de 2 000 habitants de la métropole et des DOM.

En plus du besoin du calcul de l'évolution de l'indice des prix, d'autres besoins sont apparus, notamment celui des comparaisons des niveaux de prix entre territoires (entre la France métropolitaine et l'outre-mer ou entre pays de l'Union européenne comme le prévoit la Commission européenne pour mesurer le bon fonctionnement du grand marché intérieur) ou celui du suivi de marchés très spécifiques (produits éco-labellisés notamment).

Ces nouveaux besoins d'une part, et l'intérêt manifesté par la Commission européenne (Eurostat) pour la nouvelle source de données que représentent les données de caisses des enseignes de la distribution pour les statistiques de prix d'autre part, ont décidé l'Insee à lancer, à la fin de l'année 2009, un projet d'exploitation de ces données. L'accès aux données de caisse est aujourd'hui un moyen d'améliorer la qualité de l'indice produit. Il doit à terme permettre le calcul d'indice directement à partir des données fournies par les enseignes, pour le champ des produits concernés. Ce projet s'inscrit dans une modernisation plus large de l'IPC qui inclut notamment la refonte de l'échantillon d'agglomérations, un renforcement du suivi de la vente par internet et une évolution des processus de collecte (notamment du matériel de collecte).

Le projet vise à calculer des indices de prix en reprenant la méthodologie actuelle de l'IPC (indice de Laspeyres chaîné annuellement). Ces indices seraient fondés sur un échantillon de produits (EAN<sup>5</sup> x magasin), chaque produit pesant dans l'indice d'ensemble en proportion de son poids dans le chiffre d'affaires total observé sur l'ensemble de l'année précédente. Les données de caisse exploitées seront celles relatives aux biens de consommation courante vendus dans les super et hypermarchés, c'est-à-dire principalement les produits alimentaires (hors produits frais), les produits d'hygiène et les produits d'entretien courant de la maison. Leur champ fait actuellement l'objet de 35 000 relevés par mois. A l'avenir, les données des supérettes pourraient faire l'objet d'une même exploitation.

Les travaux méthodologiques, qui ont pu débiter grâce à la mise en place d'un groupe de travail entre l'Insee et la distribution, démontrent la faisabilité statistique du projet : la qualité des données reçues a été validée et les travaux de simulation d'indices sont concluants pour leur précision et leur comparaison avec les indices actuels.

Dans un second temps, les travaux ont été menés sur les données de caisse quotidiennes transmises par plusieurs enseignes depuis novembre 2012. Ces données sont transmises à l'Insee à titre gracieux dans le

<sup>5</sup> European article number, c'est-à-dire un code barre

cadre de conventions triennales conclues avec les enseignes volontaires pour ce deuxième temps. Ces conventions limitent l'utilisation des données à l'élaboration d'un indice de prix.

L'étude préalable du projet s'est conclue à la fin du premier trimestre 2015. Cette étude préalable a permis à l'Insee d'affiner la méthodologie statistique de calcul d'indice à partir des données quotidiennes, de construire un processus de traitement des données et de dessiner l'infrastructure informatique nécessaire à sa réalisation (combinaison de technologies SGBDR<sup>6</sup> et « big data »).

L'Insee a également réfléchi collectivement au statut des informations numériques du secteur privé, réflexion qui dépasse le champ du projet et s'inscrit désormais dans le cadre du groupe de travail ad hoc créée au sein du Cnis<sup>7</sup> et de l'Insee.

---

<sup>6</sup> SGBRD : système de gestion de base de données relationnelles

<sup>7</sup> CNIS : Conseil national de l'information statistique

## 4.2 Tracfin et ses équipes refondent l'ensemble de leurs activités sur une stratégie de projet collaborative, itérative et une gestion du changement forte



Meilleures pratiques pour le « big data » et

l'analytique.

### Le contexte Tracfin

Les missions de Tracfin ne consistent pas en la détection de comportements atypiques parmi une masse de données homogènes.

La base de travail repose sur un flux d'informations diverses et très hétérogènes qui enrichissent des analyses basées sur des déclarations ou informations de soupçons reçues.

Le flux de données reçues par Tracfin connaît un fort accroissement ces dernières années. L'introduction des Communications Systématiques d'Informations concourt également à la constitution d'un stock de donnée volumineux.

Le besoin de Tracfin repose sur une capacité d'apprentissage de son stock pour faciliter la priorisation des déclarations de soupçons, optimiser la capacité d'enrichissement et la qualité des externalisations effectuées par Tracfin et réévaluer en continu son stock pour développer des thématiques stratégiques.

Le service Tracfin est composé de 112 agents.

### Le support technologique

Progiciel interne et logiciel libre : Tracfin a fait le choix dans les années 2006 – 2007 de produire son propre Système d'Information adapté à ses besoins

Aujourd'hui la volonté est de construire un dispositif basé sur des technologies issues du logiciel libre (univers Hadoop) en assurant l'intégration totale du dispositif et en développant à façon les interfaces.

Il y a peu de retours d'expérience significatifs

Technologie naissante orientée vers le marketing : le benchmarking réalisé par Tracfin sur le choix des technologies pour des organismes éprouvant des besoins similaires s'est relevé peu significatif. La plupart des outils adresse soit des mécanismes de détection parmi un stock volumineux d'informations homogènes, soit des outils de connaissance client orienté vers le marketing.

### Une expérience unique basée sur l'analyse de la valeur

Les différentes structures ayant mis en œuvre des dispositifs basés sur de l'analytique ou des traitements « big data » ont réalisé des prototypes en marge de leur SI actuel sur des besoins métier isolés permettant de rapidement dégager vis-à-vis des utilisateurs, la valeur du patrimoine informationnel dont ils disposent.

Aucune expérience similaire à TRACFIN de refonte complète du SI basée sur ce type de technologie n'a été portée à notre connaissance.

La logique retenue est donc éloignée des standards habituels.

### Des préceptes CNIL à s'approprier

Le principe fort de la loi Informatique et liberté de 1978 instituant que les données doivent être collectées et traitées selon les modalités et les finalités fixées a priori, s'oppose de fait à la logique analytique à déduire une stratégie et des objectifs à partir du stock d'informations collectées. Ce type de projet doit donc être accompagné en amont avec la direction conformité de la CNIL pour mesurer les risques intrinsèques de croisement massif des données et de définir les meilleures modalités de maîtrise des libertés publiques et individuelles.

#### Des conceptions projets risqués

La philosophie de ce type de projet se base sur une intuition de résultat positif (voir un engagement agnostique) et non sur une analyse de la valeur a priori. Dès lors, ce type de projet constitue nécessairement un risque et nécessite d'adapter les études d'opportunité (ROI, objectifs à atteindre). Cela rend délicat le dialogue avec les sponsors pour qui l'investissement doit être cadré.

#### Des ressources spécifiques

La mise en œuvre de projet d'analytique ou de « big data » nécessite une approche moins conventionnelle du projet informatique. Cela nécessite généralement une adaptation des ressources humaines (formation spécifique, coaching) ou un recrutement d'experts de ce type de traitement (ressource rare et coûteuse).

Le recours à la prestation externalisée rencontre la même problématique de spécificités des ressources. La plupart des grands industriels habituellement positionnés sur les marchés publics ne disposent, pas encore du moins, des ressources adaptées et font appel à une sous-traitance coûteuse refacturée pour l'administration.

Les sociétés spécialisées sont généralement de taille réduite (PME ou TPE) et ne sont pas forcément adaptées ou privilégiées dans les mécanismes d'appel d'offre des marchés publics.

Les solutions par marché multi-éditeurs restent l'option la moins onéreuse mais nécessitent généralement des discussions complexes pour intégrer les modèles économiques proposés.

Les solutions déployées supposent également une sensibilisation forte des utilisateurs in fine et imposent généralement un bilan de compétence et une réorientation des missions et des objectifs des agents impliqués.

#### La stratégie de projet

TRACFIN a opté dans le cadre de son projet de refonte sur une stratégie de projet itérative. De nombreux démonstrateurs ont été lancés pour éviter une stratégie de big bang, l'objectif est de faciliter la transition et éviter ainsi le désastre industriel.

Le découpage en sous-projets correspond à la réalisation de trois sous-ensembles techniques distincts en forte interface.

La coordination transversale et de la gestion des interfaces internes/externes sont internalisées.

Un expert au profil de « data scientist » a été recruté, il a reçu un double positionnement Direction/Equipe projet.

#### La conduite du changement

La part du projet dédiée à la conduite du changement est importante (25% du financement et des ressources)

Une équipe industrielle d'AMOA est intégrée dans l'équipe projet.

Par ailleurs, le projet se caractérise par la mise en place de comités de terminologie, la constitution d'un groupe de référents issus des départements opérationnels, mobilisables à temps partiel au profit du projet (expression de besoins, recette) et la recette continue des livraisons, afin de s'assurer d'une implication permanente des utilisateurs finaux et de la hiérarchie en vue de garantir d'une part, l'adéquation des besoins en analyse et d'autre part, la bonne prise en compte par la structure des résultats/recommandations issus des traitements algorithmiques.

#### Des préconisations de l'équipe projet de Tracfin pour une généralisation étatique.

Ce type de besoin et de technologie rompt avec le processus classique de déroulement d'un projet SI dans l'administration (cycle en « V »), nous souhaiterions donc : un guide méthodologique voire une directive interministérielle pour fixer la méthodologie d'un tel projet en mettant en avant une logique de développement « en spirale », voir pérenniser et renforcer autour de l'entité interministérielle ETALAB un centre d'expertise de la « data science » susceptible d'appuyer les nouveaux projets, la mise en œuvre d'une filière de formation interministérielle autour des métiers et technologies de « big data ».

### 4.3 Le projet interministériel de « big data » collaboratif VITAM se distingue par une politique de réutilisation et une gouvernance partagée pluriministérielle

VITAM, un projet « big data » collaboratif

Il y a peu de temps encore, toutes les démarches et décisions publiques étaient sous forme papier, domaine pour lequel l'administration a une expertise certaine. La conservation de ces documents papier a permis jusqu'à présent d'assurer la continuité de l'État. On ne citera que certains domaines où l'importance pour le citoyen et pour l'État de cette conservation est manifeste : l'état-civil, le cadastre, l'urbanisme, les décisions de justice, la gestion des dossiers individuels et de la paie des personnels civils et militaires, la correspondance diplomatique, les opérations militaires,...

Dans la dynamique actuelle de développement numérique, de plus en plus de décisions concernant directement les citoyens et plus largement toutes les actions de l'État n'existent plus que sous forme numérique. Par conséquent, il est impératif d'accompagner la transition numérique de l'État voulue par le gouvernement par la garantie de sa continuité numérique.

Jusqu'à présent les acteurs du secteur public ont utilisé des techniques de conservation classiques, utilisables pour les volumes modérés qu'ils ont eu à gérer. Il faut désormais faire face à la croissance exponentielle du numérique.

Dans ce contexte, deux nécessités font loi :

- il faut, grâce à un saut technologique, se mettre en capacité de traiter les milliards d'objets de toutes natures qui devront être conservés. Le passage aux technologies du « big data », associées à une approche sémantique sur une base cloud, est nécessaire pour relever le défi que représentent les 3 V : volumétrie et variabilité des données et de leurs métadonnées, et vitesse d'accès ;
- il faut apporter une réponse collective à un problème commun, en évitant la dispersion des énergies et des compétences. La préservation numérique de masse est un sujet pour tous les grands acteurs publics, qui commencent pour beaucoup à chercher des solutions. Le moment est venu pour la réutilisation large d'une solution commune.

Trois ministères (affaires étrangères et développement international, culture et communication, défense), ayant mission d'archivage définitif et l'expertise archivistique associée, ont décidé d'unir leurs efforts, sous le pilotage de la DISIC, pour développer une brique logicielle réutilisable dans des contextes différents permettant la conservation et la consultation de très gros volumes de documents numériques dans le respect des obligations légales et réglementaires. Cette brique constituera le cœur de leurs propres projets de développement de services d'archivage numérique, mais pourra aussi s'intégrer à tous les systèmes d'information des acteurs publics de l'État, dont plusieurs ministères qui ont déjà exprimé leur intérêt, ou les collectivités territoriales qui le souhaiteront.

S'insérant dans la logique du PIA, le programme VITAM associe transition numérique de l'État et développement d'activité économique. Il s'appuiera sur des PME innovantes, dont l'expertise sera précieuse pour la réussite du programme. De nombreux acteurs du monde économique, grands groupes bien sûr mais aussi PME, suivent d'ailleurs avec attention le lancement du programme.

Ce programme, lancé le 9 mars 2015, sera mené sur un mode agile pour assurer l'adéquation au besoin et permettre des livraisons rapides et régulières, avec enrichissement fonctionnel, sur sa durée de 5 ans. Outre les cycles courts propres à l'agilité, il est prévu comme grands jalons une première version d'intégration diffusée en 2016, et une première version de production mise à disposition mi-2017, puis une version majeure tous les ans.

L'association de nombreux acteurs pour en faciliter la réutilisation sera une constante du programme. Les travaux préparatoires, commencés depuis 2013, notamment sur les problématiques d'archivage des contenus numériques les plus foisonnants (messageries, outils collaboratifs), ont déjà associé de nombreux acteurs du monde des archives au niveau de l'État, des collectivités territoriales et des établissements publics.

#### **4.4 La mission pluridisciplinaire de lutte contre la fraude de la DGFIP s'articule étroitement avec les services de contrôle**



Une nouvelle mission pour une meilleure détection de la fraude fiscale: le datamining

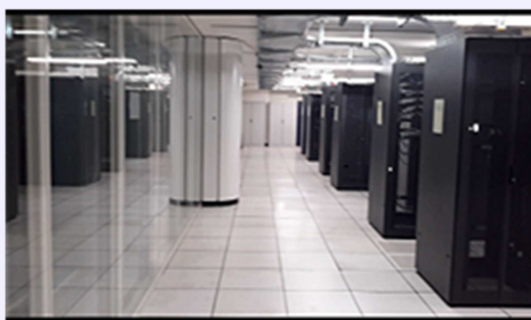
La lutte contre la fraude fiscale est une priorité absolue de la DGFIP. Pour mieux repérer la fraude, une mission a été mise en place à l'été 2013.

Placée sous la double responsabilité du chef du service du contrôle fiscal et du chef de service informatique, elle met en œuvre un nouveau processus de détection de la fraude fiscale, au travers d'une méthode d'exploration des données appelée « datamining ».

Le datamining consiste à explorer un très grand nombre de données aux fins d'y repérer des profils de fraude.

Les travaux ont débuté par un repérage de la fraude en matière de remboursements de crédits de TVA et se poursuivent désormais sur la TVA de façon plus générale.

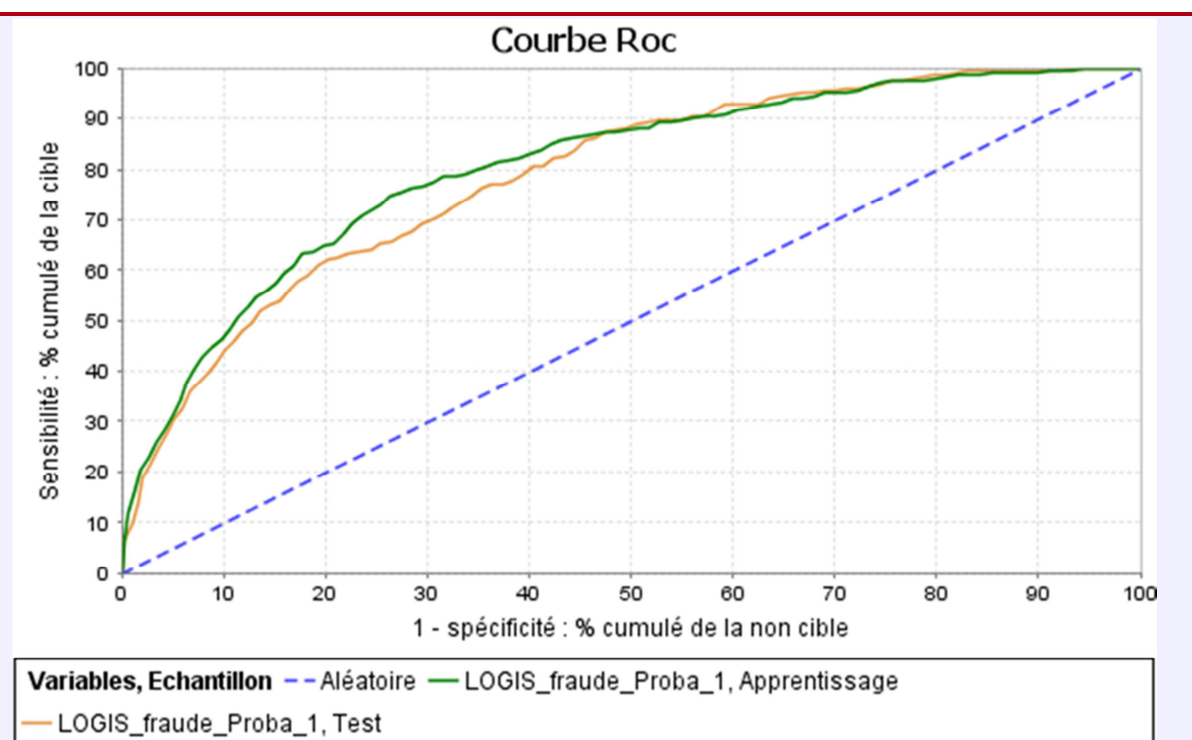
L'expérimentation, validée par la CNIL en janvier 2014, utilise les données des applications métiers de la DGFIP qui, une fois mises à disposition sur une plate-forme dédiée et sécurisée, exploitée totalement en interne par les personnes du service des systèmes d'information, sont analysées et étudiées par des statisticiens recrutés spécifiquement par la mission.



Crédit photo DGFIP

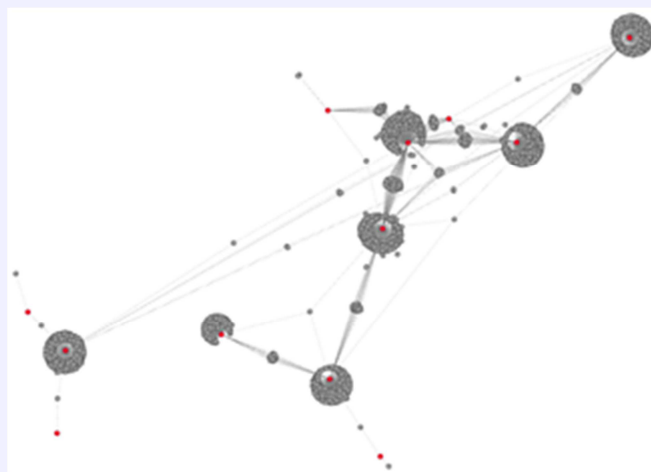
L'intérêt est d'avoir désormais pour l'exploration de ces données une approche statistique qui permet de révéler des corrélations jusqu'alors non connues du métier.....





Source DGFIP

....mais également une visualisation de liens entre entités. Cette nouvelle offre va permettre aux services de connaître des relations entre différentes données d'identification pour une meilleure connaissance du tissu fiscal.



Une expertise métier est cependant indispensable sur les résultats issus des algorithmes pour n'en retenir que les plus pertinents au regard des enjeux ou des pistes étudiées et ceux qui peuvent le plus possible mettre en évidence une fraude.

Le datamining est en effet une démarche méthodologique rigoureuse, développée en vue de « faire parler » l'information contenue dans les systèmes informatiques, en mettant en exergue d'éventuelles relations entre les données observées.

Des premiers contrôles ont été engagés sur un échantillon d'entreprises réparties sur la quasi-totalité des directions du réseau et les résultats sont prometteurs.

La mission, d'une dizaine de personnes, présente la caractéristique d'être très majoritairement composée de ressources internes (7 agents DGFIP, 2 agents de l'Insee et une universitaire diplômée en MASTER 2 datamining). Cette équipe bénéficie ponctuellement d'un transfert de compétence par un spécialiste.

La quantité, la variété et des nouveaux besoins de visualisation des données, entre pleinement dans le cadre du « big data ».

L'explosion quantitative (parfois redondante) et la nécessité de pouvoir l'analyser dans un processus totalement sécurisé à la fois techniquement mais également juridiquement (sous la vigilance de la CNIL notamment), rend nécessaire la constitution de ce type de structure et son fonctionnement en mode agile. Echanges, restitutions rapides et lisibles, nouvelle façon de penser quant à l'exploration des données sont les objectifs de la mission.

Mission requêtes et valorisation

DGFIP

#### **4.5 Le succès du « big data » à la direction générale des douanes et des droits indirects repose sur la qualité des relations avec le terrain**

Le datamining à la DGDDI, les relations avec les services de terrain par Damien THILLOU

L'expérimentation datamining a été officiellement lancée mi-2014 au sein de la DGDDI par la création d'un chargé de mission en charge de son développement au sein de l'administration, avec un objectif à terme d'une cellule dédiée à cette activité innovante dans le cadre de la création d'un service national d'analyse de risque et de ciblage.

L'objectif de cette nouvelle activité est de valoriser la dématérialisation des documents douaniers pour en faire une exploitation informatique à des fins de lutte contre la fraude. Si les données accessibles aujourd'hui à la cellule sont loin d'être exhaustifs, cela permet de lancer de premières expérimentations pour éprouver ce genre de méthodes.

C'est un nouvel outil au sein de l'administration, qui a pour vocation à apporter un soutien opérationnel et un complément aux activités de ciblage traditionnelles déjà industrialisées (« règles » métiers introduites dans le système électronique de ciblage). Cette nouvelle activité repose sur trois piliers sans lesquels les retombées opérationnelles ne pourront avoir lieu : un outillage et/ou investissement informatique bien entendu, des profils de compétences adaptés à l'analyse de données, mais aussi et surtout une relation forte avec les services opérationnels, et c'est sur ce dernier pilier que nous allons à présent insister.

On peut grossièrement découper les services métiers en deux ensembles : les services de ciblage, et les services de contrôle sur le terrain, les deux sont essentiels au succès d'une nouvelle activité comme le data-mining : sans les premiers, pas de participation essentielle au développement ou à l'affinage des analyses que peut produire le datamining au travers de leur apport « métier » ; sans les seconds, pas de contrôle sur le terrain pour confirmer ou infirmer les analyses faites, et in fine, obtenir ce pour quoi cette nouvelle activité se développe : détecter de nouveaux courants de fraudes et améliorer le ciblage.

Ainsi l'acceptabilité de ce genre de nouvelles méthodologies par les services est essentielle pour emporter l'adhésion de ces derniers et pouvoir apporter un soutien concret par cette nouvelle activité au ciblage douanier.

Cette acceptabilité se construit et s'entretient sur le long terme par différents axes de travail :

- Maintenir un effort de communication auprès notamment des services de ciblage pour démystifier l'activité et bien rappeler les moyens et enjeux : la machine ne remplacera pas l'humain – sujet assez sensible dans un contexte préexistant d'augmentation de prescriptions de contrôles entre services, exposer l'aspect complémentaire de ce genre de méthode, ainsi que son aspect soutien (effectuer beaucoup plus rapidement certaines tâches basiques d'analyses), et enfin rappeler l'importance de l'apport de leurs connaissances métier dans le processus d'analyse des données ;
- Commencer « petit », en privilégiant un nombre de signalements très réduit mais très méticuleusement choisi afin d'avoir des retombées positives rapides permettant de susciter l'adhésion des services et enclencher un cercle vertueux ;
- N'utiliser, au moins dans un premier temps, que des analyses qui sont facilement communicables et compréhensibles dans leur angle d'attaque ou leur esprit, éviter des analyses boîtes noires (réseaux de neurones, etc.) qui ne feront qu'obtenir une méfiance des services et perdra l'avantage des analyses interprétables de pouvoir guider les enquêteurs lors des contrôles ;
- Enfin, présenter aux agents de ciblage différentes pistes d'analyses, achevées ou en cours, afin leur montrer les bénéfices potentiels et susciter une émulation positive au sein de ces services.

#### 4.6 AF/KLM associe des compétences technologiques et fonctionnelles pour le « big data »

Le « big data » requiert un mélange de compétences technologiques et fonctionnelles

Le « big data » nécessite de nouvelles compétences sur des technologies émergentes, mais il s'agit d'une évolution qui s'inscrit dans la continuité de l'histoire, du savoir-faire et des missions du Systèmes d'Information. Chez AF/KLM, l'option retenue a été de s'appuyer en grande partie sur l'évolution des personnes déjà présentes dans l'organisation.

Concernant l'axe fonctionnel, nos « data scientis » ont été identifiés dans les équipes de Recherche opérationnelle et travaillent en binômes avec les équipes métier lorsque l'expertise sur les données est requise. C'est une condition nécessaire pour générer de la valeur pour le Groupe. Enfin, il est également essentiel de savoir présenter les résultats pour qu'ils soient compris et exploitables.

Une vision du « big data »

« Le « big data » est un ensemble de nouvelles technologies permettant d'améliorer considérablement les performances de traitement de grands volumes de données, structurées ou non, à un coût raisonnable »

Le choix d'Air France-KLM s'est porté en priorité sur un framework HADOOP<sup>8</sup> et tout un écosystème autour de cette plate-forme : bases de données non structurées (NoSQL), utilisation massivement distribuée de la mémoire des serveurs (in-memory computing), traitement temps réel d'importants flux de données (complex event processing), visualisation des données.

Puis des techniques d'analytique et de statistique ont été mises en œuvre pour analyser ces données stockées de façon brute, sans pré-traitement, dans un unique espace de stockage, le data-lake.

Il a fallu créer de l'appétit pour l'ouverture sur les données extérieures à l'entreprise

De nombreux champs d'application potentiels

Une fois le dispositif en place, de nombreux champs d'application sont apparus, et d'autres apparaissent encore : une vue du client à 360° (au-delà de « flying blue »), un ciblage des offres grâce à l'analyse du comportement des internautes, une analyse prédictive des pannes ou de la maintenance.

La réussite opérationnelle du projet KARMA de revenue management.

Ce traitement quotidien sur la base de deux années d'historique de données en tout genre fournit les éléments utiles à une centaine d'analystes qui interprètent les résultats et valident ou non la stratégie quotidienne. L'implication de l'humain dans la boucle est le point le plus délicat, car il faut que les analystes s'approprient le système et lui fassent confiance : le « big data » perturbe la manière de fonctionner des personnes !

---

<sup>8</sup> HADOOP est un framework libre pour le développement d'application « big data »

#### 4.7 Le HUB de BPI France décloisonne l'innovation

##### BPIFRANCE LE HUB : NOUS DECLOISONNONS L'INNOVATION

Bpifrance « Le Hub » est une nouvelle offre de services dédiée aux Grandes Entreprises et aux Entreprises de Taille Intermédiaire.

Nous capitalisons sur notre connaissance fine et large du tissu entrepreneurial français, pour augmenter le nombre de collaborations commerciales ou d'acquisitions innovantes par des grands groupes ou ETI français.

Le cœur de notre métier est d'accompagner les dirigeants dans la détection et la rencontre avec leurs futurs partenariats innovants ou acquisitions de sociétés innovantes.

Notre métier est de rapprocher les cultures et accroître le ROI de nos clients, dans une stratégie gagnant-gagnant pour les grandes organisations et les startups.

##### UN POSITIONNEMENT UNIQUE DE TIERS DE CONFIANCE

Véritable tiers de confiance, nous connectons au plus juste les besoins des dirigeants aux PME les plus innovantes de leur secteur avec un seul parti-pris : la réussite de la relation d'affaires.

A ce titre, notre approche consiste à nous concentrer sur le facteur principal de réussite ou d'échec d'une relation partenariale ou d'un rapprochement : le facteur humain. Nous traitons donc des aspects « extra-financiers » des opérations sur lesquelles nous accompagnons nos clients.

Notre métier consiste à accompagner nos clients grandes entreprises et ETI sur toutes leurs opérations en innovation externe (achats innovants, partenariats business, investissements minoritaires, fusions acquisitions), non seulement sur leur exécution, mais également en amont (phase de définition/maturation) et en aval (post-exécution/intégration) afin d'en maximiser les chances de succès.

##### UNE EQUIPE PLURIDISCIPLINAIRE ET UN RESEAU D'EXPERTS

Pour mener à bien les missions que vous nous confiez, Bpifrance « Le Hub » a réuni une nouvelle équipe qui intègre des profils spécifiques correspondant à vos enjeux en innovation externe (M&A, stratégie, financement, numérique, économie de l'innovation).

Notre équipe s'appuie à la fois sur l'expertise interne de Bpifrance (plus de soixante ans dans l'innovation et 2200 collaborateurs) et sur un pool externe d'experts de terrain associés à chacune de nos missions.

##### NOTRE COLONNE VERTEBRALE : LA DONNEE

Fort de notre expérience et de notre réseau, nous avons construit une base de données riche et complète, couvrant plusieurs dizaines de milliers de sociétés innovant sur tous les secteurs : du textile, à l'industrie en passant par les biotechnologies et le numérique. Notre base est constamment mise à jour et nous permet de qualifier les PME qui correspondent aux besoins de nos clients.

Afin de servir notre usage dans les meilleures conditions nous travaillons avec deux partenaires issus du monde des start-ups, afin d'enrichir et d'exploiter nos bases, dans le respect de la confidentialité, au service des affaires industrielles entre grandes entreprises et startups.

Taro UGEN

Directeur Adjoint du Hub

Tél : 01 53 89 72 43

Mob : 07 85 05 80 11

6-8 bvd Haussman 75009 Paris

#### **4.8 Les réalisations des ministères couvrent dès à présent une gamme élargie de circonstances et dégagent des lignes de force**

Les réalisations des ministères exposées ci-dessus, celles évoquées dans les réponses aux questionnaires et lors des entretiens, notamment dans le secteur de la santé, montrent que trois des champs principaux de l'application des « big data » dans les administrations sont dès à présent l'objet de travaux et de réalisations :

- L'utilisation de tierces données ;
- L'optimisation de processus ;
- L'amélioration de l'efficacité et de l'efficience de politiques – notamment grâce à un meilleur criblage ou ciblage.

D'autres champs restent et devraient être abordés ou utilisés plus fréquemment :

- La territorialisation de politiques ;
- La visualisation des situations et des résultats, notamment pour l'aide à la décision ou la relation avec les personnes, les publics ;
- L'usage de données issues directement ou indirectement des objets.

Les champs déjà abordés sont donc vastes, porteurs d'enjeux et de création de compétences, les champs qui restent à aborder le sont tout autant.

Ces expériences ont dégagé des lignes de force qui sont autant d'atouts des ministères qui se sont engagés dans la démarche :

- L'identification de problématiques métiers prioritaires pour les décideurs ;
- La mobilisation de ressources internes et le recrutement de spécialistes – en petit nombre mais dont la présence peut être cruciale ;
- L'insertion d'équipes innovantes dans des administrations de tradition.

**Dans ces conditions, et compte-tenu des enjeux importants et des interrogations qui perdurent, deux mouvements doivent être envisagés de façon positive :**

- **Le premier consistant à élargir la posture « big data » à l'échelle des directions ou des ministères où des réalisations pionnières se développent avec succès ;**
- **La deuxième à initialiser des opérations pionnières dans des directions ou des ministères nouveaux et où existent des circonstances favorables – problématiques, équipes motivées, etc.**

## 5 LES EXPERTS APPORTENT DES A PRESENT DES VISIONS CONSTRUITES DANS LES DOMAINES MANAGERIAUX, HUMAINS, FINANCIERS, TECHNIQUES ET JURIDIQUES

### 5.1 Politique de l'AGD en matière de « big data » et d'analytique

La création d'une fonction d'administrateur général des données de l'Etat ouvre des perspectives pour un développement prometteur et organisé du « big data » au sein de l'administration.

Le développement des données numériques produites et reçues par l'État, ainsi que le développement des nouveaux outils et des nouvelles stratégies de traitement de ces données, ouvrent de très grandes perspectives d'amélioration des politiques publiques : meilleure allocation de ressources, meilleur pilotage de politiques, recherche d'économies au plus proche des réalités, meilleure organisation des services.

Pour favoriser la diffusion de ces pratiques, le Premier ministre a créé, sous son autorité et rattaché au SGMAP, la fonction d'administrateur général des données (AGD), par le décret du 16 septembre 2014. Ce décret confère à l'AGD un premier ensemble de pouvoirs et de responsabilités, nécessaires au démarrage de cette nouvelle activité, proche de celle de « Chief data officer », fonction de plus en plus fréquente dans les grandes entreprises et dans les grandes villes américaines. Les États-Unis ont ainsi créé une fonction similaire en février 2015 en instaurant un « Chief data scientist », puis le Royaume-Uni en mars 2015.

Pour réussir sa mission, l'AGD doit s'impliquer dans la qualité des données détenues par le service public, de la bonne gouvernance de ces données et de leur bonne circulation. Il devra créer la compétence nécessaire à l'utilisation de ces données par les administrations (compétence qui intègre les « big data », le « machine learning » et la conduite du changement, il devra forger les méthodes permettant de mettre ces compétences au service du changement, et diffuser largement cette compétence au sein de l'administration.

Le noyau de l'équipe AGD (4 personnes) a été créé fin décembre 2014 et a engagé un travail dans trois directions :

=> gouvernance de la donnée

- ☐ Coopération avec la DISIC sur les bonnes règles de gouvernance de la donnée ;
- ☐ Préparation de l'interface de saisine sur les questions sur la circulation de la donnée et réponse aux premières saisines : <https://agd.data.gouv.fr/saisines-de-lagd/> ;
- ☐ Organisation du processus de saisine pour avis de l'AGD et remise d'un premier avis qui a nourri le projet de loi « Santé » : <https://agd.data.gouv.fr/2015/04/02/avis-portant-sur-la-publication-la-rectification-et-la-reutilisation-des-informations-portant-sur-les-professionnels-de-sante/> ;
- ☐ Préparation de la mission confiée par le Premier ministre à M. Antoine Fouilleron, magistrat à la Cour des comptes, sur les « ventes de données entre administrations » et accompagnement de cette mission : <https://agd.data.gouv.fr/2015/06/01/saisine-pm-cc/> ;
- ☐ Préparation du premier rapport annuel sur la gouvernance de la donnée prévu pour septembre 2015.
- ☐ Début de structuration d'un réseau de datascientists dans les administrations.

=> qualité de la donnée

② Conduite et présidence de l'accord entre IGN, La Poste, le DGFIP et OpenStreetMap permettant de développer en libre accès la base nationale de toutes les adresses géolocalisées : <http://adresse.data.gouv.fr/> ;

② Soutenu le développement en open source du logiciel de microsimulation Open Fisca : <http://www.openfisca.fr/> ;

=> Utilisation des datasciences

② Approche de différentes administrations pour identifier des problèmes pouvant être résolus par les approches fondées sur la donnée et engager un travail avec les plus motivées (SAE, Pôle emploi, Banque publique d'investissement, CNAM, Brigade des sapeurs-pompiers de Paris, gendarmerie...) ;

② Conduite avec le SAE un travail permettant d'appuyer la négociation des achats d'électricité et d'économiser entre 100 et 200 millions d'euros de dépense publique : <https://agd.data.gouv.fr/2015/05/17/analyser-les-consommations-energetiques-des-batiments-publics/> ;

② Organisation de partenariats de recherche avec l'École polytechnique, l'université Paris- Dauphine, la chaire big data de l'Essec mais aussi le cabinet Office britannique.

## **5.2 La gouvernance des projets « big data » trouvera son utilité à la croisée de celles des systèmes d'information et du décisionnel selon des modes différenciés selon les types de projet et les contingences locales**

La gouvernance d'un projet « big data » par Gilles Fournier

Trois éléments me paraissent essentiels :

### 1) La nécessité d'un sponsor

Les projets pionniers plus encore que les autres ont besoin d'un sponsor politico-administratif situé au bon niveau des organisations (élevé dans la hiérarchie) pour leur permettre de résister aux aléas qui ne manquent pas de les affecter, s'inscrivant dans la durée, capable de donner du sens, de résister aux conjonctures défavorables, parer aux résistances, au doute voire à l'hostilité.

### 2) Un mode de gouvernance adapté à l'ambition fixée au projet

Les choix d'organisation de la conduite du projet, le niveau de formalisme méthodologique retenu doivent s'adapter aux enjeux, au périmètre des acteurs concernés, au niveau de complexité des coopérations inter services à organiser, structurer, piloter, et aux moyens (budgets, RH) mobilisés.

S'il s'agit de réaliser une « preuve de concept » dans un champ et une durée limités, mettant en jeu peu d'acteurs et/ou peu de services, avec l'attente d'un retour sur investissement rapide, le mode de gouvernance à mettre en place s'apparente à du « développement agile » reposant sur des coopérations directes et peu formalisées entre donneurs d'ordre et équipes chargées de la mise en œuvre. S'il s'agit de viser des « victoires rapides », c'est la réactivité qui sera le principal levier du succès des projets.

A l'inverse, des projets plus ambitieux, de plus long terme, mobilisant beaucoup d'acteurs et de moyens appelleront des modes de gouvernance plus structurés, mettant en jeu des comitologies très formalisées.



### 3) L'articulation de la gouvernance des projets « big data » avec la gouvernance stratégique des SI et du décisionnel

Aujourd'hui il est difficile de déterminer les particularités de la gouvernance des projets « big data ». Les travaux du groupe réuni à l'occasion de la mission n'ont pas vraiment permis de dégager des traits caractéristiques, en partie du fait du trop faible nombre d'expériences analysées.

Cependant il est probable que les projets « big data » s'inscriront à la croisée de la gouvernance stratégique des systèmes d'information et de celle du décisionnel, pour autant que celles-ci existent réellement.

Dès lors, la mise en place d'une gouvernance spécifique du « big data » devra tenir compte des éco système en place, des contextes ministériels marqués par des historiques différents, en évitant l'écueil de la redondance. Un dispositif de gouvernance n'est efficace qu'à la condition qu'il soit vraiment utile, c'est-à-dire que des décisions y soient prises qui ne peuvent être arrêtées et appliquées autrement. Dans le cas contraire, souvent constaté, la démobilisation ou le désintérêt des vrais décideurs aboutissent à des délégations de responsabilité formelles, peu ou pas assorties d'un mandat clair et encore moins contrôlées.

Le centre de gravité de cette gouvernance stratégique, comme pour les SI ou le décisionnel, doit se situer du côté des métiers, même si la mobilisation des DSI dans les projets « big data » s'imposera nécessairement. Mais l'objectif de traiter des volumes de données, structurées ou non, de plus en plus importants, requérant des investissements en infrastructures et outils spécifiques et nécessairement soumis à des arbitrages budgétaires ne doit pas conduire à les analyser comme des projets techniques.

L'enjeu est bien de mettre ces opérations reposant certes sur de nouvelles méthodes et de nouveaux outils au service de meilleures décisions opérationnelles. L'objectif est donc que ces démarches – « *business intelligence, data mining, analytique, big data* » – sortent des mains des spécialistes (qui restent néanmoins encore largement à trouver dans la sphère publique) pour se rapprocher des opérationnels et se mettre au service des managers.

## **5.3 Faire se développer vigoureusement le « big data » dans les administrations de l'État nécessitera des talents variés, animés d'un esprit de communauté, gérés et formés de façon spécifique au sein des métiers mêmes**

Les opérations pionnières de « big data » ont été initiées, construites par des personnes motivées, innovatrices et compétentes, des « talents » au sens le plus récent de ce mot. L'ambition, nécessaire, d'une massification raisonnée et progressive du « big data » nécessitera un nombre croissant de talents.

L'enquête, le dialogue avec les responsables de services, les chefs de projets, les experts, l'observation des réalisations à l'étranger et dans le secteur privé menés par la mission permettent d'éclairer les sujets les plus cruciaux :

- i) les compétences nécessaires et les emplois ;
- ii) les prérequis à la formation et la constitution d'une communauté ;
- iii) les conditions de la gestion des personnes ;

Ils permettent aussi d'éclairer concrètement ces sujets par des exemples :

- iv) de la gestion des compétences à l'Autorité de sûreté nucléaire – ASN ;
- v) et chez Air France / KLM.

### 5.3.1 L'activité « big data » créera – et nécessitera - directement un nombre significatif d'emplois, exprimé en milliers, notamment de développeurs, avec des salaires supérieurs à la moyenne des emplois homologues

La Commission européenne a réalisé une étude qui s'appuie plus particulièrement sur le marché du travail du « big data » au Royaume-Uni, à l'horizon 2017. Elle est annexée à sa communication sur la « data driven economy » de juillet 2014.

Le nombre d'emploi concerné serait de près de 140 000 au Royaume-Uni, ce qui est cohérent avec la projection faite dans le plan « big data » de la nouvelle France industrielle : environ 130 000 en France.

La projection du nombre d'emplois pour les administrations de l'Etat et ses contractants est loin d'être évidente. Cependant, dans l'hypothèse où le « big data » serait utilisé aussi intensément par les administrations que par les autres acteurs, l'Etat aurait besoin, dans ses services ou chez ses contractants, de milliers de personnes dans peu d'années, ce qui montre l'importance numérique des moyens humains à mobiliser dans les services de l'Etat ou par la commande publique.

En termes de spécialités et plus précisément de nombre de personnes par spécialités, le résultat de l'étude de la Commission est intéressant dans la mesure où elle montre l'importance en volume des besoins en développeurs – 42 % du total – et relativise l'importance toujours en nombre des besoins en recrutement de « data scientists » : 1 % du total. Les analystes métiers et les administrateurs informatiques formant deux contingents importants. **Ces éléments montrent que la quasi-totalité des ressources humaines peuvent venir des personnels existants ou au travers de filières existantes**, dans ou au profit des directions métiers et de la DSI sans avoir à procéder à des recrutements importants proportionnellement au besoin total, comme le montre l'exemple d'Air France / KLM.

Cependant, cette bonne nouvelle est pour partie compensée par une moins bonne nouvelle pour les employeurs, singulièrement les ministères : l'étude constate et prévoit des salaires plus importants – à hauteur de 21 % - pour ceux travaillant dans le « big data » que pour ceux travaillant sur d'autres sujets – analystes métiers, architectes, project managers, etc.

Ces éléments montrent aussi les enjeux attachés aux choix d'internalisation et/ou d'externalisation des différentes fonctions et donc de la cohérence à avoir à moyen-long terme entre la prévision d'activités liée au « big data », la politique de gestion des ressources humaines, notamment de formation interne, et celle d'internalisation / externalisation.

A ce stade de développement des « big data » dans les ministères et chez leurs opérateurs, la définition cohérente de ces trois politiques est une priorité et une urgence : la prévision d'activités « big data », la politique de gestion des ressources humaines, notamment de formation interne, et celle d'internalisation / externalisation.

De plus, l'Etat doit envisager à moyen terme une communauté des professionnels du « big data » qui comprendra plutôt un millier de personnes plutôt que quelques dizaines ou centaines. Les plans de mobilisation, de formation, d'animation doivent être pensés à cette échelle dès aujourd'hui.

En dernier lieu, l'étude précise les technologies dont les spécialistes devraient être les plus demandés – cf. encadré.

Les spécialités les plus recherchées dans le domaine du « big data » - Extrait de la communication de la Commission européenne relative à la « data-driven-economy ».

An analysis of skills requirements for different big data roles showed the specific/related technical knowledge and skills currently most in demand in each case were as follows:

- o For big data Developers: NoSQL, Java, JavaScript, MySQL and Linux together with TDD, CSS and Agile development knowledge.
- o For big data Architects: Oracle, Java, SQL, Hadoop, and SQL Server and Data Modelling, ETL, Enterprise Architecture, Open Source and Analytics.
- o For big data Analysts: Oracle, SQL and Java together with Data Modelling, ETL, Analytics and Data Analysis.
- o For big data Administrators: Linux, MySQL, Puppet, Hadoop and Oracle along with Configuration Management, Disaster recovery, Clustering and ETL.
- o For big data Project Managers: Oracle, Netezza, Business Objects and Hyperion together with ETL, and Agile Software Development – PRINCE2 and Stakeholder Management skills are also a common specified requirement in this case.
- o For big data Designers: Oracle, SQL, Netezza, SQL Server, Informatica, MySQL and Unix plus ETL, Data Modelling, Analytics, CSS, Unit Testing, Data Integration and Data Mining.
- o For Data Scientists: Hadoop, Java, NoSQL and C++ along with Artificial Intelligence, Data Mining and Analytics. A high proportion of adverts were noted also to make reference to a need for Statistics and Mathematics skills.

### 5.3.2 La constitution et l'animation d'une communauté tonique sont des prérequis au déploiement de formations de qualité

La mission considère que la démarche la plus prometteuse éprouvée dans plusieurs contextes, consiste à recruter des personnels des métiers et de les former à ces nouvelles spécialités (- analytique ou techno données) plutôt que de créer une filière de spécialistes dont on ne sait pas gérer le plan de carrière ni l'intégration des métiers, compte tenu de la résistance des gens du terrain à intégrer de nouvelles approches qui sont contreproductives.

#### IGPDE

Les prérequis pour une formation de qualité et le marché du travail

Pour qu'une formation soit efficace et fonctionne bien, il faut en premier lieu qu'il y ait des stagiaires et qu'elle corresponde à leurs besoins. Cette lapalissade n'en est pas tout à fait une.

L'expérience de l'IGPDE semble prouver que lorsque l'on veut monter une formation sur des métiers nouveaux ou innovants, la difficulté est de la faire connaître aux stagiaires auxquels elle pourrait être utile. La clé du succès dépend alors en premier lieu de l'existence d'une communauté bien identifiée, animée de façon dynamique, avec des objectifs clairs, notamment en matière de compétences attendues.

Ainsi, la moitié du sujet consiste à réfléchir aux moyens de faire connaître au sein de l'Etat, des ministères et des directions, le produit, en l'occurrence les data comme support des métiers. Il faut un leader, un chef de file de la communauté du « big data ». Il faut également constituer un cercle de personnes légitimes sur le sujet, qui partagent les bonnes pratiques, la connaissance des opérations lancées ou projetées tant en interne qu'en externe. Cela peut reposer sur différents outils : création d'événements, mise en place de cycles (par exemple : les conférences récurrentes comme les matinales de l'achat public, des petits déjeuners thématiques, des rencontres thématiques avec déclinaisons en cycles plus spécialisés, ...), constitution de groupes de travail.... Ce que l'on recherche, c'est l'effet boule de neige : colloques + cycles + leader + ... C'est aussi créer une culture commune.

Lorsque la communauté vit et identifie un chef de file, il devient alors bien plus facile d'élaborer un cahier des charges des modules de formation – cadre indispensable à la conception d'une formation, de diriger le travail avec un prestataire, d'organiser des tests de la formation et de déployer progressivement. La co-conception d'un module de formation est une entreprise redoutable si le groupe n'est pas bien conduit et il peut être plus efficace de déléguer ce travail (ou une première version) à un seul membre de la communauté.

En matière de formation, un écueil courant est d'avoir de bonnes idées, de bons produits, mais que les personnes qui tireraient profit de ce module ne connaissent pas son existence ! Le gros problème reste donc d'assurer une bonne information des stagiaires. Lorsque cette formation a été conçue selon le mode indiqué, sa notoriété est déjà bien en marche.

Enfin, pour qu'une opération de formation dans un nouveau domaine puisse être efficace à grande échelle, il est indispensable qu'elle reçoive le soutien d'autorités qui vont la porter auprès des responsables. Si votre sujet acquiert une forte visibilité, la formation pourra se faire connaître. Mais la formation ne peut pas assigner une priorité à un sujet qui n'est pas perçu comme prioritaire !

### 5.3.3 La gestion des spécialistes du « big data » gagne à être une composante de celle des personnels des métiers.

Deux approches coexistent dans les ministères et les grandes directions en ce qui concernent la gestion des ressources humaines participant à la transformation numérique.

La première approche consiste à recruter, le cas échéant à former, des personnels des métiers aux spécialités de la transformation numérique. Cette solution est mise en œuvre notamment par la Gendarmerie, la DGFIP, les douanes, etc.

La deuxième consiste à créer des filières de spécialistes, souvent renforcées par des contractuels.

La deuxième solution se heurte, plus que la première, à des écueils, devenus évidents, de gestion des carrières, des compétences et à une difficulté récurrente de dialogue avec les métiers.

De plus, l'existence de deux approches différentes se révèle être à l'expérience une source de limitations, de freins pour la création d'une communauté, de ressources ou de solutions trans-ministérielles ou interministérielles de la transformation numérique.

La mission considère donc :

- d'une part, qu'il est de l'intérêt commun de choisir une approche homogène dans tous les ministères et directions, notamment pour converger et, le cas échéant mutualiser, des problématiques, des ressources, à élargir les parcours professionnels, les formations, etc. ;
- d'autre part, que la première solution – consistant à s'appuyer sur les personnels et les corps des métiers - est préférable à la seconde, non seulement parce qu'elle a prouvé ses avantages mais aussi parce que le big data est tout particulièrement partie prenante des métiers tant en terme de problématiques et de prise de décision qu'en termes de processus et de changement des métiers.

### 5.3.4 La gestion des multiples compétences nécessaires à l'Autorité de sûreté nucléaire constitue un exemple de professionnalisation de haut niveau de fonctionnaires issus de corps différents



#### **LA GESTION DES COMPETENCES A L'AUTORITE DE SURETE NUCLEAIRE (ASN) : UN ENJEU DE CREDIBILITE POUR L'ASN ET D'EFFICACITE POUR SON CONTROLE** (Luc Chanial, secrétaire général)

L'Autorité de sûreté nucléaire (ASN), autorité administrative indépendante depuis 2006, est chargée du contrôle de la sûreté nucléaire et de la radioprotection en France. Ses principales missions portent sur le contrôle des activités et installations nucléaires, l'élaboration de la réglementation, la délivrance d'autorisations, l'information du public et la participation à la gestion des situations d'urgence.

Le champ de compétence de l'ASN regroupe aujourd'hui la sûreté des installations nucléaires (centrales électronucléaires, installations du cycle du combustible, installations en démantèlement, installations industrielles et de recherche...) et du transport des substances radioactives, la gestion des déchets radioactifs et des sites et sols pollués par la radioactivité, la radioprotection dans les secteurs médical, industriel et de recherche, la radioprotection du public et de l'environnement ainsi que la sécurité des sources radioactives.

Les missions de l'ASN dans un domaine aussi étendu que celui-ci et présentant des enjeux sociétaux, environnementaux, techniques et médiatiques croissants nécessitent des compétences permanentes et adaptées. Pour répondre à cet objectif, gage de sa crédibilité et d'un contrôle efficace, l'ASN a mis en place, depuis de nombreuses années, une politique de développement et de gestion des compétences à la fois ambitieuse, structurée et évolutive.

Cette politique repose tout d'abord sur une dynamique de recrutements qui conduit l'ASN à accueillir, chaque année, entre 60 et 70 nouveaux agents. Administration d'Etat, l'ASN recrute prioritairement dans les viviers de corps de fonctionnaires. Qu'ils soient issus de corps techniques (mines, industrie et mines, ponts et forêts...) ou, depuis 2002 et l'élargissement des missions de l'ASN au domaine de la radioprotection, de corps de santé (génie sanitaire, pharmaciens ou médecins...), l'ASN puise dans ces viviers l'essentiel de ses compétences. Pour répondre au mieux à ses besoins, elle recrute par ailleurs des agents contractuels qui renforcent ses compétences dans des domaines aussi variés que la radioprotection des patients, les facteurs sociaux, organisationnels et humains, le juridique, l'informatique... Elle bénéficie enfin de la présence de salariés mis à disposition contre remboursement par divers établissements publics (IRSN, CEA, AP-HP...) qui apportent leur expertise acquise au sein de leur propre structure.

La plupart des nouveaux arrivants intègrent ensuite un cursus de formation structuré, obligatoire, adapté à leur domaine d'activité et qui leur permettra d'exercer le métier principal de l'ASN : celui d'inspecteur. Les formations sont dispensées soit par l'ASN elle-même, soit par des prestataires externes. Un socle commun, identique pour l'ensemble des agents nouvellement recrutés, permet de présenter l'ASN, ses métiers, son organisation. Il permet également de détailler le processus d'inspection, ses différentes étapes ainsi que les responsabilités associées. Complété par d'autres formations dans le domaine de la communication générale ou de la protection contre les risques associés aux rayonnements ionisants, ce socle permet également aux nouveaux arrivants de s'intégrer dans et de développer la culture ASN. Il est complété par des formations techniques ou réglementaires dédiées à l'activité, à la connaissance du domaine contrôlé et aux missions attendues. Orienté vers le domaine « sûreté nucléaire » ou « radioprotection », chacun bénéficiera d'un plan de formation adapté à ses futures missions et ajusté en fonction de ses compétences. Ainsi, l'ingénieur recruté pour exercer, au sein de la Direction des centrales nucléaires de l'ASN, le contrôle des centrales exploitées par EDF aura-t-il un cursus de formation différent de celui qui a été recruté pour rejoindre la Direction des rayonnements ionisants et de la santé pour contrôler les centres de radiothérapie. A l'issue de

ce parcours, qui s'échelonne en moyenne sur une année (les formations les plus courtes durant de 1 à de 2 jours et les plus longues pouvant atteindre 4 semaines), et qui se complète par une expérience « terrain » en compagnonnage, une habilitation formelle « inspecteur » est alors délivrée permettant à l'agent recruté d'exercer pleinement son métier d'inspecteur.

La formation délivrée lui permet également d'exercer, au sein de son entité, toutes les activités complémentaires à celles de l'inspection qui contribuent également à l'amélioration du niveau de sûreté des installations et constituent même, dans certains cas, le cœur de métier des agents : analyse de dossiers techniques, demandes de modification d'installation, demande d'autorisation, suivi d'arrêt d'installation, analyse d'incident...

En complément de ces formations, obligatoires pour l'habilitation, des formations obligatoires après l'habilitation et optionnelles sont également prévues.

Par ailleurs, et afin de reconnaître les compétences et expérience de ses inspecteurs, l'ASN a mis en place un processus lui permettant de désigner les inspecteurs à qui elle peut confier des inspections plus complexes ou à plus forts enjeux. Les candidats, auditionnés par une commission spécifique, peuvent ainsi être désignés et reconnus « inspecteurs confirmés ».

L'ASN consacre ainsi des moyens importants au développement et au maintien des compétences de ses agents. Chaque année, près de 400 000 € de frais pédagogiques sont affectés à la formation des quelques 500 agents qui composent l'effectif de l'ASN et plus de 3 000 jours. agents sont dispensés.

L'ASN considère cet investissement indispensable pour assurer un haut niveau de contrôle et asseoir sa crédibilité dans un environnement toujours plus exposé et dont les enjeux ne cessent de croître.

Elle veille à ce que son dispositif soit, à tout moment, adapté à ses besoins et au contexte dans lequel s'effectuent ses missions. Elle veille également à favoriser les évolutions professionnelles de ses agents pour, d'une part leur offrir des perspectives de carrière attractives et, d'autre part pouvoir valoriser pleinement, dans la durée, l'investissement qu'elle a consacré en matière de développement des compétences.

C'est, dans le domaine des ressources humaines, deux des enjeux principaux de l'ASN.

### 5.3.5 Air France / KLM a construit ses activités et ses compétences de haut niveau dans le domaine du « big data » et des « data sciences » sur la base de ses personnels et de ses équipes existants

#### Les compétences

L'entreprise bénéficiait au départ d'une bonne expérience sur des technologies utiles au « big data » : développeurs (R, Python, Java, IHM, BI & DWH, SOA à fond) et production (experts Teradata mais aussi experts BD, experts Linux).

Elle a acquis des compétences R&D mathématiques et statistiques pointues pour concevoir de nouveaux outils adaptés au « big data ». Le rôle de « data scientist » s'est déployé grâce à une parfaite connaissance des données et des processus métier.

Les équipes de Recherche Opérationnelle ont très naturellement évolué vers la « data science » (profils ingénieurs généralistes, master « recherche opérationnelle », ENSTA). Des binômes mathématiciens-développeurs ont été mis en place selon les nécessités. En tout, ce sont 50 personnes qui jouent le rôle de « data scientists ».

Concernant la gestion des grands volumes de données hétérogènes et provenant de sources multiples, l'acquisition de compétences a été aisée au départ, mais est devenue rapidement complexe lorsque l'on a voulu optimiser les traitements et profiter pleinement de la puissance du framework. Ainsi, de nouveaux rôles sont apparus : « data management, data governance, data integration, data privacy ».

#### Le support externe

Air France KLM a pleinement profité de la communauté Open Source, très active autour des technologies « big data » et s'est appuyé sur de petites entreprises de services numériques (ESN) spécialisées. Elle a utilisé à fond les supports de formation et documentation mis en ligne par les distributeurs HADOOP.

L'entreprise a aussi la volonté d'intensifier les relations avec les universités et la recherche et construit un cursus « big data » avec un partenaire.

### **5.4 Les questions juridiques multiples et complexes constituent un point d'attention pour les porteurs de projet et justifient la mise en place d'un réseau d'experts**

De façon générale le droit relatif aux données est encadré, à la fois en tant que données, le cas échéant en tant que données personnelles, en tant que données publiques ou en tant que bases de données. Par ailleurs les bases de données sont soumises au titre de la propriété littéraire et artistique et le droit propre aux bases de données.

Les sources de droit sont européennes et françaises. D'autres domaines s'appliquent au droit sur les données : droit la responsabilité, droit pénal relatif à la fraude informatique modifiée<sup>9</sup>, ainsi que des dispositions plus sectorielles relatives par exemple la statistique, la santé, l'environnement.

Deux exemples d'évolutions vont intervenir :

- l'anonymisation et la ré-identification vont changer,
- Le droit d'auteur va évoluer avec une exemption texte et « data-mining ».

Confrontés à ces évolutions, les projets doivent s'appuyer sur des juristes compétents sur tous les champs.

Pour illustrer ces questions, trois éclairages des problématiques sur le droit figurent ci-dessous :

- La question des données personnelles est un point d'attention des ministères. En effet, les réalisations menées dans les administrations ont intégré la protection des données personnelles et ont permis de construire une méthode de travail avec la CNIL. Cependant, de nombreux sujets sont à venir, les modalités d'application du règlement européen en cours d'adoption, les bonnes pratiques en matière d'anonymisation et de garantie de la temps de cette propriété ;
- La responsabilité est un sujet important et qui pose de nouvelles questions en particulier lorsque les directions ont acquis des données de tiers ou qu'elles aient cédé des données ;

---

<sup>9</sup> n°88-19 du 5 janvier 1988

Enfin,, la constitution de lacs de données nécessite une gestion du risque adaptée et une capacité de gestion de crise renforcée en cas d'intrusion. Certains projets de « *big data* » conduiront à l'ouverture de droits d'accès plus larges et /ou à la constitution de lacs de données plus vastes. L'adaptation des politiques de sécurité à ces deux aspects doit s'accompagner de la mise en place de dispositifs de surveillance permanente de détection d'intrusion de réaction et de gestion de crise, notamment en cas de vol de données.

L'Etat se situe déjà par rapport au « *big data* » comme un fournisseur de données.

« Etalab, depuis sa création, a mis en ligne le portail national [data.gouv.fr](http://data.gouv.fr) qui permet l'accès à plus de 13.000 informations publiques gratuites et réutilisables.

En rendant, par exemple, accessibles les dépenses du budget de l'Etat à partir d'un seul fichier brut dans un format réutilisable ou la liste des biens immobiliers propriété de l'Etat, [data.gouv.fr](http://data.gouv.fr) contribue à rendre des comptes aux citoyens sur le fonctionnement de l'Etat et de ses administrations en permettant une plus grande transparence de leur fonctionnement. » (Source portail ETALAB

#### **5.4.1 Les « *big data* » créent de nouvelles responsabilités qu'il convient d'anticiper**

Porteurs de nouvelles pratiques, confrontés à de nouveaux risques, les acteurs du « *big data* » doivent s'y préparer activement en mettant en place des politiques – gestion du risque, assurance, gestion de crise, ...





## La responsabilité des « auteurs » du big data

### Le mot auteur ici sera plus pris comme générateur de données massives

On ne peut passer à côté du Big Data qui va produire entre 3,6 % et 7% de PIB pour la France. Cependant, au-delà de la structure technique qu'il faudrait mettre en place, il est intéressant d'envisager les diverses responsabilités encourues des différents acteurs possibles, générateurs de données.

Tout d'abord, nous décelons quatre principaux acteurs :

- 1) L'individu à la base de la pyramide qui peut souvent transmettre ses données à son insu ;
- 2) Les FAI : étendent leurs services au-delà d'internet: la "connectique" appliquée à tous les secteurs: santé, famille, travail loisirs, argent, maison etc.. (Orange, la poste);
- 3) Les API (Application Programming interface) interface de programmation: plateforme par laquelle un logiciel offre des services à d'autres logiciels;
- 4) GAFA: sont collecteurs de données et font le commerce de données: leur business model est fondé sur le commerce de données, centré sur le client.

Le droit français positif offre tout un arsenal juridique pour engager la responsabilité de chacun :

#### 1) L'individu :

Avec la montée de l'internet des objets et de l'internet de tout, il aura la possibilité de collecter ses propres données. Il est en bas de l'échelle et devient la mine d'or du système. Il faut donc le protéger plus que tout. Cette protection peut se faire par un système d'opt in sur la collecte, le traitement et la finalité des données sur la base du consentement du code civil (article 1109 et suivants du code civil).

Sa responsabilité est de fournir des données justes en adéquation avec son identité physique. A l'inverse, on pourra le poursuivre pour dol, erreur ou violence. Verra-t-on une violence numérique ? ou un harcèlement numérique ?

#### 2) Les FAI

Les FAI vont développer leur propre système de Big data. Dans ce cadre là, il faut s'attacher aux conditions générales d'utilisation et de vente et créer à défaut d'un modèle européen un modèle français de CGU/V. Leur responsabilité pourra être recherchée sur la base des articles 1134, 1135 et 1147 du code civil pour irrespect des CGU/V.

#### 3) Les API

*Isabelle LANDREAU, Avocate à la cour, 25, rue du général Foy, 75008 Paris,  
isabelle.landreau.avocat@droitdanslemil.fr*



### La responsabilité des « auteurs » du big data

Les API seront des acteurs incontournables et générateurs de données puisqu'elles seront souvent à l'origine de la transformation de la donnée et de l'exploitation de la donnée.

Leur responsabilité pourra être recherchée sur les fondements juridiques suivants : responsabilité contractuelle de l'article 1147 du code civil et alternativement responsabilité délictuelle de l'article 1382 dès lors qu'une cyber-attaque, une coupure d'électricité ou autre négligence auront engendré un dommage corporel ou intellectuel sur l'individu. Les API auront une obligation de résultat d'avoir un Plan de Reprise d'Activité (PRA) sur les données traitées et exploitées.

#### 4) Les GAFA

Les GAFA collectent déjà de nombreuses données du citoyen européen. Au delà des mesures et sanctions prises par la Commission Européenne, les tribunaux français sont engagés dans un contrôle plus strict de leur activité quant aux données personnelles grâce notamment à la remise en cause de la clause attributive de compétence. Il faut aussi prévoir leur mise en cause sur le fondement de la lésion pour déséquilibre du contrat entre le citoyen européen et leur entité. De plus, il faut utiliser l'article L. 120-1 du code du commerce sur les pratiques commerciales trompeuses en évoquant que le consommateur est trompé sur les conditions d'utilisation de la donnée, la portée de son engagement, le stockage de ses données et la réutilisation massive de ses données sans son consentement par rapport à l'usage premier des données.

Enfin, il ne faut pas négliger la soft Law qui dans les pays anglo-saxons ont une efficacité importante et permettent des avancées pratiques considérables.

### **5.5 La définition et la mise en œuvre d'un cadre organisationnel des données stimulerait l'usage des données**

L'usage des données peut être rendu plus fréquent, plus aisé, moins coûteux ou plus rapide si l'on crée un environnement favorable à leur usage.

Thomas Saint-Aubin propose, sur la base de son expertise, trois mesures pour stimuler l'usage des data de l'administration par les services de l'Etat.

Par Thomas Saint-Aubin

Responsable du pôle stratégie et coordinateur du Dilab à la DILA, Services du Premier Ministre

thomas.saint-aubin@dila.gouv.fr / 01-40-58-76-77

#### Proposition 1

- Recenser les données grises de l'administration et constituer un référentiel des métadonnées

La constitution d'un référentiel de métadonnées est un pré-requis indispensable à l'exploitation des données collectées disponibles pour les politiques publiques.

La plateforme data.gouv.fr est une plateforme de recensement des données publiques ouvertes par les administrations.

Les administrations sont parfois réticentes à ouvrir certaines données, notamment celles qui ne répondent pas à la qualification juridique des informations publiques, à l'ensemble du secteur privé.

Sur ces " données grises", on observe une multiplication des échanges dématérialisés de données entre les administrations.

La mise en place d'une plateforme interne aux administrations de recensement des métadonnées sur l'ensemble des données produites ou collectées dans les SI de l'Etat pourrait favoriser le partage de connaissance de cet actif informationnel stratégique, permettre les décroisements entre les bases de données publiques et dans certains cas, constituer un préalable à l'ouverture de certaines de ces données.

Ce référentiel pourrait également, à terme, s'étendre au recensement des données collectées et produites par le secteur privé susceptibles de présenter un intérêt pour la transformation numérique de l'Etat et susceptible de faire l'objet de commande publique groupée.

#### Proposition 2

- Réaliser des expérimentations autour de l'interopérabilité des données grises

Le recensement de l'ensemble des actifs informationnels des SI de l'Etat doit favoriser les expérimentations des administrations autour du « big data ».

En particulier, la mise en place d'un travail de catégorisation juridique de cet actif informationnel interministériel et la définition d'un cadre juridique ad hoc pour l'exploiter dans le cadre d'expérimentation pourrait constituer l'une des missions affectées à l'équipe pluridisciplinaire ( voir paragraphe 5) dédiée et constituer un préalable à la réalisation de prototypage de services utiles au secteur public.

A terme, ce travail de catégorisation juridique de l'actif informationnel public pourrait permettre de le valoriser dans le cadre de partenariat avec le secteur privé (via les partenariats d'innovation notamment pour la co-crédation publique-privée des futurs services publics numériques autour des données).

#### Proposition 3

- Publication et alignement des principaux référentiels et ontologies utilisés dans l'administration pour l'interopérabilité des « big data »

Avec le soutien de l'équipe pluridisciplinaire affectée au projet, il est proposé que les différents départements ministériels recensent et partagent les principaux référentiels et ontologies utilisés dans le cadre de leur activité.

Ceci constitue un préalable à leur publication éventuelle sur le web sémantique, ce qui représente également un enjeu géo-stratégique pour la France.

Il est proposé de développer l'approche collaborative publique-privée, nationale et locale, pour la publication, la maintenance et l'alignement de grands référentiels nationaux, en priorité autour des principales thématiques interministérielles.

La complétude de ces données pivots dépend de la mobilisation d'une multitude d'acteurs publics et privés, nationaux et locaux.

Au-delà de l'ouverture des données pivots, il faut donc maintenant également organiser la gouvernance des référentiels publics par la constitution d'ontologies<sup>10</sup> sur le Web sémantique<sup>11</sup> et la publication au format SKOS des vocabulaires.

La constitution d'une gouvernance de ces données pivots, véritable nouveau "biens communs du numérique", doit permettre de mutualiser des ressources et des investissements pour favoriser à la fois le développement de l'administration numérique et de l'économie numérique autour de l'interopérabilité des « big data »

## 5.6 Des incitations financières françaises et européennes sont mobilisables pour développer des innovations en matière de « big data » administratives

Les autorités françaises et européennes ont mis en place des incitations à la transformation numérique des organisations, nombreuses. Les activités « big data » des administrations de l'Etat sont éligibles à certaines.

Les porteurs de projets ont tout intérêt à procéder à l'identification des incitations qu'ils pourraient utiliser, en lien avec les services compétents : SGMAP, DGE, etc.

Le dispositif 'Challenges « big data »' fait partie des systèmes qui doivent tout particulièrement retenir l'attention des porteurs de projets. La DGE – direction générale des entreprises, le présente ci-dessous, ainsi que le dispositif 'R&D « big data »', plus ciblé.

DGE

Challenges « big data »

Un challenge « big data » est un concours à l'attention de startups, qui utilise des jeux de données mis à disposition par de grandes entreprises ou des entités publiques. En exploitant ces jeux de données, les startups peuvent développer des solutions technologiques innovantes et de nouveaux cas d'usage et services, ou encore aider à la création de nouveaux modèles économiques. Les challenges sont généralement d'une durée de 3 à 12 mois et permettent de sélectionner un ou plusieurs lauréats. Certains challenges peuvent prendre la forme d'hackathons.

Le lancement de challenges « big data » permet de favoriser des démarches d'innovation ouverte entre, d'une part, des entreprises ou entités publiques détenant des données exploitables, et d'autre part des startups ou PME fournisseurs de technologies innovantes. Ces challenges représentent une opportunité pour faire émerger des solutions radicalement nouvelles pour tous les acteurs.

Cinq initiatives ont bénéficié en 2014 d'un soutien de l'Etat pour le lancement de challenges. Ces initiatives sont portées par des animateurs, qui sont des entités ayant vocation à animer des écosystèmes

<sup>10</sup> Ensemble structuré des termes et concepts représentant le sens d'un champ d'informations (source wikipédia)

<sup>11</sup> Vise à aider l'émergence de nouvelles connaissances en s'appuyant sur les connaissances déjà présentes sur Internet (source wikipédia)

d'innovation (pôle de compétitivité, accélérateur, etc.) dans la mise en place de challenges « big data ». Plusieurs entités publiques ont lancé des challenges dans ce cadre.

Un nouvel appel « Challenges numériques » va être lancé durant l'été 2015, dans le cadre du Programme des Investissements d'Avenir.

#### Projets de R&D « big data »

Le « big data » induit une transformation majeure dans l'usage du numérique par les entreprises de tous les secteurs économiques. A titre d'exemple, de nombreux usages du « big data » apparaissent dans des domaines aussi variés que la santé, le tourisme et la grande distribution. En termes d'emplois, l'enjeu du « big data » est de créer ou consolider de l'ordre de 137 000 emplois, soit directement dans l'industrie informatique, soit dans les fonctions technologiques au sein des entreprises, soit dans les fonctions métiers (par exemple les fonctions de vente) des entreprises.

Le caractère disruptif de l'innovation introduite par la diffusion du « big data » dans un champ toujours plus large de domaines industriels ouvre enfin d'importantes opportunités de création de valeur et de différenciation, notamment pour les startups et les PME du numérique. Le marché des activités liées au « big data » a représenté en France environ 1,5 milliards d'euros en 2014, et atteindra près de 9 milliards d'euros d'ici 2020.

Face à ces enjeux, les pouvoirs publics mobilisent l'ensemble des leviers dont ils disposent pour positionner la France comme le pays de référence dans le domaine du « big data ». L'Etat a notamment mobilisé près de 60 M€ pour soutenir des projets de R&D :

- Plusieurs de ces projets correspondent à des initiatives sectorielles concrètes, qui visent à accélérer le développement de la demande, i.e. des usages du « big data », dans des domaines variés : assurances, smart city, santé...
- Ces projets peuvent également viser à soutenir le développement d'entreprises porteuses de solutions particulièrement innovantes et capables de saisir ces nouvelles opportunités économiques : ces entreprises sont souvent des startups ou des PME engagées dans une croissance rapide. Les projets retenus couvrent un large éventail de problématiques technologiques : conception conjointe d'architectures avec les applications du « big data », outils et méthodes pour l'exploitation de grandes masses de données, réponses apportées à de grands défis applicatifs...
- L'Etat a également soutenu la mise en place d'une plate-forme d'expérimentation mutualisée, à l'attention des PME et startups du secteur du « big data ». Celles-ci peuvent ainsi s'appuyer sur des moyens de test, comparaison et vérification à l'échelle des problèmes traités.

les appels à projets sont publiés à l'adresse :

<http://www.entreprises.gouv.fr/secteurs-professionnels/appels-a-projets-economie-numerique>

## 6 LES PERSPECTIVES DU « BIG DATA » CONDUISENT A RECOMMANDER L'ENGAGEMENT D'UNE PHASE COURTE ET ACTIVE DE CONSTRUCTION DES CONDITIONS DE MASSIFICATION DE LEUR USAGE

La mission sur la base de ses observations estime que le « *big data* » est prometteur pour améliorer l'efficacité et l'efficience des politiques publiques mais n'en a pas trouvé la preuve. Ce constat est largement partagé par les pionniers.

Nous constatons qu'il existe actuellement une dynamique de pionniers avec trois caractéristiques :

- En interrogeant les responsables de programmes LOLF, nous avons identifié quelques de pionniers, il faut souligner leur dynamisme et leur attitude innovante ;
- Cependant ils sont peu visibles et assez isolés entre eux, et de leurs homologues européens ;
- Les coûts engagés par ces opérations pionnières représentent des parts toutes à fait négligeables des moyens des organisations concernées.

Cependant, le « *big data* » s'inscrit à la rencontre de deux mouvements plus vastes, celui de la transformation numérique de l'Etat et celui de la valorisation et de l'ouverture des données : la transformation numérique de l'Etat touche les métiers, les processus, les organisations et leur management, les relations entre services et avec les publics ; l'ouverture et la valorisation des données se développent et doivent trouver leur forme et leur point d'équilibre.

Le « *big data* » est donc à la fois une opportunité nouvelle et un prolongement de réalisations et de cultures bien établies. Par contre, les services de l'Etat souffrent d'un grand cloisonnement entre services pour les données et les pratiques. De plus, si l'Etat est ouvert à l'acquisition de données externes, cela ne résulte pas d'une politique organisée.

Un état des lieux des réalisations en matière de « *big data* » est en cours au niveau de la commission européenne, il sera disponible en septembre prochain. Nos contacts avec les services n'ont pas fait apparaître à ce stade de pays qui se distinguent.

Les administrations de l'Etat et leurs opérateurs sont rentrés dans une phase d'appropriation du « *big data* » qui doit être élargie et consolidée. En conséquence, la mission considère que si l'étape exploratoire est maintenant bien établie, il est nécessaire de consolider les savoir-faire par des opérations pour décider de l'opportunité d'une industrialisation et d'une planification des opérations.

Durant cette phase pourront être identifiés et évalués les enjeux, les conditions de la mise en œuvre et les domaines d'excellence du « *big data* ».



Dans ces conditions, il nous semble raisonnable de lancer une démarche construite d'appropriation du « big data » reposant sur des réalisations, des apprentissages et la constitution d'un collectif pour une culture commune afin d'établir d'une part, une preuve globale que les « big data » peuvent améliorer les politiques publiques et d'autre part, les conditions de ces améliorations. Cette démarche pourrait s'appuyer sur la dynamique des pionniers que les propositions visent à étendre se renforcer et s'accélérer en termes de nombre d'opérations, de diversité, de champs d'application et de taille des réalisations.

Les recommandations devraient favoriser l'émergence d'idées, la communication entre les acteurs des différentes origines et statuts, ainsi qu'un meilleur appui ministériel et interministériel, tant décisionnel qu'opérationnel, au niveau national et déconcentré.

Cette politique peut s'appuyer sur deux orientations générales pendant cette période de deux ans et avant une évaluation :

- La première consistant en la création d'une communauté du « big data » dans les administrations animée et ouverte ;
- La deuxième consistant à étendre à l'échelle d'un petit nombre de politique publique et en susciter dans des secteurs nouveaux.

Cette période de deux ans devrait se clore par une évaluation qui permettra en toute connaissance de causes de définir les conditions de la mise en œuvre de la politique de « big data » pour l'Etat et ses opérateurs : gouvernance, ressources.

Concrètement, ces orientations se traduisent par sept groupes de mesures concernant chacun des niveaux de l'Etat.

Les moyens afférents restent modestes et peuvent s'appuyer sur les différentes incitations françaises et européennes comme le PIA (voir ci-dessus). Le schéma proposé, qui fait une part importante à la mutualisation sous différentes formes, apparaît comme économe en ETP, considérant que tous les services à terme seront engagés dans le « big data ».

Les mesures ci-dessous ne nécessitent pas de textes législatifs pour leur mise en œuvre.

**En ce qui concerne le portage politique et dans le cadre de la politique gouvernementale du numérique :**

#### **Recommandation n° 1.**

Engager un programme de deux ans de consolidation en vue de la massification des « big data »

Identifier concrètement cinq programmes LOLF dans lesquels on tirera profit largement du « big data » et de l'analytique

Cela permettra de convaincre un nombre significatifs de responsables de programmes LOLF de l'opportunité et de l'urgence à explorer les apports potentiels du « big data » d'une part et d'autre part de décider des modalités d'une industrialisation après cette période probatoire de deux ans. La sélection des programmes LOLF concernés par cette opération pourrait prendre une forme d'un appel à projets de transformation. Le choix pourrait être organisé par le SGMAP ou à sa création par le DINSIC et s'appuyer sur un jury indépendant. L'objet est de sélectionner les programmes LOLF sur la base du volontariat et de la motivation, mais aussi dans des secteurs variés pouvant démontrer une diversité dans l'intérêt du « big data ».

#### En ce qui concerne les ministères :

**Recommandation n° 2.** Encourager des POC de découverte de différentes natures (tierces données, objets, visualisation, etc...) notamment dans des domaines pionniers à conduire en mode projet

Evaluer et mobiliser les ressources humaines et financières existantes et nécessaires à la réalisation des chantiers

Pour préparer le moyen terme, introduire le « big data » dans la GPEEC et dans les formations initiales et continues, avoir recours de façon préférentielle au personnels présents et parmi eux entreprendre une action de formation en fonction des besoins futurs, en complément s'appuyer sur les entreprises innovantes d'informatique et de data sciences

#### En ce qui concerne les services déconcentrés :

**Recommandation n° 3.** Mandater des préfets pour initier une politique régionale de « big data » (par exemple deux régions), en lien en tant que de besoin avec les collectivités territoriales

Leur affecter une équipe pluridisciplinaire et les ressources budgétaires idoines

Cette équipe pluridisciplinaire pourrait être dimensionnée à une dizaine d'ETP avec des compétences technologiques, statistique et datascientists. Elle pourrait être rattachée à un service du Premier ministre en charge de la transformation numérique.

#### Au niveau interministériel

**Recommandation n° 4.** Constituer une communauté ouverte des décideurs et des professionnels de l'administration du « big data » et l'animer

Cette communauté, outre des agents de l'Etat devrait associer des acteurs du monde de la recherche et en tant que de besoin des experts issus de PME innovantes. L'animation de cette communauté doit créer un ciment tel qu'une dynamique du big data puisse se construire pendant les deux années consacrées à la consolidation.



**Recommandation n° 5.**

**Créer l'incubateur bug data de l'Etat en constituant une équipe d'appui et de compétences dont les missions sont :** Favoriser l'utilisation des leviers d'incitation français et européens.

Créer les conditions à une demande de formations et mettre en valeur des parcours de formation

Promouvoir des modèles (juridique, économique et technique) d'accès aux données externes (secteur privé, collectivités publiques, délégataires de service public)

Offrir l'accès à une plateforme équipée et armée d'une équipe

Accompagner les ministères dans la gestion et l'assurance des nouveaux risques

Animer les relations avec la CNIL, la CADA

Faire une revue des cadres juridiques pour lever les obstacles au « big data » et proposer les amendements

Documenter au fil de l'eau les réalisations ainsi que les réalisations homologues en Europe

L'équipe est constituée pour 2 ans, ses compétences seront revues suite à l'évaluation.

**Recommandation n° 6.**

**Faciliter les échanges de données entre les services de l'Etat, ses opérateurs**

Disposer d'un cadre contractuel pour des cessions de droit d'usage de données et d'un médiateur pour traiter les conflits entre services d'Etat et leurs opérateurs

Disposer d'un comité de déontologie pour l'utilisation des données

Identifier les charges afférentes à la mise en forme, la complétude, la qualité, à la livraison des données

Constituer un répertoire des modèles de données

Cette recommandation est de première importance pour les pionniers rencontrés par la mission.

**... et pour préparer la phase suivante dans les meilleures conditions**

**Recommandation n° 7.**

**Evaluer les réalisations ministérielles, régionales et interministérielles pour proposer une politique de « big data » de l'Etat pour ses politiques publiques**

Retenir une méthode (par exemple MAREVA) et définir des critères sur la base desquels l'évaluation sera faite

Cette méthode comme méthode d'analyse de la valeur pourrait reposer sur les cinq axes suivants :

- Retour sur investissement mesuré soit par des rentrées supplémentaires (lutte contre la fraude ou économies de fonctionnement) ;
- Les externalités par une amélioration de l'atteinte des objectifs LOLF et par là même des politiques publiques ;

- Les internalités en permettant par exemple de meilleurs choix pour les décideurs de l'Etat ou un meilleur fonctionnement des opérationnels sur le terrain grâce aux résultats des analyses de ces données (criblage, ciblage pour les douaniers) ;
- Une meilleure adéquation aux décisions des politiques ;
- La maîtrise des risques : maîtrise des technologies innovantes sous-jacentes, corpus juridique, risque d'image, sociaux,...

### **Les moyens**

Le programme de deux ans de consolidation évoqué à la recommandation 1 sera doté d'une quarantaine d'ETP affectés au SGMAP mis à disposition :

- Des programmes LOLF lauréats du concours (4 ou 5 par programme, soit 20 à 25 ETP en tout) ;
- De chacune des deux préfectures (2 ETP par préfecture), Cette organisation pourrait préfigurer un centre de compétences commun aux services déconcentrés de l'Etat pour la phase postérieure à la consolidation ;
- D'une équipe interministérielle d'appui d'une dizaine d'ETP identifiée et logée au SGMAP et au service de la communauté.

## **ANNEXES**



## Annexe 1 : Lettre de mission



Paris, le 24 DEC. 2014

Le Ministre de l'Economie, de l'Industrie  
et du Numérique

Le Secrétaire d'Etat chargé de la Réforme de l'Etat  
et de la Simplification

La Secrétaire d'Etat chargée du Numérique

à

Monsieur le Vice-président du Conseil général de  
l'Economie, de l'Industrie, de l'Energie et des  
Technologies

Objet : Mission relative à l'usage des technologies « big data » dans l'administration

Monsieur le Vice-Président,

Aux yeux du monde de l'entreprise, les données des systèmes d'information ont changé de nature : les données ne sont plus des charges, elles sont devenues des ressources potentielles. Cette évolution radicale est à la fois culturelle, managériale et technique. Elle est désignée sous le terme de « big data » ou mégadonnées en français, terme ambigu parce que l'on désigne ainsi à la fois la capacité et la volonté de rassembler, de traiter et d'agréger non-seulement des volumes de données importants mais aussi des informations évolutives ou non-structurées (textes, images ...) issues de sources hétérogènes, parfois lacunaires. Ces opérations peuvent s'effectuer à des échelles de temps variables, sur l'historique comme en temps réel.

Le « big data » est déjà une réalité concrète, notamment aux Etats-Unis, dans un petit nombre de secteurs d'activités, essentiellement le marketing et les services commerciaux en ligne, ou encore les services de renseignements des Etats. Dans de nombreux autres secteurs, son développement est encore embryonnaire : les activités des administrations publiques civiles, les assurances, les secteurs de la santé et de la dépendance, l'enseignement scolaire, universitaire et professionnel, les transports et l'énergie, etc.

En termes de volume, les services de l'Etat sont relativement pauvres en données, pour diverses raisons : la collecte et le traitement des données sont encadrés législativement et réglementairement, sous la contrainte du respect de la vie privée et du principe de finalité principalement. Inversement, des acteurs privés possèdent de grandes quantités de données d'intérêts divers, notamment les réseaux sociaux, les moteurs de recherche, les fournisseurs d'accès Internet, etc.

Le recours par l'administration à des données externes, qu'elles soient structurées ou non-structurées, en complément de ses propres données pourrait révéler dans certains cas, les promesses du « big data » en faveur de l'amélioration de l'efficacité des services et de leur réactivité. De façon réciproque, l'« open data » peut être un facteur d'innovation externe.

Certains axes d'améliorations sont cités dans les publications, sans être, en général, fondés sur des opérations pilotes ou des expériences observées : organisation de la sécurité publique et prédiction criminelle, ciblage et criblage des contrôles, parcours de soin et validation de protocoles, analyse de la réussite scolaire ou universitaire, connaissance des entreprises, statistique publique, gestion de la ville, fraude fiscale et sociale, etc.

De plus, au-delà de la recherche de l'efficacité et de l'efficience, le recours au « big data » permettrait, dans certains cas et dans le respect des libertés publiques, de remédier à l'asymétrie d'information qui s'est créée, notamment par l'effet des réseaux sociaux, entre le public et les agents publics, par exemple les forces de l'ordre.

Le volet « big data » de la « Nouvelle France industrielle » mobilise dès à présent les compétences industrielles et scientifiques dans ce domaine.

Cependant le déploiement de cette nouvelle démarche est encore très embryonnaire surtout au sein des administrations publiques. Un consensus se dégage aujourd'hui sur la nécessité d'expérimenter sans délais mais posément certaines des opportunités imaginées.

Ces expérimentations amorceraient le processus collectif d'apprentissage des administrations et permettraient d'identifier les prérequis, les points-clés et les bonnes pratiques d'un développement plus important du recours à l'approche « big data ».

Différents points méritent un examen : l'acquisition des compétences, en particulier les conditions de recrutement et de formation des personnels des administrations, l'origine et la variété des données, les conditions de collecte des données externes, le respect de la vie privée, l'adaptation des infrastructures des systèmes d'information, la gestion du changement dans les services opérationnels ainsi que la création d'un climat de confiance. Les secteurs particulièrement prometteurs semblent être la sécurité, la détection de la fraude, les parcours de soin.

Nous vous demandons donc d'explorer avec les services compétents de l'administration l'opportunité et la faisabilité d'une démarche collective des administrations de l'Etat pour la réalisation d'une plate-forme ouverte de définition et de prototypage de services utilisant les approches du « big data », par exemple sous la forme d'un Lab.

Vous pourrez notamment constituer un groupe de travail avec les ministères les plus concernés et le SGMAP.

Vos travaux pourront s'appuyer sur les services du SGMAP, d'ETALAB et de la DISIC. Ils ont également vocation à nourrir la réflexion relative à la construction de la fonction d'administrateur général des données instituée par le décret du 16 septembre 2014.

Vous nous remettrez une note d'étape sous trois mois, et votre rapport nous sera remis sous six mois.



Emmanuel MACRON



Thierry MANDON



Axelle LEMAIRE

**Annexe 2 : Liste des acronymes utilisés**

CGE	Conseil général de l'Economie
LOLF	Loi organique pour les lois de finances
RPROG	Responsable de programme
DGFiP	Direction générale des finances publiques
AGD	Administrateur général des données
GED	Gestion électronique de documents

**Annexe 3 : Liste des personnes ayant participé aux ateliers**

Mme Brigitte BACCAÏNI,	Commissariat général à l'égalité des territoires
M. Stéphane BOST	Préfecture de la région Alsace
M. Philippe CHATAIGNON	Ministère de la justice
Mme Agnes CLANCIER	Secrétariat général du Gouvernement
M. Pierre FOULQUIER	Tracfin
M. Gilles FOURNIER	Ministère de l'éducation nationale
Mathieu GARNIER	Commissariat général à l'égalité des territoires
Mme Pauline GIVORD	Insee
M. Louis HUMEAU	Tracfin
M. Michel ISNARD	Insee
Mme Marine JARA	Ministère des affaires étrangères
M. Jean-Séverin LAIR	Ministère de la culture
Mme Maryse LE BRAS	Ministère éducation nationale
M. Florian LEZEC	Ministère enseignement supérieur et recherche
M. Fabrice MATTATIA	Ministère de l'éducation nationale
M. Thomas SAINT-AUBIN	Secrétariat général du Gouvernement
M. Damien THILLOU	DGDDI
M. Thierry VALLAT	Ministère des affaires étrangères
Mme Cécile VANDAMME	DGFIP
M. Benoît ZEDET	Ministère de la jeunesse et des sports



#### ***Annexe 4 : Liste des intervenants aux ateliers***

M. Henri VERDIER, Administrateur général des données

Mme Cécile VANDAMME, DGFIP

M. Michel ISNARD, Insee

M. Luc CHANIAL, Secrétaire général ASN

M. Perica SUCEVIC, SGMAP

Mme Isabelle LANDREAU, Avocate

M. Jean Christophe LALANNE, Air France

M. Damien THILLOU, DGDDI

M. Jean-Séverin LAIR, Programme VITAM

M. Louis HUMEAU, Tracfin

M. Gilles FOURNIER, Ministère de l'éducation nationale

## Annexe 5 : Programme des ateliers

### Première séance le 16 avril 2015, 14h00-18h.

NECKER - salle 5043R

- 14h00-14h15	- Ouverture par Mme Laure de la BRETECHE Secrétaire générale pour la modernisation de l'action publique
- 14h15-15h00 Présentation de la mission tour de table et diaporama, document de dépouillement	- Mission
- 15h00 15h30 Présentation lutte contre la fraude en utilisant le data mining et discussion	- DGFIP Mme Cécile VANDAMME
- 15h30 15h45 pause	-
- 15h45 16h15 Présentation Insee (tickets de caisse)	- M. Michel ISNARD Insee
- 16h15-17h15 Différents modèles : acheteurs, producteurs de données, captation « boîtes noires », analytique...	- Mission

### 2<sup>ème</sup> séance le 6 mai 2015 14h00-18h

PMF – salle 5

- 14h-14h45 Propriété intellectuelle et responsabilité - 14h45h-15h30 Les cadres juridiques (CADA et CNIL) - 15h30-15h45 Echanges	- Propriété intellectuelle, droit d'auteur, droit des bases de données (Avocat Mme Isabelle LANDREAU) - Accès aux documents administratifs (CADA) Données personnelles.
- 15h45-16h Pause	-
- 16h-16h45 Les compétences : une expérience externe, - 16h45-17h15 La gestion des ressources	- Jean-Christophe LALANNE (EVP CIO Group Air France KLM) - M. Luc CHANIAL Secrétaire général ASN
- Discussion 17h15 - 18h	-

**3<sup>ème</sup> séance le 21 mai 2015 14h00-18h**

PMF – salle 5

- 14h-14h30 Douane : Une expérience de data mining, le lien avec les métiers	- Damien THILLOU, chargé de mission DGDDI
- 14h30-15h00 :VITAM : une expérience à trois	- Jean-Séverin LAIR Sous-directeur des systèmes d'information Ministère de la culture et de la communication
- 15h-15h30 TRACFIN : Une expérience de conduite du changement	- Louis HUMEAU, chef de projet Tracfin
- 15h30-15-45 Pause	
- 15h45–17h30 Gouvernance des projets (centre de décision, centre de commandement), projet, conduite du changement, organisation Politique (policy), mutualisation, compétence, apprentissage	- animateur : Gilles FOURNIER adj dir DEPP Education nationale - Tour de table
- 17h30-18h Et maintenant : programme	- Mission

**4<sup>ème</sup> séance le 16 juin 2015 14h30-18h**

PMF – salle 5

- Relecture et discussion des projets de recommandations

## **Annexe 6 : Liste des personnes rencontrées ou interrogées**

### **Organismes publics et parapublics**

#### **Commission européenne**

- Mme Marta NAGY-ROTHENGASS, chef de l'unité « Chaîne de valeur des données », DG CONNECT
- M. Szymon LEWANDOWSKI
- M. Jiri PILAR

#### **Commission nationale informatique et libertés**

- Mme Isabelle FALQUE-PIERROTIN, Présidente
- M. Édouard GEFFRAY, Secrétaire général

#### **Autorité de la statistique publique**

- M. Paul CHAMPSAUR, Président,
- Mme Claudine GASNIER (ASP) Rapporteur Général,
- M. Franck COTTON, Expert informatique auprès du Secrétaire général de l'Insee,
- Mme Pauline GIVORD, Chef de division Insee.

#### **Conseil d'Etat**

- M. Laurent CYTERMANN, Maître des requêtes

#### **Premier Ministre**

- **Cabinet du Secrétaire d'Etat chargé de la réforme de l'Etat et de la simplification**
  - M. Boris JAMET-FOURNIER , Conseiller numérique et participation
- **Secrétariat général pour la modernisation de l'action publique**
  - Mme Laure de la BRETECHE, Secrétaire générale
  - M. Serge BOSSINI, Directeur, adjoint à la Secrétaire générale
  - M. Henri VERDIER, Directeur d'ETALAB, Administrateur général des données
  - M. Jacques MARZIN, Directeur de la DISIC
  - M. Jean-Jacques LEANDRI, Conseiller *e-government*
  - M. Perica SUCEVIC, Conseiller juridique, mission synthèse et soutien

#### **Ministère de l'écologie, du développement durable et de l'énergie**

- **Direction générale de la prévention des risques**
  - M. Jean-Marie DURAND, Adjoint à la Directrice générale
  - M. Jean-Jacques RICHARD, Chargé de mission systèmes d'information

**Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche**

- M. Guy WAÏSS, Adjoint au Directeur général de l'enseignement scolaire
- M. Philippe THURAT, Sous-directeur service du budget, de la performance et des établissements
- Mme Brigitte BRUSCHINI, Service du budget, de la performance et des établissements
- M. Erwan COUBRUN, Bureau de la synthèse budgétaire, des études et du contrôle de gestion

**Ministère de la justice**

- M. Benjamin CAMUS, Sous-directeur de la statistique et des études
- M. Philippe CHATAIGNON, adjoint

**Ministère des finances et des comptes publics, Ministère de l'économie de l'industrie et du numérique**

- **Cabinet de la Secrétaire d'Etat chargée du numérique**

- M. Bertrand PAILHES, Directeur du cabinet
- M. Matthieu GUERLAIN, Conseiller auprès du Ministre
- M. Alexandre TISSERANT, directeur adjoint du cabinet
- M. Florent de BODMAN

- **Institut national de la statistique et des études économiques (Insee)**

- M. Michel ISNARD, Secrétariat général, affaires juridiques et contentieuses
- Mme Françoise DUPONT, méthodologue, Direction de la méthodologie et de la coordination statistique et internationale

- **Direction générale des finances publiques**

- M. Alain ISSARNI, Chef de service, Service des systèmes d'information

- **Tracfin**

- M. Louis HUMEAU, Mission SI Tracfin,
- M. Fabien POPOFF, Responsable sécurité,
- M. Fabien DUSCHER Projet SYRANO

- **Direction générale des douanes et droits indirects**

- M. Jean-Michel THILLIER, Chef de service, adjoint à la directrice générale
- M. Damien THILLOU, Chargé de mission, sécurisation de la chaîne logistique

- **Direction générale des entreprises**

- Mme Cécile DUBARRY, chef de service économie numérique
- M. Fabien TERRAILLOT, chef du bureau du logiciel
- M. Jean-Rémi GOUZE, sous-directeur de l'action territoriale et du développement économique
- Mme Martine BAUDOIN, chef du bureau pilotage et action territoriale

- **IGPDE**

- Mme Mylène ORANGE-LOUBOUTIN, Directrice générale
- M. Fabrice DEMAISON, Directeur des actions de modernisation

**Ministère des affaires sociales, de la santé et des Droits des femmes**

- M. Christian POIRET, Secrétaire général
- M. Alain FONTAINE, Chef de la mission prospective et recherche
- Mme Nathalie BOULARD, Directrice de projet, Cheffe de la division de l'aide à la prise de décision
- Mme Agnès RAMZI,
- M. Thierry DUGENETAY, Chef du bureau systèmes d'information et informatique de proximité

**Ministère du travail, de l'emploi, de la formation professionnelle et du dialogue social**

- M. Olivier TOCHE, Chef de service, Direction générale du travail
- M. Philippe MOREAU, Chef du département des affaires générales et des prud'hommes
- Mme Christine N'GUYEN, Bureau des relations collectives du travail
- M. Jean-Marc LAOUENAN, Direction des systèmes d'information
- M. Henri FAGEBAUME, Direction des systèmes d'information

**Ministère de l'intérieur, Ministère chargé de l'Outre-mer**

- M. Régis BAUDOUIN, Responsable du programme 344 - Fonds de soutien relatif aux prêts et contrats financiers structurés à risque, directeur du service de pilotage du dispositif de sortie des emprunts à risque
- **Direction générale des collectivités locales**
- Mme Françoise TAHERI
- M. Adrien BAYLE Adjoint au chef de bureau des concours financiers de l'Etat

**Ministère de l'agriculture**

- Béatrice SEDILLOT, Chef du service de la statistique et de la prospective
- Laurent PAVARD, Président du conseil des systèmes d'information

**Ministère de la culture**

- M. Jean-Séverin LAIR, Sous-directeur des systèmes d'information
- Mme Mélanie REBOURS, Responsable du partenariat VITAM
- M. Baptiste MICHEL, Service interministériel des archives de France

**Préfecture de la région Alsace, Services de l'Etat en région**

- M. Stéphane BOUILLON, Préfet de la région Alsace, Préfet du Bas-Rhin
- M. Jacques GARAU, Secrétaire Général pour les Affaires Régionales et Européennes
- M. Stéphane BOST, chargé de mission ^politiques et équipements structurants dont les TIC
- Mme Valérie TRUGILLO, Secrétaire Générale Adjointe du Rectorat de Strasbourg
- M. Eric LOUVEL, chef de division de la direction du système informatique du Rectorat de Strasbourg
- M. Olivier BORELY, Chef des Risques technologiques de la DREAL Alsace
- M. Daniel GALLISSAIRES, Directeur Adjoint chef du Pôle 3E de la DIRECCTE Alsace
- M. Serge KAUFFMANN, Délégué régional à la recherche et à la technologie
- M. Jean-Charles SPECHT, Responsable informatique de la DRAC Alsace

## Organisations professionnelles et entreprises

### **BPI France**

- M. Taro HUGEN Directeur adjoint le Hub

### **Air France/KLM**

- M. Christophe LALANNE, CIO

### **Alliance big data**

- M. Charles HUOT, co-fondateur de TEMIS

**Annexe 7 : Liste des programmes interrogés**

185 - Diplomatie culturelle et d'influence, 209 - Solidarité à l'égard des pays en développement ;  
204 - Prévention, sécurité sanitaire et offre de soins ;  
335 - Conseil supérieur de la magistrature ;  
110 - Aide économique et financière au développement ;  
305 - Stratégie économique et fiscale ;  
117 - Charge de la dette et trésorerie de l'État (crédits évaluatifs) ;  
114 - Appels en garantie de l'État (crédits évaluatifs) ;  
145 – Épargne ;  
168 - Majoration de rentes ;  
336 - Dotation en capital du Mécanisme européen de stabilité ;  
338 - Augmentation de capital de la Banque européenne d'investissement ;  
156 - Gestion fiscale et financière de l'État et du secteur public local ;  
200 - Remboursements et dégrèvements d'impôts d'État (crédits évaluatifs) ;  
201 - Remboursements et dégrèvements d'impôts locaux (crédits évaluatifs)  
154 - Économie et développement durable de l'agriculture et des territoires ;  
149 – Forêt ;  
182 - Protection judiciaire de la jeunesse ;  
151 - Français à l'étranger et affaires consulaires ;  
203 - Infrastructures et services de transports ;  
198 - Régimes sociaux et de retraite des transports terrestres ;  
164 - Cour des comptes et autres juridictions financières ;  
340 - Haut Conseil des finances publiques ;  
102 - Accès et retour à l'emploi ;  
103 - Accompagnement des mutations économiques et développement de l'emploi ;  
101 - Accès au droit et à la justice, 310 - Conduite et pilotage de la politique de la justice ;  
140 - Enseignement scolaire public du premier degré ;  
141 - Enseignement scolaire public du second degré ;  
230 - Vie de l'élève ;  
217 - Conduite et pilotage des politiques de l'écologie, du développement et de de la mobilité durables ;  
337 - Conduite et pilotage des politiques du logement et de l'égalité des territoires ;  
124 - Conduite et soutien des politiques sanitaires, sociales, du sport, de la jeunesse et de la vie associative ;  
155 - Conception, gestion et évaluation des politiques de l'emploi et du travail ;  
214 - Soutien de la politique de l'éducation nationale ;  
195 - Régimes de retraite des mines, de la SEITA et divers ;  
152 - Gendarmerie nationale ;  
178 - Préparation et emploi des forces ;  
146 - Équipement des forces ;  
146 - Équipement des forces ;  
191 - Recherche duale (civile et militaire)  
139 - Enseignement privé du premier et du second degrés ;  
302 - Facilitation et sécurisation des échanges ;



107 - Administration pénitentiaire ;  
163 - Jeunesse et vie associative ;  
166 - Justice judiciaire ;  
220 - Statistiques et études économiques ;  
176 - Police nationale ;  
113 - Paysages, eau et biodiversité ;  
109 - Aide à l'accès au logement ;  
135 - Urbanisme, territoires et amélioration de l'habitat ;  
165 - Conseil d'État et autres juridictions administratives ;  
167 - Liens entre la nation et son armée ;  
169 - Reconnaissance et réparation en faveur du monde combattant ;  
212 - Soutien de la politique de la défense ;  
207 - Sécurité et éducation routières ;  
180 – Presse ;  
334 - Livre et industries culturelles ;  
313 - Contribution à l'audiovisuel et à la diversité radiophonique et action audiovisuelle extérieure ;  
134 - Développement des entreprises et du tourisme ;  
218 - Conduite et pilotage des politiques économiques et financières ;  
174 - Énergie, climat et après-mines ;  
161 - Sécurité civile ;  
341 - Conférence 'Paris Climat 2015' ;  
170 – Météorologie ;  
159 - Information géographique et cartographique ;  
190 - Recherche dans les domaines de l'énergie, du développement et de la mobilité durables ;  
303 - Immigration et asile ;  
224 - Transmission des savoirs et démocratisation de la culture ;  
186 - Recherche culturelle et culture scientifique ;  
148 - Fonction publique ;  
112 - Impulsion et coordination de la politique d'aménagement du territoire ;  
147 - Politique de la ville  
307 - Administration territoriale ;  
232 - Vie politique, culturelle et associative ;  
216 - Conduite et pilotage des politiques de l'intérieur ;  
162 - Interventions territoriales de l'État ;  
131 – Création ;  
143 - Enseignement technique agricole ;  
142 - Enseignement supérieur et recherche agricoles ;  
104 - Intégration et accès à la nationalité française ;  
309 - Entretien des bâtiments de l'État ;  
105 - Action de la France en Europe et dans le monde ;  
343 - Plan 'France Très haut débit' ;  
192 - Recherche et enseignement supérieur en matière économique et industrielle ;  
181 - Prévention des risques ;  
206 - Sécurité et qualité sanitaires de l'alimentation ;  
144 - Environnement et prospective de la politique de défense ;  
205 - Sécurité et affaires maritimes, pêche et aquaculture ;

- 197 - Régimes de retraite et de sécurité sociale des marins ;
- 344 - Fonds de soutien relatif aux prêts et contrats financiers structurés à risque ;
- 172 - Recherches scientifiques et technologiques pluridisciplinaires ;
- 193 - Recherche spatiale ;
- 177 - Prévention de l'exclusion et insertion des personnes vulnérables ;
- 304 - Inclusion sociale, protection des personnes et économie sociale et solidaire ;
- 157 - Handicap et dépendance ;
- 137 - Égalité entre les femmes et les hommes ;
- 158 - Indemnisation des victimes des persécutions antisémites et des actes de barbarie pendant la seconde guerre mondiale ;
- 129 - Coordination du travail gouvernemental ;
- 308 - Protection des droits et libertés ;
- 333 - Moyens mutualisés des administrations déconcentrées ;
- 119 - Concours financiers aux collectivités territoriales et à leurs groupements ;
- 122 - Concours spécifiques et administration ;
- 231 - Vie étudiante ;
- 219 – Sport ;
- 138 - Emploi outre-mer ;
- 123 - Conditions de vie outre-mer ;
- 183 - Protection maladie ;
- 215 - Conduite et pilotage des politiques de l'agriculture ;
- 175 – Patrimoines ;
- 111 - Amélioration de la qualité de l'emploi et des relations du travail ;

**Annexe 8 : Questionnaire aux responsables de programme LOLF  
Innovations pour tirer parti des données, « big data ».**

Questionnaire à l'attention des responsables de programme LOLF

1. Le big data vous paraît-il prometteur pour votre programme ? Avez-vous commencé à l'intégrer ?
2. Quelles sont vos attentes vis-à-vis des big data (économie, qualité de service, efficience, pilotage de politiques publiques ...) ?
3. Combien de temps vous donnez-vous pour vous approprier le big data ? Le porter à maturité ? Avec quels atouts et quelles faiblesses ?
4. Echangez-vous ou aimeriez-vous échanger avec des pairs sur le sujet ?
5. Etes-vous prêts à vous lancer dans un projet de big data ?
6. Quelles sont les questions que vous souhaiteriez aborder par le biais du big data ?
  - a. Avec quelle posture - expérimentation, lab, externalisation, ... ?
7. Avez-vous nommé un (ou des) responsable(s) de la politique big data ? de chacun des projets big data ?
8. Qu'attendez-vous des services du Premier ministre dans le cadre de la politique big data ? De la coopération interministérielle ? Européenne ou internationale ?
9. Qui nous conseillez-vous de rencontrer ?

Conseil général de l'économie

30 janvier 2015

Mission meilleures pratiques pour le « big data » et l'analytique dans l'administration

