Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ambassade de France au Royaume-Uni
Service Science et Technologie

N° 80

Science & Technologie au Royaume-Uni

Septembre/Octobre 2016

La révolution de la donnée au Royaume-Uni

1. Big Data & Data Science

Éditorial

par Cyrille Van Effenterre, Conseiller pour la science et la technologie

Il y a tout juste 25 ans, le 23 août 1991, l'ingénieur britannique Sir Tim Berners-Lee ouvrait l'accès du « *world wide web* » aux internautes du monde entier. Cet « *Internaut Day* », fêté comme il se doit ici, fait partie des hauts-faits en matière informatique du Royaume-Uni, qui se pose également comme berceau de l'informatique moderne grâce aux travaux d'Alan Turing, pionnier de la discipline et père de la machine révolutionnaire capable de décrypter le code nazi Enigma pendant la seconde guerre mondiale, comme le raconte le film récent *The Imitation Game*, et comme l'illustrent également plusieurs expositions à Bletchley Park, à Manchester ou au *Science Museum* de Londres.

Mais le Royaume-Uni est-il à la hauteur de sa réputation, de ses ambitions en matière de « *computing science* », et des enjeux économiques et sociaux qui y sont liés ? C'est la question que se posent de manière insistante les pouvoirs publics à Londres. C'est aussi celle à laquelle ce service a essayé d'apporter des éléments de réponse, en cherchant à identifier les principales institutions actives dans ce domaine, et à faire le point des évolutions en cours.

Si l'école britannique de mathématiques pures et appliquées est de niveau mondial, notamment à Cambridge, le Royaume-Uni connait en revanche une faiblesse récurrente de la formation mathématique générale dans le secondaire comme dans le supérieur. Il fait face à une crise des vocations scientifiques, et donc à de graves lacunes en matière de compétences dans ces domaines, l'ingénierie informatique en particulier. Il doit par conséquent fortement faire appel à des ressources humaines étrangères, comme l'illustrent par exemple la présence de nombreux ingénieurs français dans la mathématique financière à Londres.

La prise de conscience politique de cette situation est récente, et comme d'autres pays européens, le Royaume-Uni a réalisé l'importance de ce domaine pour son développement économique. Depuis 2010, la thématique des compétences informatiques, notamment sur le secteur émergent et très compétitif de la « science des données » (*data science*) constitue une vraie priorité nationale.

Dans le domaine de l'éducation, on a ainsi vu affiché le projet de positionner l'apprentissage du « code » comme première langue étrangère dès le primaire ! La mise en œuvre d'une telle politique n'aura pas de conséquences effectives avant dix ans au plus tôt, mais elle traduit la prise de conscience de l'enjeu par la classe politique, généralement plutôt issue de formations classiques, juridiques ou économiques.

De même, dans le domaine scientifique et technique, un grand nombre de documents stratégiques nationaux ont été établis avec l'aide des milieux scientifiques et économiques, et des plans d'actions volontaristes ont été mis en place et financés par le gouvernement ; les diverses structures d'innovation et de transfert créées depuis peu (notamment les centres Catapult) leur accorde ainsi une place prépondérante.

Le Royaume-Uni a ainsi à la fois investi dans la recherche universitaire, contribuant à faire de la *data science* une discipline académique à part entière, dans les superordinateurs et dans des infrastructures de réseau.

Le présent dossier aborde les questions relatives aux *big data*, *data science* et infrastructures (notamment dans le domaine du calcul haute performance). Un deuxième tome traitera de secteurs étroitement associés à la question des données, notamment l'internet des objets, la cybersécurité ou encore des questions de gouvernement numérique.

Eva Legras, volontaire internationale affectée au SST comme attachée scientifique adjointe en charge des nouvelles technologies de l'information et de la communication, a rédigé ce dossier de synthèse sur la base d'une importante documentation collectée depuis près de deux ans, sur un secteur en bouillonnement et dans un paysage en pleine évolution.

Science & Technologie au Royaume-Uni
Journal d'information du Service pour la Science et la Technologie de l'ambassade de France à Londres
ISSN 2042-7719

Septembre-Octobre 2016
Numéro 80
Directeur de la publication et rédacteur en chef
Cyrille van Effenterre
Responsable de la publication
Eva Legras

Dossier rédigé par :
Eva Legras

En couverture :
Network Analysis Vizualisation
Crédits : Martin Grandjean, CC-BY-SA 3.0

Nous contacter
Service Science et Technologie
Ambassade de France au Royaume-Uni
6 Cromwell Place
Londres SW7 2JN

Téléphone: (44) 207 073 13 80
Télécopie: (44) 207 073 13 90
info@ambascience.co.uk

Pour consulter la version électronique
www.ambascience.co.uk
Taux de change de la livre le 19/09/16
1 GBP = 1,31 USD
1 GBP = 1.17 EUR

www.ambascience.co.uk

Sommaire

1. Introduction 4

2. Le paradigme de la donnée au Royaume-Uni : opportunités et enjeux 5

2.1 Les données numériques : une galaxie complexe 5

2.2 Opportunités 6

2.3 Un enjeu sensible : les données personnelles 7

2.4 Stratégies publiques génériques 7

3. Stratégies de soutien au développement des e-infrastructures et aux *big data* 8

3.1 Acteurs publics concernés 8

3.2 E-Infrastructures 9

3.3 *Big Data* 11

4. Opérateurs de recherche : entre émergence d'un nouveau champ disciplinaire et pluralité des applications scientifiques 12

4.1 Acteurs généralistes en *big data* et *data science* 12

4.2 Acteurs en *big data* et *data science* appliqués à des secteurs spécifiques 15

5. E-infrastructures et calcul haute performance : un réseau en cours de consolidation 19

5.1 Diversité des équipements institutionnels 20

5.2 Consortia et logiques de regroupement 21

5.3 Equipements nationaux 22

5.4 Projets européens 25

5.5 Connectivité : le réseau académique Janet 26

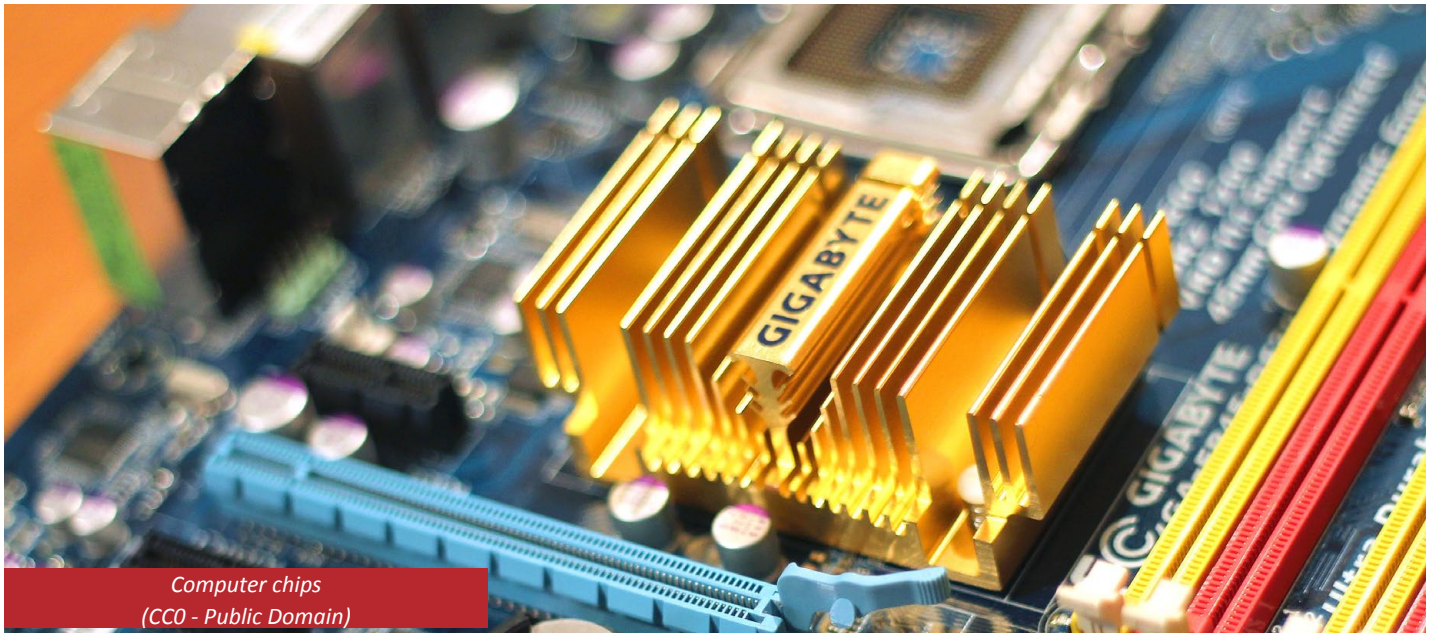
6. Du monde académique au monde économique : soutien au transfert de technologie 26

6.1 Accompagner l'innovation technologique : les centres Catapult 26

6.2 Développer l'écosystème écossais : les *Innovation Centres* 27

6.3 Créer des communautés de transfert : les réseaux KTN 28

7. Conclusion 28



Computer chips
(CC0 - Public Domain)

La révolution de la donnée au Royaume-Uni : Stratégies et acteurs de la recherche publique en *Big Data* & *Data Science*

90% des données disponibles ont été créées il y a moins de deux ans ; 2,5 milliards de gigabits de données sont générées chaque jour ; 200 millions d'emails sont envoyés, 4 millions de requêtes Google entrées, 2,46 millions d'articles Facebook partagés chaque minute. Depuis la création du World Wide Web en 1989 par Sir Tim Berners-Lee au CERN et la démocratisation des micro-ordinateurs dans les foyers à partir des années 1990, nos sociétés connaissent une profonde et rapide mutation liée à la multiplication des outils numériques et à la diversification des usages. De plus en plus petits, de plus en plus divers, de plus en plus courants, les périphériques stockant, partageant, analysant des données entièrement dématérialisées font désormais partie de notre quotidien.

1. Introduction

En sciences de l'information, la donnée[1] constitue la base du modèle DIKW (*Data Information Knowledge Wisdom*), fréquemment illustré sous forme de pyramide : la donnée est considérée comme un élément neutre en soi (chiffres, symboles, etc.), dont la valeur réside dans la mise en contexte et les usages qui en sont fait pour en extraire des informations (première étape d'interprétation). Les informations permettent à leur tour de créer des connaissances, qui influencent la prise de décision, et conduisent à l'acquisition d'une capacité de projection et de réflexion représentée par le dernier étage de cette pyramide.

L'explosion des données numériques, couplée à l'émergence de nouvelles techniques innovantes permettant de les exploiter massivement et à un moindre coût, confronte nos entreprises, sociétés et gouvernements à une nouvelle manne porteuse de promesses ambitieuses sous réserve de disposer des outils, méthodes et compétences permettant d'en retirer des informations utiles et exploitables. Ce potentiel lui a ainsi valu le qualificatif régulier de « nouveau pétrole » par les grands médias – bien que très loin de se tarir contrairement à celui-ci. Dans ce domaine comme dans d'autres, le Royaume-Uni entend imposer son *leadership* sur la scène internationale, et travaille activement depuis plusieurs années à se doter d'un écosystème public de recherche et développement favorable à de tels objectifs.

Figure 1 : Modèle DIKW.
Crédits : SST-Londres.

[1] Les données se définissent comme « principalement des faits ou des nombres, collectés pour être examinés, considérés et utilisés pour aider la prise de décision », ou d'éléments « sous une forme électronique pouvant être stockés et utilisés par un ordinateur » : http://dictionary.cambridge.org/dictionary/english/data

Le présent dossier ouvre une étude en deux parties visant à brosser un portrait synthétique de cet écosystème ; il ne s'agira pas de couvrir l'intégralité du sujet – très vaste ! – mais de s'attarder sur plusieurs secteurs étroitement liés à la donnée : nous traiterons ici des données massives ou *big data*, de *data science* ainsi que du calcul haute performance. Un deuxième dossier explorera des sujets connexes : l'internet des objets, en tant que nouvelle source de données promise à une croissance exponentielle à brève échéance, et la cyber-sécurité, cheville ouvrière de ce nouvel environnement. Nous y évoquerons également un domaine d'application en plein essor et appelé à profondément modifier nos structures politiques, le gouvernement numérique.

2. Le paradigme de la donnée au Royaume-Uni : opportunités et enjeux

2.1 Les données numériques : une galaxie complexe

2.1.1 Big Data et data science

L'expression « ***big data*** » (données massives ou méga-données en français) désigne à la fois des volumes de données gigantesques, devenues difficiles ou impossibles à traiter avec des systèmes de bases de données classiques, et les méthodes de traitement informatique qui permettent de les exploiter et d'en tirer des informations signifiantes en temps réel[2].

On associe couramment aux *big data* la règle des 3V :

- **Volume** de données produites et pouvant être traitées par les ordinateurs.
- **Variété** : agrégation de données hétérogènes, issues de nombreuses sources différentes : structurées ou non, ouvertes ou non, etc.
- **Vélocité** : il s'agit à la fois de la rapidité à laquelle les données sont à présent générées, et du temps de traitement des données, qui se rapproche du temps réel pour en maximiser les bénéfices.

D'autres caractéristiques lui sont associées, comme la **véracité** (exactitude de la donnée) et la **variabilité** (donnée susceptible d'être modifiée au cours du temps).

Le principal enjeu des *big data* est donc de parvenir à en gérer toute la complexité, que ce soit en termes de stockage, d'administration (collecte, traitement, etc.), ou encore de techniques et technologies permettant d'en extraire des informations utiles.

De nouvelles méthodes et techniques d'analyse (qui rejoignent la famille des *data analytics*) croisant des caractéristiques très diverses aussi bien sur le fond que sur la forme, permettent en effet de faire apparaître de nouvelles significations à partir de sources qui, prises isolément, auraient eu un intérêt plus limité. Ces techniques ont trois fonctions principales : (i) **descriptive** : comprendre un phénomène ; (ii) **prédictive** : anticiper les évolutions futures ; (iii) **prescriptive** : définir les évolutions que l'on souhaite voir se produire, via des simulations, des tests de scénarios, etc.[3]

Les techniques et technologies associées aux *big data* relèvent de plusieurs catégories[4] :

- **Techniques d'analyse de données :** *data mining*, clusterisation, apprentissage automatique (*machine learning*), statistiques, simulation, etc.;
- **Technologies spécifiquement associées aux enjeux des données de masse** : informatique en nuage (*cloud computing)*, systèmes distribués, bases de données relationnelles, etc.;
- **Techniques de visualisation**.

On désigne aujourd'hui couramment sous le terme de *data science*[5] (science des données) l'ensemble des compétences permettant d'extraire des informations à partir de l'analyse de données aussi bien structurées que non structurées. Les techniques employées relèvent de nombreuses disciplines, incluant *data mining*[6], statistiques, visualisation, programmation informatique, algorithmes ou encore probabilités.

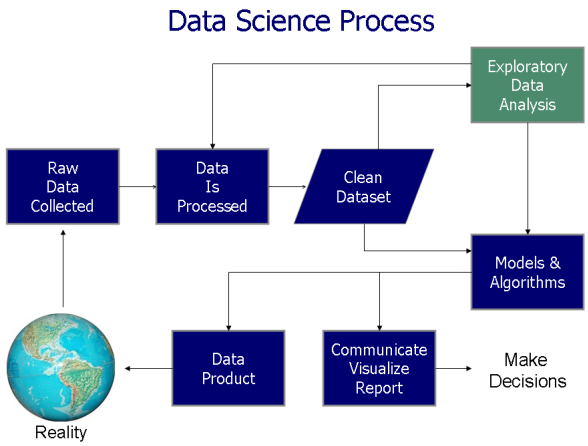


Figure 2 : The Data Science Process.
Crédits : Farecaster at English Wikipedia – CC-BY-SA.

2.1.2 E-infrastructures

Le terme d'e-infrastructure, très présent dans les stratégies publiques britanniques, désigne les réseaux, répertoires de données, ordinateurs, logiciels et compétences autorisant la collaboration et le calcul distribués, des simulations et analyses à grande échelle, et un accès rapide à de vastes jeux de données, des services d'analyse et de visualisation, ainsi que des infrastructures[7].

[2] HM Government Horizon Scanning Programme, *Emerging Technologies: Big Data*, 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf

[3] *Ibid.*

[4] McKinsey & Company, *Big Data : The next frontier for innovation, competition, and productivity,* 2011.

[5] Rappelons toutefois qu'il ne s'agit pas d'un concept nouveau. L'analyse statistique des données possède une histoire remontant au milieu du XXe siècle. Pour un rapide aperçu chronologique, cf. Gil Press, « A very Short History of Data Science », *Forbes*, mai 2013, http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#6e97412369fd

[6] Processus permettant de découvrir des corrélations signifiantes ou des tendances dans des jeux de données massives : http://www.gartner.com/it-glossary/data-mining/

[7] Prof. Dominic Tildesley, *A strategic vision for UK e-infrastructure,* 2011 ; UK Government, *E-Infrastructure: the ecosystem for innovation,* 2013.

L'accroissement exponentiel du volume de données génère plusieurs enjeux :

- **Stockage** : accès à des *data centers* et à des applications de *cloud computing* qui permettent un accès à distance aux capacités de stockage de ces derniers ;
- **Connectivité** : accès à un réseau suffisamment performant afin de transmettre et collecter des données depuis n'importe quelle localisation. Cela implique ainsi une politique de développement du réseau haut débit 4G ainsi qu'un support à la recherche dans le domaine de la 5G ;
- **Capacité de traitement :** les *big data* s'appuient sur les progrès réalisés en matière de calcul haute performance (*high performance computing* ou HPC). Indispensable à la réalisation de simulations complexes, le HPC est de plus en plus mobilisé pour l'analyse de larges quantités de données brutes. Il intègre notamment l'évolution des supercalculateurs, ces ordinateurs entièrement conçus pour optimiser leurs performances de traitement et dotés d'une architecture et de composants dédiés à cet objectif[8].

2.2 Opportunités

Le Royaume-Uni a identifié depuis plusieurs années les données comme une source de valeur sans précédent pour le pays. Plusieurs rapports et études parlementaires s'attardent sur les bénéfices économiques et en matière de gouvernance publique.

2.2.1 Economie

L'explosion des données constitue une immense source de valeur pour l'économie britannique : le *Centre for Economic and Business Research* a estimé que les *big data* pourraient générer la création de 58 000 nouveaux emplois entre 2012 et 2017, et contribuer à hauteur de 216 Mds£ à l'économie britannique. Le secteur manufacturier pourrait en être le plus important bénéficiaire, à hauteur de 45 Mds£[9].

Les applications sont multiples : modélisation de nouveaux produits dans l'ingénierie (cela se vérifie particulièrement dans le secteur automobile) ; suivi en temps réel des commentaires et réactions des clients, mais aussi des chaînes de production, ou encore des performances d'appareils pourvus de capteurs ; analyse des comportements à l'achat et exploitation pour la création de publicités ciblées.

L'exploitation des données produites par les entreprises elles-mêmes est également porteuse de valeur : la *charity* dédiée à l'innovation Nesta estime qu'un gain de productivité de 10% récompense les sociétés engagées dans ce processus[10].

Industrie	Bénéfices entre 2012 et 2017
Manutention	45,3 Mds£
Commerce	32,5 Mds£
Services professionnels	27,6 Mds£
Gouvernement central	20,4 Mds£
Santé	14,4 Mds£
Télécommunications	13,7 Mds£
Transports et logistique	12,4 Mds£
Banque de détail	6,4 Mds£
Energie	5,4 Mds£
Banque d'investissement	5,3 Mds£
Assurance	4,6 Mds£
Autres activités	27, 9 Mds£
Total économie britannique	216 Mds£

Tableau 1 : Bénéfice économique des big data par secteur d'activité. Source : Centre for Economic and Business Research, « Data equity, unlocking the value of Big Data », 2012.

2.2.2 Gouvernance

Le *Centre for Economic and Business Research* estime en 2012 que le gouvernement de Westminster pourrait générer un bénéfice de 20 Mds£ sur une période de 5 ans grâce à l'exploitation des données massives. La détection des fraudes et la réduction des coûts engendrés sur le secteur public pourraient générer 2 Mds£ d'ici 2017, et l'analyse des données de performance pourrait améliorer les institutions, en particulier le système de santé, selon la même étude[11].

La transparence des actions publiques, une rationalisation des services via par exemple la réutilisation de données individuelles déjà enregistrées par ailleurs, ou encore le développement d'algorithmes automatiques pour l'analyse de données massives sont autant de bénéfices potentiels discutés[12]. L'essor de l'ouverture des données publiques (*open data*), afin de les rendre librement réutilisables sans contrepartie, est étroitement associé à ces derniers.

2.2.3 Recherche

La recherche scientifique bénéficie de l'analyse massive des données sur de nombreux secteurs, à des fins de simulation, modélisation, etc. Le rapport parlementaire *The Big Data Dilemna* identifie trois exemples significatifs : la santé, l'analyse de données médicales massives permettant d'améliorer les traitements délivrés et de développer des protocoles plus efficaces et personnalisés ; la physique expérimentale, pour traiter les énormes quantités de données générées par des infrastructures telles que les accélérateurs de particules ; ou encore la météorologie, via notamment le travail du *Met Office*, afin de modéliser le changement climatique et son impact.

[8] Yan-Tarō Clochard, *Calcul haute performance au Japon*, rapport d'ambassade, 2016.

[9] House of Commons Science and Technology Committee, *The Big Data Dilemna,* 2016.

[10] *Ibid.*

[11] Centre for Economic and Business Research, *Data equity, unlocking the value of Big Data,* 2012.

[12] HM Government Horizon Scanning Programme, *Emerging Technologies: Big Data,* 2014.

2.2.4 *Société*

Sous l'essor conjoint des services numériques personnalisés, des applications mobiles et de l'internet des objets qui petit à petit vise à connecter et « faire parler » les accessoires de notre quotidien, le nombre de données personnelles potentiellement exploitables est en pleine explosion.

Elles offrent ainsi une ressource de premier plan pour l'application de techniques de segmentation, permettant d'identifier les habitudes et préférences de groupes d'individus en fonction de nombreux critères et ainsi de proposer des offres de service extrêmement ciblées. Si l'on pense en priorité aux plateformes telles qu'Apple, Amazon ou Google, les applications seront amenées à se diversifier. Le secteur public est ainsi susceptible de s'en emparer, comme par exemple dans le domaine de l'énergie avec les compteurs électriques « intelligents »[13].

2.3 Un enjeu sensible : les données personnelles

Les opportunités identifiées au paragraphe précédent sont également porteuses de risques, notamment en matière de sécurité et de confidentialité des données. Il s'agit d'un enjeu sensible aussi bien pour les organisations[14], cibles potentielles de cyberattaques, que pour les individus.

Les données personnelles alimentent de plus en plus le mouvement des *big data,* bien qu'encore minoritaires eu égard à l'ensemble des données dont il est question[15]. Des études ont néanmoins mis en évidence les préoccupations du grand public britannique à ce sujet, notamment le manque de contrôle sur leurs propres données ou encore les risques de discrimination (par exemple par des recruteurs ayant obtenu un accès à des données médicales)[16]. Le nombre de plus en plus important de données relatives à la vie privée augmente en outre le risque d'infraction au respect de l'intimité des individus en cas d'exploitation dévoyée, qu'elles soient collectées ou vendues sans l'accord de leurs propriétaires, ou encore croisées avec d'autres sources en dehors des usages initialement prévus.

Au Royaume-Uni, le *Data Protection Act* de 1998, qui implémente la directive européenne 95/46/EC, constitue toujours le cadre officiel pour la gestion des données personnelles. Selon ce texte, toute donnée pouvant être utilisée pour identifier un individu relève de cette catégorie. Les entreprises détenant de telles données ont l'obligation de les tenir exactes et à jour, de ne pas les conserver plus longtemps que le délai requis par le service offert, et d'obtenir le consentement des usagers pour l'utilisation qui en est prévue. Elles ont également l'obligation de fournir toutes les données personnelles qu'elles détiennent sur un individu si celui-ci en fait la demande[17].

Le récent accord européen « *General Data Protection Regulation* » (abrégé GDPR)[18] devait mettre à jour cette réglementation à horizon 2018 au Royaume-Uni afin de prendre en compte les nouvelles réalités nées de l'essor des technologies numériques. Le résultat du référendum en faveur d'une sortie du pays de l'Union Européenne entraîne néanmoins une incertitude sur l'implémentation future de cette directive dans le droit britannique.

L'organe en charge de vérifier le respect des réglementations en matière de protection des données est l'***Information Commissioner's Office***. Indépendant du gouvernement, il s'est tout particulièrement intéressé à la question des *big data* dans le cadre d'une étude publiée en 2014[19] où il a examiné les différents enjeux relatifs à la protection des données personnelles dans ce contexte particulier, et la façon dont ils s'insèrent au sein de la réglementation en place.

Ces questions d'éthique ont fait l'objet d'un intérêt particulier au sein de la récente enquête sur les *big data* de la commission parlementaire de la Chambre des Communes pour la science et la technologie, qui a recommandé au gouvernement d'établir un conseil dédié ; le gouvernement a répondu favorablement à cette recommandation et annoncé l'organisation de cette entité au sein de l'*Alan Turing Institute*, porte-drapeau de la recherche britannique en *data science* (cf. §4.1.1).

2.4 Stratégies publiques génériques

Le gouvernement définit couramment ses ambitions technologiques au sein de documents stratégiques orientés autour d'un secteur particulier. Les technologies numériques ont fait l'objet de plusieurs rapports depuis de nombreuses années, tels que *Digital Britain*, qui en 2009 annonçait l'intention du *Department for Business, Innovation and Skills* (BIS) de moderniser les infrastructures réseaux du Royaume-Uni et développer un environnement favorable à l'innovation sur ce secteur[20].

Les stratégies les plus récentes pour le sujet qui nous intéresse sont aujourd'hui les suivantes :
- La ***Cyber Security Strategy*** (2011)[21] définit les objectifs du gouvernement pour la protection des infrastructures et des données numériques jusqu'en 2016.
- L'***Information Economy Strategy*** (2013)[22] définit les objectifs du gouvernement pour favoriser le marché des infrastructures, services et données numériques.
- La ***Government Digital Strategy*** (2013)[23] définit les objectifs du gouvernement en matière de transformation numérique de ses services.

[13] *Ibid*.

[14] HM Government, *Cyber Security Breaches Survey 2016, main report*, 2016 : 65% des grands groupes et 24 % de l'ensemble des entreprises du pays ont détecté de telles failles ou attaques en 2015, pour des dommages pouvant s'élever jusqu'à 3 M£.

[15] House of Commons Science and Technology Committee, *The Big Data Dilemna,* 2016.

[16] Parliamentary Office of Science and Technology, *Big Data : an overview,* 2014. Voir également : http://www.sciencewise-erc.org.uk/cms/public-views-on-big-data/

[17] *Id., Big Data in Business,* 2014.

[18] Le GDPR est approuvé par la Commission Européenne en décembre 2015 et acté par la régulation 2016/679 et la directive 2016/680. Cf. http://ec.europa.eu/justice/data-protection/reform/index_en.htm

[19] Information Commissioner's Office, *Big Data and data protection,* 2014 https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

[20] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228844/7650.pdf

[21] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60961/uk-cyber-security-strategy-final.pdf

[22] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206944/13-901-information-economy-strategy.pdf

[23] https://www.gov.uk/government/publications/government-digital-strategy/government-digital-strategy

- L'***Open Data Strategy*** (2014)[24] définit les objectifs du gouvernement pour l'ouverture des données publiques.

Il est à noter que plusieurs documents doivent être définis ou renouvelés d'ici la fin de l'année 2016 pour ce qui relève du numérique et des données.

La très attendue ***National Digital Strategy*** aura pour but de définir les orientations et prochaines étapes pour le développement de l'économie et de l'innovation numérique au Royaume-Uni. Les données sont bien entendu appelées à y occuper une place de premier choix (non publiée au moment de la rédaction de ce dossier).

Le ***National Innovation Plan*** aura pour but de définir comment le Royaume-Uni peut améliorer son système de soutien à l'innovation. Une vaste consultation publique a été lancée entre avril et mai 2016, au sein de laquelle les données ont fait l'objet d'un axe d'enquête spécifique[25] (non publié au moment de la rédaction de ce dossier).

La ***National Cyber Security Strategy***[26] présente une nouvelle feuille de route pour les cinq prochaines années en matière de cyber sécurité afin de faire suite à la dernière en date, publiée en 2011.

Sur le plan législatif, le gouvernement a dévoilé début juillet 2016 son projet de ***Digital Economy Bill***[27], qui vise à légiférer sur la consommation de services numériques, le développement de services publiques en ligne, ou encore l'amélioration de la couverture réseau du pays.

3. Stratégie de soutien au développement des e-infrastructures et aux *big data*

3.1 Acteurs publics concernés[28]

Le **Department for Business, Energy and Industrial Strategy** (BEIS, anciennement *Department for Business, Innovation and Skills* ou BIS) est le ministère en charge des investissements publics pour le développement économique, la recherche, l'innovation et l'énergie. A ce titre, il supervise plusieurs structures d'accompagnement et de financement, dont un certain nombre sont amenées à intervenir sur le secteur de la donnée. Le Secrétaire d'état aux Universités, à la Science, à la Recherche et à l'Innovation, Jo Johnson, se situe sous la double tutelle du *Department for Education* pour ce qui relève de l'enseignement supérieur, et du BEIS pour ce qui relève de la recherche et des sciences.

Innovate UK (ex *Technology Strategy Board*) est l'agence publique de financement placée sous la tutelle du BEIS, dont la mission est de soutenir le développement de projets innovants et leur accès au marché. Parmi ses priorités figurent les technologies émergentes et habilitantes (*enabling technologies*) intégrant les technologies numériques. Innovate UK est ainsi notamment en charge des Centres Catapult[29], établissements créés sur le modèle des instituts Fraunhofer, parmi lesquels 4 interviennent à des degrés divers sur le sujet de ce dossier (cf. §6 pour plus de détails).

Research Councils UK rassemble les 7 conseils de recherche finançant la recherche britannique : *Arts and Humanities Research Council* (AHRC) ; *Biotechnology and Biological Sciences Research Council* (BBSRC) ; *Economic and Social Research Council* (ESRC) ; *Engineering and Physical Sciences Research Council* (EPSRC) ; *Natural Environment Research Council* (NERC) ; *Medical Research Council* (MRC) et *Science and Technology Facilities Council* (STFC). La thématique des données, de par son caractère transdisciplinaire, est traitée sous un aspect ou un autre par l'ensemble de ces agences au travers des projets et infrastructures qu'elles financent[30]. On peut toutefois citer l'EPSRC et le STFC parmi les agences les plus engagées sur ce terrain (cf. §4 et 5).

Le **Government Office for Science** (GO-Science) est un département rattaché au Cabinet Office. Dirigé par le conseiller scientifique en chef du gouvernement (*Government Chief Scientific Adviser)*, il a pour mission de produire des études prospectives et d'animer des communautés d'experts afin de conseiller le gouvernement sur ses orientations générales en matière de science et technologie. Il anime également le réseau des conseillers scientifiques (*chief scientific advisers*) déployés dans les divers ministères du gouvernement. Sur le secteur du numérique, il a joué ces dernières années un rôle prescripteur de premier plan à propos de l'Internet des Objets[31], de la FinTech[32] ou encore des technologies de registre distribué[33] (dont fait partie la fameuse *blockchain* derrière la monnaie virtuelle *bitcoin*).

Co-présidé par le conseiller scientifique en chef du gouvernement et un expert indépendant, le **Council for Science and Technology** est un organe consultatif non-gouvernemental qui a pour mission de conseiller le premier ministre sur des problématiques de politique scientifique et technologique. Il rassemble une vingtaine d'experts et chercheurs de très haut niveau (présidents de sociétés savantes, vice-chanceliers d'universités, etc.). Le CST a joué un rôle consultatif et prescripteur de premier plan dans la définition de la politique du gouvernement en faveur des e-infrastructures et des *big data* (cf. §3.2 et 3.3).

Le parlement britannique s'est emparé de ces questions par l'intermédiaire de ses commissions parlementaires (*committees*), qui produisent des enquêtes et des recommandations au gouvernement sur une grande variété de thèmes incluant la science. La commission pour la Science

[24] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/330382/bis-14-946-open-data-strategy-2014-2016.pdf

[25] https://www.gov.uk/government/consultations/national-innovation-plan-call-for-ideas

[26] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf

[27] https://www.gov.uk/government/collections/digital-economy-bill-2016

[28] Pour plus de précisions sur le paysage général public de la recherche et développement au Royaume-Uni, nous vous invitons à consulter la fiche de présentation publiée sur notre site à l'adresse suivante : http://www.ambafrance-uk.org/La-Recherche-au-Royaume-Uni-13433

[29] SST Royaume-Uni, *Les Centres Catapult*, rapport d'ambassade, 2015, ,http://www.ambafrance-uk.org/Dossier-Les-centres-Catapult

[30] Tous les projets financés par les agences publiques britanniques sont accessibles via le portail *Gateway to Research* : http://gtr.rcuk.ac.uk/

[31] Sir Mark Walport, *The Internet of Things. Making the most of the second Digital Revolution,* 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/409774/14-1230-internet-of-things-review.pdf

[32] Sir Mark Walport, *FinTech Futures, The UK as a World Leader in Financial Technologies,* 2015, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/413095/gs-15-3-fintech-futures.pdf

[33] Sir Mark Walport, *Distributed Ledger Technology : Beyond blockchain,* 2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf

et la Technologie de la Chambre des Communes (*House of Commons Science and Technology Select Committee*) a publié récemment plusieurs rapports sur les nouvelles technologies : « *Responsible use of data* » (novembre 2014), se penche sur la question des données récoltées par les réseaux sociaux ; « *The big data dilemna* » (avril 2016), examine les enjeux et risques des *big data* pour l'économie britannique ; « *Digital skills crisis* » (juin 2016), revient sur la crise des compétences professionnelles de haut niveau en *data science* au Royaume-Uni.

La Commission pour la Science et la Technologie de la Chambre des Lords, quant à elle, s'est intéressée à l'avenir numérique du pays dans son « *Make or Break: Britain's digital future* » (février 2015).

La Commission pour la Culture, les Médias et le Sport de la Chambre des Communes a pour sa part récemment enquêté sur la sécurité des données personnelles en ligne dans « *Cyber Security: Protection of Personal Data Online* » (juin 2016).

3.2 E-Infrastructures

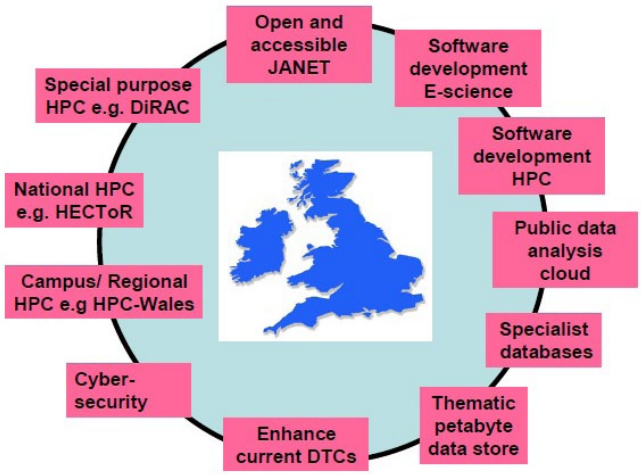


Figure 3 : Proposition d'e-infrastructure pour le Royaume-Uni
Source : Prof. Dominic Tildesley, A strategic vision for UK e-Infrastructure, 2012.
Crédits : Open Government Licence.

3.2.1 Gouvernement

Un nombre significatif d'études sur les e-infrastructures britanniques ont été produites au tournant des années 2010[34], témoignant de l'intérêt soutenu des pouvoirs publics pour ces technologies.

Le focus actuel du gouvernement sur ces équipements fait toutefois plus directement suite à un rapport intitulé *A Strategic Vision for UK e-Infrastructure*[35], commandé par le Secrétaire d'état pour les universités et la science de l'époque, David Willetts, et remis en 2011. Ecrit par le professeur Dominic Tildesley, ce document pointait notamment du doigt le retard pris par le Royaume-Uni par rapport à des pays comme le Japon, l'Allemagne ou la France, et proposait de mettre en place une feuille de route étalée sur 10 ans afin de définir précisément les divers éléments des infrastructures requises : réseaux, données et stockage, calcul, logiciels et algorithmes, sécurité et authentification, personnes et compétences (Figure 3).

Faisant suite à ces diverses recommandations, le gouvernement décide d'investir massivement dans le développement d'équipements de calcul haute performance : 160 M£ sont débloqués en 2011 pour la mise en place de 5 centres régionaux autour de superordinateurs hébergés par les universités locales, en coordination avec le supercalculateur national ARCHER localisé à l'université d'Edimbourg. Ce financement permet également de soutenir la fondation du *Hartree Centre*, un autre supercalculateur administré par le conseil de recherche STFC (*Science and Technology Facilities Council*), ainsi que la création de consortia régionaux (cf. partie 5).

Projet HPC	Financeur	Montant (M£)
National Service	EPSRC, NERC	43
Hartree Centre	STFC	30
DiRAC	STFC	15
GRIDPP	STFC	3
The Genome Analysis Centre (auj. Earlham Institute)	BBSRC	8
Monsoon	NERC/Met Office	1
JASMIN2 et CEMS	NERC et UKSA	7,75
Regional Centres : N8, SES5, MID+, HPC Midlands, ARCHIE-WeSt	EPSRC	6,5
Réseau JANET et Authentification Moonshot	Jisc	31
HPC Data Storage	EPSRC, STFC	15
Total		160

Tableau 2 : Ventilation de la première tranche de financements du gouvernement en faveur des infrastructures HPC (2011).
Source des données : UK e-infrastructure for research, JISC, 2015.

Une deuxième vague de financements à hauteur de 189 M£ est annoncée en 2012 pour le HPC et l'informatique à faible consommation d'énergie ; cette enveloppe a permis de financer davantage d'infrastructures, ainsi que des projets de *big data* en sciences humaines et sociales ou encore dans les sciences du vivant (Tableau 3).

[34] http://www.rcuk.ac.uk/research/xrcprogrammes/otherprogs/einfrastructure/

[35] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32499/12-517-strategic-vision-for-uk-e-infrastructure.pdf

Projet HPC	Financeur	Montant (M£)
Digital transformations in arts and humanities	AHRC	8
E-infrastructures for biosciences	BBSRC	13
Research data facility and software development	EPSRC	8
Administrative data centres	ESRC	36
Understanding populations	ESRC	12
Business datasafe	ESRC	14
Biomedical informatics	MRC	55
Environmental virtual observatory	NERC	13
Square Kilometre Array	STFC	19
Energy Efficiency Computing Hartree Centre	STFC	19
Total		189

Tableau 3 : Ventilation de la deuxième tranche de financements du gouvernement en faveur des infrastructures HPC (2012). Source des données : UK e-infrastructure for research, JISC, 2015.

Un organe consultatif, l'**E-Infrastructure Leadership Council** est mis en place la même année[36]. Co-présidé par le Secrétaire d'état aux Universités et à la Science, ce groupe rassemble des professionnels issus des milieux de la recherche académique, du gouvernement et de l'industrie. Il a vocation à conseiller le gouvernement sur toute question relative aux e-infrastructures : réseaux, ordinateurs, applications, etc. Il publie plusieurs rapports dont un état des lieux en 2013 des progrès réalisés et des prochaines étapes envisagées[37].

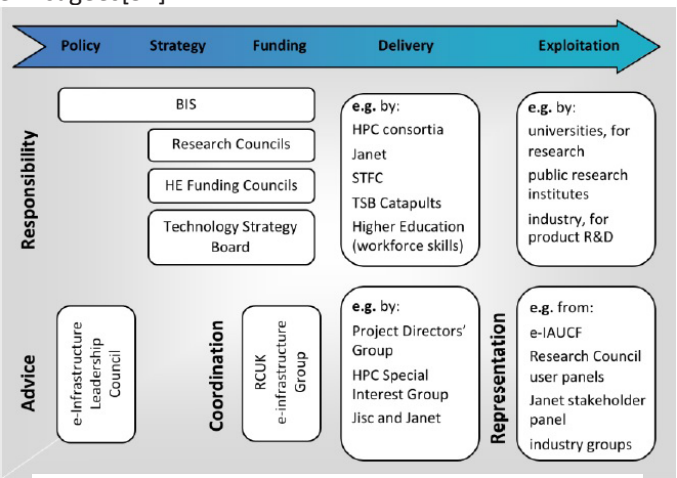


Figure 4 : The e-infrastructure pipeline. Source: HM Government, E-infrastructure : The ecosystem for innovation, octobre 2013. Crédits : Open Government Licence.

[36] HM Government, *E- infrastructure: The ecosystem for innovation*, 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249474/bis-13-1178-e-infrastructure-the-ecosystem-for-innovation-one-year-on.pdf

[37] HM Governement, *E-Infrastructure: the ecosystem for innovation. One Year On*, 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249474/bis-13-1178-e-infrastructure-the-ecosystem-for-innovation-one-year-on.pdf

En 2014-2015, une troisième vague de financements publics, à hauteur de 257 M£ alimente trois projets majeurs : la création du *Centre for Cognitive Computing* au *Hartree Centre* (115 M£ complétés par 230 M£ d'IBM) ; un nouveau supercalculateur pour le *Met Office* (100 M£) ; et pour finir, le centre de recherche en *data science* Alan Turing Institute (42 M£)[38].

3.2.2 Agences, groupes & réseaux

D'autres acteurs jouent un rôle de premier plan dans la définition d'orientations stratégiques.

Le **groupe d'intérêt spécialisé HPC-SIG** (*HPC Special Interest Group*) a été lancé en 2005 pour soutenir le développement du calcul haute performance au Royaume-Uni (lobbying, partage de bonnes pratiques, communication, etc.). Il rassemble entre autres des professionnels issus des services informatiques d'universités, ainsi que des représentants de conseils de recherche.

LE RCUK dispose d'un **National e-Infrastructure Group** (*NeI Group*) coordonnant la stratégie e-infrastructure entre les conseils de recherche, Innovate UK, le Jisc[39] et le *Met Office*[40].

RCUK[41] et **l'EPSRC**[42] ont publié deux feuilles de route en 2014, organisant leurs objectifs autour de 4 thèmes : (i) infrastructure, regroupant logiciels, infrastructures de données et équipements informatiques ; (ii) connections, à savoir réseaux utilisés et protocoles ou standards assurant la sécurité des systèmes ; (iii) capacités de recherche et développement, incluant compétences professionnelles, développement d'expertise et enseignement ; (iv) intégration des divers acteurs, communautés et éléments de l'écosystème.

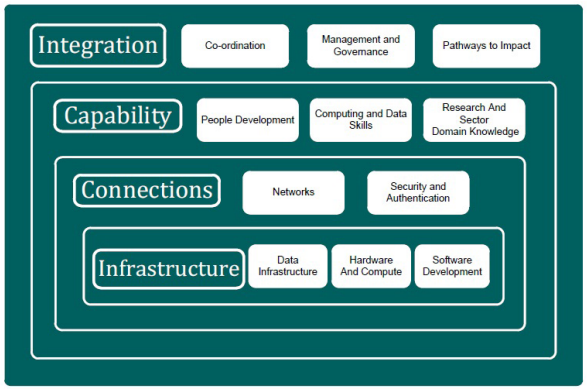


Figure 5 : E-Infrastructure Themes and Elements. Source: HRCUK, E-Infrastructure Roadmap, 2014.

[38] Michael Ball et al., *UK e-Infrastructure for Research*, Jisc, 2015, http://www.slideshare.net/comth/uk-einfrastructure-for-research-ukusa-hpc-workshop-oxford-july-2015

[39] Anciennement *Joint Information Systems Committee*, le Jisc est une *charity* qui délivre des services numériques aux établissements d'enseignement supérieur et de recherche.

[40] http://hpc-sig.org.uk/2016/03/08/2015-national-e-infrastructure-survey-report/

[41] http://www.rcuk.ac.uk/research/xrcprogrammes/otherprogs/einfrastructure/

[42] https://www.epsrc.ac.uk/research/ourportfolio/themes/researchinfrastructure/subthemes/einfrastructure/strategy/roadmap/

Les objectifs visés incluent : soutien au développement d'une offre de formation efficace dans les compétences requises (formation continue, PhDs) ; promotion de perspectives professionnelles claires et attractives dans ce champ disciplinaire ; promotion de l'établissement d'un réseau national de centres en *data analytics* ; soutien aux montées en charge des équipements et amélioration des infrastructures réseau ; standardisation des politiques de sécurité ; favoriser l'appropriation des technologies HPC par les entreprises britanniques ; poursuite des investissements dans les équipements publics.

Les conseils de recherche BBSRC, MRC, NERC et STFC, s'ils n'ont pas produit de feuille de route spécifique comme l'EPSRC, ont néanmoins engagé divers investissements pour le développement d'infrastructures adaptées à l'analyse de données massives. Ces initiatives sont détaillées dans les paragraphes suivants.

3.3 Big Data

Au-delà des infrastructures, le gouvernement s'est penché plus spécifiquement sur la question des *big data* et de leur potentiel pour l'économie britannique.

3.3.1 Financement de la recherche et de l'innovation : 8 + 2 grandes technologies

La politique de soutien à la recherche et à l'innovation sur le secteur du numérique s'appuie au Royaume-Uni sur « huit grandes technologies » identifiées en 2013 par le Secrétaire d'état aux Universités et à la Science de cette époque, David Willetts, comme spécifiquement porteuses de progrès et de croissance économique pour le pays. Dotée d'un budget initial de 600 M£, cette orientation, inscrite dans la stratégie industrielle édictée par le gouvernement en 2013, vise à soutenir financièrement le lancement de projets ou d'infrastructures de grande envergure afin de positionner le pays en situation de *leadership* sur ces secteurs[43]. Les *big data*[44] et l'informatique à économie d'énergie se positionnent en tête de cette liste, dans laquelle on trouve également la robotique et les systèmes autonomes (également générateurs de nombreuses données numériques), mais aussi l'internet des objets et les technologies quantiques, ces deux dernières ayant été ajoutées *a posteriori*.

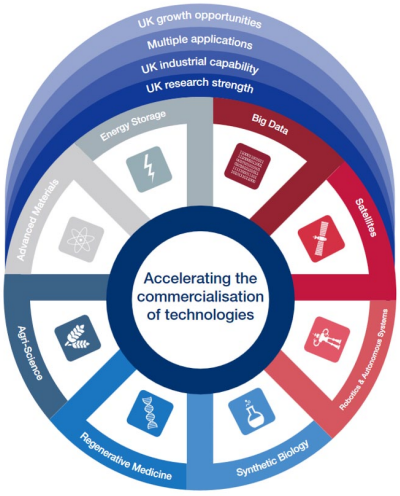


Figure 6 : 8 Great Technologies.
Source : HM Government, 2013.
Crédits : Open Government License.

3.3.2 « The data opportunity » : s'emparer de la révolution de la donnée

Le gouvernement définit en outre la même année une stratégie spécifiquement dédiée à la question des données[45] organisée autour de 11 objectifs opérationnels. Allant jusqu'à parler de « *data revolution* », on y retrouve les ingrédients fondamentaux de l'approche publique autour de la donnée :

- **Formation de professionnels hautement qualifiés** : rappelant plusieurs études identifiant un déficit de professionnels en *data science* et l'augmentation prévue de la demande au cours des années à venir, le gouvernement affirme notamment sa volonté de travailler en étroite collaboration avec divers acteurs (*Universities UK*, Nesta, *Open Data Institute*, etc.) afin d'identifier précisément la nature des compétences en déficit et définir des espaces de collaboration avec l'industrie.
- **Soutien aux infrastructures pour les maintenir à la pointe des dernières évolutions technologiques** : le gouvernement affirme également son intention d'attirer des investissements étrangers et de favoriser l'accès aux e-infrastructures nationales, notamment dans le domaine du HPC, pour les entreprises présentes sur une série de secteurs stratégiques ainsi que pour les PMEs (Petites et Moyennes Entreprises).
- **Soutien à la recherche et développement** par l'intermédiaire de l'EPSRC notamment.
- **Partage et exploitation de données** issues de sources diverses pour la recherche ou le secteur privé, ainsi que l'autorisation légale du *text and data mining* pour des applications non-commerciales.

Le ***Council for Science and Technology*** a joué un rôle incitatif majeur en amont de cette stratégie, écrivant aux ministres concernés par la *data science* une lettre intitulée « *The Age of Algorithms* » dans laquelle il énonce une série de recommandations parmi lesquelles on trouve la première référence à l'établissement d'un institut de recherche national dédié aux données – le futur *Alan Turing Institute* (cf. § 4.1.1)[46].

Le CST a également produit en 2015 une cartographie des principales institutions de recherche et infrastructures spécialisées dans le cadre de sa série « *Science Landscape Seminar Series* »[47], destinée à alimenter un projet de cartographie générale du paysage de la recherche scientifique britannique[48]. L'aperçu des acteurs présenté à la partie suivante s'appuie sur ce travail.

Rappelons pour finir les investissements consentis par plusieurs conseils de recherche afin de développer des infrastructures adaptées à l'administration et à l'analyse des données massives[49]. On peut ici signaler la contribution de l'AHRC, dédié aux arts et aux humanités, qui identifie les

[43] https://www.gov.uk/government/publications/eight-great-technologies-infographics

[44] Le bureau de la propriété intellectuelle du gouvernement a publié une étude de l'activité R&D britannique sur ce secteur à partir de l'analyse des brevets: Intellectual Property Office, *Big Data : A patent overview,* 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325024/informatics-bigdata.pdf

[45] HM Government, *Seizing the data opportunity, a strategy for UK data capability,* octobre 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

[46] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224953/13-923-age-of-algorithms-letter-to-prime-minister__1_.pdf

[47] https://www.gov.uk/government/publications/science-landscape-seminar-big-data-e-infrastructure-and-supercomputing

[48] https://www.gov.uk/government/collections/uk-knowledge-landscape

[49] http://www.rcuk.ac.uk/research/infrastructure/big-data/

big data comme un objet de financement majeur au sein de son « *Digital Transformation Theme* ». Il reçoit ainsi en 2013 une enveloppe de 4,6 M£ pour financer une série de 21 projets de recherche incluant cette dimension dans le domaine des arts et des humanités[50].

4. Opérateurs de recherche : entre émergence d'un nouveau champ disciplinaire et pluralité des applications scientifiques

Le paysage de la recherche britannique dans le domaine des *big data* et *data science* repose en grande partie sur l'héritage d'instituts et centres de recherche spécialisés de longue date en informatique et statistique, tels que les départements universitaires d'informatique ou encore des structures à vocation transdisciplinaire comme l'*E-Research Centre* à Oxford, qui intègrent donc les nouvelles méthodes et technologies appliquées aux données numériques.

On note néanmoins l'émergence récente de plusieurs infrastructures davantage spécialisées sur ce secteur. Celles-ci bénéficient à la fois des nouvelles priorités stratégiques du gouvernement, et de l'intérêt croissant du monde industriel se traduisant par l'ouverture de laboratoires ou le développement de projets en partenariat avec des universités, afin de développer des solutions adaptées aux dernières problématiques du marché.

Il convient ainsi de faire la distinction entre deux grandes familles d'acteurs dédiés à la recherche dans le domaine des *big data* et *data science* : les progrès technologiques de ces dernières années suscitent la création de départements et laboratoires spécifiquement dédiés à la science des données, qu'il s'agisse de l'aborder sous un angle généraliste ou bien, considérant son caractère transdisciplinaire, de s'y consacrer sous un angle plus spécifique. Cette dernière caractéristique implique également que de plus en plus de structures incluent désormais à leurs activités le traitement et l'analyse de données massives afin de soutenir leurs projets de recherche, et, de fait, s'imposent aussi comme des acteurs à part entière dans ce domaine.

Il ne s'agira donc pas ici de présenter une liste exhaustive, mais de s'attarder sur quelques instituts et organismes emblématiques afin d'illustrer la diversité des applications existantes.

4.1 Acteurs généralistes en *big data* et *data science*

4.1.1 A l'échelle nationale : l'Alan Turing Institute, catalyseur de l'expertise britannique en data science

Annoncé par le chancelier de l'Echiquier dans le budget du gouvernement en mars 2014, l'*Alan Turing Institute* a vocation à centraliser l'expertise britannique sur la science des données.

Il s'agit d'une *joint venture* entre 5 universités sélectionnées à l'issue d'un appel à candidatures lancé à l'été 2014 : Cambridge, Edimbourg, Oxford, UCL, Warwick, et le conseil de recherche *Engineering and Physical Sciences Research Council* (EPSRC).

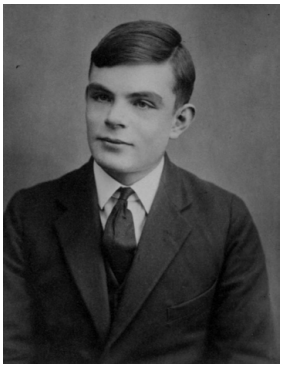


Figure 7 : Alan Turing à 16 ans. Crédits: Domaine Public.

Son nom rend hommage au célèbre mathématicien qui déchiffra le code de la machine nazie Enigma au cours de la seconde guerre mondiale, et est considéré comme l'un des pères de l'informatique contemporaine.

Localisé au sein de la *British Library*, au cœur du *Knowledge Quarter* londonien[51], ce nouvel établissement se donne quatre grands objectifs : (i) développer son activité de recherche fondamentale en *data science* ; (ii) établir des partenariats stratégiques afin de soutenir les applications de ses travaux de recherche ; (iii) assumer une activité d'enseignement afin de former la prochaine génération de *data scientists* ; (iv) intervenir dans les grands débats publics et de société relatifs aux données.

Le gouvernement a annoncé un investissement de 42 M£ sur cinq ans, auxquels s'ajoutent les contributions des universités partenaires (5 M£ chacune), et des partenaires privés recherchés par l'Institut.

Il accueille un total de 87 chercheurs associés issus des 5 universités fondatrices (13 de Cambridge, 13 de UCL, 32 d'Oxford, 12 de Warwick et 17 d'Edimbourg)[52].

S'il a été inauguré officiellement en novembre 2015, il a démarré ses activités plusieurs mois en amont en organisant des ateliers et séminaires thématiques destinés à explorer les divers aspects des *data sciences* et à élaborer son plan d'action[53]. A l'issue de ce travail, l'ATI a rendu publiques ses orientations qui s'organisent autour de quatre domaines thématiques : représentations mathématiques, inférence et apprentissage, systèmes et plateformes, compréhension du comportement humain ; et 6 secteurs prioritaires d'application de la recherche : ingénierie, technologie, défense et sécurité, villes intelligentes, services financiers, santé et bien-être.

La priorité accordée aux services financiers a d'ores et déjà été soulignée par la volonté du gouvernement de soutenir la recherche sur les monnaies numériques et les technologies associées, telles que la *blockchain* et les technologies de registre distribué. Un budget de 10 M£ a été annoncé en octobre 2015 pour que l'*Alan Turing Institute* collabore

[50] http://www.ahrc.ac.uk/documents/publications/the-challenges-of-big-data/

[51] Le Knowledge Quarter est un programme visant à valoriser les quartiers londoniens de King's Cross et Euston où se concentrent plusieurs dizaines de prestigieuses institutions scientifiques, culturelles, médiatiques et de recherche. Il rassemble 74 partenaires parmi lesquels figurent, outre la British Library, le British Museum, la University College of London, le Wellcome Trust, ou encore le quotidien Guardian : http://www.knowledgequarter.london/

[52] http://turing.ac.uk/faculty-fellows/

[53] Les conclusions de ce travail de prospective sont disponibles à l'adresse suivante : https://turing.ac.uk/content/uploads/2016/06/The-Alan-Turing-Institute-Scoping-Programme.pdf

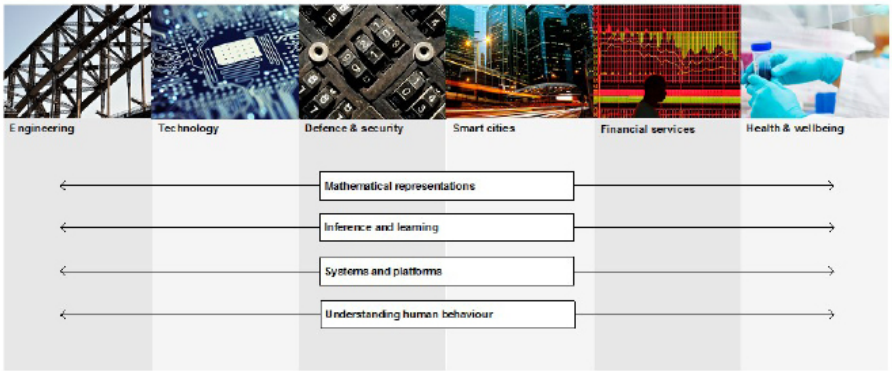


Figure 8 : Priorités stratégiques de l'ATI.
Source : Alan Turing Institute strategy.
Crédits : Alan Turing Institute.

avec les conseils de recherche britanniques, le Digital Catapult et le secteur privé sur ce sujet[54].

L'Institut a en outre signé plusieurs partenariats stratégiques afin de soutenir ses premiers travaux de recherche avec :

- **La Lloyd Register Foundation** : cette prestigieuse *charity* dédiée au développement de la recherche et de l'éducation en ingénierie investit 10 M£ sur 5 ans dans un programme de recherche conjoint avec l'ATI afin d'explorer le rôle et les applications des données en ingénierie.
- **Intel** : annoncé en novembre 2015, ce partenariat vise à développer un programme de recherche permettant aux chercheurs des deux organisations de collaborer dans le domaine du HPC. Intel met à disposition de l'ATI ses solutions en HPC tandis que l'ATI lui permet d'utiliser ses recherches en algorithme pour le développement des futures générations de microprocesseurs ; une équipe Intel sera hébergée à cet effet au sein de l'Institut. Le groupe apporte également son support au programme de doctorants et de chercheurs de l'ATI.
- **L'agence de renseignement britannique GCHQ** (*Government Communications HeadQuarters*) : l'ATI et le GCHQ ont annoncé au cours de l'été 2015 leur intention de collaborer sur la diffusion de bonnes pratiques pour l'exploitation des *big data*, ainsi que de travailler à l'éducation et la recherche sur les méthodes d'analyse pouvant s'appliquer aux données ouvertes et commerciales. Le département du GCHQ en charge de la sécurité de l'information, le CESG (*Communications Electronics Security Group*), apportera également un soutien consultatif sur la sécurité des politiques de données mises en œuvre au sein de l'Institut.
- **Cray Inc.** : le fabricant américain de superordinateurs collaborera avec l'EPSRC afin de mettre ses ressources à disposition d'ARCHER, le superordinateur britannique localisé à l'université d'Edimbourg, afin d'améliorer sa puissance de calcul. Le partenariat implique également des travaux de recherche sur les conditions technologiques nécessaires à l'élaboration des futures générations de systèmes *big data*.

[54] https://www.gov.uk/government/consultations/digital-currencies-call-for-information

A noter

Le gouvernement a par ailleurs annoncé en mars 2016 un investissement à hauteur de 15 M£ pour développer un *National Institute for Smart Data* au sein de l'université de Newcastle, dont l'objectif sera de créer un centre d'excellence régional assemblant l'expertise des secteurs privé, public et académique pour l'exploitation des *big data* sur des secteurs spécifiques (santé, transport, etc.).

4.1.2 A l'échelle d'une université : Les Data Science Institutes *et* Data Science Laboratories

Le développement de ce nouveau champ disciplinaire depuis le début de la décennie incite de plus en plus d'universités britanniques à ouvrir des départements ou laboratoires dédiés offrant une double activité de formation (de niveau master et doctorat) et de recherche académique. De par leur caractère transdisciplinaire, ces structures rassemblent fréquemment des chercheurs issus de multiples spécialités : informatique, statistiques, biologie, physique, sciences sociales, etc.

Il existe actuellement une dizaine de ***Data Science Institutes (DSI)*** au Royaume-Uni au sein des universités suivantes : Imperial College London ; University of Warwick ; University of Manchester ; Lancaster University ; Bournemouth University ; University of Essex (*Institute for Analytics and Data Science*) ; University of Leeds (*Institute for Data Analytics*); Cardiff University (*Data Innovation Research Institute*); St Andrews University (*Institute for Data-Intensive Research*); University of Loughborough (*Centre for Data Science*).

A un échelon inférieur, on assiste également au développement récent de *Data Science Laboratories* ; les universités d'Oxford, de Bristol, de Warwick[55], d'Essex (*Data Science Research Group*) et de Cardiff sont d'ores et déjà pourvues de telles équipes. Le laboratoire d'Oxford, lancé en 2016, s'inscrit dans le cadre d'un partenariat avec la compagnie aérienne *Emirates* et vise à développer des méthodes et

[55] Le *Data Science Lab* de Warwick dépend de sa Business School et ne fait pas partie de son *Data Science Institute*.

Focus
Le *Data Science Institute* d'Imperial College London

Ouvert en 2014 afin de répondre aux choix stratégiques d'Imperial College London identifiant les données comme un chantier prioritaire, le *Data Science Institute* se positionne sur un champ disciplinaire similaire à celui de l'*Alan Turing Institute*.

Figure 9 : KPMG Data Observatory.
Crédits : Imperial College London / KPMG.

Ses objectifs sont de coordonner la recherche associée à la *data science* menée à Imperial College, former des *data scientists*, développer des technologies de gestion et d'analyse des données, soutenir le transfert de technologie et d'innovation au travers de ses collaborations et partenariats, et enfin, promouvoir la science des données auprès du grand public et des décideurs politiques. Il vise ainsi à travailler sur des disciplines aussi variées que les statistiques, l'apprentissage automatique (*machine learning*), la visualisation ou encore l'informatique en nuage (*cloud computing*). Les domaines d'applications sont tout aussi divers : de l'astrophysique à la finance, de la sociologie à la météorologie, en passant par la santé ou encore les sciences biologiques.

Doté d'une équipe d'environ 50 chercheurs, il supervise deux laboratoires transdisciplinaires lancés à la fin de l'année 2015 : le *Social and Cultural Analytics Lab*, sur les données massives culturelles et sociales, et le *Data Assimilation Lab* sur les méthodes et technologies d'assimilation de données (c'est-à-dire l'ensemble des processus par lesquels les observations sont collectées sur une certaine période de temps et incorporées dans des modèles informatiques).

Il est également impliqué dans un partenariat avec l'université chinoise Zhejiang University incluant un laboratoire conjoint. Les projets de recherche en cours portent notamment sur l'analyse et la visualisation de données sur les migrations chinoises, l'exploitation de données pour des bases volumineuses de littérature scientifique (*data mining*) ou encore des simulations associées à la dynamique des fluides.

Le DSI supervise par ailleurs 8 projets de recherche collaboratifs s'inscrivant sur les secteurs des villes intelligentes, de l'environnement et de la santé.

Parmi les partenariats stratégiques noués avec des acteurs privés, l'un des plus emblématiques est certainement celui signé avec le groupe de consultance KPMG. Doté d'un investissement de la firme à hauteur de 20 M£, ce partenariat a notamment permis l'ouverture du ***KPMG Data Observatory,*** inauguré en novembre 2015. Cette installation se présente sous la forme d'un mur circulaire couvert de 64 écrans, alimentés par 32 ordinateurs et offrant une vision à 313 degrés, conçu et hébergé par le *Data Science Institute*. Cet observatoire s'adresse aussi bien aux universitaires qu'aux entreprises afin de leur permettre de travailler à l'analyse et la visualisation des *big data*, permettant de se pencher sur des thèmes aussi variés que la fraude en ligne, le changement climatique ou encore le système solaire.

Les autres partenariats du DSI sont très fortement tournés vers l'écosystème chinois :

- ***Huawei Data Science Innovation Lab*** : ce laboratoire fait suite à la signature d'un protocole d'accord (*memorandum of understanding*) entre Imperial College et le groupe informatique Huawei en 2013. Intégré au DSI, il vise à participer aux objectifs interdisciplinaires de l'Institut. Les thèmes de recherche priorisés pour le moment sont : l'apprentissage profond (*deep learning*) à haute performance ; les modèles de propagation de l'information dans les réseaux ; les algorithmes pour le minage de données de haute dimension ou haute fréquence ; les réseaux de capteurs pour le corps ; l'application des données massives pour la science ("*big data for big science*").

- **Jaywing** : annoncée en mars 2016, cette collaboration avec l'un des acteurs majeurs du marketing vise à appliquer les *big data* à l'étude des comportements des consommateurs et leurs réponses émotionnelles aux campagnes marketing et de publicité.

- **Thomson Reuters** : annoncé en juillet 2016, ce partenariat s'inscrit dans le domaine de la FinTech (*Financial Technologies*) et de la RegTech (*Regulatory Technologies*) afin de comprendre les risques, enjeux et opportunités du secteur financier par l'analyse et l'exploitation de vastes jeux de données.

outils d'analyse pour exploiter plus efficacement les donnée détenues par l'entreprise et affiner son offre de services.

Plus rarement, on trouve également un modèle d'organisation reposant sur une logique similaire de capitalisation des compétences à l'échelle d'une même institution, mais sur un modèle plus souple que la création d'un département ou d'un laboratoire dédié : c'est le cas de la ***Big Data Strategic Research Initiative*** de l'université de Cambridge[56].

Les initiatives et réseaux de recherche stratégique (*Strategic Research Initiatives and Networks*) sont des programmes développés par l'université de Cambridge afin de catalyser l'expertise de chercheurs répartis dans ses différents départements, sur des thématiques jugées innovantes. Il s'agit aussi bien de faciliter le montage de projets de recherche et les collaborations multidisciplinaires que de formaliser une communauté d'experts reconnue comme acteur à part entière sur la scène nationale et internationale.

Les *big data* font ainsi partie des thèmes sélectionnés. Lancée en 2013, la *Big Data Strategic Research Initiative* rassemble un réseau fort d'une centaine de chercheurs. Ses activités de recherche s'organisent autour de cinq thèmes : (i) fondations théoriques ; (ii) gestion et traitement des données ; (iii) imagerie ; (iv) éthique, accès et impact ; (v) mise en application des *big data* (« *Making Big Data Work* »).

4.1.3 A l'échelle d'un département : insertion des big data *et* data science *à un portefeuille plus large*

Si des structures institutionnelles commencent à émerger depuis peu, la plupart des activités académiques en *big data* et *data science* sont inclues pour de nombreuses universités dans les portefeuilles de départements, instituts et laboratoires dédiés à des champs disciplinaires plus traditionnels tels que l'informatique, les statistiques ou encore les mathématiques.

On peut citer l'exemple emblématique d'un organisme oxfordien dédié à la recherche multidisciplinaire : l'*E-Research Centre* de l'Université d'Oxford. Fondé en 2006, il s'agit d'un département dépendant de la division pour les sciences mathématiques, physiques et de la vie. Impliqué dans les technologies numériques au sens large, il a vocation à catalyser l'expertise et les projets de tout chercheur souhaitant intégrer cette dimension à ses travaux, à l'intérieur et à l'extérieur de l'université. Fort d'une équipe d'une cinquantaine de chercheurs, ses activités s'organisent autour de 5 grandes thématiques : calcul scientifique ; logiciels, répertoires et gestion des données ; internet, information et interaction ; e-infrastructure ; informatique visuelle. Il fournit également aux chercheurs un accès à des infrastructures informatiques distribués et à haute performance au sein de son *Advanced Research Computing Facility*[57] (cf. la partie 5 sur les e-infrastructures pour une vision générale du paysage britannique en HPC).

4.1.4 Formation doctorale: les Centres for Doctoral Training

Outre le financement de la recherche au travers des appels à projet, les conseils de recherche soutiennent également la formation doctorale via plusieurs outils parmi lesquels figurent les centres de formation doctorale (*Centres for Doctoral Training* ou CDTs). Financés pour beaucoup par l'EPSRC, il s'agit de structures hybrides à mi-chemin de l'enseignement et de la recherche organisées autour de grandes problématiques scientifiques transdisciplinaires, associant des experts issus de divers horizons ; l'accent est tout particulièrement mis sur les collaborations avec le secteur privé. Les doctorants retenus sont financés pendant quatre ans. Bien qu'il en existe depuis la fin des années 2000, une vague importante de nouveaux CDTs (115) financés à hauteur de 500 M£ par l'EPSRC et 450 M£ par des partenaires extérieurs (industrie, universités, etc.) a été annoncée en 2013.

Big data et *data science* font partie des thématiques prioritaires identifiées. Citons notamment :

- CDT *Data Science* de l'université d'Edimbourg ;
- CDT *Cloud Computing for Big Data* de l'université de Newcastle ;
- CDT *High Performance Embedded and Distributed System* d'Imperial College London ;
- CDT *Financial Computing and Analytics* de UCL ;
- CDT *Web Science Innovation* de l'université de Southampton ;
- CDT *Next Generation Statistical in Science: the Oxford-Warwick Statistics Programme* des universités d'Oxford et de Warwick.

4.2 Acteurs en *big data* et *data science* appliquées à des secteurs spécifiques

4.2.1 Recherche médicale

Les techniques d'analyse des données massives sont aujourd'hui devenues indispensables dans le champ de la recherche médicale. Qu'il s'agisse des données des patients, du fruit de nouvelles méthodes d'analyse des organismes tel que le séquençage du génome, ou encore des informations recueillies par de plus en plus d'appareils connectés (*wearable devices*), les chercheurs sont aujourd'hui confrontés à de réels défis en termes de stockage, administration et traitement d'une ressource perçue comme un véritable eldorado : le croisement à vaste échelle de ces sources variées suscite en effet de nombreux espoirs depuis le développement de la médecine personnalisée, qui vise à répondre aux spécificités et antécédents de chaque patient, jusqu'à la compréhension plus fine des logiques de développement et de propagation des maladies.

La plupart des institutions britanniques de recherche dans le domaine de la santé se sont ainsi dotées à plus ou moins grande échelle d'équipes, voire d'infrastructures permettant de développer leurs compétences en bio-informatique.

[56] http://www.bigdata.cam.ac.uk/

[57] http://www.arc.ox.ac.uk/content/home

Focus
Génétique et Epidémiologie :
***Big Data Institute / Li Ka Shing Centre for Health Information and Discovery* – Université d'Oxford**

L'Université d'Oxford est en cours de développement d'un *Big Data Institute* spécialisé dans le domaine de la santé, qui doit ouvrir ses portes à la fin de l'année 2016.

Il s'agit de la seconde phase du déploiement du *Li Ka Shing Centre for Health Information and Discovery*. Soutenu à hauteur de 20 M£ par la fondation Li Ka Shing dédiée à l'éducation et à la santé, et par le HEFCE (*Higher Education Funding Council for England*) à hauteur de 10 M£, ce centre de recherche ambitionne à terme de s'imposer comme un acteur majeur de la recherche médicale et de l'analyse de données massives dans les domaines de la génétique, de l'épidémiologie et des maladies humaines.

La première phase a consisté en l'ouverture en 2013 du *Target Discovery Institute*, spécialisé dans l'identification de nouvelles cibles médicamenteuses en collaboration étroite avec l'industrie pharmaceutique.

Le *Big Data Institute*, quant à lui, aura pour mission de développer de nouvelles méthodes et techniques afin de générer, stocker et analyser des jeux de données massives issues de sources très diverses : dossiers électroniques de patients, séquençage d'ADN, tests cliniques, etc. L'objectif est de mieux comprendre le développement des maladies humaines. Quatre thèmes de recherche ont d'ores et déjà été identifiés : (i) génomique ; (ii) santé des populations ; (iii) surveillance des maladies infectieuses ; (iv) méthodologie (informatique, statistiques, ingénierie, etc.).

Le *Big Data Institute* devrait à terme accueillir une équipe d'environ 350 chercheurs, organisés en une trentaine d'équipes de recherche.

Outre le HEFCE et la fondation Li Ka Shing, il bénéficie du soutien financier du MRC, de la *British Heart Foundation* et de la *Robertson Foundation*.

- **Instituts en bio-informatique financés par le MRC**

Le MRC a annoncé en 2014 un financement de 39 M£ afin de soutenir 6 centres de recherche destinés à travailler sur les données cliniques, de santé et de bio-informatique. Les centres retenus sont : le *Leeds MRC Medical Bioinformatics Centre* ; l'*University of Oxford Big Data Institute* (cf. encadré) ; le *MRC/UVRI Medical Informatics Centre* au sein de la *MRC/UVRI Uganda Research Unit* ; le *MRC Consortium for Medical Microbial Bioinformatics* dirigé par Warwick ; le *Medical Bioinformatics partnership* dirigé par Imperial College London, et l'*University College London Partners*, incluant le *Francis Crick Institute*[58].

- **Analyse des données médicales de patients : *Farr Institute of Health Informatics Research***

Le *Farr Institute* est un réseau de centres d'excellence dédiés à l'analyse des données médicales massives produites par le système de santé britannique à des fins de recherche scientifique.

Annoncé en juillet 2013 par le Secrétaire d'état pour les Universités et la Science de cette époque, David Willetts, et lancé en avril 2014, il rassemble pas moins de 21 partenaires académiques et du secteur de la santé. Un consortium de 10 financeurs ont avancé 17,5 M£ en 2013 pour développer quatre centres d'excellences en e-santé à l'issue d'un appel à candidatures, complétés par une enveloppe de 20 M£ du MRC. Ces quatre centres, dirigés respectivement par les universités de Swansea (CIPHER, *Centre for Improving Population Health through E-Health Research*), Manchester (HeRC, *Health e-Research Centre*), University College London (CHAPTER, *Centre for Health service and Academic Partnership in translational E-Health Research*) et Dundee (HIC, *Health Informatics Centre*) constituent les quatre piliers historiques du *Farr Institute*.

Les activités de l'Institut s'organisent autour de plusieurs thèmes :

- **Recherche scientifique** en *data science*, éthique des données et e-infrastructure appliquée aux domaines suivants : compréhension des maladies chroniques ; santé de la mère et de l'enfant ; santé environnementale et publique ; services de santé ; médecine de précision et essais cliniques. Les activités de recherche incluent également le développement de nouvelles méthodes et techniques statistiques pour l'exploitation des données massives.

- **Formation et compétences** : offre de formation initiale à partir du niveau master dans les universités partenaires et de formation continue en *data science* et e-santé.

- **Engagement avec le public** et encouragement au développement de liens avec les chercheurs : organisation d'évènements, ateliers, rencontres, implication du public dans les programmes de recherche, etc.

- **Partenariats institutionnels et activité de consultance**.

[58] http://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/initiatives-in-informatics-research/

Focus
***Agrimetrics* – le premier centre *big data* dédié aux données agricoles**

Ce centre de recherche s'inscrit dans la stratégie britannique en agritech (*agriculture technologies*) publiée en 2013, dans laquelle a été annoncé un investissement de 90 M£ afin de créer 4 centres pour l'innovation agricole (*Centres for Agricultural Innovation*)[1].

Inauguré en octobre 2015, Agrimetrics est le premier de ces centres à voir le jour. Financé à hauteur de 11,8 M£ par le gouvernement à travers Innovate UK, il s'agit d'une *joint-venture* entre Rothamsted Research (centre de recherche dédié à l'agriculture financé par le BBSRC), l'université de Reading, le *National Institute of Agricultural Botany* (NIAB), et *Scotland's Rural College* (SRUC, établissement d'enseignement supérieur et de recherche en agriculture).

L'objectif de ce centre est d'exploiter les *big data* afin de soutenir l'industrie agro-alimentaire et contribuer à la définition d'un système alimentaire durable et efficace. Il se donne plus particulièrement pour mission de : (i) créer un réseau d'organisations académiques et industrielles afin de mettre en œuvre les *data sciences* appliquées à l'agriculture (ii) construire une plateforme de données et un écosystème favorisant leur partage et leur exploitation dans un environnement sécurisé pour les entreprises (iii) aider les entreprises à développer leurs capacité à exploiter les *data sciences* (iv) développer des standards et des indicateurs métriques pour l'industrie agro-alimentaire (v) soutenir et intégrer l'activité des trois autres centres pour l'innovation agritech en mettant à disposition ses compétences et infrastructures en *data science*.

[1] https://www.gov.uk/government/publications/uk-agricultural-technologies-strategy/uk-agricultural-technologies-strategy-executive-summary#what-we-will-do

L'Institut entend se rapprocher d'un certain nombre d'organismes de recherche impliqués dans les *big data* et *data science* à partir de 2018, notamment l'*Alan Turing Institute*, le *Wellcome Trust Sanger Institute* ou encore l'ESRC A*dministrative Data Research Network*[59].

Pour plus d'informations sur la e-santé et la paysage britannique complet en matière de collecte, stockage et analyse des données médicales, nous vous invitons à vous reporter à notre dossier consacré à ce sujet[60].

- **Génomique : *Wellcome Trust Sanger Institute***

Fondé en 1992 et localisé sur le *Wellcome Genome Campus* près de Cambridge, cet organisme attaché à la prestigieuse *charity* médicale Wellcome Trust se spécialise dans la recherche sur le génome, et est reconnu comme le contributeur unique le plus important au programme de séquençage du génome humain *Human Genome Project*. Il bénéficie d'un financement quinquennal récurrent du Wellcome Trust, ainsi que d'enveloppes accordées par d'autres organismes (conseils de recherche, *charities*, Union Européenne, etc.), qu'il obtient sur compétition.

Ses activités de recherche s'organisent autour de 7 grands programmes dont l'un, intitulé *Computational Genomics*, est consacré à l'informatique appliquée à la génomique. L'objectif de ce programme est de développer de nouvelles méthodes informatiques pour la gestion et l'analyse des vastes quantités de données générées par la recherche sur l'ADN. Deux angles d'approche sont adoptés : l'étude des variations du génome et l'analyse des fonctions du génome.

Le *Wellcome Trust Sanger Institute* abrite également un important *data centre* en sciences du vivant, contenant environ 30 PB de données, ainsi qu'un superordinateur[61].

4.2.2 Environnement

Le secteur de l'environnement est également générateur de larges quantités de données, notamment dans le domaine du climat ou encore de l'observation de la terre via les technologies spatiales. Le *Met Office* constitue ainsi un opérateur de recherche central sur le changement climatique, et dispose d'équipements en conséquence (cf. partie 5). Le conseil de recherche dédié à l'environnement NERC reçoit en 2013 une enveloppe publique de 13 M£ afin de soutenir les *big data* sur son secteur. Ces fonds seront largement affectés à la création ou l'amélioration d'infrastructures informatiques à haute performance, détaillées plus bas. Au-delà de ces secteurs emblématiques, d'autres, tel l'agriculture, ont également identifié l'exploitation des données massives comme désormais indispensables (cf. encadré).

4.2.3 Sciences Humaines, Economiques et Sociales

- **Edition scientifique : *Big Data Institute* – University College London**

L'université londonienne UCL et le groupe Elsevier ont annoncé en décembre 2013 leur intention d'ouvrir un institut de recherche conjoint sur le thème des *big data* appliquées au secteur de l'édition scientifique[62]. Cette collaboration s'appuie sur l'acquisition du réseau social académique Mendeley par Elsevier, dont les locaux sont situés à Londres. Reposant sur une équipe de 13 chercheurs, il anime actuellement un portefeuille de 8 projets orientés pour la majorité autour de l'analyse des réseaux académiques. UCL appartenant aux universités fondatrices de l'*Alan Turing Institute*, cette équipe est étroitement associée aux activités de ce dernier.

[59] http://www.farrinstitute.org/about/what-we-do

[60] SST Royaume-Uni, *La e-santé au Royaume-Uni : les données médicales personnelles et le secteur de la recherche*, rapport d'ambassade, 2013, http://www.ambafrance-uk.org/La-e-sante-au-Royaume-Uni

[61] http://www.sanger.ac.uk/science/groups/information-communications-technology

[62] https://www.ucl.ac.uk/big-data/bdi

Focus
Environnement urbain : le *Urban Big Data Centre* Université de Glasgow

Lancé en septembre 2014[1], l'UBDC se donne comme objectif de « relever les enjeux sociaux, économiques, et environnementaux auxquels les villes font face ». Il rassemble une équipe de 45 chercheurs issus de l'université de Glasgow, mais également de 6 établissements partenaires : Edimbourg, Bristol, Cambridge, Reading, Sheffield, et l'université américaine de l'Illinois située à Chicago.

L'application des *big data* aux problématiques urbaines constitue le socle de ses activités, qui s'orientent aussi bien du côté des méthodes de recherche (développement de logiciels de calcul et d'analyse) que des applications à de nombreux secteurs : transport, logement, environnement, etc.

L'UBDC offre par ailleurs un service d'accès à des jeux de données exploitables par les chercheurs travaillant sur ces questions. Il s'agit aussi bien de données ouvertes (*open data*), que de données à usage restreint (*safeguarded data*) pour lesquelles un accord doit être signé par le chercheur afin de répondre aux exigences des propriétaires de ces données (pas d'exploitation commerciale, exploitation au sein d'un espace sécurisé uniquement, etc.).

[1] http://www.esrc.ac.uk/research/our-research/urban-big-data/

- **Données administratives et sociétales : l'ESRC *Big Data Network***

Les données collectées par les administrations publiques, les entreprises ou tout autre type d'organisation constituent un gisement de ressources précieux pour les chercheurs ; leur accès est toutefois très souvent rendu difficile par les nombreuses restrictions qui peuvent s'appliquer à leur communication à de tierces parties.

Afin d'abaisser ces barrières, le conseil de recherche ESRC a lancé en 2011 un ambitieux programme en plusieurs phases intitulé ***Big Data Network***, financé jusqu'en 2019, et destiné à créer un vaste réseau de partenaires académiques, publics et industriels.

La première phase, lancée en 2013 et financée à hauteur de 64 M£, vise à développer un réseau dédié aux données administratives : l'***Administrative Data Research Network* (ADRN)**[63]. Rassemblant universités, structures gouvernementales, agences de statistiques, partenaires privés, pourvoyeurs de fonds et chercheurs, son but est de faciliter l'accès et le croisement de données administratives anonymes collectées par le secteur public. Quatre centres de recherche ont été créés à cet effet dans chacune des régions du Royaume-Uni : Angleterre, Ecosse, Pays de Galles et Irlande du nord.

La deuxième phase, lancée en 2014 et financée à hauteur de 14 M£, vise à créer 3 centres de recherche spécialisés sur les entreprises et les autorités locales : les ***Business and Local Government Data Research Centres***[64]. Ces organismes sont les suivants : (i) le *Urban Big Data Centre*, situé à l'université de Glasgow (cf. encadré) ; (ii) l'ESRC *Business and Local Governement Data Research Centre*, au sein de l'université d'Essex, associant également les universités d'East Anglia, du Kent et la London School of Economics (LSE) ; (iii) le *Consumer Data Research Centre*, rassemblant les universités de Leeds, Oxford, Liverpool et UCL. L'ESRC finance par ailleurs sur cette tranche le UK *Data Service Big Data Network Support* (BDNS)[65], hébergé à l'université d'Essex, dont la mission est d'apporter un support transversal pour l'accès et l'exploitation effective des données concernées.

La troisième phase[66], lancée la même année, s'attache davantage aux données issues de la société civile et des médias sociaux. Contrairement aux deux précédentes, elle vise moins à soutenir la création d'instituts de recherche[67] qu'à supporter des structures déjà existantes[68] et à financer des projets de recherche, au nombre de quatre, s'inscrivant dans cette thématique[69].

- **Statistiques : les projets *big data* de l'*Office for National Statistics***

L'office national des statistiques (ONS) britannique s'est engagé depuis plusieurs années dans des initiatives visant à soutenir la recherche en *big data* et *data science* appliquées à ses vastes ressources en données, et s'est doté depuis janvier 2014 d'une équipe dédiée à cette thématique[70].

Elle pilote notamment un *Big Data Project* depuis 2013. Etalée sur une période de quinze mois, la première phase de ce programme s'est organisée autour de 4 projets pilotes visant à travailler sur des jeux de données spécifiques collectés par l'ONS : (i) les données de prix sur internet ; (ii) les messages échangés sur le réseau social Twitter ; (iii) les données issues des compteurs dits « intelligents » ; (iv) les données de géolocalisation issues de téléphones mobiles.

Par ailleurs, le chancelier de l'Echiquier a annoncé en mars 2016 un investissement de 10 M£ en faveur de l'ONS afin de soutenir la fondation d'un ***Data Science Campus*** au sein des quartiers généraux de l'*Office* à Newport (sud du Pays de Galles) dans les deux ans à venir[71]. Cette déci-

[63] https://adrn.ac.uk/

[64] http://www.esrc.ac.uk/research/our-research/big-data-network/big-data-network-phase-2/

[65] https://bigdata.ukdataservice.ac.uk/

[66] http://www.esrc.ac.uk/research/our-research/big-data-network/big-data-network-phase-3/

[67] Un projet d'établissement transdisciplinaire dédié aux nouvelles formes de données était à l'étude mais ne s'est pour le moment pas encore concrétisé.

[68] 1,4 M£ ont été accordés à un réseau de recherche dédié à l'économie numérique, les *Digital Economy Hubs,* afin de développer leurs activités sur ce secteur : http://www.rcuk.ac.uk/research/xrcprogrammes/digital/research/digital-economy-hubs-and-research-centres/. A noter que l'équipe *Horizon* (université de Nottingham) présentée plus bas en fait partie.

[69] http://www.esrc.ac.uk/research/our-research/civil-society-data-partnership-projects/

[70] https://gss.civilservice.gov.uk/blog/2016/03/big-data-at-ons/

[71] https://www.ons.gov.uk/news/news/newonseconomicscentreofexcellenceanddatasciencecampusannounced

sion fait suite à un rapport indépendant commandé par le gouvernement à Charles Bean[72], ancien gouverneur adjoint de la banque d'Angleterre, dans lequel il préconisait l'établissement d'une telle structure afin de favoriser le développement de l'économie numérique. Le *Data Science Campus* aura pour double mission d'améliorer les techniques d'exploitation et d'analyse des données collectées par l'ONS, mais également d'améliorer ses méthodes d'évaluation de l'économie numérique britannique[73].

4.2.4 Science de l'internet et science du web

Les données massives produites chaque jour sont en très grande partie générées par les usages que nous faisons d'internet, du web et des applications toujours plus nombreuses nous incitant à y déposer ou créer des données numériques. De fait, il s'agit ainsi d'un objet d'études qui aujourd'hui ne peut plus se passer des compétences des *data scientists* au sein des programmes de recherche.

- **Internet (1) : L'*Oxford Internet Institute* – Université d'Oxford**

Fondé en 2001, l'*Oxford Internet Institute* (OII) est un département multidisciplinaire de la division des sciences sociales (*Social Sciences Division*) de l'Université d'Oxford, dont la mission est de comprendre les pratiques du web à travers le développement d'une science sociale de l'internet.

Etroitement associé à l'*Alan Turing Institute* dont l'université d'Oxford est l'un des membres fondateurs (cf. §4.1.1), ses activités de recherche s'organisent autour de 8 grandes thématiques faisant intervenir *data science* et *big data* dans leur mise en œuvre : économies numériques ; connaissance et culture numérique ; connectivité, inclusion et inégalités ; politique et gouvernement numérique ; éducation, vie numérique et bien être ; éthique et philosophie de l'information ; gouvernance et sécurité de l'information ; *social data science*. L'OII accueille 35 professeurs et 60 étudiants de 2e et 3e cycle.

S'il a bénéficié lors de sa fondation d'un don de 10 M£ de la *Shirley Foundation* ainsi que d'un soutien financier du HEFCE (*Higher Education Funding Council for England*), il s'agit aujourd'hui d'une structure indépendante financièrement.

Parmi ses nombreux projets, l'OII anime depuis 2015 un réseau de chercheurs dédié à l'utilisation des *big data* pour les questions de développement et de réduction des inégalités. Intitulée ***« Big Data and human development »***[74], cette initiative se donne pour mission d'analyser les usages potentiels des données massives afin de traiter les problèmes de discrimination et d'exclusion au sein de nos sociétés. Il vise notamment à développer un observatoire numérique prenant la forme d'une plateforme web dédiée à ces questions, informant et donnant accès à des jeux de données.

- **Internet (2) : Le *Web Science Institute* – Université de Southampton**

Fondé en novembre 2013 et lancé officiellement en juin 2014, le *Web Science Institute* (WSI) est un département de l'université de Southampton intervenant sur un champ disciplinaire voisin de l'OII. Ses activités de recherche s'organisent autour des thèmes suivants : analyse des réseaux sociaux ; cyber sécurité ; machines sociales ; santé en ligne ; et observatoire du web.

Bénéficiant du prestigieux parrainage de Sir Tim Berners-Lee et de Sir Nigel Shadbolt, co-fondateurs de l'*Open Data Institute*[75], l'objectif de cet institut est d'étudier les interactions entre le web et nos sociétés tout en aspirant à une fonction d'expertise auprès des décideurs et dans les débats publics.

L'Université de Southampton héberge le *Web Science Trust*[76], une *charity* établie en 2006 rassemblant un réseau mondial de laboratoires visant à soutenir le développement d'une science du web (*web science*) (on y retrouve également l'*Oxford e-research centre* et l'*Oxford Internet Institute*).

Au sein de ce réseau, le WSI anime un projet d'observatoire du web (*web observatory)*[77]. L'objectif est de créer un réseau décentralisé coordonnant des communautés de *web scientists* et des « observatoires du web », tels que le *Southampton Web Observatory* (SUWO)[78]. Aujourd'hui au nombre de 13, ces projets académiques ont pour but d'identifier, de donner accès et de faciliter l'analyse de ressources existantes sur le web, permettant d'en comprendre les logiques et les usages[79]. L'accent est tout particulièrement mis sur l'interopérabilité des outils d'analyse et de visualisation.

- **Données personnelles : *Horizon Digital Economy Research* – Université de Nottingham**

Fondé en 2009 grâce à un financement initial à hauteur de 40 M£ de RCUK, de l'université de Nottingham et de plusieurs dizaines de partenaires académiques et industriels, *Horizon Digital Economy Research* consiste à la fois en un centre de recherche académique et en un centre de formation doctorale (*Centre for Doctoral Training*).

Fort d'une équipe de plus de 30 chercheurs et professeurs, *Horizon* s'intéresse à l'impact des technologies numériques sur nos sociétés sous l'angle des données personnelles : exploitation, bénéfices sociaux et économiques, enjeux et problématiques, etc. Les *big data* y sont donc ici étudiées et exploitées dans cette perspective.

[72] Prof. Sir Charles Bean, *Independent review of UK economic statistics*, mars 2016, https://www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-final-report

[73] https://www.ons.gov.uk/aboutus/whatwedo/datasciencecampus

[74] http://bigdatadevelopment.ox.ac.uk/

[75] http://theodi.org/

[76] http://www.webscience.org/

[77] http://wstweb1.ecs.soton.ac.uk/?page_id=969

[78] https://webobservatory.soton.ac.uk/ ; cf. Wendy Hall *et al.*, "The Southampton University Web Observatory", in *1st International workshop on Building Web Observatories, ACM Web Science 2013, 1-3 May 2013*, 2013, http://eprints.soton.ac.uk/352287/1/bwebobs13_The%20Southampton%20University%20Web%20Observatory.pdf

[79] Thanassis Tiropanis, "The web observatories", in *Web Science Trust, A celebration of 10 years of the science of the web*, 2016, http://www.webscience.org/wp-content/uploads/sites/117/2016/07/Web-Science-Trust-Brochure-Final.pdf

5. E-infrastructures et calcul haute performance : un réseau en cours de consolidation

Comme nous avons eu l'occasion de le voir plus haut, les pouvoirs publics ont manifesté un intérêt soutenu pour les infrastructures informatiques au Royaume-Uni. On s'attachera ici plus spécifiquement à présenter succinctement le paysage britannique en matière de superordinateurs et de réseaux mis à disposition des acteurs académiques travaillant sur le secteur des *big data*.

Les conseils de recherche EPSRC et STFC y jouent un rôle de premier plan. L'EPSRC, outre l'administration du superordinateur ARCHER, l'une des machines les plus puissantes du pays à disposition de l'ensemble de la communauté britannique (cf. §5.3.1), assume également la coordination des initiatives dans le domaine du calcul haute performance au sein de RCUK, et est à l'origine de la création de plusieurs centres régionaux venant compléter le tissu des infrastructures offertes à la communauté scientifique. Le STFC, quant à lui, gère également plusieurs des principaux équipements de calcul haute performance du pays, tels que le *Hartree Centre*, le DiRAC ou encore Jasmin (cf. §5.3.2).

Pour la plupart des équipements dont il sera ici question, un accent tout particulier est mis sur leur ouverture aux acteurs industriels afin de renforcer le potentiel de ces technologies pour l'économie du pays. On notera également le rôle moteur joué par plusieurs groupes internationaux – IBM, Fujitsu – dans le développement de plusieurs de ces équipements.

Le paysage britannique des équipements en calcul à haute performance s'organise en 4 niveaux (figure 9)[80] :

- **Niveau local :** ordinateurs localisés dans les universités et les centres de recherche, offrant des ressources limitées prioritairement destinées aux équipes de l'institution hôte pour des projets de faible ampleur.
- **Niveau régional :** plusieurs collaborations associant différentes universités et organismes de recherche se sont développées au début de la décennie sous l'impulsion des conseils de recherche afin de mutualiser les ressources et offrir des services plus performants, et plus diversifiés que les équipements localement détenus par les organismes de recherche (voir figure 10). S'ils ont été accessibles à l'ensemble de la communauté académique du pays, ils ont visé en priorité à consolider les tissus R&D régionaux. Ce niveau est actuellement en cours de restructuration.
- **Niveau national :** plusieurs infrastructures de haut niveau s'adressent à l'ensemble de la communauté scientifique britannique parmi lesquelles on peut citer : le superordinateur national ARCHER, hébergé à l'EPCC de l'université d'Edimbourg ; le Hartree Centre, situé à Daresbury ; le réseau DiRAC, qui associe 5 superordinateurs répartis entre Edimbourg, Cambridge, Durham et Leicester.
- **Niveau européen :** le Royaume-Uni fait partie de l'infrastructure européenne PRACE (*Partnership for Advanced Computing in Europe*), qui lui donne accès à des superordinateurs situés dans les pays hôtes que sont l'Allemagne, l'Espagne, l'Italie et la France. Il met également à disposition des partenaires un faible pourcentage de ressources de son superordinateur ARCHER dans le cadre d'un programme financé par ce partenariat.

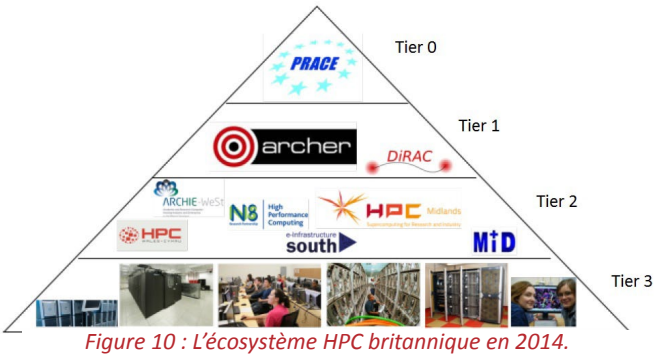


Figure 10 : L'écosystème HPC britannique en 2014. Crédits : EPSRC.

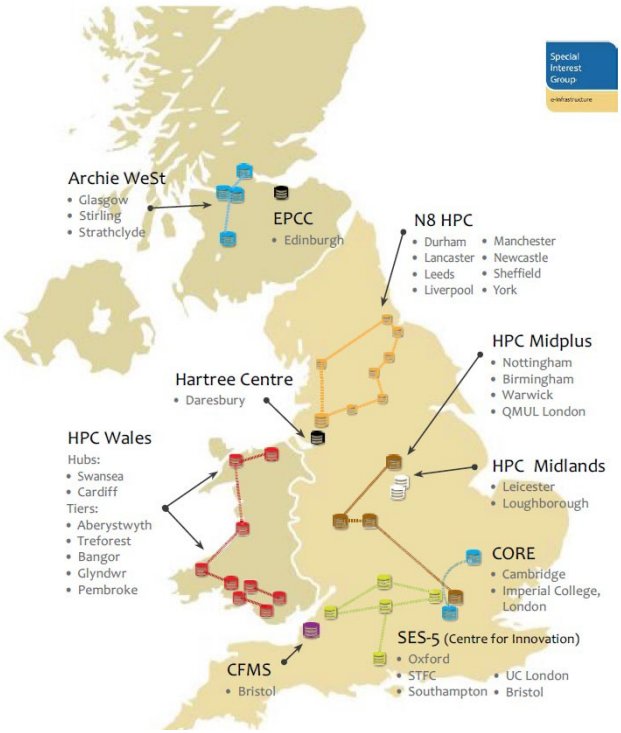


Figure 11 : Carte des équipements britanniques en HPC disponibles pour un usage académique et commercial en 2014. Source : Craig Kirkwood, "UK High Performance Computing, An Overview of Capability", KTN e-Infrastructure special interest group, 2014.

Notons toutefois que les superordinateurs ne constituent pas les seuls équipements à soutenir la recherche britannique dans le domaine des *big data*. De même, les infrastructures réseaux sont essentielles pour assurer efficacement le partage de données et l'utilisation des applications proposées à distance. Soulignons pour finir que la gestion des données générées et collectées nécessite également des infrastructures de stockage tout aussi performantes afin d'accueillir l'ensemble des ressources voulues[81].

[80] Ces niveaux correspondent globalement aux différentes échelles de service Tier 3 (équipements locaux) à Tier 1 (national) et Tier 0 (européen).

[81] Le lecteur intéressé pourra se reporter à la liste très complète disponible dans Council for Science and Technology, *Science Landsape Seminar Series …*, *op. cit.*, 2015.

5.1 Diversité des équipements institutionnels

5.1.1 Universités

Pas moins de 37 établissements d'enseignement supérieur offrent de tels services.

Leurs performances sont très variables, depuis l'EPCC de l'université d'Edimbourg hébergeant le meilleur superordinateur public du pays (relevant de fait plutôt de la catégorie des services nationaux), jusqu'aux services plus modestes permettant aux chercheurs de l'institution hôte de pallier aux limites de leurs postes de travail pour des projets de faible ampleur. Il s'agissait pendant longtemps du niveau le plus accessible aux chercheurs souhaitant travailler en calcul haute performance, jusqu'à la constitution d'équipements régionaux permettant de compléter l'offre de services disponibles (cf. §5.2).

5.1.2 Organismes de recherche publics

Outre les universités, un certain nombre d'organismes publics offrent à la communauté académique des services de calcul haute performance, en général spécialisés dans un secteur spécifique. On peut ici citer l'exemple de l'***Earlham Institute***, institut de recherche financé par le BBSRC, dont la mission est de développer des outils de génomique et de bio-informatique. Ses six installations dédiées au calcul haute performance sont utilisées par les équipes de l'Institut afin d'analyser les larges volumes de données générés par ses équipements de séquençage[82].

5.1.3 Autres organismes

Plusieurs autres institutions exerçant une activité de recherche disposent également d'infrastructures spécialisées en calcul haute performance afin de supporter leurs travaux. C'est particulièrement vrai dans le domaine de la climatologie, demandeuse de ressources informatiques considérables afin de générer les complexes simulations permettant de surveiller et d'anticiper les variations du climat, ainsi que pour le stockage et l'administration des quantités énormes de données produites par les outils d'observation et de mesure.

Le ***Met Office***, l'agence de météorologie britannique, constitue un bon exemple. Doté depuis 1990 d'un centre dédié à la recherche scientifique sur le changement climatique, financé par le ministère en charge de l'énergie (jusqu'à récemment le DECC, aujourd'hui le BEIS) et le ministère en charge de l'environnement, de l'alimentation et des affaires rurales (Defra), il s'est imposé comme un acteur de premier plan dans cette thématique hautement stratégique. Outre la surveillance des variations et l'étude des processus physiques, chimiques ou biologiques impliqués, ce centre développe également des simulations informatiques de haut niveau afin de représenter et d'anticiper les modifications climatiques analysées, et dispose pour ce faire d'équipements dédiés au calcul haute performance.

Le gouvernement a annoncé fin 2014 un investissement à hauteur de 97 M£ pour un nouveau superordinateur dédié au *Met Office,* basé sur le site d'*Exeter Science Park*. Il s'agit d'un Cray XC40, plus puissant qu'ARCHER, capable d'opérer près de 3000 milliards d'opérations par seconde[83].

On peut également citer l'ECMWF, propriétaire des plus puissantes machines du Royaume-Uni. L'***European Centre for Medium-Range Weather Forecasts*** évolue sur le même champ disciplinaire, et dispose de deux superordinateurs Cray. Il s'agissait jusqu'en 2015 d'un modèle semblable à celui d'ARCHER, le service de calcul haute performance national (cf. §5.3.1), qui a toutefois connu une montée en niveau ces derniers mois, les positionnant sur les modèles plus performants du fournisseur (Cray XC40), similaire à celui du *Met Office.*

Dans un autre domaine, Le ***Wellcome Trust Sanger Institute***, présenté plus haut, est un exemple intéressant : avec des machines équipées de près de 17 000 processeurs (*cores*) et 30 PB de données stockées, il s'agit de l'un des plus importants centres de données du pays sur son secteur[84].

5.2 Consortia et logiques de regroupement

5.2.1 Les Centres d'excellences régionaux de l'EPSRC

En 2011-2012, l'EPSRC a lancé un appel à candidatures afin de créer cinq centres d'excellence régionaux pour le calcul à haute performance, dans le but de compléter des équipements historiquement déjà présents.

Financés grâce à une enveloppe conjointe du conseil de recherche (8 M£ complétés par 2 M£ de financements récurrents), des organismes hôtes et des partenaires, ces centres régionaux proposent des ressources intermédiaires entre les infrastructures locales hébergées par les universités et ARCHER tout en offrant une diversité d'architectures informatiques. Ils ont également vocation à créer des communautés académiques d'experts travaillant sur le calcul haute performance, tout en favorisant les collaborations en R&D avec les écosystèmes industriels régionaux, en leur ouvrant notamment l'accès des superordinateurs. Pour finir, ces centres ont également une vocation de formation, afin d'accompagner l'adoption de ces nouvelles technologies par les équipes qui en expriment le besoin[85].

Il faut toutefois signaler qu'un certain nombre des financements dont il est ici question sont déjà terminés ou bien arrivent à expiration au cours de cette année[86]. L'EPSRC a lancé en juin 2016 un nouvel appel à candidatures afin de renouveler ce réseau d'équipements régionaux, offrant un capital de 20 M£ pour mettre en place entre 5 et 8 nouveaux centres. Les propositions retenues seront annoncées à l'automne 2016 pour une mise en œuvre début 2017[87].

[82] http://www.earlham.ac.uk/high-performance-computing

[83] http://www.metoffice.gov.uk/about-us/who/sustainability/environment/supercomputers

[84] http://www.sanger.ac.uk/science/computational-and-information-technology

[85] EPSRC, *Strategy for the developing landscape of Tier-2 HPC in the UK,* 2014, https://www.epsrc.ac.uk/files/research/tier2hpcstrategy/

[86] *Ibid.*

[87] https://www.epsrc.ac.uk/funding/calls/tier2capitalinvestments/ (résultats finaux non connus au moment de la finalisation de ce dossier).

Les premiers consortias qui ont été financés par l'EPSRC sont les suivants:

- **HPC Midlands**[88] : associant les universités de Leicester et de Loughborough, ce centre est fondé en 2013 grâce à un financement de 1 M£ de l'EPSRC. Les projets délivrés s'inscrivent dans les domaines de l'ingénierie et manufacture avancées, de l'énergie, des *big data* et des sciences de la vie.
- **MidPlus**[89] : cette collaboration, associant les universités de Warwick, Birmingham, Nottingham et Queen Mary University of London (QMUL), est lancée début 2012 grâce à un financement de 2 M£ de l'EPSRC complété par 1,5 M£ des institutions partenaires.
- **N8 HPC**[90] : le *N8 Research Partnerhsip* est une collaboration de recherche fondée en 2007, associant les universités de Durham, Lancaster, Leeds, Liverpool, Manchester, Newcastle, Sheffield et York, dans le but de promouvoir des activités de R&D à l'échelle régionale (nord de l'Angleterre), incluant un service de HPC. Les thèmes de recherche prioritaires sont l'agroalimentaire, et les transformations urbaines et communautaires.
- **Archie WeSt**[91] : cette collaboration entre les universités de Glasgow, Glasgow Caledonian University, Strathclyde, Stirling et West of Scotland est établie en 2012 grâce à un financement de 1,6 M£ de l'EPSRC. Le superordinateur, Archie, est situé à l'université de Strathclyde. Les projets s'inscrivent dans une grande variété de disciplines : énergie, chimie, matériaux, sciences de la terre, etc.
- **SES**[92] **et CORE**[93] : *Science and Engineering South* est un consortium entre les universités de Cambridge, Oxford, Southampton, Imperial College London, University College London, et depuis cette année, King's College London qui vise à faciliter l'accès aux équipements et infrastructures scientifiques détenues par les universités partenaires, incluant les équipements HPC. Le centre d'innovation correspondant, SES-5, lancé en 2012, a été financé à hauteur de 3,8 M£ par l'EPSRC et est arrivé à son terme en 2015[94].
  Au sein de ce consortium, le dispositif CORE, lancé en 2012, mutualise les équipements HPC des universités de Cambridge et d'Imperial College London afin de créer une infrastructure d'une capacité de 300 TFLOPS[95], destinée aux communautés académiques et industrielles ; une partie des infrastructures de CORE fait également partie du réseau national DiRAC (cf. §5.3.2).

Citons également pour compléter ce paysage le **HPC Wales**[96] : cette collaboration entre le gouvernement gallois, les universités de la région et le groupe international Fujitsu est établie en 2011 grâce à un financement conjoint des différents partenaires à hauteur de 40 M£. Il s'agit d'un réseau distribué dont les différents nœuds se situent au sein des universités de Cardiff, Swansea, Bangor, Aberystwyth, Glamorgan et Glyndwr. L'ensemble offre une capacité de traitement de 320 TFLOPS. Ce projet est désormais lui aussi arrivé à son terme.

5.2.2 Autres partenariats

D'autres logiques de regroupement voient le jour dans des écosystèmes plus spécifiques, incluant l'analyse de données massives et complexes à leurs services.

C'est tout particulièrement le cas au sein de la recherche biomédicale, et on peut noter l'effort des principales organisations britanniques engagées sur ce secteur. Le **eMedLab** est un *data centre* partagé, administré par un consortium composé de UCL, *Queen Mary, University of London,* L*ondon School of Hygiene and Tropical Medicine*, le *Francis Crick Institute,* le *Wellcome Trust Sanger Institute* et l'*European Bioinformatics Institute*. Financé à hauteur de 8,9 M£ par le gouvernement en 2014, il offre aux organismes partenaires un service de partage et d'analyse de larges jeux de données[97].

Cet équipement est hébergé dans le JISC *Shared Data Centre* avec les équipements de King's College London et Imperial College financés par le MRC et le *National Institute for Health Research* (NIHR). Lancé en 2014, ce site partagé est administré par un consortium composé de UCL, *King's College London*, *Sanger Institute, Francis Crick Institute, London School of Economics and Political Science*, et *Queen Mary, University of London*. Bénéficiant d'une enveloppe initiale de HEFCE, il se situe à l'ouest de Londres, dans la ville de Slough[98].

Dans le domaine de la physique, citons le **GRIDPP**[99] (*UK Computing for Particle Physics*). Il s'agit d'une collaboration financée par le STFC entre physiciens des particules et informaticiens du Royaume-Uni et du CERN. Cela inclut un réseau informatique distribué entre 17 institutions britanniques, contribuant au réseau associé à l'accélérateur de particules LHC (*Large Hadron Collider*), dédié à l'analyse des données qui en sont issues.

5.3 Equipements nationaux

Le Royaume-Uni dispose de plusieurs équipements de calcul à haute performance s'adressant à l'ensemble de la communauté britannique. Il s'agit soit d'infrastructures de très haut niveau généralistes, donnant accès à des ressources informatiques étendues, soit d'infrastructures dédiées à un champ disciplinaire circonscrit afin d'en soutenir le développement en proposant des services avancés d'analyse[100].

[88] http://www.hpc-midlands.ac.uk/about/

[89] http://www2.warwick.ac.uk/fac/cross_fac/midplus/

[90] http://n8hpc.org.uk/

[91] http://www.archie-west.ac.uk/

[92] http://www.ses.ac.uk/

[93] http://www.ses.ac.uk/high-performance-computing/core/

[94] Oxford Advanced Research Computing (ARC) ; UCL Research IT Services ; Iridis (superordinateur hébergé à l'université de Southampton); CORE; Cambridge et Imperial College London. Le consortium bénéficie également d'un partenariat stratégique avec le superordinateur Emerald, administré par le STFC et hébergé au Rutherford Appleton Laboratory. Cf. Martin Hamilton *et al., Report of 2015 UK National e-Infrastructure Survey of HEIs and Research Institutes, JISC,* 2016, http://hpc-sig.org.uk/wp-content/blogs.dir/sites/63/2016/03/NeI-PDG-UK-National-E-Infrastructure-Survey-2015-Report.pdf

[95] Le FLOPS (*FLoating-point Operations Per Second*) est une unité mesurant le nombre d'opérations réalisées par une machine en une seconde et permet d'évaluer sa performance. 1 TFLOPS représente ainsi 1000 milliards d'opérations par seconde.

[96] http://www.hpcwales.co.uk/

[97] http://www.emedlab.ac.uk/

[98] https://www.jisc.ac.uk/shared-data-centre

[99] https://www.gridpp.ac.uk/

[100] HM Government, *E-Infrastructure : One year on,* 2013.

5.3.1 *Equipements génériques*

- **EPCC**

L'EPCC (anciennement *Edinburgh Parallel Computing Centre*) est un acteur historique de la R&D britannique en HPC. Fondé en 1990 au sein de l'université d'Edimbourg et fort d'une équipe de 75 personnes, il héberge aujourd'hui trois équipements nationaux :

- **ARCHER** : succédant à HecTOR en 2013, ARCHER est le service national britannique de calcul haute performance. Financé à hauteur de 43 M£ par l'EPSRC et le NERC, prévu pour une durée de 5 ans, il est accessible aux chercheurs par l'intermédiaire des conseils de recherche britanniques, ainsi qu'aux entreprises. La technologie d'ARCHER est fournie par le fabricant de superordinateurs américain Cray Inc. Il s'agit du superordinateur le plus puissant du pays détenu par une institution de recherche publique.

- **DIRAC BlueGene/Q** : il s'agit d'un des nœuds de l'infrastructure DiRAC, intégrant un réseau de 5 superordinateurs répartis dans plusieurs universités (cf. §5.3.2).

- **UK Research Data Facility :** Financé par l'EPSRC et le NERC, le RDF vise à fournir un service de stockage (d'une capacité de 23 PB) et d'analyse de données de recherche associées à l'informatique. Il est entré en service en 2012 et est conçu pour rester en activité au-delà des cinq ans prévus pour ARCHER.

Les chercheurs de l'EPCC exercent par ailleurs une activité de recherche en informatique, sur le développement de logiciels et dans le domaine des *big data*. Pour ce qui relève de ce dernier thème, l'EPCC a ainsi été associé à près d'une cinquantaine de projets de recherche depuis sa création.

- ***Hartree Centre***

Fruit d'une collaboration entre le STFC et le géant américain IBM, le *Hartree Centre* est établi en 2012 grâce à un financement initial de 37, 5 M£ du BIS, et se situe sur le campus national pour la science et l'innovation Sci-Tech Daresbury dans le Cheshire (nord de l'Angleterre).

Il est très tourné vers le secteur privé, et met à disposition quatre équipements HPC à des fins de modélisation, simulation, visualisation, analyse de données massives et développement d'une informatique à économie d'énergie[101].

Les activités du centre s'organisent autour de 4 grandes familles de services : (i) **Recherche et développement** (mise à disposition des ressources et experts du centre); (ii) **Développement et optimisation de logiciels** (soutien à l'optimisation ou l'écriture de codes s'appuyant sur du calcul à haute performance); (iii) « ***Platform as a service »*** (accès monétisé aux équipements HPC du centre); (iv) **Formation et compétences** (organisation d'ateliers et évènements en *data science* et HPC).

[101] http://www.hartree.stfc.ac.uk/Hartree/Solutions/45426.aspx

Le *Hartree Centre* a par ailleurs reçu une enveloppe de 19 M£ du gouvernement en 2013 pour développer un programme de recherche sur l'informatique à économie d'énergie (*Energy Efficient Computing programme*), actuellement toujours en cours[102].

Partenariat stratégique avec IBM

Faisant suite à un nouvel investissement public à hauteur de 113 M£ annoncé par le gouvernement à l'automne 2014, un partenariat stratégique d'une valeur de 313 M£ est signé en juin 2015 avec le groupe IBM, afin de développer les capacités du centre dans le domaine des *big data.* L'accord inclut notamment un engagement d'IBM à mettre à disposition du Centre ses ressources technologiques – dont un accès privilégié à Watson, système informatique de renommée internationale dans le domaine de l'apprentissage automatique et du traitement des données – et son expertise – une vingtaine de chercheurs d'IBM sont affectés au *Hartree Centre* afin de travailler avec ses équipes. Par ailleurs, dans le cadre de ce partenariat stratégique, le site de Daresbury doit accueillir un nouveau centre de recherche dédié à la modélisation, simulation et analyse de données massives : le POWER *Acceleration and Design Centre* (PADC)[1].

[1] http://www.hartree.stfc.ac.uk/Hartree/Info+hub/45482.aspx

5.3.2 *Equipements thématiques*

- **Astronomie et cosmologie : DiRAC**

DiRAC[103] (*Distributed Research utilising Advanced Computing*) est une infrastructure HPC spécialisée dans les secteurs de la physique des particules, l'astronomie et la cosmologie. Bénéficiant de deux vagues de financement successives du STFC (12,3 M£ en 2009 et 15 M£ en 2011), DiRAC repose sur une architecture distribuée permettant de faciliter la diversité des projets de recherche soumis. Les cinq superordinateurs composant cet équipement sont les suivants :

- ***Cambridge HPC Service Data Analytic Cluster :*** l'université de Cambridge dispose d'équipements HPC organisés autour de deux clusters, dont le premier, « Darwin », entré en service en 2012, fournit ses ressources à la fois en interne aux chercheurs de l'université, et vers l'extérieur dans le cadre de DiRAC[104].

- ***Cambridge COSMOS Shared Memory Service :*** lancé en 1997 par un consortium de chercheurs britanniques sous l'impulsion du célèbre Stephen Hawking, ce superordinateur se spécialise dans les mêmes domaines disciplinaires que DiRAC. Les principaux contributeurs du consortium COSMOS sont aujourd'hui issus des universités de Cambridge, Central Lancashire, Portsmouth, Sussex, et Imperial College London[105].

[102] http://www.hartree.stfc.ac.uk/Hartree/Initiatives/45431.aspx
[103] http://dirac.ac.uk/
[104] http://www.hpc.cam.ac.uk/services/darwin/introduction
[105] http://www.damtp.cam.ac.uk/cosmos/science/

Université	Nom	*Cores*	Pic de performance théorique (TFLOPS/sec)
Oxford	ArcusB	5440	538
Leeds	Arc2	3040	316
Cambridge	Wilkes	1536	256
Bristol	BlueCrystal	9000	240
Oxford	Arcus-GPU	12	146
UCL	Legion	7816	115
Sheffield	Iceberg1	3440	112
Manchester	Computational Shared Facility	6288	111
Cardiff	Raven	4352	110
Leicester	ALICE	3972	101

Tableau 4 : Top 10 des superordinateurs des HEI britanniques en fonction de leur pic de performance théorique (TFLOPS /sec).
Source des données : Martin Hamilton et al., op.cit., 2016.

Nom	Institution hôte	Système	*Cores*	Performance (TFLOPS)	Fournisseur
EPCC	université d'Edimbourg	ARCHER	118080	1643	Cray
DiRAC	université d'Edimbourg	Dirac Blue Gene	98000	800	IBM
	Université de Cambridge	Cambridge HPC Service	9600	200	Dell
	Université de Cambridge	Cambridge COSMOS SHARED MEMORY Service	1856	42	Altix
	Université de Durham	Durham ICC Service	6740	140	IBM
	Université de Leicester	Leicester IT Services	4352	95	unknown
Archie West	Université de Strathclyde	Archie	3408	38	Dell
N8 HPC	Université de Leeds	Polaris	5312	110	SGI
Hartree Centre	Daresbury Laboratory	Blue Joule	98000	1025	IBM
	Daresbury Laboratory	Blue Wonder	11520	223	IBM
HPC Midlands	Université de Loughborough	Hera	3000	48	Bull
MidPlus	Queen Mary University of London	Apocrita	1800	36	Intel
	Université de Warwick	Minerva	2500	unknown	Intel
CORE	Université de Cambridge ; Imperial College London	CORE Cloud	220000	300	unknown
CFMS	Université de Bristol	Main System	288	20	IBM
SES	STFC	Emerald	1008	114	Hewlett-Packard
	Université de Southampton	Irdis	11088	108	IBM
HPC Wales	Université de Cardiff	Cardiff Hub	17000	128	Fujitsu
	Université de Swansea	Swansea Hub	2064	43	Fujitsu

Tableau 5 : Spécifications techniques des principales infrastructures nationales et régionales de calcul haute performance au Royaume-Uni.
Sources : Craig Kirkwood, UK High Performance Computing, An Overview of Capability, KTN e-Infrastructure special interest group, 2014 ; sites web des infrastructures.

- ***Durham ICC Service Data Centric Cluster :*** entré en service en 2011, le superordinateur de l'université de Durham, s'appelle COSMA, et est administré par une équipe de 3 chercheurs [106].

- ***Edinburgh 6144 node Blue Gene/Q :*** co-développé avec IBM, ce superordinateur est hébergé au sein de l'EPCC de l'université d'Edimbourg[107].

- ***Leicester IT Services Complexity Cluster :*** outre deux services de calcul haute performance destinés à ses équipes, l'université de Leicester accueille le 5e nœud de DiRAC, construit selon une architecture innovante organisée autour de commutateurs interconnectés[108].

- **Climat : Jasmin et MONSooN**

Jasmin[109] est une e-infrastructure intégrant un service de calcul haute performance. Bénéficiant d'une enveloppe de 7,75 M£ annoncée par le gouvernement en 2011, Jasmin consiste en un ensemble d'infrastructures informatiques financées par le NERC et l'agence spatiale britannique *UK Space Agency* (UKSA). Pilotée par le STFC *Centre for Environmental Data Archival* (CEDA), elle a pour double mission de fournir des services de stockage et d'administration de données, et des services d'analyse et de calcul scientifique intensif à destination de la communauté scientifique britannique travaillant sur le secteur de l'environnement, en particulier la recherche sur l'atmosphère, la modélisation du climat et l'observation de la Terre. Si le cœur de Jasmin se situe sur le campus d'Harwell au STFC *Rutherford Appleton Laboratory,* divers équipements de stockage et de calcul sont hébergés au sein des universités de Leeds, Bristol et Reading.

MoNSooN[110] (*Met Office and NERC Superco(o)mputer Nodes*) est, comme son nom l'indique, une initiative conjointe entre le *Met Office* et le conseil de recherche dédié à l'environnement NERC dans le cadre de leur partenariat stratégique *Joint Weather and Climate Research Programme* (JWCPR), visant à améliorer les collaborations entre les chercheurs financés par les deux institutions. Financé à hauteur de 1 M£ par le gouvernement en 2011, MONSooN vise à créer une plateforme de calcul haute performance s'appuyant sur les équipements de l'agence de météorologie.

5.4 Projets européens

Après avoir été identifié à plusieurs reprises comme une technologie stratégique au niveau européen au cours des années 2000, le HPC est aujourd'hui pleinement intégré au programme-cadre pour la recherche et l'innovation Horizon 2020 comme secteur d'investissement. La stratégie européenne sur ce sujet, publiée en 2012, identifie trois volets d'action : (i) développer la prochaine génération de technologies HPC, d'applications et de systèmes en vue d'atteindre l'échelle exaflopique[111] ; (ii) fournir un accès aux meilleurs supercalculateurs et services à la fois pour les acteurs privés et académiques ; (iii) mettre en place des centres d'excellence pour l'application des technologies HPC sur des secteurs identifiés comme stratégiques[112].

Cette stratégie se concrétise dans plusieurs initiatives auxquelles le Royaume-Uni est aujourd'hui associé.

5.4.1 Infrastructure européenne en HPC : PRACE

Initialement fondé en 2012 sous le 7e Programme Cadre de Recherche et Développement Technologique européen (PCRD), PRACE[113] (*Partnership for Advanced Computing in Europe*) rassemble 25 partenaires. Il permet aux chercheurs et industriels européens d'accéder à 6 supercalculateurs localisés dans 4 pays hôtes : Allemagne, Espagne, Italie et France.

Le Royaume-Uni est impliqué dans PRACE par l'intermédiaire du conseil de recherche EPSRC, qui finance le supercalculateur national ARCHER. Dans le cadre du projet *Distributed European Computing Initiative* porté par PRACE, ARCHER met notamment à disposition 5% de ses ressources aux chercheurs européens[114].

5.4.2 Communauté européenne de soutien au développement des technologies HPC : ETP4HPC

L'ETP4HPC est la plateforme technique européenne (*European Technical Platform*) dédiée au calcul haute performance. Au nombre de 38, les ETPs sont des communautés animées par des acteurs industriels reconnus par la Commission Européenne comme stratégiques sur les secteurs concernés. Les ETPs rassemblent ainsi grands groupes et acteurs de la R&D afin de développer des stratégies, nourrir les opportunités de collaborations et assumer un rôle de conseil auprès de l'Union Européenne pour l'implémentation de son programme Horizon 2020.

Plusieurs acteurs britanniques sont représentés parmi les membres de l'ETP4HPC : l'EPCC de l'université d'Edimbourg, le *cluster* régional HPC Wales, le conseil de recherche STFC, et l'université Queen's University Belfast. L'ETP4HPC a publié en 2013 un agenda stratégique de recherche, mis une première fois à jour en 2015 et destiné à être une nouvelle fois actualisé en 2017. L'ETP accorde une importance toute particulière à la recherche pour le développement de machines exaflopiques[115].

Notons pour conclure que l'ETP4HPC est le partenaire de la Commission Européenne au sein du Partenariat Public Privé (PPP) pour le calcul haute performance (*PPP for high performance computing*), lancé en janvier 2014. Une série

[106] https://www.cosma.dur.ac.uk/home

[107] https://www.epcc.ed.ac.uk/facilities/dirac

[108] http://www.dirac.ac.uk/Resources.html

[109] http://www.jasmin.ac.uk/

[110] http://collab.metoffice.gov.uk/twiki/bin/view/Support/WhatIsMONSooN

[111] C'est-à-dire des systèmes capables de traiter au minimum un milliard de milliards de calculs par seconde, soit 1 exaflops. Il s'agit de l'objectif aujourd'hui recherché par les meilleurs ordinateurs mondiaux.

[112] http://ec.europa.eu/programmes/horizon2020/en/h2020-section/high-performance-computing-hpc. Cf. Yan-Tarō Clochard, *Calcul haute performance au Japon*, rapport d'ambassade, 2016, pour une synthèse plus complète des politiques européennes sur ce sujet : http://diplomatie.gouv.fr/fr/IMG/pdf/fr-2016-04_hpc_note_ytc_cle8f8eb7.pdf

[113] http://www.prace-ri.eu/

[114] https://www.epcc.ed.ac.uk/facilities/other-facilities/prace

[115] http://www.etp4hpc.eu/

Global Rank	UK Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
17	1	ECMWF	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect	126,468	3,944.70	4,249.30	
18	2	ECMWF	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect	126,468	3,944.70	4,249.30	
29	3	United Kingdom Meteorological Office	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect	89,856	2,801.80	3,019.20	
30	4	United Kingdom Meteorological Office	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect	89,856	2,801.80	3,019.20	
50	5	EPSRC/University of Edinburgh	ARCHER - Cray XC30, Intel Xeon E5 v2 12C 2.700GHz, Aries interconnect	118,080	1,642.50	2,550.50	
56	6	Science and Technology Facilities Council - Daresbury Laboratory	Blue Joule - BlueGene/Q, Power BQC 16C 1.60GHz, Custom	131,072	1,431.10	1,677.70	657
82	7	University of Edinburgh	DiRAC - BlueGene/Q, Power BQC 16C 1.60GHz, Custom	98,304	1,073.30	1,258.30	493
101	8	AWE	Spruce A - SGI ICE X, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR	44,520	958.7	997.2	855
127	9	AWE	Spruce B - SGI ICE X, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR	35,640	767,5	798.3	685
400	10	University College London (UCL)	Grace - Lenovo NeXtScale nx360M5, Xeon E5-2630v3 8C 2.4GHz, Infiniband QDR	10,944	341.3	420.2	

Tableau 6 : Top 10 des meilleurs superordinateurs britanniques sur la scène internationale.
Source : « Top500 List », 47th Edition, June 2016, https://www.top500.org/lists/2016/06/.

de 8 centres d'excellence pour les applications HPC scientifiques ou industrielles sur des secteurs stratégiques[116], sélectionnés en 2015, font partie de ce partenariat.

5.5 Connectivité : le réseau académique Janet

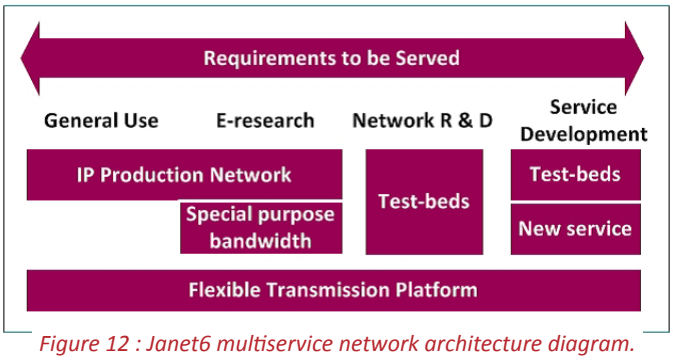


Figure 12 : Janet6 multiservice network architecture diagram.
Source : EPSRC E-Infrastructure roadmap, 2014.

Lancé pour la première fois en 1984, Janet est le réseau national pour la recherche et l'éducation (*UK's National Research and Education Network*). Administré par le Jisc (anciennement *Joint Information Systems Committee*), *charity* qui délivre des services numériques aux établissements d'enseignement supérieur et de recherche, Janet alimente les infrastructures publiques HPC ainsi que de façon plus générale les établissements publics de recherche au Royaume-Uni. C'est donc par ce réseau que les chercheurs britanniques partagent leurs térabits de données massives et accèdent aux applications de calcul haute performance. Janet est constituée de plus de 5000 km de fibre optique d'une capacité de 10 GB/seconde. Elle a subi sa dernière mise à jour au cours de l'été 2016.

Par ailleurs, Janet bénéficie d'une enveloppe publique de 4 M£ pour développer ses relations avec le secteur industriel afin de leur offrir des conditions d'accès à son réseau et ses services, dans le but d'encourager les collaborations avec les universités.

6. Du monde académique au monde économique : soutien au transfert de technologie

Ainsi que nous venons de le voir, le Royaume-Uni est doté d'un écosystème de recherche publique dynamique dans le domaine des *big data*, tout particulièrement en tant que technologie appliquée au sein d'une grande diversité d'organismes thématiques.

Si le secteur privé s'engage déjà résolument avec les universités et organismes de recherche que ce soit pour le financement de laboratoires ou d'infrastructures, la signature de partenariats scientifiques ou encore l'accès aux infrastructures informatiques à haute performance, la valeur accordée aux données par les acteurs économiques implique la présence de cette thématique dans le portefeuille d'un certain nombre d'opérateurs publics s'attachant à faciliter les rapprochements entre milieu académique et industriel.

6.1 Accompagner l'innovation technologique : les centres Catapult

Les centres Catapult sont des centres physiques, indépendants et à but non lucratif, dont la finalité première

[116] **EoCoE** pour le secteur de l'énergie ; **BioExcel** pour la recherche biomoléculaire ; **NoMaD** pour les nouveaux matériaux ; **MaX** pour le design des matériaux ; **ESIWACE** pour la simulation du climat ; **E-CAM** pour la simulation et la modélisation ; **POP** pour l'optimisation des performances et la productivité ; **COEGSS** pour les systèmes globaux. Cf. http://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computing-applications

est de connecter les universités et les entreprises dans un domaine technologique spécifique. Plus largement, les missions des centres Catapult sont : (i) de réduire le risque de l'innovation ; (ii) d'accélérer le rythme du développement commercial ; (iii) de créer de l'emploi ; (iv) de promouvoir une croissance durable ; (v) de développer compétences et connaissances au Royaume-Uni[117].

Sur les 11 centres existant actuellement, plusieurs offrent des services (portails de données, systèmes de visualisation, etc.) ou bien sont impliqués dans des projets pleinement inscrits dans les problématiques soulevées par ce dossier.

On ne peut manquer de citer en premier lieu le **Digital Catapult** qui se positionne précisément sur le secteur de l'économie numérique. Il a pour mission de débloquer les contraintes dans la « chaîne de valeur des données », et plus généralement, de catalyser la valeur contenue dans des jeux de données identifiés comme stratégiques pour les startups et PMEs qui constituent sa principale audience. Ses quatre orientations sont les données fermées, les données personnelles, les contenus créatifs et l'internet des objets.

Le Digital Catapult a notamment noué des partenariats avec l'*Oxford Internet Institute* pour mener des projets dans le domaine de l'économie numérique ; le *Web Science Institute* de l'université de Southampton, afin de travailler en *data science*, incluant le thème des *big data,* et la *School of Informatics* de l'université d'Edimbourg, également en *data science.*

Considérant son champ d'activités, de nombreux projets en cours ou achevés du centre Catapult visent donc à faciliter l'exploitation de données, qui, si elles ne relèvent pas toujours de la catégorie des *big data*, s'inscrivent néanmoins dans le champ de la *data science*. Citons notamment :

- **Des portails d'accès à des jeux de données spécialisées**, issues de sources aussi bien ouvertes que fermées, visant également à créer des communautés d'échange : l'*Environmental Data Exchange*[118], pour des jeux de données environnementales (pollution de l'air à Londres, climat, etc.) ; le *Building Data Exchange*[119] pour des jeux de données sur le secteur de la construction.
- **Le *Data Catalyser***[120] : plateforme de *cloud* sécurisée, destinée aux entreprises, afin de leur offrir un espace d'échange et d'exploitation de données fermées. Parmi les services proposés figure une suite d'outils d'analyse *big data*.
- **Le *Personal Data and Trust Network***[121] : conçu en collaboration avec plusieurs acteurs académiques et institutionnels parmi lesquels KTN, Innovate UK, le conseil de recherche EPSRC ou encore le groupe de recherche Horizon de l'université de Nottingham, ce projet vise à créer une communauté d'acteurs publics et privés afin de supporter les applications innovantes utilisant les données personnelles de façon responsable.

Le **Future Cities Catapult**, évolue, comme son nom l'indique, sur le secteur des technologies innovantes appliquées à l'espace urbain.

Parmi les équipements présents dans ses locaux londoniens figure le ***Cities Lab***, qui fournit des services de visualisation et de modélisation de données. Le centre Catapult offre également des services aux entreprises dans le domaine des « *urban analytics* » (analyse, visualisation et modélisation). Ses projets intègrent pour beaucoup la problématique des données, incontournable dans les visions actuelles des villes du futur[122]. Citons notamment : **Whereabouts London**[123], qui à partir de 235 sources d'informations différentes réinvente la notion de quartier sur la base de similarités entre les modes de vie des Londoniens ; **Bristol Data Dome**[124]**,** projet d'aménagement d'une infrastructure permettant de visualiser des données urbaines sur la ville de Bristol ; **Sensing London**[125], visant à placer des capteurs en divers endroits de la capitale afin de collecter des données sur plusieurs variables (qualité de l'air, etc.), ensuite analysées par une équipe de *data scientists* pour comprendre l'utilisation des infrastructures urbaines.

Le **Satellite Application Catapult** évolue sur le secteur des services satellitaires. S'agissant là aussi d'un domaine profondément générateur de données, plusieurs de ses infrastructures et projets s'inscrivent dans notre sujet. Ce centre Catapult propose ainsi un service de visualisation de données scientifiques, mais aussi une plateforme de données dénommée ***Data Discovery Hub***[126], ouvertes ou sous licence, relatifs à la climatologie, l'environnement, le relief, l'agriculture, l'urbanisme, etc.

Le **Transport Systems Catapult** a vocation à soutenir le développement d'une mobilité intelligente, et notamment le secteur emblématique des systèmes de transport autonome. Les locaux de ce centre Catapult, situés à Milton Keynes, abritent un laboratoire de visualisation[127] offrant une vaste gamme de services s'échelonnant jusqu'à la réalité virtuelle. Il est utilisé pour créer et cartographier des jeux de données issues du terrain afin d'améliorer les applications de mobilité intelligente. Le Catapult pilote en outre, dans le cadre d'une collaboration impliquant l'université de Nottingham, un projet intitulé **« Sentiment mapping »**[128] : le but est d'analyser les réactions des passagers sur les réseaux sociaux (notamment Twitter) afin de caractériser le sentiment général sur un trajet et parvenir *in fine* à améliorer l'expérience générale de ces transports (suggérer des alternatives en temps réel, etc.).

[117] SST Royaume-Uni, *Les centres Catapult*, rapport d'ambassade, 2015, http://www.ambafrance-uk.org/Dossier-Les-centres-Catapult

[118] https://www.environmentaldataexchange.org.uk/

[119] http://www.buildingdataexchange.org.uk/

[120] https://datacatalyser.org.uk/

[121] https://pdtn.org/

[122] Pour en savoir plus sur ce sujet au Royaume-Uni, nous vous invitons à consulter le dossier que nous lui avons consacré : SST Royaume-Uni, L'*innovation publique dans les smart cities au Royaume-Uni*, rapport d'ambassade, 2014, http://www.ambafrance-uk.org/L-innovation-publique-dans-les

[123] https://futurecities.catapult.org.uk/project/whereabouts-london/

[124] http://futurecities.catapult.org.uk/project/bristol-data-dome/

[125] https://futurecities.catapult.org.uk/project/sensing-london/

[126] http://data.satapps.org/

[127] https://ts.catapult.org.uk/imovation-centre/the-visualisation-laboratory/about-the-visualisation-laboratory/

[128] https://ts.catapult.org.uk/current-projects/sentiment-mapping/

6.2 Développer l'écosystème écossais : les *Innovation Centres*

Dans une démarche similaire à celle ayant présidé à la naissance des centres Catapult, le *Scottish Funding Council* (agence publique de financement écossaise) a engagé en 2012 un vaste programme visant à installer huit centres d'innovation sur son territoire. Bénéficiant d'une enveloppe du SFC de plus de 120 M£ étalés sur cinq ans (2013-2018), et complétée par des partenaires privés, ces structures, fortement tournées vers le tissu industriel de la région, ont vocation à rapprocher universités et entreprises afin de stimuler les collaborations en R&D dans une série de secteurs identifiés comme stratégiques pour l'Ecosse. Les centres sont les suivants : le *Data Lab* ; *Scottish Aquaculture Innovation Centre* (SAIC) ; *Construction Scotland Innovation Centre* (CSIC) ; *Innovation Centre for Sensor and Imaging Systems* (CENSIS) ; *Industrial Biotechnology Innovation Centre* (IbioIC) ; *Oil and Gas Innovation Centre* (OGIC) ; *Stratified Medicine Scotland Innovation Centre* (SMS-IC) ; *Digital Health and Care Institute* (DHI Scotland).

Le ***Data Lab***[129] est donc le centre le plus directement associé au thème qui nous intéresse. Financé à hauteur de 11,3 M£ et localisé dans les villes d'Edimbourg, d'Aberdeen et de Glasgow, il a trois grandes missions : (i) délivrer des ressources (consultance, coordination de candidatures aux appels à financement extérieurs au Royaume-Uni et en Europe) et des financements par le biais d'appels à projets[130] afin d'accompagner le lancement de propositions jugées innovantes ; (ii) proposer des formations et des modules de développement des compétences ; (iii) contribuer au développement d'une communauté rassemblant chercheurs, *data scientists* et acteurs industriels locaux.

Le *Data Lab* travaille également sur des projets en collaboration avec d'autres centres d'innovation comme CENSIS (données issues de capteurs utilisés dans les véhicules urbains), DHI et SMS-IC (traitement du cancer) ou encore SAIC (mesure de l'impact environnemental des fermes piscicoles).

Par ailleurs, le CENSIS[131], localisé à Glasgow et spécialisé sur les capteurs et systèmes d'imagerie, identifie l'analyse et la visualisation de données parmi les cinq thématiques autour desquelles s'organise sa stratégie. Ces thématiques lui permettent entre autres d'orchestrer son portefeuille de projets et de coordonner les opportunités de partenariat entre le secteur industriel et les institutions d'enseignement supérieur et de recherche de la région écossaise.

6.3 Créer des communautés de transfert de technologies : les réseaux KTN

Le *Knowledge Transfer Network Ltd*[132] est une entreprise pilotée par Innovate UK qui vise à créer un espace d'échanges entre une grande diversité d'acteurs (chercheurs, entreprises, startups, etc.) autour des secteurs technologiques soutenus par le gouvernement. Il anime notamment des communautés d'intérêt thématiques, organise des évènements – incluant des sessions consacrées aux appels à candidature d'Innovate UK ou européens (Horizon 2020) – et se concentre sur la mise en relation d'individus, d'entreprises ou d'organismes qui ne coopèrent pas (encore).

KTN a identifié 14 communautés centrales : agro-alimentaire ; environnement construit ; défense et sécurité ; énergie ; TIC et mathématiques industrielles ; transport ; matériaux et nanotechnologies ; biotechnologies ; chimie ; électronique, capteurs, photonique ; santé ; espace ; durabilité et efficacité des ressources ; création, numérique et design.

La plateforme «_Connect»[133] offre un espace de communication numérique pour l'ensemble des communautés et réseaux d'intérêt animés par KTN. Notons pour le sujet qui nous intéresse les communautés « ICT » et « *Industrial Mathematics* », ainsi que les groupes d'intérêt (*special interest group*) « *e-infrastructure* » et « *energy efficient computing* ». Chacun de ces groupes dispose d'un espace d'échange d'informations, d'articles et de ressources diverses, d'un agenda pour les évènements en lien avec le sujet, etc.

7. Conclusion

L'entrée récente des *big data* et plus largement de la « science de la donnée » dans l'univers de la recherche en tant que discipline ne constitue que la dernière étape au sein d'une histoire des technologies informatiques plus longue, forgée au cours du XXe siècle.

Le présent rapport avait pour objectif de présenter un panorama des acteurs publics impliqués sur ce secteur. Si on a pu les évoquer à plusieurs reprises, il convient toutefois de ne pas oublier l'importance de la R&D pratiquée au sein des entreprises privées, qu'il s'agisse de géants tels qu'IBM, Google, Microsoft ou Facebook présents sur le territoire britannique, ou encore de startups et PMEs désireuses d'exploiter les multiples opportunités offertes par ces nouvelles technologies.

Nous avons donc fait le choix de nous attarder sur les acteurs publics de la recherche, mais aussi sur les principaux opérateurs facilitant les transferts de technologie depuis le milieu académique vers le secteur économique, pleinement demandeur de cette ressource aujourd'hui perçue comme indispensable. Les décideurs publics ne s'y trompent d'ailleurs pas : alors que l'accent avait longtemps été mis sur les infrastructures, les stratégies et plans d'action intègrent désormais pleinement la dimension de la donnée comme ressource à forte valeur ajoutée.

Le paysage actuel de la recherche britannique témoigne néanmoins de la dualité du concept de *data science* ; en phase de gagner ses galons de champ disciplinaire auquel sont consacrés des laboratoires, des départements, voire des établissements entiers, il n'en conserve pas moins ses caractéristiques profondément fonctionnelles pour de nombreuses disciplines, qui de fait s'équipent de plus en plus à

[129] http://www.thedatalab.com/

[130] Cette fonction le différencie donc des centres Catapult, qui eux, n'apportent pas de financements directs par appels à projets à la R&D.

[131] http://censis.org.uk/

[132] http://www.ktn-uk.co.uk/

[133] https://connect.innovateuk.org

la fois en personnels et en machines permettant de mener des projets de pointe. Mais quel que soit l'angle d'approche, le pays affirme sa volonté de construire un écosystème performant et une communauté de *data scientists* confirmés capable de répondre aux nombreux besoins identifiés.

Cheville ouvrière des capacités de travail dans ce domaine, la question des infrastructures d'administration et de calcul n'en est pas pour autant occultée ; le Royaume-Uni a mené ces dernières années un effort soutenu afin de développer un écosystème varié et surtout adapté à la diversité des demandes de sa communauté scientifique. Il s'agit néanmoins d'un processus toujours en cours, qu'il conviendra de surveiller avec attention dans les années à venir.

Nous nous situons donc aujourd'hui à un moment charnière pour l'avenir de la recherche britannique et le développement de son écosystème économique sur le thème de ce dossier. Entre émergence d'acteurs ambitieux et diversification des applications, renouvellement des infrastructures et définition de nouvelles stratégies au plus haut niveau politique, le pays est en cours de construction des fondations sur lesquelles reposent ses ambitions de *leadership* dans un des secteurs les plus compétitifs à l'échelle internationale.

Dossier rédigé par Eva Legras.

Bibliographie indicative :

Gouvernement :

- Prof. Dominic Tildesley, « A strategic vision for UK e-infrastructure », 2012, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32499/12-517-strategic-vision-for-uk-e-infrastructure.pdf

- HM Government, « E-Infrastructure: the ecosystem for innovation. One Year On », 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249474/bis-13-1178-e-infrastructure-the-ecosystem-for-innovation-one-year-on.pdf

- Council for Science and Technology, « The Age of Algorithms: algorithms, analytics, modelling, and data for growth and public sector efficiencies », juin 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224953/13-923-age-of-algorithms-letter-to-prime-minister__1_.pdf

- HM Government, « Seizing the data opportunity, a strategy for UK data capability », octobre 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

- HM Government Horizon Scanning Programme, « Emerging Technologies: Big Data », décembre 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf

- Council for Science and Technology, « Science Landscape Seminar Series: Representative Big Data, Supercomputing and E-Infrastructure », juin 2015, https://www.gov.uk/government/publications/science-landscape-seminar-big-data-e-infrastructure-and-supercomputing

Parlement :

- Parliamentary Office of Science and Technology, « Big Data : an overview», juillet 2014, http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-468

- Parliamentary Office of Science and Technology, « Big Data and Business », juillet 2014, http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-469

- Parliamentary Office of Science and Technology, « Trends in ICT », octobre 2015, http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0510

- House of Lords Select Committee on Digital Skills, « Make or Break: the UK's Digital Future », février 2015, https://www.publications.parliament.uk/pa/ld201415/ldselect/lddigital/111/111.pdf

- House of Commons Science and Technology Committee, « The Big Data Dilemna », janvier 2016, http://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf

- House of Commons Science and Technology Committee, « Digital Skills Crisis », juin 2016, http://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf

Régulation :

- Information Commissioner's Office, « Big Data and data protection », 2014, https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

Recherche & innovation :

- Centre for Economic and Business Research, « Data equity, unlocking the value of Big Data », 2012, http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf

- Research Councils UK, « RCUK and Big Data », 2013, http://www.rcuk.ac.uk/documents/documents/big-data-timeline-web-pdf/

- Research Council UK, « e-infrastructure roadmap », 2014, http://www.rcuk.ac.uk/documents/documents/roadmapforelc-pdf/

- EPSRC, « Strategy for the developing landscape of Tier-2 HPC in the UK », 2014, https://www.epsrc.ac.uk/files/research/tier2hpcstrategy/

- Craig Kirkwood, « UK High Performance Computing, An Overview of Capability », KTN e-Infrastructure special interest group, 2014, https://connect.innovateuk.org/documents/3108763/0/UK+HPC+capability+report/87d65dd0-82da-42da-9e7b-18a97ea06003

- Imperial College London, « Data Science Institute », 2014, https://www.imperial.ac.uk/media/imperial-college/data-science-institute/DSI_booklet_Web.pdf

- Imperial College London, « Data Science Institute Strategy 2016-2021 », 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224953/13-923-age-of-algorithms-letter-to-prime-minister__1_.pdf

- Michael Ball et al., « UK e-Infrastructure for Research - UK/USA HPC Workshop », Oxford, July 2015, http://www.slideshare.net/comth/uk-einfrastructure-for-research-ukusa-hpc-workshop-oxford-july-2015

- Martin Hamilton et al., « Report of 2015 UK National e-Infrastructure Survey of HEIs and Research Institutes », JISC, 2016, http://hpc-sig.org.uk/wp-content/blogs.dir/sites/63/2016/03/NeI-PDG-UK-National-E-Infrastructure-Survey-2015-Report.pdf

- The Alan Turing Institute, « Our roadmap for science and innovation », novembre 2015, https://turing.ac.uk/content/uploads/2015/12/SIS_for_ATI_WebPage.pdf

- The Alan Turing Institute, « Shaping our strategy », juin 2016, https://turing.ac.uk/content/uploads/2016/06/Shaping-our-Strategy-The-Alan-Turing-Institute.pdf

- The Alan Turing Institute, « Scoping Programme », juin 2016, https://turing.ac.uk/content/uploads/2016/06/The-Alan-Turing-Institute-Scoping-Programme.pdf

Fiches et Rapports d'ambassade :

- SST Royaume-Uni, « La e-santé au Royaume-Uni : les données médicales personnelles et le secteur de la recherche », 2013, http://www.ambafrance-uk.org/La-e-sante-au-Royaume-Uni

- SST Royaume-Uni, « L'innovation publique dans les smart cities au Royaume-Uni », 2014, http://www.ambafrance-uk.org/L-innovation-publique-dans-les

- SST Royaume-Uni, « Les Centres Catapult », 2015, http://www.ambafrance-uk.org/Dossier-Les-centres-Catapult

- SST Royaume-Uni, « Fiche Pays Recherche Royaume-Uni », 2016, file:///C:/Users/VI-IT/Downloads/20160805_fichecurie.pdf

- SST Japon, « Calcul haute performance au Japon », 2016, http://diplomatie.gouv.fr/fr/IMG/pdf/fr-2016-04_hpc_note_ytc_cle8f8eb7.pdf

Consultants & Presse :

- McKinsey & Company, « Big Data : The next frontier for innovation, competition, and productivity », 2011, http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

- Daniel Zeichner, « The big data explosion sets us profound challenges - how can we keep up? », The Guardian, juillet 2016, https://www.theguardian.com/science/political-science/2016/jul/02/the-big-data-explosion-sets-us-unprecedented-challenges-how-can-we-keep-up?CMP=share_btn_tw

www.ambascience.co.uk

Les articles de Science et Technologie au Royaume-Uni publiés par le service Science et Technologie de l'Ambassade de France au Royaume-Uni sont diffusés selon les termes de la licence Creative Common Paternité-Pas d'utilisation commerciale-Partage des Conditions Initiales à l'Identique 4.0 France. Les autorisations au-delà du champ de cette licence peuvent être obtenues via une demande par courriel à info@ambascience.co.uk