



HAL
open science

Open Data en éducation : enjeux, développement et stratégie pour la bibliothèque de l'IPE

Yves Michel Thomas

► **To cite this version:**

Yves Michel Thomas. Open Data en éducation : enjeux, développement et stratégie pour la bibliothèque de l'IPE. domain_shs.info.docu. 2021. mem_03710368

HAL Id: mem_03710368

https://memic.ccsd.cnrs.fr/mem_03710368v1

Submitted on 30 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



le **cnam**
intd

Open Data en éducation : enjeux, développement et stratégie pour la bibliothèque de l'IPE

Mémoire
pour l'obtention du Titre professionnel
« Chef de projet en ingénierie documentaire
et gestion des connaissances »
Niveau 7 – Bac +5

présenté et soutenu par :

Yves Michel THOMAS

Date et lieu de la soutenance

- Mardi 29 juin 2021
- Soutenance en visioconférence

Membres du jury

- Chartron, Ghislaine, directrice, professeure du CNAM, HDR
- Bitoun, Corinne, spécialiste de l'information à la bibliothèque de l'IPE

Promotion 50 (2019-2020)



Paternité - Pas d'Utilisation Commerciale - Pas de Modification

Remerciements

Je voudrais remercier Mme Chartron pour sa patience et son accompagnement bienveillant.

Je remercie Mme Bitoun pour sa disponibilité et ses encouragements.

Mes remerciements à toute l'équipe de la bibliothèque de l'IPE, et aux collègues qui ont répondu positivement à mes sollicitations.

Aux formateurs du CNAM, et particulièrement à Mme Raïs qui, souvent, a su trouver les mots justes pour rassurer.

Notice

Yves Michel THOMAS. Open Data en éducation : enjeux, développement et stratégie pour la bibliothèque de l'IIEP. Mémoire professionnel INTD, Titre I, Chef de projet en ingénierie documentaire. Conservatoire national des arts et métiers – Institut national des Sciences et Techniques de la Documentation, 2021, 72 p. Promotion 50.

Ce mémoire, réalisé dans le cadre d'un stage de formation, explore la problématique de l'Open Data, et particulièrement des données ouvertes de l'éducation. À travers un survol historique de ce mouvement, il présente l'évolution, le développement et les enjeux de l'Open data de l'éducation. L'analyse de quelques plateformes permet d'illustrer les critères, les enjeux, et les difficultés inhérentes à la mise en œuvre d'une politique d'Open data. Principalement en ce qui concerne la diffusion sur internet de données ouvertes de qualité. Et, dans l'hypothèse d'une implication de la bibliothèque de l'IIEP dans une mission d'assistance technique à l'ouverture des données, l'étude propose quelques éléments de définition d'une stratégie d'ouverture des données de l'éducation.

Descripteurs

- Open data
- Données ouvertes
- Données publiques
- Comparaison internationale
- Education
- Guide stratégique
- Bibliothèque
- IIEP-UNESCO

This thesis, produced as part of a training course, explores the issue of Open Data, and particularly Open Data in education. Through a historical overview of this movement, it presents the evolution, development and challenges of Open Data in education. The analysis of a few platforms illustrates the criteria, the challenges and the difficulties inherent to the implementation of an Open data policy. Mainly with regard to the dissemination of quality open data on the Internet. And, in the hypothesis of the involvement of the IIEP library in a technical assistance mission for open data, the study proposes some elements for the definition of an open data strategy for education.

Keywords

- Open data
- Public data
- International comparison
- Strategic guide
- Education
- Library
- IIEP-UNESCO

Table des matières

Remerciements	2
Notice	3
Table des matières	4
Listes des illustrations	6
Introduction	7
I. Première partie : L’Open Data et le secteur de l’éducation	10
1. L’Open Data, une question contemporaine	10
1.1. Les origines du mouvement : recherche d’une généalogie.....	10
1.2. La construction de la donnée publique ouverte.....	11
1.3. Des barrières notables.....	14
2. Les données ouvertes de l’éducation	15
2.1. Les données ouvertes « dans » l’éducation	15
2.2. Autour des données ouvertes de l’éducation	16
2.3. Les enjeux de l’Open Data de l’éducation.....	17
2.4. Les données ouvertes de l’éducation dans le baromètre Open Data	19
2.5. Les classes de réutilisation	19
3. De quelques plateformes et de la qualité des données de l’éducation.....	21
3.1. Le corpus : les plateformes du Brésil, de la France, de l’Inde et du Kenya	21
3.2. Fonctionnalités des portails.....	23
3.3. La qualité des données ouvertes de l’éducation	26
3.3.1. Contexte d’ouverture : le cadre légal	27
3.3.2. Licences.....	28
3.3.3. Formats.....	28
3.3.4. Des métadonnées.....	29
3.4. Conclusion de la première partie.....	32
II. Deuxième partie : Le cadre d’accueil : l’IPE	33
1. L’IPE et ses activités	33
1.1. Une mission pour une organisation statutairement autonome	33
1.2. Organigramme et fonctionnement de l’IPE	33
1.3. Le cœur de métier de l’institut : développement des capacités, production de connaissances, diffusion	34
1.3.1. Formation	35
1.3.2. Coopération technique.....	36
1.3.3. Recherche et Développement.....	36
1.3.4. Partage des connaissances	36
1.4. La bibliothèque de l’IPE	37
1.4.1. Mission, personnel et fonctionnement.....	37

1.4.2. Fonds documentaire et principales bases de données	38
1.4.3. Une attention au contexte informationnel.....	39
2. Les données vues des Nations unies : un aperçu.....	41
2.1. L'ONU et les « data »	41
2.1.1. Nécessaire « révolution » pour les ODD.....	41
2.1.2. « Une stratégie des données pour une action de tous partout »	42
2.2. Les « data » du côté de l'IPE et contexte du stage	43
2.2.1. Données de la recherche et politique de libre accès	43
2.2.2. Le projet « Les données géospatiales dans la planification et la gestion de l'éducation ».....	44
2.2.3. Données ouvertes sur l'école	44
2.2.4. Open Data et SIGE, quels effets bénéfiques ?.....	46
2.2.5. Contexte du stage.....	46
III. Troisième partie : Éléments pour un guide stratégique pour l'Open data de l'éducation	48
1. La bibliothèque de l'IPE face à l'Open data	48
2. Élaborer une stratégie d'ouverture des données de l'éducation	50
2.1. Cerner les enjeux, clarifier les objectifs	50
2.2. Pilotage opérationnel.....	51
2.3. Diagnostic de l'existant	51
2.4. Stratégie de publication	54
2.5. Préparation des données.....	58
2.6. Accompagner les données publiées.....	61
Conclusion.....	64
Bibliographie	66

Listes des illustrations

Tableau 1. Le modèle stellaire de T. Berners-Lee.....	12
Tableau 2. Formats de fichiers utilisés par les plateformes du corpus	29
Figure 1. Catégories et jeux de données suggérés par le G 8	13
Figure 2. Freins du point de vue des fournisseurs d'open data.....	15
Figure 3. Capture d'écran data.gouv.fr	17
Figure 4. Capture d'écran data.gob.es.....	17
Figure 5. Organisation utilisant données ouvertes par secteur. Open data impact map.....	18
Figure 6. Évolution de la qualité des données ouvertes.....	19
Figure 7. Réutilisation des données ouvertes dans le secteur de l'éducation (2018)	20
Figure 8. Capture d'écran data.gouv.fr.....	24
Figure 9. Capture d'écran data.gov.in. Sector	24
Figure 10. Capture d'écran opendata.go.ke	25
Figure 11. Capture d'écran data.gouv.fr.....	29
Figure 12. Capture d'écran data.gov.in	30
Figure 13. Capture d'écran dados.gov.br.....	31
Figure 14. Capture d'écran planipolis.iiep.unesco.org/fr	39
Figure 15. Capture d'écran. undp.org.	41
Figure 16. Capture d'écran. Open data Corsica. État de l'art	53
Figure 17. Capture d'écran data.gouv.fr.....	59
Figure 18. Capture d'écran. Freins à l'Open data du point de vue des réutilisateurs	61

Introduction

L'entrée dans l'ère numérique, souvent comparée à une révolution à l'instar de l'invention de l'imprimerie à caractères mobiles, a entraîné une explosion du volume de données collectées. Via les différents capteurs, nos ordinateurs et nos téléphones en sont parmi les plus communs, le volume de données générées et échangées quotidiennement atteint environ 2,5 quintillions d'octets. Ainsi on en vient à évoquer une « révolution des données » pour désigner notamment leur inflation exponentielle¹ et la capacité de traitement et de calcul avec des machines de plus en plus performantes. Données (volumineuses et variées) et performance computationnelle (vitesse de calcul) définissent ainsi la *règle des 3 v* du *Big data*. Les données publiques, mises à disposition dans le cadre de l'Open data, font partie de cette variété de données collectées et traitées sans en constituer l'essentiel. Au croisement du numérique, il s'agit plutôt de deux dynamiques parallèles : le big data renvoie à une méthode de collecte et d'analyse des données, et l'Open data s'inscrit dans la problématique de l'accès aux données et aux informations publiques. L'Open data est un mouvement impulsé par des acteurs issus de la société civile animés de l'idéal démocratique de la participation citoyenne. Aussi est-il très éloigné des motivations marchandes des géants de l'économie numérique qui ont recours au big data². D'une certaine manière, il faut plutôt rapprocher le mouvement en faveur de l'ouverture des données des initiatives *civic tech*, et l'Open data de l'*Open gouvernement*.

Ayant pris naissance aux États-Unis d'Amérique, les premiers acteurs de l'Open data se situent dans le sillage des pionniers de l'internet qui ont porté l'idée des communs du numérique. En effet, sensibilisés aux enjeux du numérique, en ce qui concerne la possibilité de rendre accessibles des quantités appréciables de données collectées par l'administration, ils y ont vu une opportunité de renforcement de la participation citoyenne dans les affaires publiques. Bien inspirés en ce sens de l'idéal démocratique américain de liberté d'opinion et de diffusion l'information publique – une des expressions de la méfiance envers le Léviathan sur les rives occidentales de l'Atlantique. Cette formule, très probablement apocryphe, attribuée à Thomas Jefferson, « *information is the currency of democracy* » est assez évocatrice des motivations du mouvement Open data. C'est en réponse à d'intenses demandes exprimées durant la campagne présidentielle de 2008, que le président Barack Obama a signé dès son entrée en fonction, en janvier 2009, le *Memorandum on Transparency and Open Government*. Il s'agit de la première initiative qui devait permettre la publication des données fédérales sur le portail *data.gov*, dès le mois de mai. L'exemple a été repris l'année suivante en Grande-Bretagne, puis dans d'autres pays anglo-saxons. Les trois principes de base qui définissent la donnée ouverte, à savoir réutilisabilité, interopérabilité, et gratuité, ont pu être concrètement mis en application. Actuellement on dénombre plus de 300 portails de données administratives ouvertes dans le monde, dont plus d'une cinquantaine plateformes gouvernementales³. Autant dire qu'il s'agit d'une initiative qui s'est internationalisée. Ces plateformes publient des données réparties en catégories, qui circonscrivent des secteurs d'intervention de l'administration, dont l'éducation, telles qu'elles ont été proposées lors du sommet du G 8 de 2013 en Irlande.

La plupart des études sur l'Open data soulignent une faible réutilisation (identifiée ou signalée) relativement au volume de données publiées sur les plateformes. Par ailleurs, lorsque l'on consulte les portails, il apparaît que les données ouvertes de l'éducation ne font pas partie de celles qui font l'objet de plus de réutilisation. Comment expliquer une telle situation ? Peut-on

¹ Le volume de données évalué à 2 zettaoctet (zo) en 2010, en atteindra 175 zo en 2025. (Un zettaoctet équivaut à mille milliards de gigaoctets).

² Cela ne signifie pas que seuls les acteurs économiques ont recours au big data

³ Il s'agit exclusivement de plateformes d'entités étatiques ou infra-étatiques. Voir [Data.gov](https://www.data.gov/open-gov/),

<https://www.data.gov/open-gov/>

Cependant, [DataPortals.org](http://datacatalogs.org/), recense au total 590 portails. Voir, <http://datacatalogs.org/>.

mettre en cause les jeux de données publiés ? Est-ce qu'il faut approcher les données « brutes » de l'éducation de la même manière que celles issues d'autres secteurs comme les transports, l'économie, la finance ? La réutilisation des données de l'éducation exigerait-elle des mesures incitatives spécifiques ? Il est difficile de suggérer des réponses à ces questions dans le cadre de ce travail. Il faut toutefois noter que lorsque les données de l'éducation font l'objet de réutilisation, il s'agit essentiellement de la fourniture de services destinés à sélectionner les meilleurs établissements ; à l'amélioration du fonctionnement du système scolaire et la rationalisation de l'offre de services éducatifs ; à la promotion de la transparence pour atténuer les effets de la corruption ; et au renforcement de la redevabilité par la participation des parents d'élèves dans la vie des établissements. Ainsi, l'Open data de l'éducation est en résonance avec l'idée des tableaux de bord dans les systèmes scolaires⁴. Les tableaux de bord ont fait l'objet d'un programme de recherche de l'Institut international pour la planification de l'éducation (IIPE) qui porte sur l'éthique et la corruption dans l'éducation. Le tableau de bord peut être considéré comme une forme de réutilisation de donnée de l'éducation destinée à fournir des informations notamment sur les établissements scolaires. Il faut néanmoins souligner qu'il n'est pas tributaire d'une politique publique d'Open data. En tant qu'institut faisant partie du système des Nations Unies, le programme de recherche de l'IIPE s'inscrit dans la quête des méthodes les plus efficaces pour arriver à une utilisation optimale de ressources allouées à l'éducation. Dans un contexte de tarissement de l'aide au développement, une meilleure utilisation des fonds est considérée comme une condition incontournable si l'on veut se rapprocher de l'*Objectif 4 de développement durable* (ODD 4).

Pour l'IIPE, il ne s'agissait pas de promouvoir l'Open data de l'éducation auprès des États, au sens où l'entendaient les initiateurs de ce mouvement. Même si les États sont de plus en plus nombreux à se lancer dans l'ouverture de leurs données administratives⁵. Cependant, nous estimons qu'il ne serait pas irréaliste, compte tenu de la reconnaissance de son expertise en planification de l'éducation et de la diffusion de ses publications sur les données ouvertes de l'école, qu'un État membre puisse solliciter son assistance sur les données de l'éducation dans le cadre d'une politique d'Open data. S'il reste évident que cette assistance ne saurait porter sur les questions proprement informatiques du projet, l'hypothèse d'un rôle de conseil n'est pas à exclure. À quelles conditions dans ce cas, une bibliothèque comme celle de l'IIPE, et compte tenu de sa mission et des modalités de la coopération de l'institut avec les États membres, pourrait trouver sa place dans le cadre d'une assistance portant sur l'Open data ? Telle est la question centrale qui sous-tend ce travail effectué dans le cadre de notre mission de stage à l'IIPE.

Il se divise en trois parties. Dans un premier temps, nous avons procédé à un état de l'art de l'Open data de l'éducation. Nous y avons effectué une brève historique du mouvement de l'Open data, de la place des données de l'éducation dans l'univers des données administratives ouvertes, et une revue d'un corpus de quatre (4) plateformes nationales de données ouvertes qui publient des données de l'éducation. Il est donc question de *plateformes orientées fourniture* pour reprendre la terminologie de Turki et Foulonneau [1, Broudoux, Chartron, sous dir. p. 117].

Dans un second temps, nous avons très succinctement présenté le cadre de l'expérience de stage. Ensuite, nous avons abordé la question des data au sein du système onusien ou de préférence présenter comment cette organisation souhaite emprunter ce tournant des

⁴ « *Tableau de bord* : document visant l'agrégation d'informations sur les ressources, les processus et/ou les performances concernant un niveau administratif du système éducatif : circonscription, district ou région, et destiné à un usage interne et/ou externe dans l'administration scolaire ». Voir Institut international de planification de l'éducation. *Information et transparence : Tableaux de bord des écoles en Afrique subsaharienne*, p. 18.

⁵ Des politiques d'Open data inscrites dans certains cas dans le cadre plus large de l'*Open government*. A l'heure actuelle, le Partenariat pour le gouvernement ouvert regroupe 78 États.

données. Finalement, nous avons soulevé la place ou l'absence des *data* au sein de l'IPE. Faut-il préciser qu'il ne s'agit pas d'une exposition de la position institutionnelle, mais de plutôt de recueil des points de vue de certains cadres.

Dans un troisième et dernier temps, nous avons proposé des éléments d'un guide stratégique dans le cadre d'une démarche ouverture des données à destination de la bibliothèque de l'IPE. En essayant autant que possible de nous en tenir aux données de l'éducation. Son ambition est surtout d'informer afin d'entrouvrir une porte sur l'Open data. Puisque l'institut ne compte pas à ce jour se lancer dans l'ouverture de ses données, et n'a pas encore participé à la mise en œuvre d'un projet idoine, nous n'avons pas pu travailler sur un cas concret de projet d'Open data. À n'en pas douter, cela constitue la principale limite de ce travail qui est de portée générale. S'il arrive à susciter de l'intérêt pour l'Open data et permettre de nourrir certaines réflexions au sein de l'équipe de la bibliothèque, ne serait-ce pour améliorer certaines pratiques, nous aurons plutôt œuvré utilement.

I. Première partie : L'Open Data et le secteur de l'éducation

1. L'Open Data, une question contemporaine

1.1. Les origines du mouvement : recherche d'une généalogie

À l'échelle de l'histoire des sociétés, l'Open data est un mouvement relativement récent, et étroitement lié aux évolutions dans le domaine des technologies de l'information et de la communication (TIC). Néanmoins, si l'on se réfère aux idéaux promus et aux finalités des porteurs de ce mouvement, il est possible, tout en évitant le piège d'une approche téléologique, de l'inscrire dans la filiation d'une dynamique plus ancienne de démocratisation des sociétés contemporaines. En effet, il est question *libérer* les données publiques en les rendant accessibles aux citoyens ordinaires afin qu'ils puissent être en mesure de se faire leur propre jugement, et comprendre les soubassements des décisions administratives. Et, bien évidemment, de pouvoir les contester de manière éclairée, les passer au crible de la critique. Nous sommes donc au cœur de l'enjeu de transparence de l'action publique. Vu sous cet angle, on peut considérer le mouvement Open data comme une des manifestations de la vieille quête du « bon gouvernement », si l'on se réfère au monde euro-américain. Des traditions philosophiques et des courants politiques ont porté avec constance la revendication de sortir les décisions des gouvernants de l'opacité caractéristique d'un certain art de gouverner, fondement de la toute-puissance du souverain. Le parlementarisme, la liberté de s'informer et de la presse constituent différentes expressions d'un mouvement historique visant la soustraction des décisions engageant la collectivité du sceau du secret.

Du fait de l'extension des domaines d'intervention de l'État, le citoyen est confronté de plus en plus à une administration publique chargée de traiter différents dossiers, ou de répondre à des besoins et demandes de nature diverse. Ainsi la bureaucratie étatique produit et est en possession d'une quantité astronomique d'informations et de données concernant maints aspects de la vie sociale, sur les citoyens. Dès lors, la possibilité d'accéder aux données publiques collectées devient un enjeu démocratique. Le renforcement du contrôle démocratique sur l'usage qui en est fait, et l'effectivité du principe de la reddition des comptes apparaissent de plus en plus comme une nécessité. Aussi, par exemple, aux États-Unis avec le *Freedom Information Act*, (1966), et en France avec la *Commission d'accès aux documents publics* (1978) tout citoyen pouvait-il obtenir la transmission de documents classifiés [2, Cardon]. Toutefois, il a fallu attendre l'entrée dans le XXI^e siècle pour assister à une accélération de la mobilisation en faveur de l'ouverture des données publiques grâce à la mobilisation de l'ONG américaine *Sunlight Foundation* qui promeut les *civic tech*, ou les technologies civiques, [3, CNIL]. La présidence Barack Obama aux États-Unis (2009-2017) a apporté une première réponse politique à ces demandes citoyennes. Dès son entrée en fonction, le président américain a mis en place le cadre réglementaire adéquat pour renforcer la transparence de l'action des autorités fédérales grâce aux technologies du web.

Il importe toutefois de souligner que la question de l'ouverture des données a d'abord émergé dans le champ scientifique. En effet, dans un article datant de 1942, *The normative structure of science*, le sociologue américain Robert Merton avait montré le bénéfice de l'ouverture des données scientifiques [4, Chignard]. La notion de « données ouvertes » a été forgée dans les milieux scientifiques principalement dans le domaine des recherches en géophysique et en sciences environnementales entre autres. Étant donné que l'objet de ces recherches ne saurait être circonscrit dans les limites des frontières étatiques, certains savants avaient estimé logique que les données produites puissent être librement échangées avec l'ensemble des communautés scientifiques de différents pays. Il s'agit également d'une réactualisation de l'idéal humaniste qui a fait dire naguère, même à un moment de passion pour le national, que

le savoir est le patrimoine de l'humanité et que la science n'a pas de patrie. La dimension collaborative du travail scientifique favorisé la constitution de communautés en réaction à l'orientation propriétaire de l'industrie informatique. Elles s'opposent au verrouillage des logiciels et posent le principe de l'ouverture en promouvant l'idée de biens communs numériques.

Au-delà d'une simple question de principe (ouverture, collaboration, participation), les communautés de l'*open source* préconisent des logiciels ouverts sensés plus être sûrs, plus efficaces et plus robustes que les propriétaires. Plus fondamentalement, l'ouverture garantirait l'innovation continue. S'inspirant de l'*open source*, et en se basant sur des critères essentiellement juridiques, l'Open Knowledge Foundation (OKFN) a proposé en 2005 une définition du savoir ouvert qui renvoie à toute œuvre à laquelle tout le monde est libre d'y accéder, d'utiliser, de modifier et de partager⁶. Samuel Goëta estime qu'en « *posant la base d'un élargissement de l'open source au savoir, l'Open Definition a constitué une ressource précieuse pour l'ouverture des données publiques* » [5, Goëta, p. 20]. En effet, le terme « savoir » tel qu'employé par l'OKFN inclut des contenus (musique, film, livre), des données scientifiques, géographiques, historiques, des données publiques des administrations [6, OKFN]. La revendication de l'ouverture des données gouvernementales résulte en conséquence de la convergence de nombreux facteurs parmi lesquels l'accessibilité aux documents administratifs, et la promotion des communs. C'est dans un tel contexte, qu'en décembre 2007, des militants des libertés numériques, des entrepreneurs et des chercheurs se réunirent à Sebastopol, en Californie, pour jeter les bases de l'ouverture des données gouvernementales – *Open government data* (OGD).

1.2. La construction de la donnée publique ouverte

Les participants aux débats de Sebastopol (*Open Government Working Group Meeting*), une trentaine, ont travaillé à la définition des principes de l'ouverture des données gouvernementales. Cette indispensable étape de clarification, de délimitation du périmètre, de précision des modalités et des procédures de l'ouverture, devrait faciliter la promotion de l'ouverture des données publiques. À la fin de cette rencontre, les huit principes de l'Open data furent adoptés. L'ouverture des données publiques est présentée comme essentielle à la démocratie, et facteur d'amélioration de l'efficacité des services publics. L'Open data se définit alors en fonction des critères suivants :

- Données complètes (si elles ne portent pas atteinte à la vie privée et à la sécurité),
- Données brutes,
- Données à jour,
- Données accessibles à tous pour tout usage,
- Données structurées pour permettre un traitement automatisé par les machines,
- Données dont l'accès est non discriminant,
- Données sont en format non-propriétaire,
- Données diffusées sous licence libre.

À la veille d'une année électorale, les initiateurs du mouvement avaient souhaité un engagement des protagonistes politiques sur cette question. Ils ont réussi à susciter l'attention de l'équipe du candidat démocrate qui a su faire un usage efficace des technologies numériques dans la campagne électorale. Ainsi, dès sa prise de fonction, en janvier, le nouveau président Obama, a répondu à leurs attentes en signant un mémorandum sur la transparence et le gouvernement ouvert. Puis à la fin de l'année 2009, une directive sur les

⁶ <https://opendefinition.org/od/1.1/fr/>

conditions de l'ouverture fut envoyée aux différents ministères et agences fédérales. Il s'agissait alors d'un développement majeur, puisque l'Open data allait trouver pour la première fois une traduction concrète avec sa mise à l'agenda qui devait entraîner des décisions émanant des autorités publiques. Forts de ces premiers succès, les militants de l'ouverture ont continué à exiger l'application intégrale des critères de l'Open data. Par exemple, dès février 2009, Timothy Berners-Lee, considéré comme le principal inventeur du web, fut en première ligne pour obtenir que les données soient publiées à l'état brut avec le fameux slogan « *raw data now!* ». L'idée étant que les données « brutes » assureraient davantage de liberté d'interprétation et d'analyse de la part des réutilisateurs d'une part ; et, d'autre part, elles permettraient de gagner un temps précieux. Puis, en 2010, afin de pouvoir mieux évaluer les initiatives d'open data tracer une perspective, il proposa en 2010 un modèle hiérarchique d'ouverture des données de cinq étoiles. Avec cette approche souple et graduée, Berners-Lee considère la question de la licence comme le critère central de l'Open data. Toute donnée la donnée distribuée, indépendamment du format, avec une licence ouverte est considérée comme une donnée ouverte. Le nombre d'étoile correspondant à chaque palier (voir *Tableau 1, infra*) est surtout un indicateur de la facilité d'utilisation desdites données. Il s'agit d'un modèle davantage tourné vers l'utilisateur, les réutilisateurs en l'occurrence.

☆	Publier des données sur le web avec une licence ouverte indépendamment du format (en Pdf par exemple).
☆☆	Publier des données structurées, lisibles par les machines
☆☆☆	Publier des données dans un format ouvert et non propriétaire
☆☆☆☆	Utiliser des URI pour désigner les choses dans les données afin de pouvoir en faire référence
☆☆☆☆☆	Lier les données publiées à d'autres données pour fournir le contexte

Tableau 1. Le modèle stellaire de T. Berners-Lee

Lors de la réunion du G 8 en Irlande (2013) les autorités politiques reprirent l'initiative sur la question des données ouvertes. Les chefs d'États et de gouvernement considérées traditionnellement comme étant les principales économies du monde adoptèrent une à l'occasion d'adopter une charte sur l'Open data. Son préambule établit, entre autres, le constat d'un mouvement mondial créé par les nouvelles technologies, au cœur duquel se trouve l'open data. De l'avis de dirigeants réunis, ce mouvement recèle un potentiel énorme qui peut améliorer la performance des gouvernements et des entreprises, renforcer la participation citoyenne, et stimuler la croissance économique. En ce sens, ils se sont entendus sur cinq principes d'ouverture des données qui s'inscrivent dans la continuité des différentes initiatives de la société civile :

- ✓ Les données sont ouvertes par défaut ;
- ✓ Les données publiées doivent être de qualité et en quantité ;
- ✓ Les données doivent être accessibles et réutilisables par tous ;
- ✓ Les données doivent être de nature à améliorer la gouvernance ;
- ✓ Les données doivent encourager l'innovation.

Par ailleurs, une liste des quatorze catégories (avec illustration de leurs jeux de données prioritaires), recoupant des secteurs d'activité ou d'intervention publique, a été publiée à l'occasion (voir *Figure 1, p. 17*). Pour les signataires de la charte, cette liste reflète l'importance de ces catégories dans la perspective de l'amélioration de la gouvernance démocratique et l'innovation. Même si elle n'est toujours reprise à l'identique dans les différents portails, elle semble constituer une référence ou une source d'inspiration pour les plateformes gouvernementales.

Si ce renfort du groupe des principales puissances du monde, à l'exception de la Chine, a été un soutien conséquent à l'ouverture des données, cela n'a pas suffi pour autant à conférer à la charte une légitimité indiscutable. Ce « club des puissants » n'est pas suffisamment représentatif d'un monde en pleine mutation géopolitique, dans lequel les acteurs privés (notamment les associations) exercent une influence non négligeable. Ainsi, en marge de la 70^e session régulière de l'Assemblée générale de l'ONU en 2015, et à l'initiative de l'*Open Government Partnership* (OGP), la Charte internationale de l'open data (acronyme anglais ODC) vit le jour. Cette charte, fruit d'une collaboration inédite entre certains États des organisations de la société civile, a été adoptée par vingt-deux États et des collectivités infra-étatiques (villes, régions... Elle reprend l'essentiel de la charte du G 8 à laquelle fut ajouté un sixième principe introduisant la notion de développement inclusif. L'ODC constitue en quelque sorte une tentative de synthèse des multiples approches de l'Open data, pas toujours identiques en termes d'exigence, des différents acteurs étatiques et de la société civile.

Néanmoins, par-delà les divergences au sein de la famille de l'Open data, l'on a pu s'accorder sur la définition des contours de la donnée ouverte en fonction trois critères : **juridique**, la loi autorise la réutilisation des données ; **technique**, la diffusion des données en format ouvert et interopérables ; **économique**, mise à disposition sans monétisation directe. Si le mouvement Open data peut mettre en avant une philosophie généreuse, il peut susciter également certaines interrogations légitimes quant à ses résultats et ses conséquences. Notamment en ce qui concerne les réels bénéficiaires de la transparence, la fragilisation des modèles économiques, la privatisation des données publiques [4, Chignard, p. 52].

Data Category* (alphabetical order)	Example datasets
Companies	Company/business register
Crime and Justice	Crime statistics, safety
Earth observation	Meteorological/weather, agriculture, forestry, fishing, and hunting
Education	List of schools; performance of schools, digital skills
Energy and Environment	Pollution levels, energy consumption
Finance and contracts	Transaction spend, contracts let, call for tender, future tenders, local budget, national budget (planned and spent)
Geospatial	Topography, postcodes, national maps, local maps
Global Development	Aid, food security, extractives, land
Government Accountability and Democracy	Government contact points, election results, legislation and statutes, salaries (pay scales), hospitality/gifts
Health	Prescription data, performance data
Science and Research	Genome data, research and educational activity, experiment results
Statistics	National Statistics, Census, infrastructure, wealth, skills
Social mobility and welfare	Housing, health insurance and unemployment benefits
Transport and Infrastructure	Public transport timetables, access points broadband penetration

Figure 1. Catégories et jeux de données suggérés par le G 8

1.3. Des barrières notables

Les indispensables questionnements sur l'Open data, ont le mérite d'inciter à y réfléchir par-delà les possibilités offertes grâce aux technologies actuelles. Il s'agit donc d'un sujet de société. Par ailleurs, il importe de ne pas sous-estimer les barrières de toute sorte qui peuvent surgir dans le cadre d'une stratégie d'ouverture des données. Ces freins peuvent relever de différents paramètres : politique, légal, technique, financier, organisationnel, entre autres. Dans sa méthodologie d'évaluation du niveau de préparation à l'ouverture des données, la Banque mondiale a recensé les principaux éléments à considérer [7, Open Government data Working Group] :

- Leadership,
- Cadre réglementaire et juridique ;
- Structures institutionnelles ;
- Données au sein du gouvernement ;
- Demande pour les données ouvertes/l'engagement citoyen ;
- Écosystème relatif aux données ouvertes ;
- Financement.

Les barrières qui sont susceptibles de se dresser en relation avec les éléments préalablement soulignés méritent d'être appréciées en fonction des contextes. Fondamentalement, étant donné qu'il s'agit de publier des données publiques, l'engagement des plus hautes autorités politiques est indispensable. L'Open data a un coût financier qu'on gagnerait à ne pas sous-estimer. Surtout en considérant le cas des pays dits « en développement » qui font face à plusieurs autres problématiques cruciales. S'y engager revient à prévoir un budget annuel régulier pour répondre aux implications financières sans s'attendre à des retours palpables. Il y a une exigence d'investissement en ressources humaines et matérielles assez conséquent. Enfin, l'ouverture des données publiques peut être difficilement considérée comme un sujet « politiquement rentable », une aune à laquelle l'on mesurera l'action d'une personnalité politique, une réalisation susceptible de faire grimper la cote de popularité dans un sondage d'opinion. Sauf s'il y a une forte demande qui émerge de la société, un engagement citoyen très affirmé.

Selon la Banque mondiale, l'impact de la dimension financière peut être atténué si l'on réfléchit en termes d'écosystème. Autrement dit si le projet est inséré dans une approche plus ample de renforcement des capacités des institutions publiques, de financement de l'innovation et de l'infrastructure informatique entre autres. Dans la pratique, l'aspect financier semble demeurer un frein sérieux au développement d'une politique d'ouverture des données. Même dans les pays à hauts revenus. Notamment parce qu'il s'agira d'offrir des données qui auparavant étaient des produits commerciaux qui contribuaient au financement des activités de certaines institutions. Ainsi, dans le rapport analytique de *l'European Data Portal* sur les barrières à l'ouverture des données publiques dans les 28 pays de l'Union européenne, le financement y figurait comme un des principaux freins [8, European Data Portal, p. 8].

Barriers for Open Data publishers

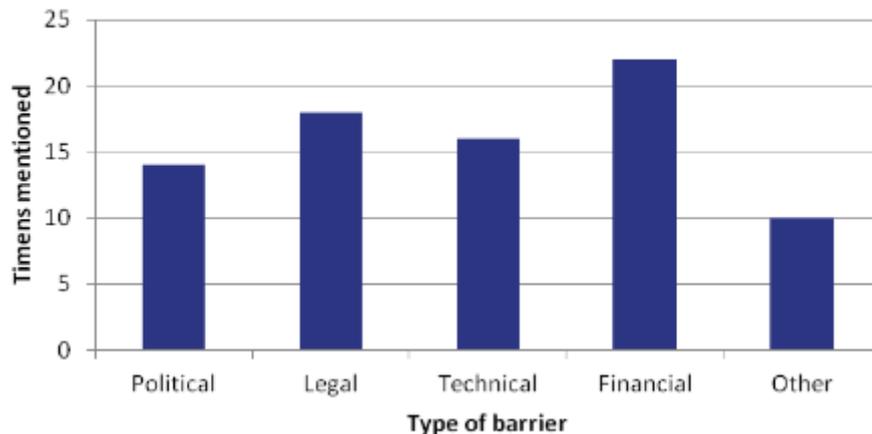


Figure 2. Freins du point de vue des fournisseurs d'open data

2. Les données ouvertes de l'éducation

2.1. Les données ouvertes « dans » l'éducation

La liste des catégories de données ouvertes proposées par le G 8 fournit des exemples de jeux de données de l'éducation qui peuvent entrer dans la catégorie « Éducation ». On y trouve : *Liste des écoles ; performance des écoles, compétences numériques*. Par ailleurs, dans la catégorie « Science et Recherche », on trouve des exemples de jeux de données issus des *activités de recherche et pédagogiques*. En toute logique, il est possible de considérer ces données comme données du secteur de l'éducation : données de recherche d'une part ; et données pédagogiques ouvertes, d'autre part. Alors, les données ouvertes dans le secteur de l'éducation peuvent englober divers types de données, et également donner lieu à différentes réutilisations. D'ailleurs les signataires de la Charte internationale de l'Open Data avaient pris l'engagement d'encourager l'intégration des données ouvertes dans les programmes scolaires afin d'inculquer la culture des données dès le plus jeune âge. Laquelle culture permettrait de profiter pleinement de l'ouverture, de renforcer la capacité d'action de ces futurs citoyens. Certains chercheurs estiment qu'il est possible d'aborder l'Open data dans le champ de l'éducation selon trois perspectives, dépendamment des acteurs considérés [9, Atenas, Havemann] :

- *Décideurs politiques* : surtout intéressés aux données pour la planification de l'éducation, la politique éducative ;
- *Parents et étudiants* : en quête d'informations sur le classement des établissements afin de faire « des choix rationnels sur marché de l'éducation » ;
- *Enseignants* : qui utilisent les données (de l'éducation et de n'importe quelle catégorie) à des fins d'enseignement, comme ressources pédagogiques.

Cependant, l'utilisation des données ouvertes à des fins d'enseignement reste peu développée. Elles ne seraient pas immédiatement utilisables, car, d'une part, elles sont issues de pratiques professionnelles réelles ; et, d'autre part, la plupart des enseignants ne sont pas encore suffisamment formés à la culture des données. Pourtant, les données ouvertes dans l'éducation sont une condition nécessaire pour une réelle exploitation de la dynamique de l'open data dans toutes ses potentialités. La *data literacy* est indispensable pour un réel approfondissement de la démocratie citoyenne si l'on pose les données comme bien commun. Il y a aussi un enjeu important de culture informationnelle dans la société dite de l'information. Selon Anne Lehmanns et Camille Capelle, « [la] *thématique des données ouvertes en éducation*

permet () de concilier une approche technique et informatique avec une approche culturelle et critique des questions informationnelles » [10, Lehmans ; Capelle, p. 13].

2.2. Autour des données ouvertes de l'éducation

Comment définir les données de l'éducation ? En s'autorisant cette tautologie, l'on peut avancer qu'une donnée de l'éducation est une donnée. Selon la définition de la Commission électrotechnique internationale, une donnée est une « *représentation d'une information sous une forme conventionnelle convenant au traitement par des moyens humains ou automatiques* »⁷. La donnée numérique « *est le résultat direct d'une mesure, soit elle est existante dans un logiciel ou un progiciel dans lequel elle a été entrée ou calculée. Elle peut être collectée dans une base de données d'une application (), par un capteur () ou numérisée depuis des données manuscrites* » [11, Kobler, p.]. Dans notre contexte, la donnée est la description la plus élémentaire d'une réalité. Elle peut se présenter sous une forme quantitative (30 °C, par exemple), ou qualitative (rouge, par exemple). Si ces éléments permettent d'avoir une idée plutôt précise de la notion de donnée, alors que serait-ce une donnée de l'éducation dans le cadre de l'ouverture des données publiques ? Il s'agit de données publiques relatives à l'éducation en tant que système éducatif. De manière non exhaustive, ces données concernent, d'une part, les élèves, les enseignants, le personnel administratif des établissements et d'autre part, les ressources matérielles, les allocations budgétaires, les structures administratives. Ainsi, les données ouvertes de l'éducation, open data de l'éducation, sont les données produites ou utilisées dans le cadre du fonctionnement d'un système éducatif et diffusées dans un contexte d'open data.

Cette acception pragmatique, évite toute approche fondationnelle qui s'évertuerait à retrouver des caractères, des propriétés, ou de l'essence d'une donnée de l'éducation. Les coordonnées géospatiales d'un établissement scolaire en contexte d'open data de l'éducation sont ici considérées comme des données ouvertes de l'éducation. Par exemple, dans le plan d'action pour le gouvernement ouvert du Kirghizistan, il est prévu de diffuser des données relatives à la structure des établissements scolaires pour signaler ceux qui répondent ou pas aux normes parasismiques [12, Kyrgyz Republic Action Plan 2018-2020]. Plus succinctement, il n'est pas toujours aisé de délimiter l'éventail des données de l'éducation en contexte d'open data. La répartition thématique des jeux de données dans la plate-forme d'open data du ministère de l'Éducation nationale en France peut être très utile pour aider à saisir en quoi peuvent consister concrètement les données ouvertes de l'éducation (*Figure 3, page 21*). Face à la diversité des données qui peuvent relever de données de l'éducation en contexte d'ouverture des données publiques, une réflexion sur leur pertinence, leur d'intérêt pour la communauté éducative, et le public en général, s'avère dès lors nécessaire. Une étude portant sur les principes du gouvernement ouvert dans les universités espagnoles a identifié les jeux de données les plus consultés (*Figure 4, page 21*). Aussi, recueillir en amont, à travers un processus participatif, les attentes potentielles du public dans une entreprise d'ouverture des données est généralement recommandé, et même nécessaire. Penser à l'exposition de données de l'éducation qui intéressent le public est un des critères de réussite d'une stratégie d'open data.

⁷ Voir, <https://www.electropedia.org/iev/iev.nsf/display?openform&ievref=171-01-02>

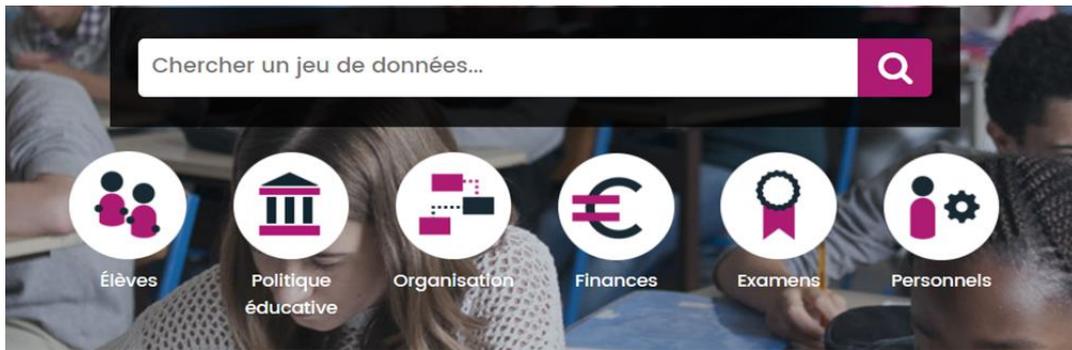


Figure 3. Capture d'écran data.gouv.fr



Figure 4. Capture d'écran data.gob.es

En dépit de leur contexte différent, ces deux illustrations donnent une idée des jeux de données de l'éducation distribués en open data. En effet, dans la plupart des cas, les jeux de données diffusés concernent :

- a Liste des établissements (quelques fois avec géolocalisation, ou par statut : public ou privé ou conventionné)
- b Dénombrement des élèves par groupe d'âge, sexe...
- c Résultats scolaires par établissements ou régions...
- d Effectifs enseignants répartition par sexe...
- e Personnel non-enseignant
- f Les infrastructures et équipements scolaires
- g Dépenses en financements bourses d'études...
- h

2.3. Les enjeux de l'Open Data de l'éducation

Pourquoi ouvrir les données de l'éducation ? Cette question renvoie également aux ambitions nourries et aux attentes formulées ? Le préambule de la Charte du G 8 est un éclairant

condensé d'optimisme quant aux opportunités offertes par le mouvement. Les acteurs engagés dans l'ouverture des données publiques s'attendent unanimement à ce qu'elle contribue à la transparence et la dynamisation de l'action publique. L'administration constamment exposée au regard examinateur des administrés deviendrait ainsi plus efficace. L'Open data renforcerait l'engagement civique et la participation citoyenne, faciliterait l'autonomisation (*empowerment*) des citoyens, encouragerait et favoriserait l'innovation comme facteur de croissance économique. On peut se demander s'il est possible de retrouver intégralement ces bénéfices dans l'ensemble des domaines de l'action publique. Est-ce que, par exemple, l'innovation pour stimuler la croissance économique sera-t-elle au rendez-vous de l'ouverture des données de l'éducation ? Autrement dit, existeraient-ils des enjeux spécifiques à l'open data dans les systèmes éducatifs ? Est-ce que l'ouverture des données de l'éducation comporte des enjeux différents des autres secteurs ? Sans perdre ces interrogations de vue, il faut garder à l'esprit que l'ouverture des données participe essentiellement d'une démarche volontariste. Il peut être question également de répondre à une obligation réglementaire, légale, ou tout simplement pour s'aligner sur ce qui se fait ailleurs, dans d'autres pays, secteurs, d'autres entreprises. Il est difficile de tirer aujourd'hui un bilan définitif de l'Open data puisqu'il s'agit d'un mouvement assez récent avec ses marées hautes, ses marées basses et voire ses ressacs.

La consultation des portails de diffusion des données ne permet pas d'isoler des motifs spécifiques de publication des données. Peut-on en conclure pour autant qu'il n'y aurait pas de raisons proprement sectorielles, d'enjeux spécifiques à chacun des secteurs ? Les enjeux peuvent-ils avoir des traductions concrètes différentes d'un secteur à l'autre ? Pour le secteur de l'éducation, l'objectif pourrait être le renforcement du contrôle citoyen sur la manière dont les autorités publiques gouvernent le système éducatif. Il peut également s'agir d'impulser une politique de la transparence conçue comme un outil efficace dans la lutte contre la corruption. Dans le cas du Kirghizistan, le gouvernement mise sur l'Open data pour réduire la corruption dans le secteur éducatif. L'ouverture des données est une manière d'adresser la question puisque dans les classements internationaux sur la corruption le pays occupe une place peu flatteuse. Dans d'autres contextes, c'est la recherche de l'efficacité des dépenses publiques grâce, entre autres, à l'identification des besoins par secteur géographique ou par établissement entre autres. Les données publiées peuvent concerner également la performance des établissements, le taux de passage, de réussite, ou l'état d'avancement de certains projets d'infrastructures. En portant une attention soutenue aux jeux de données exposés, il est également possible de se faire une idée des objectifs fixés par les producteurs. En effet, les jeux de données libérés permettent dans une certaine mesure d'orienter les classes de réutilisation pour faciliter l'appropriation des données sectorielles.

Cependant, les données de l'éducation ne paraissent pas être très demandées ou susciter suffisamment d'attention de la part des développeurs. En effet elles ne sont pas légion, les organisations, qui les réutiliseraient, comparativement à d'autres secteurs : environ soixantaine selon l'*Open data impact map* (Figure 5).

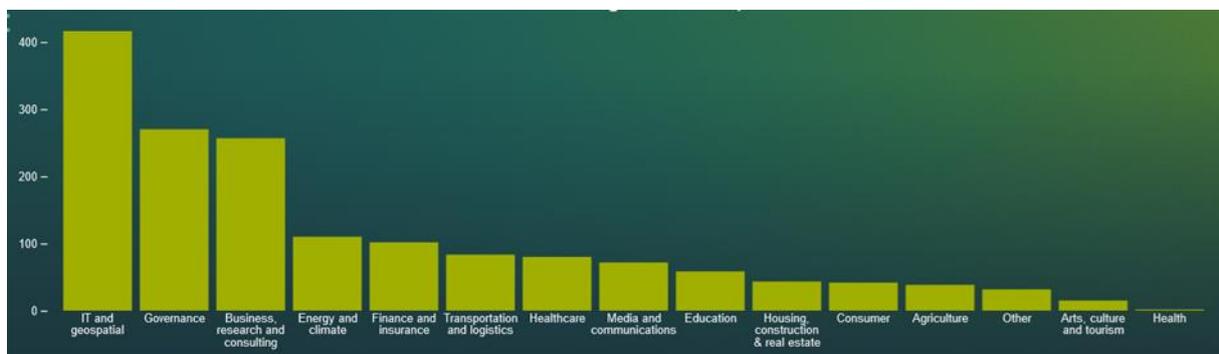


Figure 5. Organisation utilisant données ouvertes par secteur. Open data impact map

Cette base de données recense les organisations (entreprises, associations à but non-lucratif, universités, et groupes de développeurs) qui réutilisent les données administratives ouvertes à travers le monde. Soulignons que lesdites organisations ne réutilisent pas que des données de l'éducation, elles peuvent également les croiser avec les données issues d'autres secteurs pour développer certains produits [12, Center for Open Data Enterprise].

2.4. Les données ouvertes de l'éducation dans le baromètre Open Data

Que représentent les données ouvertes de l'éducation dans l'univers de l'Open data ? Quelle est sa part volumétrique ? Dans la quatrième édition du Baromètre de l'Open Data de la *World Wide Web Foundation* (qui couvre la période allant de juillet 2015 à juin 2016, et qui porte 115 pays), l'on peut constater une régression de la disponibilité des données ouvertes du secteur social auquel fait partie l'éducation. Sur la période, les données de l'éducation représentent 8 % des bases de données publiques ouvertes. De plus, les données de seulement neuf États (Argentine, Brésil, Danemark, États-Unis, France, Géorgie, Jamaïque, Malaisie, Royaume-Uni) peuvent être considérées comme étant réellement ouvertes [14, World Wide Web Foundation]. Et, en ce qui concerne les trente pays les mieux classés sur la période juillet 2016 – septembre 2017, il n'y a que 13 % des bases de données de l'éducation publiées qui peuvent être réellement considérées comme étant ouvertes. Néanmoins, l'on peut constater une amélioration de la qualité des données ouvertes de l'éducation par rapport aux années précédentes (*Figure 6*).



Figure 6. Évolution de la qualité des données ouvertes

Par ailleurs, parmi les trente pays leaders du baromètre, seulement les données réellement ouvertes de l'éducation Brésil, Canada, France, et du Royaume-Uni ont obtenu un score supérieur ou égal à 70/100 selon les critères d'analyse de la *Web Foundation*. Il est important de signaler ce fait notable qu'il s'agit surtout de pays membres du G 20, dont 3 du G7. De même dans si l'on se réfère au 4^e Baromètre (2015 – 2016), sur les neuf États avec les données de l'éducation réellement ouvertes, six sont membres du G 20. Il en ressort une certaine distribution géopolitique de la qualité des données ouvertes de l'éducation qui reflète globalement la distribution de la puissance dans la géopolitique mondiale.

2.5. Les classes de réutilisation

L'un des principes cardinaux de l'open data consiste en la diffusion des données « brutes », lesquelles peuvent dans certains cas s'avérer difficilement compréhensibles pour le grand public. D'où l'intérêt de bien prendre en compte la réutilisation lorsque l'on publie des données ouvertes. Même si, en termes de degré de complexité, les données de l'éducation ne sont pas en général les plus difficilement intelligibles. Dans l'état actuel du développement de l'Open data, la notion de réutilisation fait l'objet de discussions. En général, les plateformes

renseignent sur les réutilisations signalées qui sont surtout l'œuvre de professionnels. Lorsqu'elles permettent de proposer de nouveaux produits, ou de nouvelles applications. Or si l'on prend le cas de la France par exemple, selon l'article 10 de l'Ordonnance de juin 2005, relative à la réutilisation des données publiques, est considérée comme réutilisation, l'utilisation par « *toute personne qui le souhaite à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été élaborés ou sont détenus* » [15, JORF]. Des spécialistes de l'Open data, distinguent en général quatre classes de réutilisation [4, Chignard]. :

- *Consultation* : les données sont directement consultées par le public.
- *Médiation* : des visualisations des données sont offertes au public.
- *Application* : Les développeurs utilisent les données pour offrir des services et des applis au public.
- *Réutilisation spécialisée* : Les données techniques complexes utilisées comme intrants dans le cadre d'autres produits.

Il faut signaler que la médiation, comme réutilisation des données, ne consiste pas forcément en l'intervention d'un tiers. En effet certaines plateformes proposent des outils de visualisation ou de prévisualisation afin de rendre les données plus facilement accessibles au public non professionnel. En ce qui concerne les données publiques ouvertes de l'éducation, elles font plutôt l'objet des trois premières classes de réutilisation précitées. Par ailleurs, selon l'*Open data impact map*, (Figure 7, *infra*), dans le secteur de l'éducation, les réutilisations des données ouvertes sont ainsi réparties : i recherche ; ii optimisation ; iii développement de produits et services ; iv) mobilisation/plaidoyer [12, Center for Open Data Enterprise]. Nous devons préciser qu'il ne s'agit pas de réutilisation de données de l'éducation, mais de réutilisation de données ouvertes (indépendamment du secteur) dans le secteur de l'éducation et de la formation.



Figure 7. Réutilisation des données ouvertes dans le secteur de l'éducation (2018)

En ce concerne les réutilisations des données de l'éducation, elles consistent le plus couramment en la localisation des établissements, l'évaluation de leur performance (résultats scolaires), la disponibilité des formations. Il s'agit aussi de réutilisation sous forme d'applications ou autres, qui permettent aussi de disposer d'informations sur les dépenses par établissements, sur la qualité de l'encadrement, et le coût des études. Les objectifs de ces réutilisations peuvent varier d'un acteur à un autre, même si la logique d'accompagnement des parents à choisir les meilleurs établissements semble prédominer. En voici quelques exemples de réutilisation :

- [London Schools Atlas](#), est une application de la municipalité de Londres. Elle vise à orienter sur le « marché de l'offre scolaire ». Elle offre à la possibilité d'accéder à un

large éventail d'informations sur les écoles publiques de Londres : degrés, élèves, types d'écoles, écoles éligibles aux repas scolaires gratuits, enfants des foyers à faibles revenus.

- [Stop Schools Cut](#), est une initiative syndicale au Royaume-Uni, (*National Education Union*). Elle propose des visualisations qui permettent de voir les conséquences des coupes budgétaires dans l'éducation en Angleterre. L'objectif visé est la sensibilisation et la mobilisation du public pour exiger un meilleur financement en faveur de l'école publique.
- [Mejora tu escuela](#), portail mexicain, combine les deux aspects. D'une part, il vise à aider les parents à faire les meilleurs choix d'établissements. D'autre part, il a une dimension de mobilisation contre la corruption et une meilleure répartition des allocations aux établissements.
- [Q Edu](#), au Brésil, s'inscrit dans une démarche plus large d'accompagnement de tous les acteurs du système éducatif brésilien (gestionnaires, directeurs, enseignants, parents) en vue de la transformation de l'éducation grâce aux données.
- L'application web [FFT Education datalab](#), au Royaume-Uni, s'inscrit dans une démarche similaire.

En somme, les données ouvertes de l'éducation permettent des réutilisations variées dépendamment des ambitions et engagements des utilisateurs. Et, comme cela a été précédemment mentionné, les classes de réutilisation sont aussi conditionnées par les jeux de données publiés, leur qualité, la fréquence de mise à jour entre autres. D'où la nécessité de s'intéresser également aux portails/plateformes.

3. De quelques plateformes et de la qualité des données de l'éducation

3.1. Le corpus : les plateformes du Brésil, de la France, de l'Inde et du Kenya

L'Open data est inconcevable sans le web et l'internet. La distribution des données en ligne qui les rend accessible à tout utilisateur est un critère important de l'Open data. Ainsi, la question de la diffusion en est centrale dans le cadre d'une politique d'ouverture des données. D'où l'intérêt de prêter attention aux plateformes ou portails, à la qualité des données publiées afin, entre autres, d'étudier l'exposition des jeux de données de l'éducation. Dans le cadre de travail, nous allons nous aborder le cas de quatre (4) portails gouvernementaux : France, Brésil, Inde, Kenya. Le choix de ce corpus repose sur l'articulation de deux critères qui nous paraissent importants :

- a) L'engagement de ces États dans la dynamique de l'ouverture des données, en témoigne leur classement dans le baromètre de la Word Wide Web Foundation.
- b) La dimension géopolitique et économique, c'est-à-dire retenir des États issus de différents continents, et de niveau de développement économique inégal.

D'abord, la France, qui occupe la 4^e place dans le classement, pays européen et qui fait partie du G 7. Ensuite le Brésil, à la 14^e place, pays du continent [sud-] américain, membre du G 20 et à revenu moyen supérieur. Après, l'Inde, État de l'Asie (du Sud) qui se situe à la 16^e place

du baromètre, et également membre du G 20, mais à revenu moyen inférieur ; Enfin Kenya, également pays à revenu moyen inférieur⁸. Même s'il cet État de l'Afrique (subsaharienne) ne figure pas dans le baromètre des trente pays leaders. Il a été retenu à la place de la République d'Afrique du Sud (22^e) et de la Sierra Leone (30^e). Les plateformes gouvernementales de ces deux États sont malheureusement inaccessibles⁹. Par ailleurs, le Kenya est le pays africain le mieux classé, à la 35^e place (sur 114) dans la quatrième édition du baromètre de la Web Foundation [14, Word Wide Web Foundation].

Nous avons opté pour des portails gouvernementaux, en mettant de côté les données publiques mises en ligne par des tiers diffuseurs. Même s'il peut s'agir de personnes morales de droit public comme les organisations ou agences internationales¹⁰. Les États retenus disposent d'une plateforme nationale de diffusion des données ouvertes, indépendamment de leur forme politique, unitaire ou fédérale :

- France : *Data.gouv.fr* : <https://www.data.gouv.fr/fr/>

Cette plateforme chargée de diffuser des données publiques de l'État et des collectivités notamment est animée par la mission Etalab de la Direction interministérielle du numérique (DINUM). Data. gouv, agrégateur de données publiques, opère sous [Udata](#), un logiciel open source maison¹.

- Brésil : *Portal Brasileiro de Dados Abertos* : <https://dados.gov.br/>

La plateforme est une composante de l'Infrastructure nationale des données ouvertes (INDA), qui est dirigée par un comité formé majoritairement d'entités publiques (présidence et ministères), et de représentants de la société civile, et des universités. Le portail est agrégateur des données publiques brésiliennes. La plateforme opère avec un logiciel open source, [CKAN](#) – Comprehensive Knowledge Archive Network. Ce logiciel, l'un des plus utilisé pour la diffusion des données ouvertes, est développé par l'OFKN.

- Inde : *Open Government Data Platform India* : <https://data.gov.in/>

Cette plateforme est établie dans le cadre de la politique de partage des données (acronyme anglais NDSAP), et diffuse des données de toutes les entités publiques du gouvernement fédéral de l'Inde. Elle est animée par le Centre informatique national (NIC). La plateforme qui fonctionne sous Drupal, repose sur une solution open source maison, fruit d'une coopération indo-américaine, Open Government Platform (OGPL)¹¹.

- Kenya : *Kenya Open Data* : <https://www.opendata.go.ke/>

Son portail est chargé de la diffusion des données publiques des entités publiques (gouvernement et collectivités). L'*ICT Authority*, une agence gouvernementale dépendant du ministère des TIC en assure la gestion. L'Initiative des Données Ouvertes du Kenya (KODI) a opté pour la solution en SaaS ArcGIS d'[Esri](#).

⁸ Sachant que même si le Brésil et l'Inde sont des pays du groupe G 20, le premier est classé parmi les pays à revenu moyen supérieur, et le second à revenu moyen inférieur comme le Kenya.

Voir, World Bank Country and Lending Groups,

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>.

Le poids de l'Inde dans la géopolitique mondiale compense le revenu par habitant dans notre approche.

⁹ Afrique du Sud, <http://data.gov.za/>. Adresse introuvable. Internet archive permet de consulter des versions archivées, <https://web.archive.org/web/20190429181106/http://data.gov.za/> [consulté le 11/12/2020]

Sierra Leone, <https://opendatasl.gov.sl/>. Erreur du serveur : 500. Versions archivées des pages, <https://web.archive.org/web/20201026014126/https://opendatasl.gov.sl/> [consulté le 11/12/2020].

¹⁰ A l'instar du portail open data du projet *Africa Information Highway* de la Banque africaine de développement (BAD), <https://dataportal.opendataforafrica.org/>

¹¹ About Open Government Data (OGD) Platform India, <https://data.gov.in/about-us>. Open Government Platform, <https://ogpl.github.io/index-en.html>

Pouvons-nous dire que ces portails possèdent ce que ce que l'on considère en général comme les « *caractéristiques désirables d'un portail Open Data* » [16, Gouigoux, pp. 35-36] ? Entendons la mise à jour régulière des données, le téléchargement en plusieurs formats réutilisables, la fourniture des données par le biais d'une API, des fonctions recherche, l'accessibilité, et la robustesse. Notre audit portera sur les aspects fonctionnels des portails, et la qualité des données ouvertes de l'éducation publiées.

3.2. Fonctionnalités des portails

L'analyse de ces quatre portails n'a pas pour objectif d'établir un classement. Elle consiste essentiellement à identifier leurs fonctionnalités principales. Bien évidemment, nous y trouverons des éléments communs et des particularités.

- Page d'accueil

La présentation des pages d'accueil de ces portails est assez classique, surtout au niveau de la *ligne de flottaison* (above-the-fold). Cependant son exploitation reste inégale d'un portail à l'autre. Les uns ont mieux réussi que les autres à afficher les informations importantes, à bien les mettre en évidence, et à organiser un haut de page d'accueil très intuitif. Le nombre de *rubriques du menu principal* est plutôt variable, et peut, dans certains cas, aller au-delà de sept onglets. Les rubriques « données » (jeux de données) et « secteurs/ organisations » (découpage administratif ou thématique) sont clairement identifiées dans les barres de navigation de ces différents portails. Les utilisateurs trouveront également une rubrique de présentation des portails qui renseigne sur leur mission et les sens de la démarche d'ouverture des données. En ce qui concerne le « *scrolling* », en trois défilements, les utilisateurs, peuvent prendre connaissance de l'organisation du contenu du portail, de l'animation du site (mise en vedette des meilleures réutilisations, des données les plus récentes, de l'actualité de l'open data, entre autres) et des services proposés (téléchargement, création de compte).

Étant donné qu'il s'agit de portails gouvernementaux, les couleurs nationales de ces États forment la base de leur *identité visuelle* en ce qui a trait aux codes couleurs. Si la sobriété de la page d'accueil est de rigueur dans le cas des portails de la France et du Brésil, ce n'est pas le cas pour le portail Kenya et surtout pour celui de l'Inde. La page d'accueil du portail de cette dernière peut paraître plutôt surchargée. Des considérations ergonomiques pourront du moins en être la raison. En effet, l'on peut constater un effort certain pour le rendre plus accessibles aux utilisateurs lambda. Le portail indien, qui est *responsive design* et de bonne facture, paraît avoir tranché en faveur de l'approche *user friendly* dans le débat sur l'exposition des données en contexte open data. Le choix indien suit une ligne contraire par rapport à celle de ceux qui estiment que la simple fourniture de données de bonne qualité répond largement aux ambitions du mouvement. *Data.gov.in* permet des visualisations en proposant une interface d'affichage distincte de la fourniture de la donnée.

- Système de navigation

Au niveau du système de navigation, les portails proposent une combinaison de deux modèles : la *navigation dirigée* et la *navigation par recherche*. Le *moteur de recherche interne* n'est pas disponible dans tous les portails, en général un formulaire de recherche à champ unique est proposé. Autrement, les portails n'offrent pas la possibilité de passer en mode recherche avancée. Toutefois, grâce à l'organisation thématique et sectorielle, l'utilisateur peut facilement accéder au secteur pour lequel il souhaite obtenir des données. Dans ce cas, au maximum, il suffit de faire deux clics (France, Kenya) ou trois clics (Brésil, Inde) pour arriver à la page recherchée, consulter les jeux de données, et leurs éventuelles réutilisations. De plus, les utilisateurs ont leur disposition des *facettes latérales* placées à gauche ou à droite dépendamment du portail. Elles permettent, entre autres, d'affiner la recherche des données, et ajuster le niveau de granularité. Par ailleurs, si l'on retrouve le *menu hamburger* dans le portail indien, et le *fil d'Ariane* dans le portail brésilien, les deux autres portails ne proposent

pas de *système de navigation dit secondaire*. Néanmoins les utilisateurs ont la possibilité d'affiner les résultats de la recherche à l'aide de facettes (mots-clés, producteurs, formats...). De plus il est possible de *retourner à la page d'accueil* en cliquant sur le logo quand la petite *icône d'accueil* n'y apparaît pas.

- Structuration hiérarchisée

Les quatre plateformes ont adopté la structuration hiérarchisée qui répond bien aux exigences des sites à fort contenu. Cependant, en ce qui a trait à la recherche de jeux de données, pour l'ensemble des portails, les rubriques du menu ne permettent pas toujours la navigation la plus intuitive. Dans le cas de *data.gouv.fr*, les étiquettes à gauche de la page d'accueil permettent d'accéder plus facilement aux jeux de données thématiques. Par exemple en cliquant sur l'étiquette « *Éducation, Recherche, Formation* », le menu apparaît avec les différents thèmes associés à ce champ thématique (*Figure 8*).



Figure 8. Capture d'écran *data.gouv.fr*.

En ce qui concerne de la plateforme indienne, il faut cliquer sur le menu hamburger en haut à droite ou sur le bouton « skip to navigation », pour voir s'afficher le menu de navigation. Ensuite, en cliquant sur la rubrique « *Sector* » apparaissent les sous-rubriques désignant les « catégories » (sous forme de tuiles) avec les éléments visuels (icônes) associés (*Figure 9*).

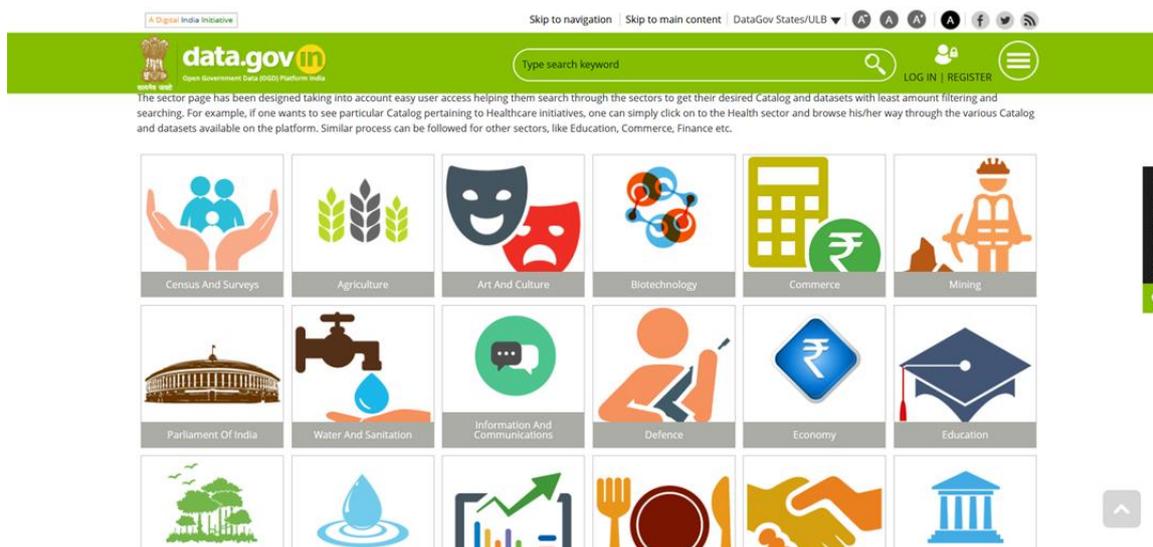


Figure 9. Capture d'écran *data.gov.in. Sector*

Il est plus facile de rechercher les jeux de données à partir de la barre de menus dans le portail du Brésil. Quand on clique sur la rubrique des données (*Dados*), s'ouvre une page de vignettes de jeux de données, et des facettes, à droite de l'écran, aident à préciser sa recherche. Le

portail du Kenya se situe entre ces deux modèles. Les tuiles des catégories thématiques avec leurs éléments visuels apparaissent dès la page d'accueil (Figure 10).

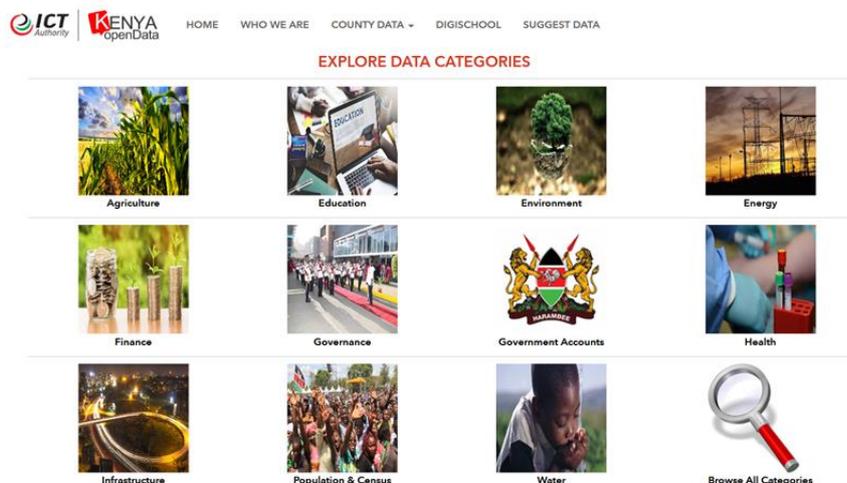


Figure 10. Capture d'écran opendata.go.ke

En somme, tous les portails permettent à l'utilisateur une navigation cohérente grâce à une hiérarchisation logique des contenus. Il est possible d'accéder aux données ou aux visualisations en deux ou trois clics comme préalablement mentionné.

- Services proposés

Les services offerts par les plateformes gouvernementales étudiées tentent de répondre à des préoccupations centrales de l'Open data, c'est-à-dire exposer, partager des données publiques, mais également promouvoir la participation citoyenne. Les portails offrent des services qui indiquent qu'ils visent à la fois des utilisateurs amateurs et professionnels. Avec ses nombreuses visualisations et les choix de ses éléments visuels, l'interface du portail de l'Inde dénote un réel effort en vue d'atteindre un public qui n'a pas de compétences techniques, informatiques confirmées. En outre, à l'instar du portail du Brésil, il répond également aux normes d'accessibilité afin de prendre en compte les utilisateurs souffrant de certains déficits sensoriels.

Tous les portails offrent des possibilités de *téléchargement* des données sous différents formats, et des *outils de visualisation* d'inégale valeur. Dépendamment de ses besoins, l'utilisateur pourra choisir de consulter les données via un tableur ou de se contenter de la visualisation. En outre, les portails permettent la récupération et le partage de données via API. Assurant des fonctions d'agrégation, les plateformes de France et du Brésil, dans une moindre mesure, exposent des données récupérées par *moissonnage*. Cependant, seule *data.gouv* utilise le format DCAT (Data Catalog Vocabulary) pour le moissonnage.

Les utilisateurs peuvent *créer des comptes* personnels sur *data.gouv.fr* et *data.gov.in* afin de publier des données, dans le cas du premier, ou de proposer des données au data manager de la catégorie concernée, dans le cas du second. Dans le cas du Kenya, il n'est pas nécessaire de créer un compte pour suggérer des données, il suffit d'ajouter son adresse électronique dans un formulaire pour pouvoir charger ses données. Sur *data.gouv.fr*, grâce au compte utilisateur, il est possible d'ouvrir une discussion sur un jeu de données ou une réutilisation. Des *communautés* Open data sont créées (groupe de discussion Google, Facebook, Telegram) autour de la plateforme brésilienne. Il est également possible de suivre l'actualité des portails via les *réseaux sociaux* (Facebook, Twitter, GitHub), et les *flux* RSS. Si, dans l'ensemble ces portails permettent d'accéder à des données publiques en open data reflètent une volonté d'informer les utilisateurs, comment y sont exposées les données de données de l'éducation ? Qu'en est-il de leur qualité ? En effet l'on peut se demander quelle

serait l'utilité de données ne répondant pas aux critères de qualité ? C. Maday juge cette question essentielle pour *respecter le contrat moral de l'Open data* à savoir la diffusion de données réutilisables [17, MADAY].

3.3. La qualité des données ouvertes de l'éducation

En contexte Open data, la réutilisabilité des données exposées est un principe fondamental. Elle est assurée par la réunion de trois critères. Le choix du format d'exposition de la donnée sur internet qui relève d'une dimension *technique*. La licence qui autorise la réutilisation qui découle de la dimension *juridique*. La fourniture des données reposant sur le principe de gratuité, ou d'absence de buts lucratifs, qui résulte de la dimension *économique*. Malgré sa centralité, la qualité des données ouvertes est-elle réductible à leur réutilisabilité ? Dépend-elle essentiellement du statut de son producteur (gouvernement, agence gouvernementale, collectivité infra-étatique) ? À quoi renvoie concrètement la qualité des données publiques en contexte open data ? Il est impossible d'apporter une réponse définitive à ces considérations dans le cadre de ce travail. Cependant, il existe par exemple la norme ISO/IEC 25012 : 2008 qui établit trois classes de critères qui sont :

- ❖ Inhérents aux données (*exactitude, complétude, cohérence, crédibilité, ponctualité*) ;
- ❖ Inhérents simultanément aux données et au système (*accessibilité, conformité, confidentialité, efficience, précision, traçabilité, intelligibilité*) ;
- ❖ Dépendant du système (*disponibilité, mobilité, et récupérabilité*).

Ces critères portent essentiellement sur le contexte de production des données, et concernent peu leur exposition en contexte open data¹². Des données peuvent remplir tous ces critères sans être une donnée ouverte. Dès ses balbutiements, la question de la qualification de la donnée ouverte a été débattue au sein de la famille de l'Open data. Les participants à la rencontre de Sebastopol de 2007 ont fixé les huit critères de qualification. Les données ouvertes doivent être, *complètes, brutes, à jour, accessibles, lisibles par une machine, non-discriminantes, fournies dans un format non-propriétaire, diffusées en licence libre*. Quant à ces critères, ils concernent essentiellement sur la diffusion. Malgré des sensibilités différentes, ils ont fait l'unanimité au sein du mouvement, et l'esprit a été sauvegardé dans la Charte internationale de l'Open data (2015). Dans un guide adressé aux diffuseurs de données ouvertes, l'*Initiativa Aporta* (Espagne) a tenté une synthèse de ces deux approches (production et diffusion) en proposant des indicateurs de qualité des données ouvertes [18, Data.gob.es].

Le simple utilisateur ne dispose pas des compétences requises et souvent ni du temps nécessaire pour procéder à une évaluation de la qualité des données en fonction de ces indicateurs. Il serait plutôt tenté de se fier à la crédibilité du fournisseur de données. Pour certains, la question de la qualité des données en contexte Open data est bien un sujet compliqué, « *un point difficile à aborder car que veut dire la « qualité » des données ? En principe, on s'attend à des données complètes, sans erreurs, bien documentées et avec les métadonnées les plus exhaustives possible d'un côté et s'appuyant sur certains des référentiels existants de l'autre. Malheureusement, cela n'est pas toujours le cas. Un travail en amont semble être alors nécessaire* » [19, Meszaros et al, p. 64]. Ce nécessaire travail en amont, qui est très bien pris en considération dans la synthèse proposée par l'initiative espagnole, est hors de portée dans le cadre de ce travail. Nous nous contenterons des critères

¹² La société Opquast a élaboré une liste de 72 bonnes pratiques, *Open Quality Standards*, à destination des producteurs de données ouvertes, <https://checklists.opquast.com/fr/opendata/> [consulté le 30/11/2020]. Il s'agit bien de « bonnes pratiques » et non de standards contraignants.

d'évaluation de la Banque mondiale s'inspirant de ceux de l'*Open data project* du gouvernement américain. Ainsi, selon l'institution de Bretton Woods, pour que la donnée ouverte soit de qualité, elle doit être :

- a) Publique
- b) Accessible
- c) Explicitée
- d) Réutilisable
- e) Complète
- f) Actualisée
- g) Gérée après publication [20, Guide pratique de l'Open data].

Ces critères d'appréciation siéent mieux au cadre de cet exercice qui ne porte pas sur les qualités intrinsèques des données statistiques ou autres. Ils peuvent être combinés avec le *Cadre commun d'évaluation de l'Open data* (en développement) de la World Wide Web Foundation et la GovLab de l'Université de New York [21, Caplan et al.]. Nous nous en inspirons pour évaluer la qualité des données de l'éducation diffusées par les plateformes de notre corpus. Ainsi, nous allons aborder les points relatifs au contexte d'ouverture ; aux formats de publication, aux métadonnées, aux licences, à l'actualisation des données et de leur pertinence, et au suivi avec le public. Par ailleurs, compte tenu du corpus, les critères d'exigence ne vont pas au-delà de la troisième ou de la quatrième étoile du modèle de Tim Berners-Lee (voir *Tableau 1, p. 12*). En résumé, nous avons exclu la question des données ouvertes liées.

3.3.1. Contexte d'ouverture : le cadre légal

Le contexte d'ouverture des données concerne fondamentalement le cadre légal qui autorise ou facilite la mise en place de politique d'Open data. Il est toujours possible d'entrevoir les premières briques qui ont permis la construction de la maison Open data dans le dernier quart du vingtième siècle. Cependant, force est de constater que les véritables cadres légaux qui ont favorisé l'ouverture des données publiques remontent à une quinzaine d'années. Les cas de notre corpus l'illustrent assez bien. En France, les publications du décret, relatif à la réutilisation des informations publiques, et de la circulaire ministérielle, portant création de mission Etalab, (26 mai 2011) ont concrètement posé les bases de l'Open data. C'est au cours de cette même décennie que l'Inde a adopté la *National Data Sharing and Accessibility Policy* (2012) dans le sillage de la *Right to Information Act* (2005). En ce qui a trait au Brésil, la *Lei de Acesso a Informação Pública* (2011) et surtout le décret n° 8.777 (2011) ont permis d'instituer le cadre légal et réglementaire de l'Open data. Dans le cas du Kenya, l'initiative d'Open data a été lancée, dès 2011. Elle s'est appuyée sur deux dispositifs légaux successifs et complémentaires. D'abord, l'article 35 de la Constitution de 2010, octroyant aux citoyens le droit d'accès à l'information publique, tout en ouvrant la possibilité à la publication de toute information qui concerne le pays. Ensuite, sur la *Kenya Access to Information Act*, (2016) qui garantissant le droit d'accès à l'information et aux données produites par les entités publiques¹³. Toutefois, il n'y a pas encore de cadre légal spécifique portant sur l'Open data. Cela a sans doute un impact sur l'engagement institutionnel à l'ouverture. En effet, même si l'on ne saurait tout ramener à l'absence d'un cadre légal, dédié, la quantité et la qualité des données diffusées peuvent être un indicateur de la faiblesse de l'engagement institutionnel¹⁴.

¹³ Le portail kenyan <https://www.opendata.go.ke/>, ne fournit aucune information concernant le cadre légal de l'ouverture des données publiques.

¹⁴ Dans la catégorie « Education », on dénombre 772 jeux de données sur le portail indien, 485 sur le portail français, 52 pour le Brésil, et 80 fichiers sur le portail kényan. Toutefois, nous n'ignorons pas la taille de la population de ces pays. Nous pouvons considérer également la variable fréquence de publication et actualisation des données.

3.3.2.Licences

Même si la question des licences ne dépend pas exclusivement d'une législation spécifique sur l'Open data, l'absence de cadre légal peut également avoir des conséquences sur la réutilisabilité des données exposées. Les données de l'éducation de *data.gouv.fr* sont, à l'instar des données des autres catégories, soumises à une licence maison, « *Licence ouverte/Open License* », conçue par Etalab¹⁵. Elle confère le droit de réutilisation sous réserve de mention de la paternité. Elle est compatible avec les licences qui obligent de mentionner au moins le nom du producteur. Comme l'*Open Government Licence* (OGL) du Royaume-Uni, la *Creative Commons Attribution* (CC-BY), et l'*Open Data Commons Attribution* (ODC-BY). L'Inde a également fait le choix d'une licence maison, compatible avec les autres licences susmentionnées, la *Government Open Data License*¹⁶. Tandis que le Brésil a fait le choix de la licence « [Open Database License](#) » (ODbL) pour la majeure partie de ses jeux de données (48) et de « [Creative Commons Attribution](#) » (CC BY-SA 3.0) pour les autres. De fait, la plateforme impose le partage dans les mêmes conditions, avec la même licence de l'œuvre originale (*copyleft*). Certaines données de l'éducation du Kenya sont diffusées sous licence « [Creative Commons Zero](#) » (CC0 1.0), néanmoins une quantité non négligeable l'est sans spécification de la licence.

3.3.3.Formats

Le format utilisé représente l'un des principaux indicateurs de la qualité des données ouvertes. Ce critère de qualité concerne tout particulièrement la problématique de l'accessibilité et de la réutilisabilité des données ouvertes. Le format doit assurer la lisibilité machine (*machine-readable*) afin de faciliter un traitement automatisé des données publiées. Les formats peuvent être appréciés selon qu'ils soient ouverts ou fermés. En consultant les données de l'éducation exposées par les différentes plateformes, on peut rapidement constater la prédominance assez marquée des formats ouverts ou des standards de fait. C'est-à-dire de formats qui garantissent l'interopérabilité, et la possibilité de modification indépendamment du logiciel de création du fichier. Étant donné la forte proportion de données de nature statistique, les formats textuels sont très utilisés pour l'exposition des données tabulaires (avec *xls* ou *csv*). Pour les données hiérarchiques c'est surtout *json* ou *xml*. Il y a également une variété de formats de fichiers : traitement de texte (*docx*, *odt*, *doc*) ; coordonnées géographiques (*GeoJSON*, *kml*, *shp*¹⁷) entre autres. Certaines plateformes comme celles du Brésil et de la France (*Figure 11, page suivante*) affichent les statistiques d'utilisation des formats¹⁸. Le portail indien se singularise d'une certaine manière en proposant par défaut le téléchargement tous ses fichiers en format *csv*. Cependant, avec la possibilité d'exportation de certains jeux de données en différents formats de fichiers (*xls*, *ods*, *json*, *ods*, *jsonp*, *xml*). Les données publiées prêtent en principe à une forte interopérabilité grâce aux différentes options de téléchargement proposées, notamment via API. Ce mode de téléchargement est proposé par les autres plateformes, par exemple le portail kenyan offre la possibilité de charger des fichiers GeoJSON via API¹⁹. Cependant, en ce qui concerne la plateforme brésilienne, les données de l'éducation sous forme de fichiers en formats *xml* et *json* ne sont pas accessibles puisque l'entrepôt de

¹⁵ Licence ouverte/Open License, <https://www.etalab.gouv.fr/licence-ouverte-open-license> . Une seconde version a été publiée en avril 2017.

¹⁶ Government Open Data License - India <https://data.gov.in/government-open-data-license-india> .

¹⁷ *Shapefile* d'Esri. Ce format de fichier géographique est utilisé dans *data.gouv.fr* (France), et dans *opendata.go.ke* (Kenya).

¹⁸ Brésil, <https://dados.gov.br/organization/ministerio-da-educacao-mec> [consulté le 04/01/2020]
France, <https://www.data.gouv.fr/fr/datasets/?sort=-created&organization=534ff8fa3a7292c64a77f22> [consulté le 04/11/2020].

¹⁹ https://opendata.arcgis.com/datasets/186f97cdce59416caea1c2ee63185aa5_0.geojson [consulté le 05/11/2020].

données, la Plateforme de gestion des indicateurs (PGI), a été désactivé en 2015. En conséquence, la qualité des données de l'éducation en sort considérablement affectée puisque seulement cinq jeux de données, en format csv, sur les 52 affichés sont réellement accessibles. Du côté du portail indien, le téléchargement en xls de quelques fichiers a également échoué. Les plateformes de France et de l'Inde exposent leurs données dans des formats accessibles et lisibles par une machine dans une proportion considérable, ce qui n'est pas le cas du portail kenyan (17 fichiers csv sur 80), et de celui du Brésil.

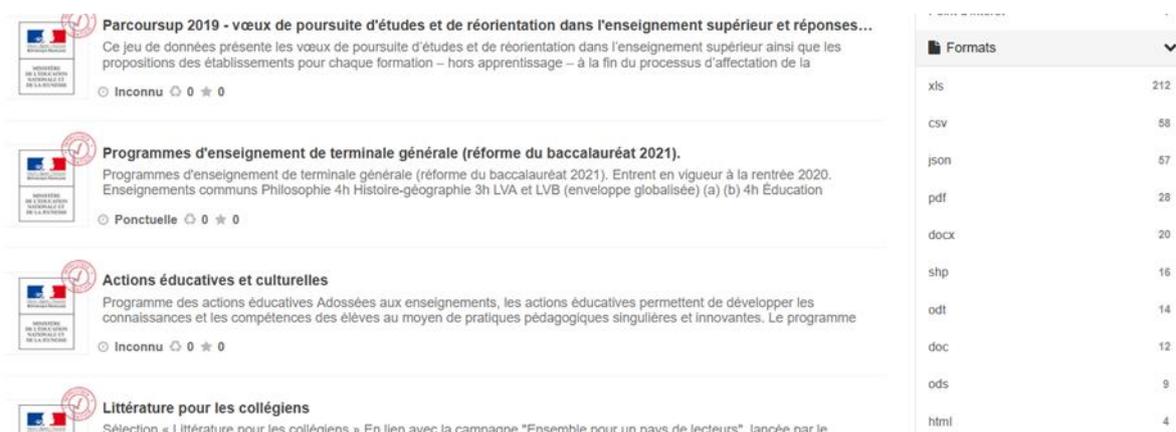


Figure 11. Capture d'écran data.gov.fr

Type de fichiers	Ouvert	Fermé	Standard
Traitement de texte	<i>Docx, odt</i>	<i>doc</i>	
Tableur	<i>csv, ods</i>	-	<i>xls</i>
Coordonnées géographiques	<i>GeoJSON, KML</i>		<i>shp</i>
Langage de description des pages	<i>xml, html, pdf</i>	-	-
Présentation	-	<i>ppt</i>	-

Tableau 2. Formats de fichiers utilisés par les plateformes du corpus

3.3.4. Des métadonnées

À l'exception de celle du Kenya, toutes les autres plateformes précisent leur schéma de métadonnées. Le standard Dublin Core, pour spécifier le contenu, la propriété intellectuelle, et l'instanciation, en sont une des sources d'inspiration. Même en l'absence de publication de son standard, quand les données kenyanes sont bien décrites, leur description s'inspire également de ce standard. Cependant aucune de ces plateformes ne semble se conformer à la norme ISO 8601 ou des recommandations *Date and Time Formats* de la W3C²⁰ portant sur

²⁰ Date and Time Formats, <https://www.w3.org/TR/NOTE-datetime>

l'inscription des dates. L'inspiration de la *Dublin Core metadata* est bien illustrée avec cette description de ce fichier (Figure 12) de données sur les statistiques scolaires en Inde.

Figure 12. Capture d'écran data.gov.in

On peut y distinguer les catégories des différentes propriétés : contenu (Title, Description, Keywords...); propriété intellectuelle National Data Sharing and Accessibility Policy [éditeur] Ministry of Human Resource. Development Department of School Education and Literacy [créateur]; instanciation (Published on data portal, format, file size, Reference url [source]). Toutes les propriétés ne sont pas formellement précisées dont la licence, la langue, le type de ressource entre autres. Dublin Core, standard générique, reste une source d'inspiration pour les plateformes. Dans certains cas, les plateformes (Inde, Brésil, France) ont défini leur propre modèle de métadonnées qui précisent les éléments de données considérés comme

obligatoires, et d'autres comme additionnels. Par exemple, la plateforme brésilienne considère la couverture spatio-temporelle comme tout simplement souhaitable. Cela n'empêche qu'on peut exceptionnellement y trouver des jeux avec de riches métadonnées témoignant d'un réel effort de description « dense » (figure 13). Globalement, pour l'ensemble des plateformes, les descriptions portent sur les métadonnées essentielles : titre du jeu de données ; sujet ; description ; créateur ; diffuseur ; licence.

Informações Adicionais	
Campo	Valor
Fonte	dadosabertos.mec.gov.br
Autor	Coordenação-Geral de Gestão Escolar (CGGE)
Mantenedor	Coordenação-Geral de Gestão Escolar (CGGE)
Última Atualização	24 janvier 2020, 20:18 (UTC+01:00)
Criado	3 juillet 2017, 14:55 (UTC+02:00)
Assunto	Educação Básica, Escolas Públicas Municipais, Escolas Públicas Estaduais e Ensino em contra turno
Autor/Mantenedor	Coordenação-Geral de Gestão Escolar (CGGE)
Catálogo origem	informacao.mec.gov.br
Cobertura geográfica	Todos os municípios brasileiros que tenham beneficiários do Programa Mais Educação.
Cobertura temporal	2014 a 2019
Descrição	Número de escolas municipais e estaduais que tiveram seus planos de atendimento do Programa Mais Educação aprovados, quantidade de alunos atendidos e repasses da adesão.
Descrição do recurso	Arquivos no formato CSV contendo dados negociais das Escolas com plano de atendimento aprovado no Programa Mais Educação no ano 2014.
Documentação	http://informacao.mec.gov.br/bilibs/PDA/PME/Dicionario_PME.pdf
Etiquetas	Mais Educação, Escolas, Contra turno, Educação Básica
Formato	CSV
Frequência de atualização	Anual
Granularidade geográfica	Municipal
Granularidade temporal	Anual
Identificador	informacao.mec.gov.br/bilibs/PDA/PME/repases_mais_educacao_adesao_2014.csv
Metodologia	Este conjunto de dados não possui origem sistêmica e por este motivo, foram fornecidos pela Coordenação-Geral de Gestão Escolar – CGGE no formato final disponibilizado para consumo no Dados.gov.
Referências	Este conjunto de dados pode ser relacionado com dados do PDDE Interativo e base do CENSO da Educação Básica (INEP)
Título	Programa Mais Educação - PME
Título do recurso	Escolas com plano de atendimento aprovado no Programa Mais Educação.
VCGE	http://vocab.e.gov.br/2011/03/vcge#educacao-infantil
Órgão - Poder	Executivo
Órgão responsável	Secretaria Executiva (SE) do Ministério da Educação (MEC).
Órgão – Esfera	Federal

Figure 13. Capture d'écran dados.gov.br

3.4. Conclusion de la première partie

L'exploration de la problématique des données ouvertes de l'éducation, dans cette première partie, a permis de scruter un mouvement (Open data) en devenir. Il se situe à la jonction des évolutions politico-sociales et technologiques. À ses débuts, il est principalement porté par des acteurs déterminés de la société civile avant l'engagement résolu des États. Dans ce nouveau régime de médiation par l'exposition des données publiques sur internet, les données de l'éducation, au sens défini dans ce travail à savoir les données diffusées par l'administration du secteur, sont bien présentes. La charte de l'open data du G 8 a renforcé la légitimité de la publication des données de l'éducation. Si elles donnent lieu à des réutilisations variées, elles font cependant l'objet d'une faible réutilisation. Les métriques des réutilisations (connues et signalées) mises en ligne par les plateformes du corpus retenu ne contredisent les tendances dégagées dans des évaluations plus globales. Il serait intéressant par ailleurs de rechercher les causes de ce faible intérêt parmi les développeurs d'applications.

Face à ce constat, nous ne saurions avancer aucune hypothèse. Même si la question de l'instauration des publics (au sens du philosophe pragmatiste américain John Dewey) des données ouvertes de l'éducation demeure d'importance. Par ailleurs, si l'on s'accroche aux recommandations des bonnes pratiques de l'Open data, la situation des pays à faibles revenus avec une moindre disponibilité de ressources technologiques interpelle. Un effort de décentrement du regard, des lieux d'élaboration de ces bonnes pratiques, vers d'autres contextes pour penser les conditions concrètes d'exposition des données ouvertes de l'éducation peut aider à approfondir davantage la question de la valeur ajoutée de l'Open data. Le modèle stellaire de Berners-Lee offre un cadre permettant la prise en compte de la réalité de ces pays si l'on veut concevoir une politique d'Open data. Finalement, en ce qui concerne plus précisément l'éducation, l'approche de la problématique des données ouvertes du secteur au niveau de l'Institut international de planification de l'éducation (IIPÉ) et de l'usage des données en général mérite qu'on s'y attarde.

II. Deuxième partie : Le cadre d'accueil : l'IIPE

1. L'IIPE et ses activités

1.1. Une mission pour une organisation statutairement autonome

Lors de sa XII^e session, en 1962, la Conférence générale de l'UNESCO (Organisation des Nations Unies pour l'Éducation la Science et la Culture) a adopté une résolution, qui donna naissance, un an plus tard à l'Institut international de planification de l'éducation (IIPE). Le mandat confié à cette nouvelle entité consistait en la promotion de l'enseignement et de la recherche « *en matière de planification de l'éducation, dans le cadre du développement social et économique* ». On peut voir dans cette formulation la prégnance d'un certain vocabulaire (planification, développement), d'une époque marquée par les décolonisations en Afrique et en Asie, et de l'idéologie développementiste portée, entre autres, par W. Rostow. Il est indéniable que la plupart des nouveaux États membres, indépendants depuis peu, avaient indéniablement besoin de cadres compétents dans différents secteurs, notamment de l'éducation. Cela explique sans doute le déploiement des activités de coopération de l'institut en général dans les pays « en voie de développement ». Pour autant, cela ne signifie pas que l'institut a été créé pour répondre aux sollicitations d'une catégorie particulière d'États membres. À travers ses activités, qui s'inscrivent dans les domaines de compétences de l'UNESCO, l'IIPE concourt à la réalisation de ses objectifs stratégiques de sa maison mère.

Les organes directeurs de l'IIPE, « Institut de catégorie 1 » du « secteur Éducation » de l'UNESCO, qui en compte sept (7) autres, relèvent directement de la Conférence générale. Étant institutionnellement rattaché à l'organisation, les membres du personnel de l'institut « *sont considérés comme fonctionnaires de l'UNESCO au sens de l'article VI, Section 18 de la Convention sur les privilèges et immunités des institutions spécialisées* » selon ses statuts. L'institut est un environnement multiculturel avec un personnel internationalisé formé de spécialistes dans divers domaines (économie, sociologie, statistique, démographie...). Les deux langues de travail sont l'anglais et le français. Cependant l'anglais en est la langue dominante, en dépit de l'implantation du siège à Paris. Le site internet de l'institut, important canal de dissémination, est bilingue. Le visiteur a la possibilité de choisir la langue de consultation (français, anglais), et permet d'accéder directement aux sites des bureaux de Buenos Aires et de Dakar. Le site de Dakar a adopté le bilinguisme français et anglais, et celui de Bureau de Buenos Aires, le bilinguisme espagnol et anglais.

1.2. Organigramme et fonctionnement de l'IIPE

Installé dans ses propres locaux dans le 16^e arrondissement de Paris, l'Institut international de planification de l'éducation, est administré par un Conseil d'administration. Un Directeur en assure la gestion :

- Le **Conseil d'administration** est composé de quatre (4) membres désignés par des instances et institutions du système onusien. Huit (8) autres membres sont élus parmi des éducateurs et économistes reconnus pour leur contribution au développement des ressources humaines. Leur mandat varie entre de 3 à 5 ans. Les membres dudit Conseil fixent l'orientation générale de l'institut, décide de l'allocation des ressources de fonctionnement. Il fait des recommandations au Directeur général de l'UNESCO pour la désignation du Directeur de l'institut, et il est consulté pour la nomination de ses hauts fonctionnaires.
- Le **Directeur** de l'institut est nommé pour un mandat de six (6) ans par le Directeur général de l'UNESCO. Chargé de la gestion de l'institut, il s'assure

de l'exécution du programme des activités. Il prépare le budget annuel à soumettre à l'approbation du Conseil d'administration. Ainsi, il est de sa responsabilité de s'assurer du respect du règlement d'administration financière.

Le personnel du siège parisien est d'environ 80 membres, il est réparti selon les unités d'activités suivantes :

- Bureau du Directeur
- Service de formation
- Coopération technique
- Recherche et Développement
- Administration
- Service d'information.

L'IIEPE dispose deux bureaux régionaux en Amérique latine, à Buenos Aires, et en Afrique à Dakar. Ce bureau était le Pôle Éducation de l'UNESCO en Afrique avant d'être rattaché à l'IIEPE en 2013. Son organigramme ne ressemble pas à celui du siège à Paris, contrairement à celui du bureau de Buenos Aires.²¹. Il s'agit peut-être d'une conséquence de son histoire institutionnelle.

Le financement des activités de l'institut provient de sources suivantes :

- Une allocation annuelle fixée par la Conférence générale de l'UNESCO ;
- Des subventions et dons d'autres institutions du système onusien et d'autres des États membres ;
- Des dons d'organismes publics ou privés, et de donateurs individuels ;
- De rémunérations de ses services.

Les ressources de fonctionnement sont versées sur un Compte spécial créé par le Directeur général de l'UNESCO. Il importe de souligner que l'allocation annuelle de l'UNESCO est relativement faible par rapport budget annuel de l'institut. Par exemple, pour l'année 2019, la dotation de l'UNESCO en représentait 9 % (2 140 000 des 23 555 039 dollars américains) [22, Rapport et états financiers vérifiés 2019]. Ainsi le partenariat avec d'autres institutions, les États et fondations s'avère-t-il crucial pour l'IIEPE, surtout dans un contexte où l'UNESCO fait face à de sérieuses difficultés financières. Ces difficultés ont malgré tout touché l'institut qui a été contraint d'adopter certaines mesures drastiques, notamment la compression de son personnel. Elles ont laissé des traces qui sont encore perceptibles parmi les salariés. Soulignons quand même que les états financiers de l'année 2019 ont été meilleurs que ceux de 2018, avec un excédent financier.

1.3. Le cœur de métier de l'institut : développement des capacités, production de connaissances, diffusion

L'IIEPE a été créé afin de répondre à des besoins en planification et gestion de l'éducation. En effet au début des années 1960, il a été constaté une « *pénurie mondiale de personnel qualifié en matière de planification de l'éducation et de la nécessité de recherches sur les problèmes fondamentaux relatifs à l'élaboration et à l'exécution de plans d'enseignement* » [23, Textes de base de l'IIEPE, pp. 52-53]. L'acte fondateur de l'institut, en l'occurrence la Résolution de la Conférence générale de l'UNESCO, a défini les activités qui devaient constituer le cœur de métier de l'IIEPE : développement des capacités (Formation), production de connaissances (Recherche). L'enjeu de la diffusion des connaissances produites a été bien souligné également. À travers un plan quadriennal dénommé « Stratégie à moyen terme » (SMT), cadre de référence de ses programmes et de ses projets, l'institut s'efforce de répondre à la mission qui lui a été attribuée par la Conférence générale.

²¹ IIEP Buenos Aires, Equipo, <https://www.buenosaires.iiep.unesco.org/es/equipo>

Dans la dixième SMT en cours, 2018 – 2021, l'institut se propose d'atteindre deux **objectifs stratégiques** :

- Les États membres planifient efficacement le développement du secteur éducatif et évaluent les performances du système ;
- Des connaissances appliquées en matière de planification et de gestion de l'éducation sont mises à la disposition des planificateurs de l'éducation et des parties prenantes.

Parallèlement, la SMT a retenu cinq **priorités thématiques** :

- Réduire les disparités au sein de l'éducation, en particulier les inégalités de genre ;
- Améliorer les résultats cognitifs et non cognitifs de l'éducation ;
- Renforcer la résilience des systèmes éducatifs par une planification adaptée aux situations de crise ;
- Améliorer la gouvernance, la transparence et la responsabilité ;
- Assurer un financement équitable et durable de l'éducation.

1.3.1. Formation

L'offre de formation de l'IPE se décline en plusieurs options afin de mieux répondre aux besoins des intervenants dans le système éducatif :

- Le *Programme de Formation Approfondie (PFA)*, en gestion et planification de l'éducation, s'adresse prioritairement aux cadres des systèmes éducatifs qui souhaitent renforcer leur expertise professionnelle. Ce programme est constitué de cours qui ont lieu à distance dans une première phase de trois mois. Puis, dans une seconde phase de six mois, ils se déroulent sur place à Paris. La formation comporte l'élaboration d'un projet de fin d'étude sur un sujet traitant d'un aspect institutionnel ou relatif au système éducatif du pays du candidat. La délivrance d'un certificat de niveau Master atteste de la réussite du parcours de formation. Ce programme a été fortement impacté par la crise sanitaire. Il doit reprendre en septembre 2021 selon un nouveau format, avec un raccourcissement de la période résidentielle (à Paris), et corollairement un allongement de la formation en ligne.
- La *Planification du secteur de l'éducation (PSE)* s'adresse aux planificateurs des ministères de l'éducation. Il s'agit de 12 semaines de cours en ligne et de 11 semaines à Paris. La formation est couronnée par la délivrance d'un Certificat international en planification du secteur de l'éducation.
- Le *Programme de cours spécialisés (PCS)* qui s'étend sur huit (8) semaines n'est pas seulement réservé aux fonctionnaires de l'éducation. La participation est soumise à une obligation d'expérience d'au moins ans dans la planification et la gestion de l'éducation. À la fin du cours, le participant reçoit une attestation de présence qui notifie également les cours suivis.
- Le *Campus virtuel de l'IPE* propose des cours en ligne, sur la plateforme Moodle. Ils sont destinés aux professionnels des ministères ou d'autres institutions souhaitant acquérir des compétences en gestion planification de l'éducation. Les participants qui auront rempli les conditions recevront un certificat à la fin des cours.

- Les *Formations sur mesure* qui sont proposées sur place à la demande des États sur des thèmes spécifiques. Ces formations peuvent également avoir lieu à la demande des organisations internationales, des organisations d'aide au développement, ou des Organisations non gouvernementales.

1.3.2. *Coopération technique*

L'IPE intervient directement auprès des ministères de l'éducation dans le cadre de la Coopération technique. Il s'agit en général d'offrir un accompagnement multiforme en fonction des besoins formulés par les États membres. La coopération peut se traduire par :

- Un appui à long terme à la planification de l'éducation. Cette planification est conçue comme un processus continu qui inclut toutes les parties prenantes. Son objectif est d'arriver à une appropriation commune de la vision et des priorités en vue du développement de l'éducation ;
- Un soutien aux centres de formation des ministères ;
- L'élaboration d'outils de gestion de l'éducation, comme le SIGE (Système d'information pour la gestion de l'éducation) ;
- Une mission d'experts pour des formations sur mesure ;
- La mise en œuvre de projets de courte durée.

La Coopération technique inclut donc une large gamme d'interventions en accord avec les États membres désireux d'améliorer leur système éducatif.

1.3.3. *Recherche et Développement*

L'IPE se définit comme un « institut apprenant », il est attentif aux évolutions des tendances qui peuvent avoir un impact sur l'éducation. La recherche est aussi un travail de défrichage effectué en amont. Elle doit précéder l'action, autrement dit l'action doit découler des résultats de la recherche. Ainsi la recherche sert à nourrir la réflexion pour les interventions dans le cadre de la coopération technique. Ce travail de fond peut également avoir une dimension de prospective en s'intéressant aux expériences de gestion et planification en cours, et aux évolutions des politiques éducatives. L'institut entend toujours appuyer ses interventions sur des bases factuelles et analytiques rigoureuses. Par exemple, certaines recherches en cours portent sur les sujets suivants²² :

- Gouvernement ouvert dans l'éducation.
- Villes et Éducation 2030
- Utilisation des données d'évaluation des apprentissages dans le cycle de l'évaluation

1.3.4. *Partage des connaissances*

Les activités de diffusion s'appuient essentiellement sur des réseaux institutionnels, notamment :

- Le Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation (SACMEQ)

²² Pour plus d'informations sur les recherches mentionnées :

Gouvernement ouvert dans l'éducation, <http://www.iiep.unesco.org/fr/gouvernement-ouvert-dans-leducation-0>

Villes et Éducation 2030, <http://www.iiep.unesco.org/fr/notre-expertise/villes-et-education-2030>

Utilisation des données de l'apprentissage, <http://www.iiep.unesco.org/fr/notre-expertise/utilisation-des-donnees-devaluation-des-apprentissages-dans-la-planification>

- Le Réseau asiatique des institutions de formation et de recherche en planification de l'éducation (ANTRIEP)
- Le réseau Éducation, Emploi et Inclusion sociale, RedEtis

Le *réseau des alumni* de l'IIPE facilite également des échanges constants sur les questions de planification et de gestion de l'éducation. Le *Forum des politiques éducatives* est un rendez-vous annuel réunissant une pluralité d'acteurs pour partager, entre autres, des expériences et connaissances. Parallèlement à d'autres outils de partage de connaissances, comme les différents portails :

- [Planipolis](#), Le portail de plans et politiques d'éducation, de cadres clés et rapports de suivi sur plus de 190 pays.
- [IIEP Learning Portal](#), Le portail de l'IIPE dédié à l'amélioration des apprentissages
- [The UNESCO HIV&Health Clearing House](#)
- [Etico](#), La plateforme sur l'éthique et la corruption dans l'éducation
- [Education à la sécurité, la résilience et la cohésion sociale](#)
- [PEFOP](#), La plateforme d'expertise en formation professionnelle
- [SITEAL](#), Le site d'information sur les systèmes éducatifs en Amérique latine

1.4. La bibliothèque de l'IIPE

1.4.1. Mission, personnel et fonctionnement

La Bibliothèque de l'IIPE était, jusqu'au mois d'octobre 2020, une unité rattachée à la Direction de l'institut. Depuis cette date elle est une entité rattachée à l'Unité Recherche et Développement, son responsable est placé sous la supervision du chef d'équipe de l'unité de rattachement. La bibliothèque assure les fonctions de centre de documentation en soutien aux équipes métiers. Elle leur fournit des informations sur des thématiques et sujets d'intérêt dans le cadre de leur travail. Elle est aussi chargée de la collection et de la diffusion de la littérature relative à la planification de l'éducation. Elle assure la gestion (contrôle qualité entre autres) de toutes les bases de données l'institut [24, Cunegati]. La taille de son personnel est assez réduite. Jusqu'au 30 septembre 2020, elle comptait un chef d'unité²³, trois documentalistes, et un vacataire à mi-temps. Ce dernier présent deux jours la semaine, s'occupe de l'accueil et du magasin. Chacune des bibliothécaires s'assure de missions spécifiques : acquisition et abonnement (ouvrages et articles imprimés ou électroniques, revues en ligne ou imprimées) ; administration et gestion des bases de données ; promotion et communication.

Le public de la bibliothèque peut être réparti en trois catégories :

- Les équipes opérationnelles de l'institut (chercheurs, enseignants, planificateurs). Elles bénéficient chacune de l'assistance d'une bibliothécaire ;
- Les stagiaires qui suivent les programmes de formation. Ils bénéficient d'un accompagnement des bibliothécaires pour les aider dans leur recherche documentaire et la découverte de la collection ;
- Le public intéressé (au profil varié) qui a la possibilité de consulter les documents en ligne, sur rendez-vous, consulter les documents sur place.

²³ La responsable de la Bibliothèque a pris sa retraite, et le poste n'a pas été pourvu. Mais un cadre de l'institut a été désigné pour « assurer la coordination ».

La bibliothèque est ouverte toute la journée de janvier à juin. Cette période correspond à celle de la présence des stagiaires en formation PFA (Programme de Formation Approfondie), PSE (Planification du secteur de l'éducation) et PCS (Programme de cours spécialisés). Le reste de l'année elle est ouverte seulement l'après-midi. Ce fonctionnement a été bousculé par la crise engendrée par le Covid-19. Jusqu'à la fin de l'été 2020, le personnel de la bibliothèque a continué à travailler à distance, notamment grâce aux réseaux sociaux d'entreprise. Puis la présence alternée a été autorisée pour l'ensemble de l'institut, dans le respect des contraintes sanitaires. Au cours des réunions hebdomadaires de l'équipe de la bibliothèque, outre les échanges plus généraux sur des questions d'ordre administratif, les bibliothécaires sélectionnent les articles à signaler de manière personnalisée, ou à indiquer sur la page de veille.

L'offre de service documentaire de la bibliothèque est assez étendue, on peut en distinguer :

- La formation à la recherche documentaire destinée aux stagiaires ;
- Le soutien aux équipes opérationnelles : **veille thématique**, notamment avec l'outil commun *Inoreader*. La veille est une fonction importante de la bibliothèque. Elle permet, entre autres, de signaler des informations d'intérêt aux équipes opérationnelles ; d'alimenter le portail Planipolis. Pour suivre l'actualité de de l'éducation les pays à veiller sont répartis entre les bibliothécaires.
- Une page de [veille thématique](#) destinée également au public internaute ;
- L'envoi de publications aux bibliothèques dépositaires ;
- Le contact avec le public internaute ;
- La mise à disposition des collègues et du grand public d'un portail [Netvibes](#) comprenant huit onglets, dont un contenant les ressources uniquement accessibles au staff de l'IIEP.

Parallèlement aux tâches sus-indiquées, les bibliothécaires assurent d'autres fonctions essentielles d'un centre de documentation : collecte, catalogage, structuration de l'information entre autres (plan de classement maison, thésaurus de l'UNESCO...). En outre, dans le cadre des activités de coopération technique, la bibliothèque a déjà, exceptionnellement, apporté sa contribution au renforcement des capacités en gestion documentaire.

1.4.2.Fonds documentaire et principales bases de données

La bibliothèque recèle un fonds documentaire multilingue, principalement en langue anglaise et en français. Il s'agit essentiellement d'ouvrages publiés, y compris des publications de l'institut et de l'UNESCO qui traitent de la planification et de la gestion de l'éducation. On y trouve également des périodiques spécialisés et des rapports de mission. La littérature grise est assez bien représentée, et elle consiste, dans certains cas, en des documents de grande valeur patrimoniale qui ne sont plus disponibles. Le cas par exemple de plans et de politiques d'éducation élaborés par certains États membres. La collection elle remonte à la création de l'institut, et la moitié en est actuellement disponible en ligne. Le personnel assure la gestion et l'alimentation de ses trois principales bases de données de l'institut :

- EPIDOC, la base du catalogue de la bibliothèque, librement accessible au public via la page web de la [bibliothèque](#) ;
- INTEPI, la base de catalogue de documents internes de l'IIEP (rapports, fiches projets, mémoire des stagiaires) non accessible au public ;
- PLANIPOLIS, qui fonctionne sous Drupal, est la base des plans et politiques d'éducation et autres. Elle est accessible au via le portail thématique éponyme. Le portail thématique [Planipolis](#) a été lancé en 2007 pour répondre à l'absence d'une base donnée mondiale sur les

politiques et plans nationaux d'éducation. La base de données du portail contient des documents d'une grande richesse. De nombreux contributeurs, y compris d'anciens stagiaires, ont facilité ce travail. Ainsi, comme cela a déjà été souligné, l'institut détient des documents et qui sont introuvables dans les archives des institutions qui les ont produits. En ce sens, la bibliothèque est « gardienne » d'une partie de la mémoire institutionnelle de certains pays qui font face à de graves lacunes en matière de conservation. Le stockage, sur ses propres serveurs, des fichiers (en format PDF) provenant des sites internet de certains ministères après sollicitation de l'accord de leurs auteurs, permet de garantir la pérennité des sources en cas de liens brisés. Depuis son lancement son succès est continu et ce portail, trilingue (anglais, français, espagnol) constitue l'un des produits phares de l'institut sur le web. Son franc et immédiat succès a incité la bibliothèque à creuser davantage la question de l'offre de service documentaire au « public distant ».



Figure 14. Capture d'écran planipolis.iiep.unesco.org/fr

1.4.3. Une attention au contexte informationnel

Au cours de ces deux décennies, la Bibliothèque de l'IIEP a accueilli des nombreux stagiaires du CNAM-INTD²⁴. Cela démontre une attention constante à l'évolution du contexte informationnel. Des thématiques diverses ont été étudiées dans le cadre de ces stages :

- *Réinformatisation* (2010). Il s'agissait de mener la réflexion afin d'identifier une solution à la suite de l'arrêt du développement de CDS/ISIS (*Computerized Documentation System/Integrated Set of Information Systems*) pour Windows (WINISIS). La réinformatisation devait permettre à la bibliothèque de répondre à ses besoins et à son souhait d'optimiser ses prestations en tenant compte de l'évolution des technologies du web (Web 2.0). Sans s'écarter de la famille « du libre », philosophie portée par l'UNESCO dans le domaine informatique, dans le monde du développement des logiciels. [25 Sobral Boyer].

²⁴ Maria SOBRAL BOYER (2010), Aurélie BERGONIER SIGNOLES (2011), Anna POLEWKA (2012), Sid-Ali CHIKH (2014).

- *FRBRisation* (2011). La bibliothèque se posait la question du décloisonnement de ses bases de données. L'enjeu étant l'amélioration de l'accès à l'information, la valorisation du fonds documentaire. Ce modèle conceptuel FRBR (*Functional Requirements for Bibliographic Records*), dans lequel le contenant l'emporte sur le contenu, et l'information bibliographique centrée sur l'œuvre, permet le dialogue entre réservoirs de données, est axé sur le web des données et facilite l'indexation par les moteurs de recherche [26, Bergonier Signoles].
- *Virtualisation* (2012). À la suite du succès du portail Planipolis, il s'agissait d'envisager l'amélioration de l'offre documentaire en ligne. Avec le développement du numérique, se posait la question du développement d'une bibliothèque hybride (simultanément public local et distant, support physique et numérique) qui permet de toucher un plus large public (local et distant). La réflexion portait sur la manière de tenir ensemble ces deux logiques. Probablement, la crise sanitaire actuelle poussera à creuser davantage cette problématique de l'extension de l'offre numérique [27, Polewka].
- *Web sémantique* (2014). L'étude portait sur la valorisation des ressources d'informations dans le cadre d'un projet de portail de l'institut, en l'occurrence *Plan4Learning*, appliquant les technologies du web sémantique. Il fallait expérimenter un référentiel standard permettant de tirer avantage du web de données [28, CHIKH].

Toutes ces thématiques soulevées dans les mémoires de stage soulignent la question de l'exploitation des opportunités offertes par le développement des technologies numériques. La capacité de réflexivité, et d'interrogations régulières sur ses pratiques, est d'importance pour une institution qui doit démontrer son aptitude à innover, à se confronter à l'exigence de répondre aux nouveaux défis, ou d'anticiper les tendances qui peuvent avoir un impact sur les systèmes éducatifs. Par exemple, l'inscription de la corruption dans le débat public mondial a incité l'IIPE à s'intéresser à ses conséquences sur la fourniture publique de services sociaux, principalement l'éducation. Aussi, l'IIPE a-t-il commencé à travailler sur la question de l'éthique dans l'éducation à partir du début des années 2000. La priorité thématique « *Gouvernance et responsabilité pour une planification transparente et participative* », de la 9^e Stratégie à moyen terme (2014-2017)²⁵, avait prévu la constitution d'une « *base documentaire sur les utilisations les plus efficaces et les plus critiques des données publiques (open data) relatives à l'éducation permettant d'améliorer la planification de l'intégrité du secteur éducatif* » [29, IIPE, p. 12]. Des recherches sur ce sujet ont été menées pendant quatre ans et ont donné lieu à de multiples publications, séminaires et rencontres²⁶. Avec la 10^e SMT (2018 – 2021), la thématique du gouvernement ouvert de l'éducation, qui implique également l'ouverture des données, représente une volonté d'élargissement de la focale. Par ailleurs, dans son mémoire de stage de 2011, Bergonier Signoles invitait à suivre les initiatives d'Open data des institutions culturelles, même s'il ne s'agissait pas d'une question prioritaire, car les « *initiatives open data auront à terme des retombées sur le positionnement de l'Institut* » [26, Bergonier Signoles, p. 119]. Plus globalement, dans un contexte où les *data* sont presque partout avec la révolution numérique, comment sont-elles

²⁵ IIEP, Stratégie à moyen terme, 2014-2017

https://unesdoc.unesco.org/ark:/48223/pf0000227245_fre?posInSet=2&queryId=33c31fe4-e0b6-4d11-90eb-a8ab215b4759

²⁶ Avec la 10^e SMT, la focale a été élargie avec la thématique du gouvernement ouvert de l'éducation, compris comme « *l'ouverture des données, processus, décisions et mécanismes de contrôle gouvernementaux pour permettre la participation et la surveillance du public, aux fins d'assurer une éducation inclusive, équitable et de qualité* ». Voir <https://etico.iiep.unesco.org/fr/gouvernement-ouvert-dans-leducation>

perçues au sein de l'institut ? Puisque l'institut fait partie du système onusien, il faut également s'intéresser aussi à la problématique des *data* dans le contexte onusien actuel. Un survol rapide destiné surtout à susciter la curiosité des bibliothécaires de l'institut, qu'à élucider de cette question au sein de l'ONU.

2. Les données vues des Nations unies : un aperçu

2.1. L'ONU et les « data »

2.1.1. Nécessaire « révolution » pour les ODD

En septembre 2015, l'Assemblée générale des Nations Unies a adopté la résolution intitulée *Transformer notre monde : le Programme de développement durable à l'horizon 2030* qui, définit les [17 objectifs de développement durable](#) (ODD).



Figure 15. Capture d'écran. undp.org.

La nécessité de disposer de données ventilées, fiables, et de qualité afin de suivre l'évolution et la mise en œuvre des politiques pour atteindre ces objectifs s'est alors imposée comme une évidence. Le Secrétaire général d'alors, Ban Ki-moon, désigna alors un panel de haut niveau afin de faire des propositions en ce sens. Les experts ont appelé à une *révolution des données*, notamment par l'ouverture des données, afin de disposer de meilleures statistiques et de données. Elles contribueront à appuyer les décisions sur des faits, et à renforcer la reddition des comptes [30, Independent Expert Advisory Group on a Data Revolution]. Deux années plus tard, le premier *UN World Data Forum*, sur le plan d'action global du développement durable, devait avancer l'idée que l'ouverture des données statistiques représente une des actions clés pour moderniser les systèmes nationaux de statistiques. Dans cette même optique, le projet « *Renforcer les capacités des pays en développement pour qu'ils s'équipent d'un accès à l'information à travers les données publiques en libre accès* », de la Division de l'administration publique et de la gestion du développement, a entrepris un travail de sensibilisation des gouvernements sur le libre accès des données publiques et la promotion des initiatives d'open data.

En ce qui concerne le secteur de l'éducation (ODD 4), l'Institut de statistique de l'UNESCO (UIS) a souligné le problème de la fiabilité des données de l'éducation dans les pays les plus concernés par cet objectif. La sous-utilisation des technologies et une carence de personnel compétent en seraient les principaux facteurs. L'UIS a également mis l'accent sur la nécessité

d'une révolution des données pour un meilleur suivi de l'éducation. Cette révolution exigerait, entre autres, un usage plus poussé des TIC. Par ailleurs, sans poser l'Open data de l'éducation, comme condition indispensable, l'UIS considère que l'ouverture des données en serait le scénario idéal : diffusion de l'information libre, ouverte et réutilisable [31, UIS].

2.1.2. « Une stratégie des données pour une action de tous partout »

En juillet 2020, l'actuel Secrétaire général de l'ONU, Antonio Guterres, a annoncé une nouvelle stratégie des données pour l'organisation internationale. Cette stratégie est présentée comme porteuse d'une vision ambitieuse dont la mise en œuvre est susceptible de transformer l'ONU. L'objectif affiché consiste à faire des Nations Unies une organisation *data-driven* : « *En construisant un écosystème de données des Nations Unies qui maximise la valeur de nos données, nous libérons tout notre potentiel : nous prenons de meilleures décisions et apportons un soutien plus important aux personnes et à la planète dans les moments les plus importants* » [32, UN Data strategy 2020-2022, p. 16] ²⁷. La stratégie définit douze principes qui doivent guider l'action de toutes les entités de la famille onusienne :

1. Traiter les données comme des actifs à l'instar des actifs corporels et incorporels de l'organisation ;
2. Viser l'excellence dans la manière de générer de la valeur avec les données ;
3. Assurer la protection et la confidentialité des données personnelles ;
4. Utiliser les données avec agentivité (*agency*) pour améliorer la décision humaine ;
5. Faire une utilisation des données avec équité ;
6. Mettre en place une gouvernance des données afin de préciser les responsabilités, les rôles, les normes et les politiques en la matière ;
7. Gérer les données avec transparence afin que les résultats puissent être compréhensibles et traçables ;
8. Les données sont la propriété de l'organisation, elles doivent être ouvertes par défaut à tous les collègues, sauf si des raisons de confidentialité l'exigent ;
9. Désigner des gestionnaires de données pour veiller à la qualité, à l'accès, à l'utilisation, à la protection et à d'autres responsabilités ;
10. S'assurer de la sécurité des données ;
11. Harmoniser les normes dans inventaires (catalogage, descriptions, classements) afin de rendre les données accessibles ;
12. Veiller à l'optimisation dans l'utilisation, la compréhension, l'accessibilité, la protection et la sécurité des données.

La politique de management des données qu'implique cette stratégie traduit une volonté d'une qui ne se limite pas à la dimension technique, informatique. Elle suggère, en quelque sorte, une division du travail, une clarification des rôles, des tâches, et des fonctions en ce qui concerne la chaîne de valorisation des données : *Data analyst, Data engineer, Data Protection officer, Data scientist, Chief Data officer*. Cette stratégie comprend également un volet formation en faveur du personnel pour doter les Nations Unies des compétences nécessaires à une meilleure exploitation des données. Comment les professionnels de l'information de famille onusienne trouveront-ils leur place dans cette nouvelle stratégie des données ? La stratégie n'apporte pas de réponse à cette question, cependant le personnel concerné devrait s'y intéresser. En tout cas, cette stratégie témoigne d'une réelle prise de conscience des enjeux liés aux données dans ce nouvel environnement technologique.

²⁷ Notre traduction

2.2. Les « data » du côté de l'IPE et contexte du stage

L'IPE, comme Institut de catégorie 1 de l'UNESCO, est aussi concernée par la stratégie proposée par le Secrétaire général. Comment cette question des « data » est-elle abordée au sein de l'institut ? Nous ne pouvons pas apporter une réponse tranchée à cette question, ni prétendre déceler ou dévoiler une ligne de l'institution sur le sujet. Cependant, nous avons pu aborder la question lors d'entretiens avec quelques cadres. Puisque la recherche constitue l'un des domaines d'activité de l'institut, la question des données de la recherche ne pouvait en être écartée.

2.2.1. Données de la recherche et politique de libre accès

Qu'est-ce qu'une donnée de la recherche ? En 2004 les pays membres de l'OCDE, auxquels s'étaient joints l'Afrique du Sud, Israël et la Russie, ont adopté une déclaration ministérielle sur les données de la recherche. Plus tard, il sera publié les *Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*. Ce document établit le principe de l'ouverture desdites données définies « *comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche* » [33, OCDE, p. 18]. En 2014, à travers le programme européen pour la recherche et le développement « Horizon 2020 », l'Union européenne lança une opération pilote d'*Open Research Data*. Cette initiative proposa d'aligner l'ouverture de l'information scientifique sur celles des données publiques [34, Commission Européenne]. En France, parmi les mesures du *Plan national pour la science ouverte* (2018), figure l'obligation d'ouverture des données issues de la recherche financée par des fonds publics. Elle est de nature à limiter certaines duplications, et favoriser la réutilisation des données collectées. Les données des sciences « dures » ont globalement servi de référence pour l'élaboration du paradigme des données de la recherche. Aussi des nombreuses interrogations peuvent surgir quant à leur définition et leur délimitation dans les sciences humaines [35, Cabrera].

L'IPE n'est certainement pas un organisme public, ni un centre d'enseignement supérieur avec des laboratoires de recherche. Mais les données de recherche ne se limitent pas exclusivement au domaine académique, ne serait-ce parce que la recherche n'est pas une activité réservée exclusivement à l'université. Les programmes de recherche peuvent également se dérouler dans le cadre d'activités gouvernementales (publiques) ou industrielles. Les activités de recherche de l'institut entrent dans le cadre de sa mission de production de connaissance en vue d'aiguiller les actions de coopération technique. Il ne s'agit pas de recherche-action à proprement parler, mais de recherche orientée vers l'action, une action qui prend appui sur de solides bases théoriques et scientifiques. Comment sont conservées les données de récoltées lors des études de terrain, même s'il s'agit d'enquêtes qualitatives ? La question, d'une éventuelle réutilisation, ou de l'ouverture des données issues de ces recherches ne paraît pas encore être à l'ordre du jour. Leur réutilisation à l'interne ne semble pas non plus être un sujet de préoccupation. En conséquence, la question du rôle de la bibliothèque pourrait jouer en ce sens n'est évidemment pas posée. En somme ces questions relatives à la science ouverte, plus précisément à l'ouverture des données de recherche paraissent constituer un non-sujet. Même si certains cadres sont plus ou moins sensibilisés à cette question. Mais, pour de multiples raisons, il existe de nombreux freins au partage qui ne découlent pas nécessairement d'une volonté de rétention de l'information.

Toutefois, il importe de souligner que l'IPE, en tant qu'institut de l'UNESCO, est engagé dans la politique de libre accès (*Open access*). En effet, depuis juillet 2013, la politique de libre accès est appliquée à toutes les publications de l'UNESCO. Ainsi, toutes les publications de l'institut sont immédiatement disponibles dans l'Archive ouverte de l'UNESCO hébergée dans le

système UNESDOC²⁸. Elles le sont sous licence Creative commons, *Attribution 3.0 Organisations Internationales (CC BY 3.0 IGO)*²⁹. Cette licence autorise, sous réserve de citer la source originale, de partager, et d'adapter l'œuvre originale pour toute utilisation, y compris commerciale.

2.2.2. Le projet « *Les données géospatiales dans la planification et la gestion de l'éducation* »

Les data de l'environnement informationnel ne sont pas complètement ignorées. À l'IIEP, la 10^e Stratégie à moyen terme (SMT) entre dans sa dernière année (2021). La préparation de la prochaine SMT a déjà bien démarré. La problématique d'un usage plus approfondi des possibilités offertes par le numérique, et plus particulièrement des possibilités de calcul et des mégadonnées, (données massives, big data) fait l'objet de réflexion stratégique. Ces données sont collectées, grâce aux moyens informatiques ou des instruments de tout ordre : des logiciels ou des capteurs variés (satellites, sondes, radiotélescopes...). Dans différents secteurs de l'activité humaine (recherche médicale, industrie, marketing, lutte contre la criminalité) il s'agit d'une question cruciale. L'enjeu est de pouvoir en tirer des informations exploitables de cette masse considérable de données, structurées ou non. Y aurait-il dans ce contexte une opportunité ? Est-ce que l'exploitation des mégadonnées, qui ne sont pas produites par les administrations de l'éducation, peut être mise à profit pour *faciliter* et *démocratiser* la planification et la gestion de l'éducation ? Quelles informations utiles peuvent-elles apporter à la planification et la gestion de l'éducation ? À raison que l'IIEP s'interroge sur l'exploitation de ces données à des fins de planification et de gestion de l'éducation.

L'institut est déjà en train d'explorer cette voie, par le croisement des données géospatiales et des données de recensement démographique et socio-économique. En effet, ces dernières ne permettent pas de saisir certaines évolutions, notamment celles relatives à la mobilité, aux migrations des populations. Parvenir à croiser données de géolocalisation avec celle des recensements classiques peut par exemple donner une idée de la sédentarité. L'IIEP ne collectera pas les données, elles seront disponibles par l'entremise d'autres entités du système onusien. Autant dire dans le respect des orientations générales sur la confidentialité, l'éthique et la protection des données du Groupe des Nations Unies pour le développement (GNUD). Le projet, intitulé *Les données géospatiales dans la planification et la gestion de l'éducation* est une première initiative qui explore le potentiel de planification et de gestion de l'éducation de ces données. Il vise à développer « de nouveaux outils et méthodologies qui repensent l'intégration des données géolocalisées dans les pratiques des ministères de l'Éducation »³⁰. L'objectif étant de permettre la micro-planification de l'éducation à travers une approche hautement contextualisée. Cette approche permettrait de mieux prendre en compte les besoins des communautés locales et favoriser une utilisation optimale des ressources.

2.2.3. Données ouvertes sur l'école

Dans le cadre de la 9^e Stratégie à moyen terme (2014-2017), l'institut a conduit des recherches sur le sujet des données ouvertes sur l'école, comprises comme « [jeu] de données agrégées qui sont publiées école par école. Ces données peuvent couvrir différentes dimensions, comme le financement de l'école, le nombre d'élèves et d'enseignants, l'équipement et les

²⁸ UNESCO, Publications en libre accès, <https://unesdoc.unesco.org/search/72a71bb0-74c9-4ef5-a26b-934dd8b90ab8/N-f0504c79-d317-4136-93b8-e8e576368761>

²⁹ Creative Commons, Attribution 3.0 Organisations internationales. <https://creativecommons.org/licenses/by/3.0/igo/deed.fr>

³⁰IIEP, Les données géospatiales dans la planification et la gestion de l'éducation. <http://www.iiep.unesco.org/fr/les-donnees-geospatiales-dans-la-planification-et-la-gestion-de-leducation> [consulté le 20/01/2021].

manuels, ainsi que les résultats aux examens »³¹. L'acception « données ouvertes » retenue dans le cadre de cette recherche ne renvoie pas exclusivement aux « données ouvertes » en contexte d'Open data. Les critères de qualification concernant l'environnement technique et le format de publication entre autres, ne sont pas applicables aux « données ouvertes » dans le cadre de cette recherche. La responsable du projet de recherche assume ce choix puisqu'il porte essentiellement sur des enjeux d'intégrité des systèmes éducatifs, la possibilité réutilisation des données publiées n'en constitue pas un enjeu majeur. Cette approche repose sur l'idée fondamentale que la diffusion de toute information publique est susceptible de contribuer à la transparence de l'action publique, encourager la participation des personnes concernées, et éventuellement constituer un bon moyen pour faire reculer la corruption. Autrement transparence de l'action de l'État par la diffusion de l'information publique a une épaisseur historique irréductible à la séquence temporelle de l'Open data.

Il s'agit là d'une démarche plutôt pragmatique, soucieuse de l'aspect opérationnel : exploiter toute opportunité de diffusion de l'information, y compris l'utilisation les nouvelles technologies pour promouvoir la transparence et l'intégrité des systèmes éducatifs. Les contextes dans lesquels intervient l'IIPE prêtent mieux à une telle acception. En effet, l'institut travaille surtout avec des pays dont les infrastructures numériques sont plutôt peu développées comparativement aux pays de l'Europe occidentale de l'Amérique du Nord. Il faut souligner également qu'il ne s'agit pas d'une réticence envers l'Open data, encore moins en éducation. L'engagement dans une politique d'Open data est un investissement coûteux en ressources humaines et financières. Or les retours, notamment par la réutilisation des données, ne sont pas toujours évidents. Les données ouvertes de l'éducation soulèvent par ailleurs d'autres considérations : leur inscription dans une stratégie plus globale d'amélioration du système scolaire ; leur opportunité, autrement une réflexion sur les conditions minimales pour en garantir des retombées positives. Nous ajouterons également la question de la sélection des données à diffuser.

L'objectif de ce projet de recherche consistait à évaluer à quelles conditions ces données permettraient d'améliorer la planification et l'intégrité de l'éducation. Plusieurs activités, notamment l'organisation d'un Forum politique à Manille aux Philippines en 2018, ont été réalisées dans le cadre de ce projet. Par ailleurs des expériences de données ouvertes sur l'école dans différents contextes régionaux ont été étudiées. Elles ont donné lieu à trois (3) publications :

- En 2016 : une étude approfondie de quatorze initiatives de tableaux de bord des écoles un peu partout dans le monde [36, Cheng, Moses] ;
- En 2018 : une revue analytique pour évaluer l'impact des tableaux de bord dans différents contextes au sud du Sahara [37, Poisson, Dupain, Nguyen] ;
- En 2019 : une exploration de la question des tableaux de bord des dans quelques pays ibéro-américains [38, Brito].

Une dernière publication propose aux planificateurs de l'éducation des éléments de réflexion approfondie sur l'ouverture des données sur les écoles. Elle fait également le tour des principaux aspects à prendre en compte : format, choix des données, engagement citoyen, les risques qui peuvent découler de l'ouverture de l'information. En dernier lieu, elle détaille les différentes étapes pour réussir une ouverture des données sur les écoles [39, Poisson].

³¹ Données ouvertes sur l'école.

<https://etico.iiep.unesco.org/fr/donnees-ouvertes-sur-lecole> [consulté 14/08/2020].

2.2.4. Open Data et SIGE, quels effets bénéfiques ?

Les plateformes de notre corpus (data.gouv.fr ; data.gov.in ; opendata.go.ke ; dados.gov.br), à l'instar des autres plateformes gouvernementales, exposent de jeux de données extraites d'un système d'information pour la gestion de l'éducation. Le SIGE est une dénomination générique d'un outil d'aide à la décision (pilotage), et plus rarement de gestion (management) des systèmes éducatifs. C'est un système d'information statistique dont les modules peuvent varier d'un système éducatif à un autre. Généralement, on y trouve ces modules concernant les élèves, les enseignants, les résultats scolaires, les infrastructures scolaires entre autres. En outre, les informations fournies par le SIGE permettent en outre d'éditionner l'annuaire statistique de l'éducation. La principale finalité de ce système est de fournir des informations aux décideurs et autres gestionnaires du système éducatif. En ce sens il est indispensable qu'il puisse fournir des données actualisées et de qualité et, condition nécessaire à une planification efficace. Ainsi, quand, à la demande d'un État membre, l'IIEP vient en appui d'un ministère pour contribuer à améliorer la planification et la gestion du système éducatif l'évaluation du SIGE est indispensable.

Les questions relatives à la fiabilité, la qualité et la disponibilité de données actualisées, figurent parmi les problèmes les plus récurrents en ce qui concerne les systèmes éducatifs. Par ailleurs, pour ce qui a trait à leur diffusion, la *Brookings institution* a mené une étude sur la disponibilité en ligne des données de l'éducation de 133 pays à faibles et moyens revenus. Elle indique que les ministères de l'éducation de trente-quatre (34) pays ne disposaient pas d'un site internet. Quand ils en disposent, soixante-douze (72) y publiaient des données statistiques de l'éducation. En termes de granularité, les données publiées de seulement vingt-neuf (29) ministères pouvaient être désagrégées par établissements scolaires. [40, Read, Atinc]. Est-ce qu'une politique d'ouverture des données aurait pu avoir un impact sur ces facteurs ? Quels bénéfices une unité SIGE tirerait-elle d'une telle politique en considérant les dimensions technologique et humaine (montée en compétences) ? Si cette politique d'ouverture des données de l'éducation arrive à plus ou moins respecter les principes de base de l'Open Data, notamment ceux qui concernent la qualité des données ouvertes et leur distribution, il est quasiment certain qu'elle y apporterait des améliorations notables. Il faut rappeler qu'en 2017 l'Institut de statistique de l'UNESCO (ISU) a souligné, en ce qui concerne la collecte des données de l'éducation, une sous-utilisation des technologies et une carence de personnel compétent.

Par ailleurs, le rapport régional (Amérique latine) rédigé dans le cadre du projet de l'IIEP sur les données ouvertes de l'école préconise, dans une de ses recommandations, une véritable politique d'ouverture des données pour éviter que la fonction sociale de l'information publique ne soit trop limitée [38, Brito]. L'ouverture des données pourrait d'un autre côté créer un effet d'entraînement en contribuant à l'automatisation des processus, à l'amélioration de la qualité des données, et le renforcement de capacité des acteurs du SIGE³². Cependant, pour les pays qui font face à ces difficultés dans leur système d'information, cela reste un vaste programme. Il faudrait intervenir sur une large palette de facteurs notamment technologique, puisque dans certains cas, la saisie est encore manuelle et quelquefois avec une graphie hésitante. D'où l'intérêt de l'approche holistique de l'Open data prônée par la Banque mondiale. En conclusion, il importe surtout d'avoir à l'esprit que si l'Open Data vise à mettre des données administratives à la disposition du grand public, le SIGE doit d'abord fournir des informations fiables aux décideurs.

2.2.5. Contexte du stage

L'IIEP a déjà accueilli, au sein de l'équipe de sa Bibliothèque, un certain nombre de stagiaires dans le cadre de leur formation à l'INTD. Les sujets des mémoires, de ces missions de stage, sont en général issus d'une problématique liée à la mission. Soit, d'une part, des

³² Entretien avec un spécialiste des SIGE de l'institut, 30/09/2020.

questionnements autour de la faisabilité d'une éventuelle évolution : réinformatisation de la gestion documentaire (2010) ; nouvelle structuration de l'information avec le modèle FRBR (2011) ; nouvelle offre documentaire en tenant compte des usagers distants (2012) ; la question de la gestion de connaissance (2013) ; ou nouvelles fonctionnalités d'un outil de veille (2015). Soit, d'autre part, de projet en cours, comme la stratégie de valorisation de ressources documentaires dans le cadre du projet de portail Plan4Learning (2014). Malheureusement, le sujet de notre mémoire n'est pas étroitement lié à la mission de stage. Il ne répond pas de l'une ou de l'autre situation susmentionnée. La portée de la mission et le contexte particulier du stage n'y sont peut-être pas étrangers. En effet, le stage a été effectué complètement à distance, sans réelle immersion (physique). Si les mesures de prévention de la propagation du Covid-19 ont été respectées, les interactions ont perdu en densité. Sans nier les impacts de certains événements d'ordre privé, le contexte n'a pas été sans conséquence sur le déroulement du stage.

En dépit de ces difficultés, l'idée de travailler sur l'Open data, et sur la contribution éventuelle de la bibliothèque a germé des échanges avec ma tutrice de stage, qui est partie en retraite à la fin du mois de septembre 2020. Des éléments factuels et analytiques ont peut-être contribué à orienter le curseur vers cette thématique. D'une part, l'institut a mené des études sur les données ouvertes sur l'école au cours du cycle périodique de la 9e SMT (Stratégie à moyen terme, 2013-2017). D'autre part, un fil ténu relie malgré tout le sujet à la mission de stage qui comportait un volet de mise à jour des annuaires statistiques de l'éducation des États membres de l'UNESCO. Sans véritable grande surprise, cette tâche s'est heurtée à la question de la disponibilité des sources sur le web. Dans certains cas en raison de liens brisés ou de sites qui ont tout simplement cessé d'exister. Par ailleurs, j'ai été confronté au problème de l'actualisation des données. Est-ce qu'une politique d'Open data aurait permis d'éviter ces difficultés ? Même si la réponse peut être loin d'être évidente, compte tenu des principes qui fondent le mouvement en faveur de l'ouverture des données, un engagement fort dans cette direction est effectivement susceptible d'impulser une dynamique positive. Cependant, l'institut n'a pas d'expérience de coopération avec les États membres sur la question de l'Open Data. En outre, l'institut n'a pas une politique d'ouverture des données, et de surcroît ne l'envisage pas pour le moment. Même si, comme entité de catégorie 1, il est soumis à la « Politique d'accès à l'information de l'UNESCO »³³. En conséquence, notre étude ne pouvait pas s'appuyer sur des éléments empiriques et factuels issus des retours de terrain, des leçons apprises. Elle ne pouvait pas non plus s'inscrire dans une approche prospective en prévision d'une ouverture des données de l'IPE. Finalement, cette absence d'éléments factuels, de sources d'information, de documents institutionnels internes à exploiter est une des limites de cette étude.

Dans ce contexte, pour surmonter ces différentes embûches, le choix retenu a été d'aborder le sujet de l'Open data de l'éducation de manière générale, et de proposer un guide pratique à l'ouverture des données, particulièrement L'ambition de cette étude est de proposer à la bibliothèque de l'institut un état de l'art et des éléments de stratégie d'ouverture. Elle peut contribuer également à alimenter la réflexion sur son apport éventuel dans le cadre d'une assistance technique à l'ouverture des données, et contribuer à identifier de nouvelles pistes d'exploration des perspectives d'évolution quant à son offre de services.

³³ Politique d'accès à l'information de l'UNESCO, <https://fr.unesco.org/this-site/access-to-information-policy> . La diffusion en Open Data n'est pas requise.

III. Troisième partie : Éléments pour un guide stratégique pour l'Open data de l'éducation

1. La bibliothèque de l'IPE face à l'Open data

Les données (data) sont présentes à l'IPE, à travers l'exploitation du « big data » pour une meilleure gestion de la planification. Elles le sont aussi avec les données ouvertes sur l'école dans le cadre de la promotion de la transparence et de la lutte contre la corruption dans les systèmes scolaires. Mais sans occuper une place particulière dans les réflexions stratégiques, même si la problématique du big data est soulevée en lien avec la planification. Plus précisément, il ne semble pas que les data vont dans un avenir proche révolutionner les pratiques de l'institut. Cela explique peut-être pourquoi la problématique des données (big data, open data, données de recherche) semble être un non-sujet au niveau de la bibliothèque. En tant que centre de documentation elle a un rôle précis de collecte, de traitement et de diffusion, destinés prioritairement aux équipes opérationnelles de l'institut. Même si, parallèlement, elle reste très engagée dans la diffusion auprès du grand public notamment sur internet. Dans, le domaine des sciences de l'information, la présence multiforme des données dans l'environnement informationnel fait l'objet de nombreuses interrogations. Nombre de travaux mettent l'accent sur les défis à venir et les évolutions nécessaires du métier. Les documentalistes de la bibliothèque de l'IPE ont intérêt à prêter une attention à ces discussions qui agitent le milieu professionnel.

L'institut a déjà effectué des recherches approfondies sur les données ouvertes sur l'école, et a publié des travaux de recherche de grande qualité sur la question. En tenant compte de sa réputation auprès des États membres, il est possible d'imaginer une demande d'assistance technique à l'ouverture des données de l'éducation. En effet, malgré certains tâtonnements et essoufflements sporadiques, l'Open data porte une dynamique qui ne va pas s'arrêter en chemin. Puisque la bibliothèque a déjà offert ses prestations, dans le cadre d'une assistance plus large en faveur des États membres, dans l'accompagnement ou la mise en place de centre de ressources, elle pourrait également apporter sa contribution sur l'Open data. Indépendamment du périmètre de l'assistance technique l'institut, la bibliothèque pourrait saisir cette opportunité pour mettre en avant la valeur ajoutée des professionnels de l'information et de la documentation dans une telle entreprise. Comme Lyne da Sylva l'a bien souligné, les « *professionnels de l'information peuvent faire deux types de contributions importantes. Du côté de la mise en ligne, ils peuvent élaborer la description des données et l'ajout de métadonnées. Du côté de l'utilisation : pour le public, les professionnels de l'information peuvent aider à faire l'évaluation nécessaire de la qualité des données ou de la pertinence pour un besoin d'information donné* » [41, Da Sylva, p. 24]. Les documentalistes de l'institut déploient déjà ces compétences dans le cadre de la bibliothèque en ligne (interface publique de la base de données EPIDOC), ou du portail Planipolis.

Alors, tant du côté de la production des données ouvertes que de la médiation, les spécialistes de l'information et de la documentation peuvent mettre en valeur certaines compétences et pratiques professionnelles. La question de la visibilité des données publiées sur le web est une des questions cruciales pour l'Open data, surtout dans la perspective du web des données. Une bonne documentation des données, critère de qualité majeur, contribue également à améliorer leur découvrabilité en complément du moteur de recherche et de l'éditorialisation du site entre autres. Des métadonnées précises, pertinentes et cohérentes facilitent la recherche des utilisateurs des plateformes. Elles permettent d'identifier, de repérer les données appropriées à leur recherche. Comme une bonne indexation du document facilite la recherche du lecteur en bibliothèque. Les professionnels de l'information et de la documentation amèneront leur compétence en matière d'indexation et de catalogage pas toujours valorisée dans les projets d'Open data. Il s'agit, là encore, d'activités récurrentes des documentalistes de l'institut, y compris dans le cadre de l'exercice de la fonction de veille.

L'Open data fait le pari de la construction d'une intelligence collective à partir de la réutilisation des données. Il s'agit d'une démarche communicationnelle qui nécessite une implication du public irréductible aux développeurs ou réutilisateurs professionnels. Aussi le professionnel de l'information doit la penser en ayant en tête la pyramide de Russel Ackoff (de la donnée à la sagesse/*compétence*). En tant que médiateur, il est interpellé par la place du citoyen ordinaire il pourra contribuer à alimenter la réflexion sur les objectifs, les périmètres de l'ouverture des données, les formes de médiation à imaginer³⁴.

En effet, il est légitime de s'interroger sur le mode de réutilisation auquel l'on peut s'attendre lorsqu'il s'agit de l'Open data dans le domaine de l'éducation ? Est-il pertinent d'appréhender les données de l'éducation strictement de la même manière que celles concernant les transports, la météo, de la fréquentation des lieux culturels entre autres ? Autrement, serait-ce prendre un trop grand risque que de faire bouger les lignes de l'Open data en attirant l'attention sur la place des données qualitatives, de la publication d'informations, de ressources documentaires pédagogiques, plus accessibles aux citoyens ou plus susceptibles de combler certaines attentes ? L'Open data de l'éducation peut ainsi être amenée à dépasser la vision des « données brutes » issues des sciences « dures » ou qui traitent essentiellement de données quantitatives. De telles interrogations ne doivent pas pourtant perdre de vue les visées premières des initiateurs du mouvement, c'est-à-dire fourniture des données administratives afin de rendre l'administration moins opaque.

Cela pourrait notamment passer par une réflexion sur une mise à disposition d'outils pédagogiques (en s'inspirant par exemple d'[Eduscol](#), des réflexions et des enjeux concernant le *Libre accès* et les *Archives ouvertes* – tout en évitant de les amalgamer) dans le cadre d'un projet d'Open data de l'éducation. En construisant une argumentation solide sur le rôle des professionnels de l'information ou en soulevant des questionnements pertinents sur l'Open data de l'éducation, la bibliothèque pourrait justifier par voie de conséquence son apport à une assistance technique à un État membre. Il pourrait consister en un appui direct au renforcement des capacités dans la structuration des données, la documentation des jeux de données, et la gestion éditoriale de portails. L'implication des documentalistes – du ministère de l'État membre – dans le projet d'ouverture des données serait alors une nécessité.

Pour la bibliothèque de l'IIEP, cela impliquerait de suivre l'évolution de la dynamique de l'Open data, particulièrement de l'éducation, sans négliger les questions liées à l'Open access, à l'Open archive, et aux données de la recherche.

Il s'agit aussi de saisir les enjeux de l'Open data, et même d'en tirer parti dans l'accomplissement de sa mission au sein de l'institut. Certaines questions sont d'intérêt pour la bibliothèque, notamment celles qui portent, par exemple, sur les **métadonnées** (standards, modèles), l'**interopérabilité sémantique**, l'**interopérabilité des catalogues** (vocabulaire DCAT), les **fonctionnalités des plateformes** de données ouvertes entre autres. Il ne s'agit pas particulièrement de sujets inconnus. En s'intéressant en priorité aux États membres avec lesquels travaille l'institut, la bibliothèque pourra documenter les questions relatives à l'accès à l'information, aux données administratives, aux données à caractère personnel, à la propriété intellectuelle, au régime juridique des licences. Ces informations pourraient par exemple figurer dans une fiche pays.

Cette veille de l'Open data de l'éducation peut être matérialisée à travers une évolution de l'offre du portail [Planipolis](#) qui pourrait prendre la forme d'un produit documentaire pour faciliter l'accès aux données ouvertes de l'éducation. Une sitographie qui recense différentes plateformes, par exemple. Ce produit pourra être intégré dans une rubrique des annuaires statistiques de l'éducation. Préalablement à une réflexion approfondie sur le positionnement de la bibliothèque, pouvant déboucher sur des actions précises, il importe d'avoir une

³⁴ Nous entendons ici les utilisateurs ne disposant pas des compétences d'un réutilisateur professionnel. Se pose ici la question du public cible, l'Open data peut-elle se contenter d'un public extrêmement réduit ?

compréhension générale d'une démarche d'ouverture des données. Autrement, comprendre les étapes clés d'une stratégie d'ouverture peut s'avérer très utile pour mieux trouver sa place dans un projet d'ouverture des données, de l'éducation en l'occurrence.

2. Élaborer une stratégie d'ouverture des données de l'éducation

2.1. Cerner les enjeux, clarifier les objectifs

Dans toute politique d'ouverture des données publiques, il s'avère indispensable de bien se représenter l'ouverture, d'une part, comme action publique ayant pour but la publicité (garante de la transparence) de l'information et des données administratives. Autrement, une action (cristalline publique, affichage, la diffusion multimédia) qui a une histoire déjà ancienne dans les relations entre gouvernants et gouvernés. Et, d'autre part, comme modalité d'ouverture portée par un mouvement plutôt récent, dans la foulée du développement des technologies du numérique. L'accent y est surtout mis sur la donnée en format numérique aisément traitable par un ordinateur. Actuellement, l'*Open Data*, tend davantage à renvoyer aux modalités de l'ouverture des données publiques, tout en préservant et en étendant la finalité qui n'est réduite à la publicité. L'ouverture devient également possibilité de réappropriation, de réutilisation à d'autres fins. Avec l'*Open Data*, dans son sens courant actuel, la donnée administrative tend à devenir bien public immédiatement disponible.

Dans l'acception du mouvement Open Data, l'ouverture est caractérisée par la réunion de ces trois conditions :

- 1) Légale : une loi qui autorise l'accès, l'exploitation, la modification des données administratives, qui de ce fait tombent dans le domaine public. Les licences spécifient le cadre de leur réutilisation.
- 2) Technique : les données sont présentées dans des formats ouverts facilitant l'interopérabilité. De manière qu'elles soient automatiquement lisibles par les ordinateurs.
- 3) Économique : les données sont offertes dans certains cas (*Loi pour une République numérique*, 2016, par exemple), ou à un coût marginal.

Trois points reviennent constamment dès lors qu'il s'agit des enjeux de l'Open Data :

- 1) La transparence et la participation citoyenne à la définition de l'action publique ;
- 2) L'amélioration et l'efficacité de l'action publique ;
- 3) Les opportunités économiques grâce aux produits innovants.

On peut admettre qu'il s'agit de formulations générales quelque peu évasives. Elles permettent néanmoins de définir des catégories d'attentes et d'objectifs en fonction de la pluralité des acteurs engagés dans l'aventure de l'Open data. Faut-il également avoir en tête que c'est un mouvement relativement récent. L'histoire nous a appris le coût souvent élevé, de l'absence de transparence et du verrouillage de l'information publique dans les sociétés. Avec le temps, il sera sans doute plus aisé de cerner tous les bénéfices de l'Open data.

L'ouverture des données administratives reste en tout cas un choix politique volontariste. Un fort engagement fort demeure l'une des clés de réussite d'un projet. Dans le cas particulier d'un ministère de l'Éducation qui s'engage dans l'ouverture de ses données, il importe également de réfléchir à l'orientation de cette action en relation avec le système scolaire :

- ❖ Diffusion d'informations statistiques sur les écoles permettant de dresser le profil du système (selon quelle granularité : découpage administratif, établissement...)
- ❖ Diffusion d'informations administratives ;
- ❖ Diffusion d'outils ou de ressources pédagogiques.

Aussi importe-t-il, dépendamment de l'ambition, de ne pas se limiter à la diffusion de données dites brutes, et de réfléchir plus largement en termes de contenus. En tout cas, un ministère de l'Éducation qui s'engage dans l'ouverture des données gagnera à intégrer la question des ressources éducatives libres dans la réflexion³⁵.

2.2. Pilotage opérationnel

L'ouverture des données mobilise différents services au sein d'une institution, elle aura un impact ou induira une charge de travail supplémentaire pour tous les collaborateurs. Il est donc nécessaire de bien prendre en compte la question du pilotage du projet. Puisque son impact ne va pas seulement concerner une direction, unité ou service, l'implication de la plus haute autorité de décision est fondamentale. Même si les aspects techniques du projet exigent une implication inégale des différentes directions, il est préférable que la coordination soit rattachée une direction générale par exemple. L'existence d'un cadre légal et réglementaire spécifique faisant obligation d'ouvrir les données aidera certainement à la légitimation du projet.

Le coordonnateur du projet qui sera amené à travailler étroitement avec les différents services doit pouvoir bénéficier d'un franc appui du sponsor parallèlement à ses compétences (pas nécessairement informatiques), sa force de conviction, sa connaissance des enjeux, des difficultés et des inévitables barrières à la démarche d'ouverture des données. Lesdites barrières ne relèvent pas seulement de l'incompréhension ou de la mauvaise foi. L'Open data implique des changements dans les pratiques, bousculera donc certaines habitudes bien ancrées. La stratégie, la formalisation, et la qualification des données en fonction des exigences normatives de l'Open Data sont du ressort du coordonnateur qui est le **réfèrent données** (data) du projet.

La dimension technique et informatique du projet est du ressort d'un **informaticien** qui maîtrise les standards techniques, des applications spécifiques, et dispose d'une vraie connaissance du système d'information et d'autres applications au sein du ministère. Il lui reviendra d'engager le dialogue quant aux dimensions techniques des changements à effectuer. Le réfèrent data, et l'informaticien, les chevilles ouvrières, du projet doivent s'assurer d'un diagnostic approfondi de l'existant en termes de données et d'infrastructure.

2.3. Diagnostic de l'existant

Une fois arrêtée la décision politique de l'ouverture des données qui fixe les objectifs, l'équipe en charge de l'élaboration de la stratégie doit procéder au diagnostic de l'existant au regard des principes de l'Open Data. Ce diagnostic portera sur les aspects suivants [voir 7, Open Government data Working Group].

- 1) *Politique* : l'ouverture des données est-elle une décision des plus hautes autorités gouvernementales (Premier ministre par exemple), ou un choix sectoriel du ministère ? Dépendamment de la situation, il sera plus ou moins facile à surmonter certains obstacles institutionnels ou des barrières techniques, juridiques, financières. Idéalement, une stratégie d'ouverture des données ambitieuse doit être portée au niveau politique le plus élevé sous la forme de politique publique ;

³⁵ Selon la définition de l'UNESCO « Les ressources éducatives libres sont les documents de recherches, d'enseignement et d'apprentissage qui sont dans le domaine public ou publiés avec une licence de propriété intellectuelle qui permet la libre utilisation, l'adaptation et la distribution ».

- 2) *Juridique* : s'informer de la loi applicable en matière d'accès, de diffusion, des droits en ce qui concerne l'information publique et les données administratives. Avec les services juridiques du ministère, il faut arriver à identifier les dispositions qui peuvent servir de bases juridiques au projet. Et celles qui peuvent éventuellement en constituer des freins. Autrement, vérifier si le cadre légal et réglementaire statue sur :
- ✓ La définition d'une information ou donnée publique. En France, la loi 78-753 (17 juillet 1978) permet de bien délimiter cette notion.
 - ✓ La diffusion des données administratives. Existe-t-il une obligation de diffusion, c'est-à-dire de mise en ligne par défaut, fournit-elle des indications concernant les demandes de communication de données publiques ?
 - ✓ La protection des données personnelles, protection de la vie privée. Il s'agit d'apprécier cette question au regard des dispositions légales. Comment traiter les données relatives à un individu, ou permettant de l'identifier ? En ce qui concerne les données de l'éducation par exemple, il est également possible de s'enquérir de l'existence ou non de dispositions particulières les concernant ;
 - ✓ Les droits concernant les données administratives. Autrement, à quelles conditions peuvent-elles être réutilisées ? Existe-il des licences de réutilisation ?
- 3) *Données*. Il s'agit de s'interroger et de comprendre l'état de la situation en ce qui concerne les données au sein de l'institution, les différents processus au cours desquels elles sont produites, collectées, utilisées, stockées (formats de fichiers). Cette évaluation permet d'effectuer l'inventaire des données au sein du ministère. Elle aidera aussi à identifier les potentiels contributeurs et les éventuels référents Open data des différents services. Ce diagnostic concerne :
- ✓ La typologie des données produites en fonction des différents services et processus (organigramme du ministère) ;
 - ✓ Le mode de gestion des données, si le système est centralisé ou géré de manière autonome en fonction des producteurs de données ;
 - ✓ Les formats de fichiers existants ;
 - ✓ Le système de collecte (informatique ou manuel), de traitement, et de diffusion des données de la statistique scolaire : rythme, format – physique ou numérique – en ligne ?
 - ✓ Les remontées des données d'autres services ou celles issues par exemple de l'institut de statistiques et qui présentent un intérêt pour le système éducatif ou utilisées par le ministère.

- 4) *Infrastructure* : puisque l'Open data est avant tout la mise en ligne des données, la question de l'infrastructure informatique revêt d'une importance cruciale. Il est nécessaire de poser les bonnes questions sur la capacité à assurer la diffusion de données ouvertes sur internet. Il est possible de se renseigner sur la bande passante mensuelle par pays (voir [Speedtest Global Index](#)). L'écosystème des ICT doit être évalué : l'infrastructure et la pénétration du réseau internet³⁶.

L'établissement de ce diagnostic doit permettre, d'une part, de bien mesurer les différentes actions à mener en vue d'atteindre les objectifs formulés. D'autre part, d'identifier les barrières à surmonter : politique, légale, technique, financière entre autres.

L'aspect financier, le coût de la politique d'ouverture, ne doit pas être sous-estimé. D'autant que les retombées financières ne seront pas clairement identifiables. La volonté politique est indispensable pour soutenir cet investissement public. Le diagnostic entraînera certainement un retour sur les ambitions initialement conçues, et permettra de mieux étayer ses objectifs spécifiques. Il appellera sans doute à une approche beaucoup plus pragmatique. Il s'agit d'une étape clé pour un bon calibrage des objectifs et l'élaboration d'une bonne stratégie d'exposition de ses données.

En résumé, le projet Open Data repose sur la maîtrise de ce processus de publication (*Figure 16*) qui doit transformer les données internes en jeu de données ouvertes à destination d'utilisateurs finaux [42, Open data Corsica, État de l'art]. Lesquels contribueront en retour à l'amélioration de la qualité des données.

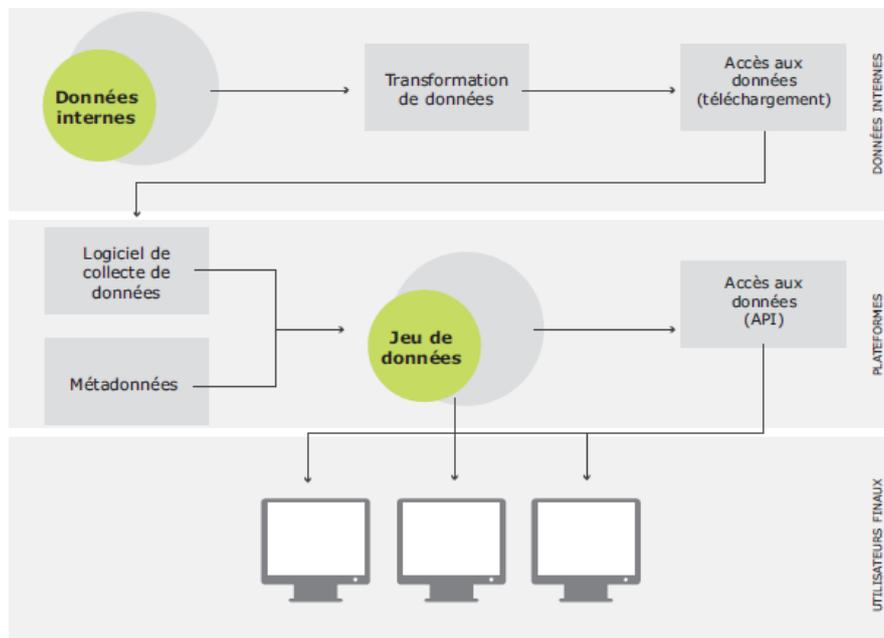


Figure 16. Capture d'écran. Open data Corsica. État de l'art

³⁶ Il faut également tenir compte de la proportion de la population qui ont des compétences de base en informatique (plus précisément savent se servir d'un ordinateur), le niveau de « culture numérique » - *digital literacy*), et de l'existence ou non d'une communauté de développeurs. Internet Users by Country (2016), <https://www.internetlivestats.com/internet-users-by-country/>

2.4. Stratégie de publication

La stratégie de publication doit porter essentiellement sur trois aspects : licence, format de données, et le canal de diffusion.

❖ Licence

La question de la licence est cruciale pour l'encadrement juridique de la réutilisation. L'Open Data ne se limite pas à la mise à disposition, dans l'unique but d'informer le public. Elle exige mais également la libération de certains droits. La licence met l'utilisateur à l'abri de l'insécurité juridique, ce qui constitue une incitation à la réutilisation. Un jeu de données publié avec une licence non renseignée n'invite pas automatiquement l'internaute à le réutiliser. Il l'oblige à contacter le producteur pour s'informer de son statut. Actuellement, il existe trois grandes catégories de licences :

1. Les licences Creative Commons

Créées en 2002 pour faciliter la diffusion de contenus numériques, elles s'articulent autour de quatre (4) clauses : la paternité, la modification, l'utilisation commerciale et le partage. Au départ, elles ne concernaient pas les bases de données, qui en Europe relèvent de droits spécifiques (sui generis) en non des seuls critères de droits d'auteur. Depuis 2013, elles les prennent en compte. Dans le cadre de l'Open Data, l'association Creative Commons recommande de ne pas utiliser certaines de ses clauses – NC (pas d'utilisation commerciale), ND (pas de modification) – afin de ne pas compromettre la réutilisation. Elle conseille de recourir à ces trois licences ouvertes :

- a)  *CC0 1.0 universel (CC0 1.0)*, qui équivaut à un transfert des données exposées dans le domaine public.

<https://creativecommons.org/publicdomain/zero/1.0/deed.fr>

- b)  *Attribution 4.0 International (CC BY 4.0)*, qui fait obligation de créditer l'œuvre, c'est-à-dire d'indiquer la provenance des données (Paternité).

<https://creativecommons.org/licenses/by/4.0/deed.fr>

- c)  *Attribution - Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0)*, fait obligation de créditer l'œuvre (Paternité) et de partager à l'identique, autrement avec la même licence.

<https://creativecommons.org/licenses/by-sa/4.0/deed.fr>

2. Les licences gouvernementales

Certains gouvernements préfèrent développer leurs propres licences, compatibles avec d'autres licences standards. L'objectif étant de les rendre le plus simple et le plus explicite possible. Il en est par exemple le cas :

- a) De la France avec la *Licence Ouverte/Open Licence*, <https://www.etalab.gouv.fr/licence-ouverte-open-licence>;
- b) De la Grande Bretagne avec l'*Open Government Licence (OGL)*, <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>;
- c) De l'Inde, avec la *Government Open Data License, (GODL)*, <https://data.gov.in/government-open-data-license-india>.

3. Open Data Commons Open Database License (ODbL)

Cette licence développée par l'Open Knowledge Foundation (OFKN) concerne spécialement sur les bases de données. Elle prend en considération la question la spécificité de leur protection juridique, qui dans certains contextes, distingue le contenant du contenu. En effet la licence qui régit la base (comme réservoir informatisé avec un « schéma ») peut éventuellement ne pas couvrir le contenu (supports visuels et audiovisuels, audio, ou autres). Si tel en est le cas, les concédants doivent indiquer les droits applicables pour les différents éléments contenus.

L'ODbL (<https://opendatacommons.org/licenses/odbl/>) constitue un standard international, et utilisée dans de nombreux pays, comme la France qui en fait une licence exclusive à côté de sa propre licence.

❖ *Formats*

Le choix du format de publication des données est un sujet important dans le cadre de la stratégie. En effet, les administrations gèrent des données de nature diverse (textuelles, géospatiales, audio, audiovisuelles...), et stockées dans différents formats de fichier. La question des formats concerne également la structuration des jeux de données (tabulaires, hiérarchiques...) et la recherche d'une plus grande interopérabilité (formats ouverts). L'enjeu du choix du format réside dans la facilité de réutilisation des données publiées qui doivent être facilement lisibles par un ordinateur. Les formats textuels sont fortement recommandés, et il est conseillé de ne pas proposer de fichiers binaires. Il est également préférable de privilégier des formats ouverts, des formats de fichier qui ne sont pas liés à un écosystème logiciel spécifique.

On peut noter une variété de formats de fichiers sur les plateformes. Toutefois, l'on admet que les fichiers *CSV* (pour les **données tabulaires**) et *JSON* (pour les **données hiérarchiques**³⁷) offrent un haut degré d'interopérabilité. Le *CSV*, qui utilise la norme de codage de caractères *UTF-8*, se présente de manière simple dans sa forme et dans sa syntaxe. Dans un fichier au format *csv*, la virgule (ou le point-virgule) généralement sépare les colonnes, et la première ligne contient les noms des champs. Aussi est-il aisément interprétable par une machine. Cependant, « ...le *CSV* fourni en France est détourné, dans le sens où le séparateur est le point-virgule, ce qui permet de garder la virgule dans les nombres décimaux. Or, la norme *RFC 4180* de l'*IANA* (*Internet Assigned Numbers Authority*) d'origine américaine, définit la virgule comme séparateur de colonnes (le point est utilisé comme séparateur décimal dans les pays anglo-saxons) » [16, Gouigoux, p. 546].

Par ailleurs, lorsqu'il s'agit de **formats d'API**, *JSON* et *XML* sont les plus courants. Ces formats dispensent du téléchargement manuel des fichiers. Ils permettent de rechercher et de récupérer automatiquement les données. Cependant, *JSON* est recommandé parce qu'il offre un plus haut degré d'interopérabilité s'agissant d'API. Il est utile de préciser que l'API est surtout intéressante pour des données volumineuses et fréquemment mises à jour. En est-il le cas pour les données ouvertes de l'éducation ? La réponse est à la discrétion des décideurs en fonction des ambitions du projet. Enfin, pour les données relatives aux **coordonnées géospatiales**, il existe des formats ouverts tels *GeoJSON* (reposant sur la norme *JSON*) ou *KML* (reposant sur le formalisme *XML*) et des standards comme *SHP* (développée par *ESRI*).

³⁷ Les données hiérarchiques, qui exigent un graphe en arbre avec une racine, ne peuvent pas être présentées sous forme tabulaire avec leur structure trop complexe. Les feuilles de calcul ne permettent pas d'établir de relations entre les données. Ainsi les formats de fichiers plats ne conviennent pas aux jeux de données à structuration hiérarchique.

En général, pour les données de l'éducation exposées sur les plateformes de données ouvertes, les formats de fichiers plats (sans arborescence) sont dominants. Cependant, puisque certains jeux de données renseignent également sur la localisation des établissements scolaires, il est préférable de s'intéresser également aux formats de coordonnées géospatiales. Toutefois, ne serait-ce qu'au début de l'implémentation du projet, il est facile de se concentrer sur le format de fichier CSV, tout en offrant autant que possible la possibilité de téléchargement dans différents formats ouverts. Cette possibilité est considérée comme la preuve, qu'en amont, il y a eu un véritable travail de formalisation et de qualification des données.

Comme les producteurs des données (tabulaires) utilisent des feuilles de calcul, les fichiers seront modifiés au format CSV. Cependant, une telle opération ne devrait pas empêcher leur publication, surtout quand il n'y a pas de collaborateurs dédiés à cette tâche. Aussi faut-il rappeler du modèle stellaire de T. Berners-Lee. Il est toujours possible dans un premier temps de se contenter des standards de faits. Les formats ne doivent pas constituer un frein à la publication des données comme précédemment indiqué³⁸. Par exemple les données de l'éducation sur les plateformes *data.gouv.fr*³⁹, et du ministère de l'Éducation nationale⁴⁰, sont pour une large part en format de fichier XLS. En effet, dans la vie réelle, le format ne va pas décourager un réutilisateur qui trouve un intérêt certain dans un jeu de données publié, même s'il s'agit d'un format Pdf.

❖ *Canal de diffusion*

Les Données Ouvertes sont pratiquement inconcevables sans publication sur le web via internet. Aussi, la stratégie de diffusion des données prendra en compte le canal à travers lequel elles seront publiées, rendues accessibles au public. Indépendamment du choix du mode de stockage des données ouvertes. Les informaticiens seront très impliqués dans les choix technologiques (catalogue de données, plateforme) puisqu'il s'agit de définir l'infrastructure technique qui doit répondre au mieux au projet⁴¹. Il existe trois (3) options s'agissant du canal de diffusion :

- ✓ Portail internet du ministère de l'Éducation : les données seraient disponibles en mode téléchargement. C'est une option possible qui dépend également du volume de données qu'on souhaiterait diffuser. C'est le choix effectué par exemple par le ministère de l'Éducation des Émirats Arabes Unis⁴².
- ✓ Plateforme ministérielle dédiée : le ministère se charge notamment de la diffusion de ses données, comme c'est le cas du ministère de l'Éducation en France⁴³.
- ✓ Plateforme interministérielle/publique qui est une des manifestations visibles d'une politique d'Open data portée au

³⁸ Le portail DoRANum (Données de la Recherche : Apprentissage Numérique) propose une intéressante liste des formats ouverts et fermés pour les différents fichiers :

https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf [consulté le 03 mai 2021]

³⁹ Voir, <https://www.data.gouv.fr/fr/datasets/?sort=-created&organization=534fff8fa3a7292c64a77f22> [consulté le 03 mai 2021]

⁴⁰ Voir, <https://data.education.gouv.fr/explore/?sort=modified> [consulté le 03 mai 2021]

⁴¹ La Banque mondiale offre une excellente revue synthétique des options technologiques à l'usage des informaticiens. Voir OpenData Toolkit, <http://opendatatoolkit.worldbank.org/fr/technology.html>.

En ce qui concerne les fonctionnalités, on peut aussi se référer à ce benchmark des logiciels/portails <https://teamopendata.org/t/benchmark-des-sofwares-portails-open-data/256> [consulté le 19 avril 2021].

⁴² United Arab Emirates, Ministry of Education, <https://www.moe.gov.ae/En/OpenData/pages/home.aspx> [consulté le 30 avril 2021]

⁴³ Data.gouv.education, <https://data.education.gouv.fr/pages/accueil/> [consulté le 30 avril 2021]

plus au sommet de l'État, le cas de *data.gouv* (France) l'Argentine⁴⁴ par exemple.

En ce qui concerne la diffusion, il est impératif de se poser la question des caractéristiques et des fonctionnalités susceptibles de faciliter une bonne expérience utilisateur (ergonomie), de créer et d'entretenir un public (communautés Open data).

La première option de publication des données ouvertes sur le portail du ministère (en téléchargement) n'est pas la meilleure option pour favoriser une visibilité optimale. Elle ne l'est pas non plus pour pouvoir disposer de fonctionnalités étendues offertes par une solution de catalogage de données ouvertes.

La plateforme ministérielle peut être une solution en l'absence de plateforme gouvernementale. L'exposition des données via une plateforme ministérielle parallèlement à une plateforme gouvernementale reste une exception en ce qui a trait aux données de l'éducation (exception faite du niveau supérieur). Il faudrait une réelle réflexion sur la valeur ajoutée d'une telle option qui doit prendre notamment en compte, la question du public cible et des objectifs. Dans le cas de la France, on se peut demander laquelle des deux plateformes (*data.gouv.fr* ou *data.education.gouv.fr*) est la plus visitée pour les données de l'éducation ?

Une plateforme nationale ou gouvernementale assurant la fonction de catalogue centrale, en moissonnant les données ouvertes des entités publiques (ministères et autres), représente un atout important. En termes d'infrastructures informatiques, de ressources, en termes d'accompagnement technique, de normalisation des jeux de données, de visibilité de la démarche Open Data.

Indépendamment du canal de diffusion retenu, dans le cadre du projet d'ouverture des données publiques, et a fortiori de l'éducation, il faut prendre en compte l'expérience utilisateur⁴⁵ :

- ✓ Recherche de jeux de données (système de navigation intuitive et ergonomique : étiquettes, moteur de recherche...)
- ✓ Présentation des jeux de données ;
- ✓ Possibilité de téléchargement ou d'exportation (sous différents formats de fichiers), API (pour les développeurs) ;
- ✓ Visualisation des données
- ✓ Dispositifs permettant de suivre l'actualité du projet.

Nous n'avons pas procédé à l'évaluation des solutions disponibles. Libres ou commerciales, elles sont nombreuses comme [CKAN](#), [Junar](#), [Opendatasoft](#), [Socrata](#)⁴⁶. Pour rester dans la philosophie de l'ouverture, il vaudrait mieux choisir un logiciel libre tel le CKAN qui est utilisé

⁴⁴ Datos Argentina, <https://datos.gob.ar/> [consulté le 30 avril 2021].

⁴⁵ Voir pour les critères détaillés (éditoriaux et techniques), OpenData Locale, Kit de ressources, Comment choisir un portail de publication des données publiques ? <https://opendatafrance.gitbook.io/odl-ressources/fiches-pratiques/premiers-pas/choisir-un-portail-open-data#comment-choisir-un-portail-de-publication-des-donnees-publiques> [consulté le 11 mai 2021].

On pourra également consulter, OpenData - 72 règles destinées aux producteurs de données ouvertes, <https://checklists.opquast.com/fr/assurance-qualite-web/download/> [consulté le 12 mai 2021].

⁴⁶ Voir la présentation de quelques solutions dans la boîte à outils de la Banque mondiale, <http://opendatatoolkit.worldbank.org/fr/technology.html>

Le projet Route-To-PA, soutenu par l'Union européenne, a également effectué une évaluation de quelques solutions disponibles. State-of-the-art Report and Evaluation of Existing Open Data Platforms, <https://project.routetopa.eu/deliverable-2-1-state-of-the-art-report-and-evaluation-of-existing-open-data-platforms-now-available/>

par nombre de portails gouvernementaux. Notons cependant que certains gouvernements ont fait le choix de produire leurs propres logiciels pour l'exposition de leurs données ouvertes. C'est le cas en France, avec *udata* développé dans par Etalab⁴⁷.

❖ *Scénario de publication des données*

En fonction de l'inventaire des données qui a été réalisé, du volume de données disponibles, des interventions nécessaires à leur qualification, et après analyse approfondie des données disponibles, il faut concevoir un scénario de publication. Deux options peuvent surgir à ce stade :

- ✓ **publication progressive** de petits volumes de données. Cette option permet de se concentrer sur un volume réduit de données afin d'effectuer un travail approfondi de qualification. Ce travail sur un nombre limité de jeux de données pourrait servir de phase expérimentale pour apprécier « in situ » les implications réelles du respect de tous les critères de qualification des données ouvertes. Ce scénario s'inscrirait plus ou moins dans l'approche agile.
- ✓ **publication massive** (histoire de frapper un grand coup) et procéder à des mises à jour régulières. Cette option exigera un temps de travail plus long avant le premier « lâcher de données », et une plus grande mobilisation des ressources. Du point de vue de la communication, l'effet pourrait être significatif. Mais il y a un risque de submersion des utilisateurs, et de trou d'air en attendant une prochaine publication.

La première option est souvent considérée comme étant plus réaliste et plus pragmatique. Surtout en l'absence d'une longue expérience d'ouverture des données, de ressources humaines limitées.

2.5. Préparation des données

La production de données au sein d'un ministère de l'Éducation a lieu le plus souvent dans le cadre d'activités propres aux unités et services. Donc avec des contraintes et objectifs parfois distincts. L'inventaire des données a normalement permis de prendre connaissance de l'éventail des données disponibles, de leur typologie, de leur origine (les données utilisées ne sont exclusivement recueillies par les services du ministère) ; de leurs formats.

Les ambitions et les objectifs permettront d'identifier les données potentiellement destinées à l'ouverture, et d'esquisser les jeux de données. S'agissant des données ouvertes de l'éducation, les jeux de données les plus courants exposés sur les différentes plateformes concernent :

- ✓ La liste des établissements (quelques fois avec géolocalisation, ou par statut : public ou privé ou conventionné)
- ✓ Le dénombrement des élèves (par groupe d'âge, sexe...)
- ✓ Les résultats scolaires par établissements ou subdivision géographique
- ✓ Les effectifs enseignants (répartition par sexe, statut, ...)
- ✓ Le personnel non-enseignant
- ✓ Les infrastructures et équipements scolaires
- ✓ Les dépenses et financement

⁴⁷ Etalab. Les logiciels libres développés par Etalab, <https://etalab.github.io/etalab/logiciels-libres.htm> [consulté le 13 mai 2021].

Il s'agit pour l'essentiel de données quantitatives. Néanmoins, on peut y trouver des données qualitatives (dans leur acception large par opposition à données quantitatives) ou les données textuelles⁴⁸. Tout dépend des ambitions du ministère.

En ce qui concerne les données quantitatives, elles sont en général issues du système d'information pour la gestion de l'éducation (SIGE). Le questionnaire d'enquête statistique, puis l'annuaire statistique de l'éducation sont de nature à faciliter l'identification des jeux de données⁴⁹. Les données du ministère ne sont pas ouvertes par défaut, il importe de les préparer à l'ouverture. En se référant constamment aux problématiques du format, de la licence, et des dispositifs légaux et réglementaires. À ce stade il est recommandé, si nécessaire, d'avoir des « référents données » pour les différentes unités et services du ministère impliqués dans la démarche.

Le référent informatique du projet envisagera la méthode d'extraction des données des bases sources, de leur transformation en données techniquement ouvertes. Il sera également chargé de la solution de stockage des données ouvertes.

❖ Structuration des jeux de données

Définir la structure des jeux de données (avec ses variables et ses valeurs) dépend des informations disponibles qui permettent de renseigner correctement. Comme le montre le jeu de données (ci-contre) sur les effectifs d'élèves et le nombre de classes⁵⁰.

Rentrée scolaire	Région académique	Académie	Département	Commune	Numéro de l'école	Dénomination principale
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0010093W	ECOLE PRIMAIRE PRIVEE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0010442A	ECOLE MATERNELLE PUBLIQUE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0010445D	ECOLE MATERNELLE PUBLIQUE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0011084Y	ECOLE PRIMAIRE PUBLIQUE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0011207G	ECOLE ELEMENTAIRE PUBLIQUE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0011253G	ECOLE MATERNELLE PUBLIQUE
2019	AUVERGNE-ET-RHONE-ALPES	LYON	AIN	AMBERIEU-EN-BUGEY	0011265V	ECOLE ELEMENTAIRE PUBLIQUE

Figure 17. Capture d'écran data.gouv.fr

Il est nécessaire également de penser à l'extraction des données présentant un intérêt mais qui ne sont pas structurées en jeux de données⁵¹. Au cours de cette étape, s'assurer que le jeu de données ne contient pas des données à caractère personnel – et le cas échéant de procéder à leur anonymisation afin de respecter le cadre légal. Ce traitement légal des données est indispensable. On veillera également à la qualité formelle des données surtout en ce qui concerne leur complétude, leur exactitude et leur cohérence. La description générale

⁴⁸ Par exemple sur data.gouv.fr :

- Sujets de philosophie du baccalauréat,

<https://www.data.gouv.fr/fr/datasets/sujets-de-philosophie-du-baccalaureat/> [consulté le 18/03/2021].

- Auteurs et autrices dans les annales du baccalauréat de français,

<https://www.data.gouv.fr/fr/datasets/auteurs-et-autrices-dans-les-annales-du-baccalaureat-de-francais/#> [consulté le 18/03/2021].

⁴⁹ Comme dans le cas de l'Afrique du Sud par exemple.

SNAP Survey. Ordinary school 2014,

<https://www.education.gov.za/Portals/0/Documents/Forms/Snap%20Ord%20Schools%202014.pdf?ver=2015-01-29-154613-467> [consulté le 29 avril 2021].

Education Statistics in South Africa 2016,

<https://www.education.gov.za/Portals/0/Documents/Publications/Education%20Statistic%20SA%202016.pdf?ver=2018-11-01-095102-947> [consulté le 29 avril 2021].

⁵⁰ Data.gouv.fr. Effectifs d'élèves par niveau et nombre de classes par école.

<https://www.data.gouv.fr/fr/datasets/effectifs-deleves-par-niveau-et-nombre-de-classes-par-ecole-date-dobservation-au-debut-du-mois-doctobre-chaque-annee/>

⁵¹ Nous ne nous arrêtons pas sur cet aspect technique dans notre travail.

s'avère importante pour bien contextualiser le jeu de données⁵². Les métadonnées sont essentielles à la description des jeux de données.

❖ Les métadonnées

L'ajout de métadonnées et leur renseignement constituent un autre aspect important du travail de formalisation des jeux de données avant leur publication. En fonction de l'outil et de la solution, les métadonnées sont fournies automatiquement. Cependant, il est possible de construire son propre modèle de métadonnées en se référant aux standards établis, et dans le respect des règles sémantiques. En outre, lorsqu'il existe une plateforme interministérielle, le schéma de métadonnées déjà bien établie, et faut se conformer à son modèle pour publier ses jeux de données. Cela facilitera le moissonnage de son catalogue⁵³. Indépendamment du choix effectué, il y a des métadonnées qui essentielles pour une bonne description de ses jeux de données :

- ✓ *Titre* du jeu de données ;
- ✓ *Sujet* (les mots-clés) du jeu de données ;
- ✓ *Description*
- ✓ *Créateur*
- ✓ *Diffuseur*
- ✓ *Licence*
- ✓ *Format*

En général, les métadonnées sont inspirées des quinze (15) éléments de description du Dublin Core simple⁵⁴. Selon les ambitions affichées notamment en termes d'interopérabilité de ses métadonnées, on peut se référer au schéma DCAT qui permet de décrire n'importe quel jeu de données exposées sur le web⁵⁵. Ce schéma repose sur six (6) grandes classes [43. Data Catalog Vocabulary] :

- ✓ dcat : Catalog
- ✓ dcat : Resource
- ✓ dcat : Dataset
- ✓ dcat : Distribution
- ✓ dcat : DataService
- ✓ dcat : CatalogRecord.

DCAT implique également le recours au modèle de données partagé (SKOS – Concept et ConceptScheme – à l'ontologie RDF – foaf : Agent). Comme pour les données, le format d'exportation des métadonnées requiert les compétences d'un informaticien surtout s'il est question d'avoir un schéma interprétable par la machine.

Le travail de préparation des données ouvertes nécessitera un investissement temporel conséquent des techniciens dédiés à cette tâche. Ce processus doit normalement favoriser une montée en compétences au niveau des ressources humaines. Et peut-être, en tout cas il est souhaitable, cette étape a permis une co-construction des données, et une meilleure fluidification de la circulation de l'information au niveau interne. Le bénéfice de l'Open data est également mesuré à cette aune.

⁵² Voir par exemple la description du jeu de données, *Formations initiales en France*. data.education.gouv.fr, <https://data.education.gouv.fr/explore/dataset/fr-en-formations-initiales-en-france/information/> [Consulté le 2 mai 2021].

⁵³ Data.gouv propose automatiquement un module pour décrire ses jeux de données. Voir, <https://guides.etalab.gouv.fr/data.gouv.fr/publier-jeu-de-donnees/#definir-qui-publie-le-jeu-de-donnees>

⁵⁴ Voir sur le site de la BNF, Dublin Core, <https://www.bnf.fr/fr/dublin-core> [consulté le 29/04/2021].

⁵⁵ Data Catalog Vocabulary (version 2), <https://www.w3.org/TR/vocab-dcat/>

Ce travail de qualification achevé, les données sont prêtes à la publication. La question de la qualité des jeux de données est cruciale, comme on peut le remarquer dans la figure ci-dessous [8, European data portal].

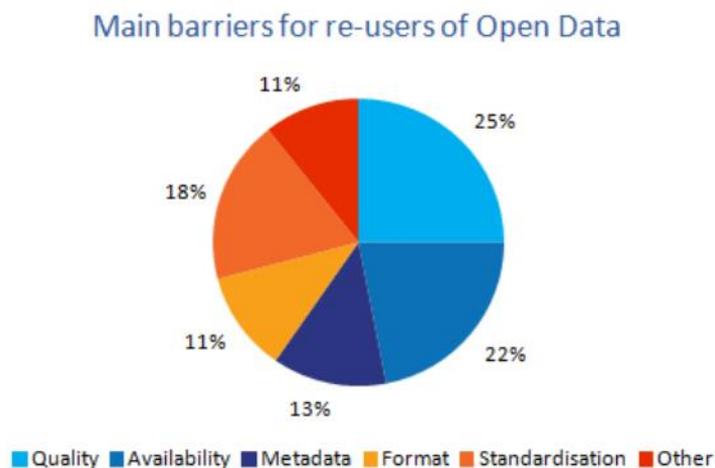


Figure 18. Capture d'écran. Freins à l'Open data du point de vue des réutilisateurs

Les retours des utilisateurs donneront l'occasion d'y revenir afin d'améliorer la qualité de ses jeux de données. D'où la nécessité de créer une passerelle d'échange avec les utilisateurs.

2.6. Accompagner les données publiées

S'engager dans une démarche d'ouverture des données représente un coût humain et matériel significatif pour une organisation. Aussi, il est indispensable de prendre les dispositions appropriées pour la réussite du projet. Il ne s'arrête pas à la publication en ligne des données. Même si l'exposition de ses données sur une plateforme constitue un point d'étape important. Puisqu'il s'agit notamment de créer un nouveau canal d'échange entre gouvernants et gouvernés, l'Open Data est un projet de communication. Il faut penser à une stratégie d'accompagnement de ses données, à la meilleure manière de les amener au public, et à la consolidation des interactions entre le public et le producteur. La stratégie d'accompagnement des données reposera sur plusieurs piliers :

- ✓ Le **portail** d'exposition, la porte accès aux données reste un site web. Le design doit donc être pris en compte. La plupart des solutions offrent la possibilité de visualisation immédiate ou de prévisualisation qui dispense éventuellement pourrait dispenser du téléchargement⁵⁶. Cela permet d'éviter de s'étendre sur la question de savoir s'il faut publier les données à l'état « brut » facilement accessible au public professionnel, ou procéder à des choix visuels pour le grand public qui n'a pas de compétences informatiques ;
- ✓ La **datavisualisation** pourrait être envisagée comme forme de médiation (hommes/données) pour les jeux de données à fort intérêt pour la communauté éducative ;
- ✓ Le renseignement des éléments de données portant sur la **mise à jour**, la mise à jour régulière des données ;

⁵⁶ data.education.gouv.fr, Etablissements fermés
<https://data.education.gouv.fr/explore/dataset/fr-en-etablissements-fermes/table/>

- ✓ Une **présentation d'introduction** des jeux de données, qui souligne éventuellement leurs intérêts et leurs limites, les finalités pour lesquelles les données ont été collectées.

Ces différents éléments concernent essentiellement l'éditorialisation du canal de diffusion. Ils ne dispensent pas d'un véritable travail de communication et de promotion de la plateforme. C'est-à-dire l'organisation d'événements à destination du grand public, de la communauté éducative, et des potentiels développeurs ou réutilisateurs professionnels (hackathon par exemple).

Même si les **réutilisations** des données publiques de l'éducation demeurent marginales en comparaison du volume de données publiées, il est malgré tout nécessaire de penser à une stratégie de valorisation et d'incitation. Tout en étant conscient que la modestie est une qualité à cultiver si l'on se réfère aux réutilisations connues des données publiées en Open Data. Leur quantité est suffisamment chétive en comparaison du volume de données ouvertes disponibles. Selon les statistiques de data.gouv, sur les 312 jeux de données de l'éducation, 289 n'ont jamais fait l'objet de réutilisation.

Néanmoins, il ne faudrait pas oublier que réutilisation n'implique pas forcément le développement d'une application par des développeurs, ou d'innovation à caractère économique ou de services marchands. C'est une question qui mérite d'être prise en compte dès le début du projet, car elle est en partie liée à l'intérêt des données pour les éventuels utilisateurs finaux ou intermédiaires. La stratégie de réutilisation, doit être considérée en amont de la publication, dans la phase préliminaire, lors des échanges avec la communauté éducative et le public en général. Ces échanges permettront d'évaluer les attentes et, a priori, d'identifier les données susceptibles de susciter l'intérêt des utilisateurs.

Dans l'accompagnement des données publiées, la stratégie de réutilisation nécessitera des actions diverses. Il s'agira par exemple de :

- ✓ Référencer, ou de recenser les réutilisations afin de leur assurer une visibilité sur la plateforme. On peut également y offrir la possibilité de contacter le développeur ou le producteur ;
- ✓ Créer un cadre d'échange permanent entre le ministère et les développeurs, pour comprendre les attentes, améliorer la qualité des données, recueillir des propositions ;
- ✓ Mettre les développeurs et les réutilisateurs professionnels en relation avec la communauté éducative. Les réutilisations de données ouvertes de l'éducation sont variées, allant, entre autres, de l'aide au choix d'établissement, au classement, à la vigilance sur l'impact du financement de l'éducation sur les établissements. Les données qui intéresseraient éventuellement les parents soucieux de la réussite de leurs enfants, ne sont pas forcément les mêmes qui intéresseraient les syndicats d'enseignants par exemple. Ces perspectives différentes (utilitaire, éthique, militante et autres) peuvent contribuer à faire germer des idées de réutilisations qu'un développeur seul face à ses données ne trouverait pas forcément ;
- ✓ Encourager, promouvoir l'utilisation des données exposées à des fins pédagogiques (travaux pratiques ou autres). Une manière d'assurer la promotion du portail ou de susciter des vocations, des idées de réutilisation à des fins diverses.

Comme cela a déjà été signalé, l'Open Data de l'administration est une entreprise de communication bidirectionnelle, dynamique, entre l'administration et le public en général. La

réussite du projet nécessitera de l'agilité, et une grande capacité d'écoute, une attention soutenue aux retours des utilisateurs. Sans ignorer la question fondamentale de l'engagement politique constant. La réussite de la stratégie en dépend fortement. Sans cet engagement, il est quasiment impossible de surmonter les freins relatifs à la mise en œuvre d'une démarche Open Data.

Conclusion

Ce travail a permis d'explorer la question de l'Open data, particulièrement de l'éducation. Nous avons suivi son développement, et abordé ses enjeux. L'Open data est confrontée à une préoccupation majeure en ce qui concerne la réutilisation. En effet, les réutilisations identifiées par les équipes des plateformes, les producteurs de données ouvertes, demeurent infimes relativement au volume de données diffusées, plus particulièrement pour ce qui a trait à l'éducation. Dymytrova et Paquienséguy ont bien souligné que rendre « *les données disponibles ne suffit pas pour générer la réutilisation et ce n'est sans doute même pas la première étape, car c'est un processus complexe qui concerne plusieurs communautés professionnelles ayant différentes finalités stratégiques, économiques et éthiques* » [44, Dymytrova, Paquienséguy]. Pour faire face à cette situation, les principaux acteurs impliqués agissent sur la qualité des données, leur découvrabilité, l'éditorialisation des portails, et s'efforcent de trouver les formes de médiation les plus pertinentes. L'essence relationnelle de l'Open data confère toute sa centralité à la question du public des données. Les producteurs de données ouvertes n'y sont les seuls acteurs, ils ne peuvent à eux seuls en garantir la réussite. Aussi, concevoir l'Open data comme une simple opération de communication politique revient à se fourvoyer. Il s'agit de se lancer dans une navigation au long cours, dans une entreprise de longue haleine. Une assistance technique de l'IIPE au profit d'un État membre devrait fortement mettre l'accent sur cet aspect fondamental et sur la nécessité de chercher constamment la réponse adéquate à la question : quelles données, pour quel public ?

Parce que les professionnels de l'information sont régulièrement interpellés par la question de la valorisation des fonds documentaires, de la qualité du service documentaire, **la bibliothèque de l'IIPE a des arguments** à faire valoir dans une éventuelle assistance à l'ouverture des données. **Notamment en ce qui concerne la gestion et l'éditorialisation de portail, la gestion de base de données, l'indexation, le catalogage, la structuration des jeux de données, la gestion des documents d'activité, et la médiation de l'information.** Assurer ce positionnement exigera une certaine montée en compétences. Concrètement le personnel de la bibliothèque devra faire preuve d'une compréhension manifeste de la philosophie qui anime le mouvement Open data et de la place des données de l'éducation. De même, il devra comprendre les étapes clés d'une démarche d'ouverture des données. Son implication dans l'assistance technique passera également par la capacité à démontrer l'importance de l'unité en charge de l'information et de la documentation du ministère dans la réussite d'un tel projet.

Dans le secteur de l'éducation, l'Open data trouve deux traductions concrètes avec des objectifs différents. D'un côté, les *données ouvertes dans l'éducation*, c'est-à-dire le recours aux données dans la formation, le développement de ressources pédagogiques pour améliorer la qualité de l'enseignement. Les exemples de données suggérées par le G 8 dans la catégorie *Science and Research* de sa charte correspondent plutôt à cette approche. Elle sied bien à l'enseignement supérieur, universitaire. De l'autre côté, les données ouvertes de l'éducation, pensées essentiellement sous la forme de mise à disposition d'informations statistiques sur le système scolaire à différentes échelles. C'est-à-dire au niveau d'un État, d'une région, d'une circonscription administrative, d'un établissement. Des données qui concernent l'éducation comme système administratif. Dans un autre secteur d'activité on aurait parlé de *tuyau* en faisant abstraction du *contenu* (contenu pédagogique). Cette approche est celle des différentes plateformes consultées. Parce qu'elles ont été pensées pour la diffusion de données quantitatives, plus aisément concevables comme « brutes ». N'y a-t-il pas lieu de pousser aussi loin possible la réflexion sur les données de l'éducation en Open data ? Autrement, quel public on risque de ne pas toucher (ou constituer) en se limitant strictement aux données quantitatives ou statistiques ? Un ministère de l'éducation ne produit-il que de l'information statistique qui puisse être distribuée en Open data ? Serait-ce impossible de

trouver d'autres informations d'intérêt pour la communauté éducative, notamment des enseignements pouvant leur être utile à l'accomplissement de leur mission. Admettons toutefois que ces interrogations peuvent faire sortir des sentiers de l'Open data notamment de l'éducation, et pousser par glissements successifs vers *l'Open content*. En effet, les pionniers de l'Open data la voyaient comme une réponse à l'exigence de transparence de l'administration, un moyen de renforcer la participation citoyenne dans la vie publique. Est-il possible d'aller plus loin en restant fidèle à l'idée de départ ? Les professionnels de l'information peuvent contribuer à cette réflexion qui fondamentalement renvoie à la médiation des données de l'éducation.

Bibliographie

Cette bibliographie analytique est présentée dans l'ordre linéaire du travail. Elle répond aux recommandations de l'INTD (norme NF ISO 690).

[1] BROUDOUX, Évelyne ; CHARTRON, Ghislaine. Big Data - Open Data : Quelles valeurs ? Quels enjeux ? *Actes du colloque « Document numérique et société » Rabat, 2015*. Louvain la Neuve, Paris : De Boeck Supérieur, 2015. 282 p. Collection « Information et stratégie ». ISBN 978-2-8073-0031-6.

Cet ouvrage confronte la question des données (data) aux enjeux géopolitiques, sociaux et économiques. L'économie numérique y est scrutée en s'interrogeant sur la création des valeurs.

[2] CARDON, Dominique. Culture numérique. Paris : Presses de Sciences Po, 2019. 428 p. Collection Les Petites Humanités. ISBN papier 978-2-7246-2365-9

Une approche de la question des TIC dans une perspective socio-historique. Il permet de saisir les enjeux et les débats des temps pionniers, et de comprendre différentes questions soulevées par la révolution numérique.

[3] Cahiers IP. Innovation & Prospective. [en ligne]. CNIL, n° 7. « Civic tech ; données, et Demos. Enjeux de données personnelles et libertés dans les relations entre démocratie, technologie et participation citoyenne ». Décembre 2019. [consulté le 10 novembre 2020]. < https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahier_ip7-web.pdf >.

Un regard synthétique qui montre comment les TIC peuvent contribuer à dynamiser la vie politique et l'engagement civique. Tout en invitant à ne pas considérer que la technique sera la panacée à la crise démocratique, et que l'essentiel doit être ramené au projet politique.

[4] CHIGNARD, Simon. Open data. Comprendre l'ouverture des données publiques CHIGNARD, Simon. Open data. Comprendre l'ouverture des données publiques. Editions Fyp, 2012. 191 p. Collection Entreprendre. ISBN 9789-2-916571-70-6.

Un travail très intéressant qui permet de comprendre les débuts du mouvement. Il explicite les contraintes et les bénéfices qu'on peut attendre d'une ouverture des données. Le travail d'un acteur de terrain bien engagé dans l'Open data, qui néanmoins en appelle à la vigilance citoyenne.

[5] GOËTA, Samuel. Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de l'open data. [en ligne]. Thèse de doctorat, Sociologie, Telecom Paris Tech, 2016. 257 p. [consulté le 10 novembre 2020] < <https://pastel.archives-ouvertes.fr/tel-01458098/document> >.

C'est une véritable approche ethnographique du travail souvent invisible pour produire les données ouvertes. La thèse analyse cette question au plus près du terrain. Elle présente une histoire très bien documentée des premiers moments du mouvement en faveur des données ouvertes. La question centrale des publics, au sens du philosophe John Dewey, est traitée avec rigueur et des exemples bien choisis.

[6] Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Open Knowledge Foundation [en ligne]. [consulté le 12 novembre 2020] < <https://opendefinition.org/od/1.1/fr/> >.

Il y est question de la définition de l'ouverture en relation avec une œuvre (de l'esprit).

7] Open Government data Working Group. Open Data readiness assesment methodology [en ligne]. The World Bank, 2015. 38 p. [consulté le 17 décembre 2020]
< http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_methodology-en.pdf >.

Il s'agit d'un cadre méthodologique d'évaluation de la préparation à l'ouverture des données. Elle a été déjà appliquée dans différents contextes.

[8] European Data Portal. Analytical Report 5 : Barriers in working with Open Data [en ligne]. European Commission. Luxembourg : Publications Office of the European Union, 2020. 37 p. [consulté le 15 décembre 2020]
< https://data.europa.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf >.

Ce rapport présente les obstacles auxquels les projets Open data font face au sein de l'Union européenne.

[9] ATENAS, Javier ; HAVERMANN, Leo. Education (Open Data sector and Communities). In DAVIES, Tim, et al. The State of Open Data : Histories and Horizons [en ligne]. Cape Town and Ottawa : African Minds, International Development Research Centre. 2019. Pp. 91-102. [consulté le 24 septembre 2020].
< <https://www.idrc.ca/en/book/state-open-data-histories-and-horizons> >.

La contribution analyse les différents usages et approches de l'Open data dans le secteur de l'éducation.

[10] LEHMANS, Anne ; CAPELLE, Camille. Gouvernance et horizon d'attente des données ouvertes pour l'éducation. In XIIe Colloque International. Réseau de recherche européen et pluridisciplinaire sur les enjeux et usages des TIC Logiques de réseaux et nouvelles gouvernances. [en ligne] Zakynthos (Grèce), 2016. [consulté le 10 octobre 2020].
< <https://hal.archives-ouvertes.fr/hal-01790492/document> >.

L'article analyse différents enjeux de l'Open data et s'intéresse à l'usage des données ouvertes dans les projets de formation en vue de garantir l'acquisition d'une culture de la donnée.

[11] KOBER, Vincent. Open data : ouverture, exploitation, valorisation des données publiques. 2^e édition. Voiron : Territorial éditions, 2017. 134 p. Collection Les essentiels. ISBN 978-2-8186-1302-3

Un travail pédagogique qui s'adresse surtout aux porteurs de projet Open data dans les Collectivités territoriales.

[12] Open Government Partnershp. National Action Plan on building of the Open Government in the Kyrgyz Republic for 2018-2020. Kyrgyz Republic. [en ligne]. 2019. [consulté le 16 novembre 2020].
< https://www.opengovpartnership.org/wp-content/uploads/2018/12/Kyrgyz-Republic_Action-Plan_2018-2020_REV.pdf >.

Il s'agit du plan d'action dans le cadre du Partenariat pour le gouvernement ouvert. Dans le cadre de ce plan d'action il a été prévu l'ouverture des données de l'éducation.

[13] Center for Open Data Enterprise. Open data map impact. Open Data Use Across Sectors. [consulté le 17 novembre 2020]
< <https://opendataimpactmap.org/sectors> >.

Un document fort intéressant sur la réutilisation des données ouvertes par secteurs (environ 15) et dans le monde réparti en huit régions.

[14] World Wide Web Foundation. Open Data Barometer. [en ligne] 4^e édition. 2017. 36 p. [consulté le 29 octobre 2020].
< <https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport-FR.pdf> >.

Le baromètre qui permet d'apprécier l'évolution de l'Open data.

[15] Journal Officiel de la République française. Ordonnance no 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques. Ministère de la Justice. [en ligne]. 2005. [consulté le 14 octobre 2020]
<
https://www.legifrance.gouv.fr/download/file/EzrW4lhD7FPW5AsxkTVC0yBTenN2gKVrvqLBncL23CaI=/JOE_TEXTE >.

[16] GOUIGOUX, Jean-Philippe. Open Data : consommation, traitement, analyse et visualisation de la donnée publique. Paris : Édition ENI, 2016. 583 p. Collection DataPro. ISBN 978-2-409-00220-5

Ce travail est un vade-mecum à l'usage de ceux qui souhaitent se lancer dans l'exploitation des données ouvertes. Un vrai guide pratique qui développe une réelle approche pédagogique pour la prise en main des logiciels. La mise en perspective et la réflexion stratégique sur l'Open data ne sont pas négligées pour autant.

[17] MADAY, Charlotte. Qualité des données et Open data. I2D - Information, données & documents. [en ligne]. 2016, volume 53, n° 4, p. 58-58. [consulté le 12 octobre 2020].
< <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2016-4-page-58.htm> >.

Ce court article met l'accent sur la qualité des données comme élément fondamental pour la réutilisation. La question de la gestion des données ne saurait être ignorée.

[18] Data.gob.es. Manuel práctico para mejorar la calidad de los datos abiertos. [en ligne]. 2017. 35 p. [consulté le 17 octobre 2020].
< https://datos.gob.es/sites/default/files/doc/file/manual_practico_para_mejorar_la_calidad_de_los_datos_abiertos_1_0.pdf >.

Un bon outil pour accompagner dans la démarche d'ouverture des données publiques.

[19] MESZAROS, Branislav, et al. Livre blanc sur les données ouvertes. Institut des sciences de l'homme. [en ligne]. Lyon : 2015. 68 p. [consulté le 11 novembre 2020]
< <https://halshs.archives-ouvertes.fr/halshs-01162692/document> >.

Les auteurs analysent les évolutions en cours concernant l'Open data. Ils fournissent des pistes de réflexion intéressantes pour bien prendre en main un projet d'ouverture des données publiques.

[20] Guide pratique de l'Open data. Data. Offre et qualité des données. The World Bank. [en ligne]. 2019. [consulté le 16 septembre 2020]
< <http://opendatatoolkit.worldbank.org/fr/supply.html> >.

Il s'agit d'une contribution importante de la Banque mondiale qui aborde plusieurs éléments essentiels d'une stratégie d'Open data. Les gestionnaires de l'information et les informaticiens pourront y trouver beaucoup d'informations.

[21] CAPLAN, Robin, et al (eds). Towards common methods for assessing open data: workshop report & draft framework. May 8th and 9th 2014, New York University. World Wide Web Foundation & Gov Lab [en ligne]. June 2014. 15 p. [consulté le 14 décembre 2020].
< <http://opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf> >.

Ce rapport reprend les différents échanges qui ont eu lieu concernant la définition d'une méthodologie commune d'évaluation des projets d'Open data. Il présente une revue synthétique des contributions de différents acteurs. Il propose une approche synthétique pour élaborer des indicateurs d'évaluation.

[22] IIEP. Rapport financier et états financiers vérifiés 2019. [en ligne]. Paris : IIEP, 2020. 72 p. [consulté le 16 janvier 2021]
< http://www.iiep.unesco.org/sites/default/files/ca_2019_iiep_fr_pub.pdf >.

[23] IIEP. Textes de base de l'IIEP. [en ligne]. Paris : IIEP, 2018. Huitième Édition. 68 p. [consulté le 9 janvier 2021].
< https://unesdoc.unesco.org/ark:/48223/pf0000372850_fre?posInSet=2&queryId=19d6ae97-8e32-49d1-bbfa-885b246ca07a >.

Ce document comprend différents textes qui ont accompagné l'évolution institutionnelle de l'IIEP depuis sa création.

[24] CUNEGATTI, Agnès. Recueil et analyse des besoins pour un site intranet. La re-création de l'intranet de l'IIEP-UNESCO. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 2016. 160 p.

Ce mémoire présente une analyse des besoins de l'IIEP en termes de fluidification de l'information interne, et explore les fonctionnalités requises pour le déploiement éventuel d'une nouvelle solution intranet.

[25] SOBRAL BOYER, Maria. Etude de solutions dans le monde du logiciel libre pour réinformatiser un centre de documentation tourné vers les pays en développement. Cas du centre de documentation de l'IIEP-UNESCO. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 2010. 145 p.

L'étude est une exploration des solutions pouvant permettre à la bibliothèque de répondre à ses besoins et son souhait d'optimiser ses prestations en tenant compte de l'évolution des technologies du web (Web 2.0).

[26] BERGONIER SIGNOLES, Aurélie. Mise en œuvre d'une nouvelle structuration des données pour valoriser le fonds documentaire de l'IIEP UNESCO : l'expérience FRBR. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 2011. 153 p.

L'étude porte sur le modèle conceptuel FRBR (*Functional Requirements for Bibliographic Records*) qui permet le dialogue entre réservoirs de données. Il est axé sur le web des données et facilite l'indexation par les moteurs de recherche. Le mémoire essaye de voir comment à partir de ce modèle, proposer une amélioration de l'accès à l'information, et la valorisation du fonds documentaire de l'IIEP.

[27] POLEWKA, Anna. Étude des enjeux de la virtualisation des collections du centre de documentation de l'IPE-Unesco. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 2012. 110 p.

Il s'agit d'envisager l'amélioration l'offre documentaire en ligne de la bibliothèque de l'IPE. L'étude aborde le cas de la bibliothèque à partir du modèle de la une bibliothèque hybride.

[28] CHIKH, Sid-Ali. L'exploitation des vocabulaires dans un portail web sémantique Le cas du portail « Plan4Learning » de l'IPE-UNESCO. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 2014. 126 p

Le travail porte sur la valorisation des ressources d'informations dans le la cadre d'un projet de portail, en l'occurrence *Plan4Learning*, qui recourt aux technologies du web sémantique. Autrement expérimenter un référentiel standard qui permet de tirer avantage du web de données.

[29] IPE. Stratégie à moyen terme, 2014-2017. Planifier l'éducation, bâtir l'avenir. [en ligne]. Paris : IIEP, s, d. 19 p. [consulté le 13 janvier 2021].
< https://unesdoc.unesco.org/ark:/48223/pf0000227245_fre?posInSet=2&queryId=33c31fe4-e0b6-4d11-90eb-a8ab215b4759 >.

[30] Data Revolution Group. A World That Counts: Mobilising The Data Revolution for Sustainable Development. Report prepared a the request of the United Nations Secretary-General, by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. [en ligne] UN Data revolution, 2014. 32 p. [consulté le 27 janvier 2021].
< <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf> >

Comment faire pour mettre « la révolution des données » au service du développement durable. Le groupe d'expert analyse la situation et formule quelques recommandations.

[31] Unesco Institute for Statistics. The Data Revolution in education. Information paper n° 39, 2017 [en ligne]. Montreal: UNESCO Institute for Statistics, 2017. 81 p. ISBN 978-92-9189-213-6. [consulté le 5 octobre 2020].
< <http://uis.unesco.org/sites/default/files/documents/the-data-revolution-in-education-2017-en.pdf> >.

C'est un plaidoyer résolu en faveur d'un soutien au renforcement des services statistiques des pays qui en ont plus besoin. L'objectif est de renforcer leur capacité de suivi et d'évaluation dans le cadre de l'ODD 4.

[32] United Nations. Data Stratégie off the Secretary-General for Action by Everyone, Everywherewith Insight, Impact and Integrity 2020-22. External version 9.1, 2020. [en ligne]. s, l. [consulté le 21 janvier 2021]
< https://www.un.org/en/content/datastrategy/images/pdf/UN_SG_Data-Strategy.pdf >.

Il y a une présentation du cadre logique de la stratégie, et une proposition de plan d'action générale que les agences onusiennes doivent s'efforcer d'adapter à leur réalité sans s'écarter des objectifs.

[33] OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Éditions OCDE, Paris, 2007. 29 p.[consulté 6 février 2021].
< <https://www.oecd.org/fr/science/inno/38500823.pdf> >

Ce document expose les principes de l'OCDE concernant les données de la recherche.

[34] Commission européenne. HORIZON 2020 en bref. Le programme-cadre de l'UE pour la recherche et l'innovation. [en ligne]. Luxembourg : Office des publications de l'Union européenne, 2014. 40 p. [consulté le 3 février 2021].
< https://ec.europa.eu/programmes/horizon2020/sites/default/files/H2020_FR_KI0213413FRN.pdf >.

[35] CABRERA, Francisca. Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques Enquête exploratoire. Mémoire pour l'obtention du Titre professionnel "Chef de projet en ingénierie documentaire". CNAM-INTD, 214. 238 p.

Ce travail réalisé dans le cadre d'un stage à Cairn. info. Il interroge entre autres la question de la donnée de la recherche dans les Sciences humaines et sociales selon leurs approches méthodologiques distinctes.

[36] CHENG, Joy ; MOSES, Kurt. Promoting transparency through information : A global review of school report cards. [en ligne]. Paris : IIEP, 2016. 151 p. [Consulté le 23 août 2020].
< <https://unesdoc.unesco.org/ark:/48223/pf0000246358> >.

Une étude approfondie de 14 initiatives de tableaux de bord des écoles développées un peu partout dans le monde.

[37] POISSON, Muriel (dir.) ; DUPAIN, Jonathan ; NGUYEN, Lena. Information et transparence : Tableaux de bord des écoles en Afrique subsaharienne. [en ligne]. Paris : IIEP, 2018. 213 p. [consulté le 27 août 2020].
< <https://unesdoc.unesco.org/ark:/48223/pf0000261740> >.

Cette étude s'attache à évaluer l'impact des tableaux de bord dans différents contextes africains. Elle analyse les deux approches les plus courantes et fournit quelques recommandations.

[38] BRITO, Alejandra. Información y transparencia: Cuadros de indicadores de las escuelas en América Latina. [en ligne]. Paris : IIEP, 2019. 186 p. [consulté le 13 octobre 2020].
< <https://unesdoc.unesco.org/ark:/48223/pf0000370474> >.

La question des tableaux de bord des écoles est analysée dans le contexte de quelques pays ibéro-américains. Du Pérou à la République Dominicaine, et du Guatemala à la Colombie entre autres.

[39] POISSON, Muriel. Open School Data : What Planners Need to Know. Ethics and Corruption in Education [en ligne]. Paris : IIEP, 2021. 202 p. [consulté le 3 juillet 2021].
< <https://unesdoc.unesco.org/ark:/48223/pf0000376705> >

Cette publication propose aux planificateurs de l'éducation des éléments de réflexion approfondie sur l'ouverture des données sur les écoles. Elle fait également le tour des principaux aspects à prendre en compte. En dernier lieu, elle énumère les différentes étapes pour réussir une ouverture des données sur les écoles.

[40] READ, Lindsay ; ATINC, Tamar Manuelyan. Information for accountability, transparency an citizen engagement for improved service delivery in education systems. Global economy and development at Brooking. Working paper n° 99. [en ligne]. S, I. 52 p. [consulté le 12 novembre 2020].
< https://www.brookings.edu/wp-content/uploads/2017/01/global_20170125_information_for_accountability.pdf >.

Cette étude analyse les bénéfices contextuels de la diffusion de l'information en ce qui a trait à l'imputabilité et la transparence. Elle met l'accent sur d'autres paramètres relatifs aux engagements citoyens.

[41] DA SYLVA, Lyne. Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information. In Documentation et bibliothèques [en ligne]. 2017. Montréal, volume 63, n° 4, pp. 5-34. ISSN (numérique) 2291-8949.
< <https://www.erudit.org/fr/revues/documentation/2017-v63-n4-documentation03290/1042308ar.pdf> >.

C'est une revue synthétique, mais bien documentée, de la problématique des données sous leurs différentes formes. Elle analyse les défis que doivent surmonter les professionnels de l'information et les évolutions qu'ils imposent à la profession.

[42] Collectivité territoriale de Corse. Open data Corsica. Etat de l'art. [en ligne]. Workpackage 5 Projet Homer. 2015. [consulté le 13 mars 2021].
< https://www.opendata.corsica/downloads/Etat-de-l-art-et-guide-technique-OPENDATA_t18507.html >.

Il s'agit d'un état de l'art réalisé dans le cadre du projet Homer auquel participent 19 partenaires européens du pourtour méditerranéen. Il présente une évaluation des projets en cours tout en formulant des recommandations.

[43] W3C Recommendation. Data Catalog Vocabulary (version 2), <https://www.w3.org/TR/vocab-dcat/> [consulté le 9 juin 2021].

Il s'agit de la version en cours (en date du 4 février 2020) des recommandations du W3C relativement au modèle DCAT.

[44] DYMYTROVA, Valentyna ; PAQUIENSEGUY, Françoise. La réutilisation et les réutilisateurs des données ouvertes en France : une approche centrée sur les usagers. Revue internationale des Gouvernements Ouverts. [en ligne]. 2017, volume 5. [consulté le 13 juin 2021].
< <https://ojs.imodev.org/index.php/RIGO/article/view/204/338> >

Cet article aborde la question de la réutilisation des données ouvertes du côté des usagers, c'est-à-dire des professionnels de la réutilisation dans leur diversité de pratiques professionnelles.