



Commission économique pour l'Europe**Conférence des statisticiens européens****Soixante-troisième réunion plénière**

Genève, 15-17 juin 2015

Point 4 de l'ordre du jour provisoire

**Modernisation de la production et des services statistiques
et gestion à des fins d'efficacité****De l'importance accrue de la «massification» des données****Note du Bureau australien de statistique***Résumé*

Le présent document renferme une description de la stratégie de diffusion adoptée par le Bureau australien de statistique, qui repose sur l'idée que la présence sur le Web ne se définit plus par le site Web principal du Bureau, mais par la visibilité et la flexibilité du Bureau sur les différents canaux de diffusion. Il y est question de la signification de la «massification» des données et de la façon dont le Bureau australien de statistique et bien d'autres organisations tirent avantage de la possibilité de communiquer avec les bases de statistiques officielles pour extraire des données vers leurs applications.

Le présent document est soumis dans le contexte du deuxième séminaire sur la modernisation de la production et des services statistiques et la gestion à des fins d'efficacité organisé pendant la réunion plénière de la Conférence des statisticiens européens.



I. Contexte

1. Le Bureau australien de statistique a pour fonction légale de faciliter au maximum l'utilisation de données à l'appui de la prise de décisions éclairées. Il travaille dans un environnement caractérisé par une interconnexion de l'information, conduisant la politique statistique de l'Australie par la création de partenariats efficaces avec les principaux acteurs. Les changements et l'innovation en cours dans le domaine de l'information incitent les pouvoirs publics, les chercheurs, les acteurs économiques et la société en général à réclamer un accès aux données grâce à des outils rapides, souples et sûrs.

2. Le présent document décrit les grandes lignes d'une stratégie qui tend à améliorer l'accès aux statistiques grâce à la massification des données. Il explique comment le Bureau australien de statistique et d'autres organisations et individus tirent avantage de la possibilité de communiquer avec les bases de données du Bureau pour en extraire le contenu vers leurs applications, que ce soit à l'aide des services Web ou des interfaces de programmation (API).

II. Nos utilisateurs

3. Le Bureau australien de statistique a toujours considéré ses utilisateurs sous l'angle de la segmentation du marché, de l'écolier à l'analyste politique, en passant par le spécialiste des modèles économiques ou par le chercheur dans le secteur de la santé. Plus récemment, il s'est intéressé plus particulièrement au comportement de ses utilisateurs, plutôt qu'à l'organisation à laquelle ils appartiennent ou au rôle qu'ils y jouent. Cette démarche lui a permis de simplifier sa façon de produire ses contenus.

4. Pour mieux répondre aux besoins des utilisateurs, on peut les diviser en quatre grands groupes de comportements. Les utilisateurs peuvent dire une des phrases suivantes:

- Connectez-moi à un numéro, pour que je puisse rapidement en parler aux autres;
- Connectez-moi à un récit intéressant pour que j'apprenne quelque chose de nouveau. Je voudrais connaître une histoire, un outil interactif ou un jeu du type Census Spotlight (spotlight.abs.gov.au);
- Connectez-moi à un résumé, car je souhaite en apprendre davantage sur l'endroit où je vis ou sur les questions de santé les plus importantes d'aujourd'hui; ou encore
- Connectez-moi à des données détaillées, pour que je puisse analyser, réutiliser et associer ces données avec d'autres données, ce qui facilitera la prise de mes décisions ou de celles de mes utilisateurs.

III. L'économie numérique

5. Pendant les dernières décennies, les organisations se sont approvisionnées en contenu en s'appuyant sur des sites Web statiques, des navigateurs sur ordinateur et des systèmes monolithiques. Notre modèle traditionnel repose sur un site Web qui livre régulièrement des publications et des données statiques à intervalles de temps fixes et prévisibles. Les utilisateurs ont accès à des lots de données détaillés au format Excel, mais uniquement en les téléchargeant manuellement depuis le site Web.

6. L'efficacité de ce modèle a été perturbée par la rapidité des progrès technologiques. La complexité de ce bouleversement dans la puissance de traitement, les supports numériques et la communauté du numérique pose un certain nombre de difficultés. Nous savons plusieurs choses:

- La quantité de données s'accroît de façon exponentielle, mais les coûts liés au stockage et à la transmission des données diminuent dans des proportions énormes, ce qui pose des problèmes de gestion et de diffusion des données;
- La chaîne logistique a été transférée des fournisseurs de données aux agrégateurs ou à des intermédiaires tels que les nouvelles plates-formes médiatiques. Le problème est de savoir comment mobiliser l'attention du client ou assurer l'imputation des données (et donc l'assurance de la qualité) dès lors que la création et la diffusion de contenu ne se limitent plus aux circuits traditionnels;
- La possibilité d'interagir avec les nouveaux supports s'est développée et dépasse aujourd'hui les modèles rigides traditionnels. Les utilisateurs peuvent, à tout moment, accéder instantanément aux contenus, les partager, les commenter ou les solliciter;
- La vitesse est essentielle dans l'économie numérique. Chaque milliseconde coûte de l'argent. Chacun veut trouver l'information au bout de ses doigts, vite et facilement;
- La communauté des développeurs d'applications est un acteur important. Les nouvelles technologies du Web permettent de créer des applications rapidement et à peu de frais et de répondre à des besoins généraux ou spécifiques;
- Il est facile de faire interpréter rapidement et automatiquement l'information par les machines, qui se chargent en partie du fastidieux travail consistant à trouver et combiner l'information sur le Web et à lui donner suite;
- Les systèmes d'intelligence artificielle, qui sont de plus en plus utilisés dans l'Internet, permettent d'automatiser les décisions.

7. Nos principaux utilisateurs sont à la fois les acteurs et les récepteurs de la révolution numérique. Les administrations, les entreprises et les chercheurs ont besoin d'accéder plus rapidement et plus facilement à l'information. Ils souhaitent accéder à des données générales détaillées et bien décrites, ainsi qu'à des microdonnées anonymes. Ils souhaitent une intégration des données comprises ou non dans le champ des domaines traditionnels de la statistique et provenant de sources diverses, pour pouvoir traiter des problèmes complexes.

8. Ils ne veulent plus du modèle traditionnel. Le téléchargement manuel de fichiers à partir d'un site Web est inefficace, lent, n'apporte rien et freine l'innovation. Nos produits constituent fréquemment un apport majeur dans leurs activités économiques.

9. Certains utilisateurs souhaitent que nous leur fournissions des données basées sur des structures prédéfinies en concertation, dans des formats tels que l'Échange de données et de métadonnées statistiques (SDMX). Cette façon de procéder, efficace pour les deux parties, permet aux utilisateurs d'extraire nos informations ou d'y accéder instantanément. Le respect de normes pour les métadonnées et les données dans l'ensemble de la production statistique permettra aux utilisateurs d'accéder plus aisément à des sources de données multiples et cohérentes, quelles que soient les sources en question.

IV. Transformation des produits et communication

10. Le Bureau australien de statistique s'attache à développer une nouvelle approche de la diffusion en s'appuyant sur les principes suivants:

- Renforcer la cohérence en stockant les données multidimensionnelles les plus détaillées et les plus complètement décrites dans un entrepôt de données qui constitue, en quelque sorte, la «source unique de vérité»;

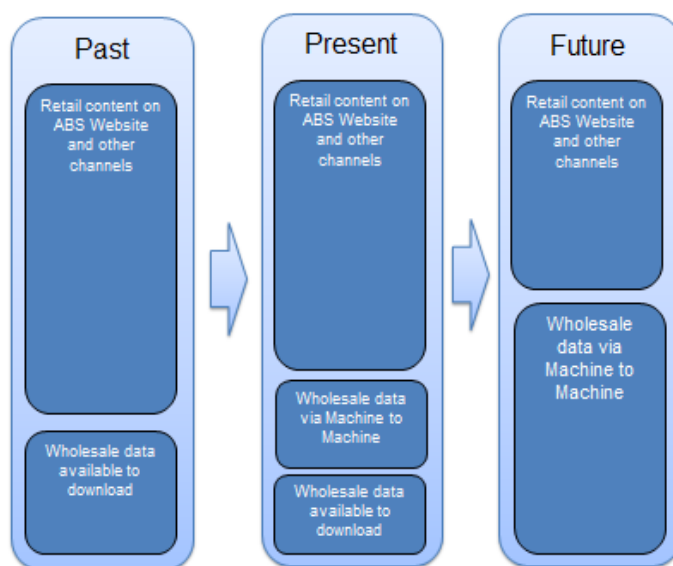
- Promouvoir et appliquer des méthodes normalisées pour les métadonnées, non seulement au sein de l'organisation, mais aussi dans tout le secteur de la statistique et de l'information;
- Collaborer avec les organisations similaires pour mettre en commun les méthodes et les outils;
- Constituer les nombreux types de produits potentiels différents à partir de la «source unique de vérité»;
- Compiler de façon dynamique les contenus de façon à répondre aux besoins des utilisateurs en fonction de leur niveau respectif de compétence statistique. Un contenu statique subsistera, mais de plus en plus, le contenu sera obtenu de façon dynamique à partir des données et des métadonnées détaillées sous-jacentes;
- Proposer différents modes d'affichage des données et des métadonnées grâce à une bibliothèque type d'outils géographiques, graphiques et visuels;
- Proposer un accès libre, actualisé et bon marché aux informations d'intérêt public;
- Permettre l'accès aux données, y compris à des volumes importants de données, par des services de machine à machine;
- Porter une attention particulière aux besoins des utilisateurs en leur proposant des statistiques de qualité, pertinentes, actualisées, faciles à consulter et à comprendre;
- Fournir aux utilisateurs un contenu adapté à tous les moyens de diffusion et à tous les dispositifs;
- Faciliter la réutilisation des données du Bureau australien de statistique par des utilisateurs externes et des portails tiers tout en assurant la visibilité du sigle ABS du Bureau et de sa propriété sur les données; et
- Faire en sorte que le mode de création du contenu nécessaire au développement de capacités statistiques soit visible en même temps que les données, sans être intrusif.

11. Une part importante de la réponse à apporter à la révolution numérique consiste dans la massification de nos données. Par massification, nous entendons le fait d'offrir l'accès à de grandes quantités de données et de contenus grâce à des solutions machine à machine susceptibles d'être réutilisées par d'autres pour dispenser des produits et des services à valeur ajoutée pour eux-mêmes ou pour leurs propres utilisateurs.

12. Cela ne signifie pas que le Bureau cessera de fournir des données, des applications et des contenus. Nous sommes en effet soumis à une obligation de service importante vis-à-vis de la société, et nous continuerons à lui fournir un éventail large et complet de données et de contenus résumés. Il existe une demande de solutions statistiques intégrées et fiables, et le Bureau australien de statistique dispose des moyens nécessaires pour répondre à cette demande.

13. En revanche, cela signifie que nous développerons de plus en plus de services machine à machine pour nous-mêmes et pour les utilisateurs, afin de pouvoir intégrer, agréger et réutiliser nos données les plus détaillées dans des formats et des solutions novatrices que le Bureau ne serait lui-même pas en mesure de proposer autrement. Cette façon de procéder permet aux utilisateurs d'adapter l'utilisation de nos données à leurs besoins propres, en particulier à leurs systèmes et à leur façon de travailler, de les utiliser de pair avec les outils d'analyse commerciale les plus répandus, et de les utiliser pour fournir leurs propres services à valeur ajoutée.

Figure 1
Transformation des produits du Bureau australien de statistique

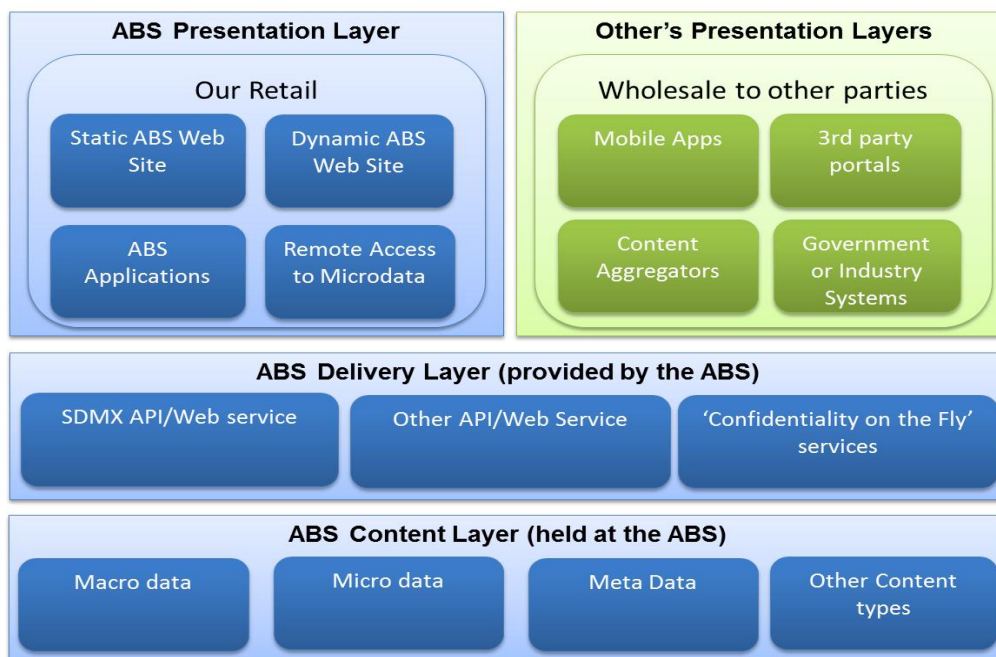


1

A. Une architecture flexible pour une publication dynamique

14. La démarche requiert une architecture (que ce soit au niveau du fonctionnement, des systèmes et de l'information) qui permette de délivrer et présenter les contenus en couches distinctes.

Figure 2
Le modèle en trois couches pour la massification des données



15. Le Bureau australien de statistique doit être en mesure de construire, répertorier, décrire, stocker, protéger et consulter les données de façon cohérente, quelle que soit la façon dont les utilisateurs choisiront par la suite de les utiliser et de les consommer. Au lieu de procéder à des publications traditionnelles réunissant de façon statique et monolithique un résumé des données, des données détaillées, des analyses et d'autres renseignements sur ces mêmes données, nous nous attacherons à réunir les éléments de contenu sous-jacents.

16. La couche fourniture («Delivery Layer» dans la figure) établit le lien entre le contenu et la présentation du contenu. Cette couche comporte deux aspects importants. Premièrement, des outils permettant aux utilisateurs et aux systèmes d'accéder au contenu à différents niveaux de granularité; deuxièmement, les services (API ou services Web), qui nous permettront de faire en sorte que notre contenu le plus détaillé puisse être transformé et rendu confidentiel avant d'être présenté à nos utilisateurs («confidentialité instantanée»). À condition de disposer des métadonnées appropriées, ces outils permettront de découper le contenu en tranches ou en cubes, puis de le combiner avec d'autres contenus dans les circuits de diffusion tiers.

17. La couche présentation («Presentation Layer» dans la figure) offre un moyen de répondre efficacement aux différents besoins des utilisateurs en présentant nos contenus de manières très diverses, dans différents formats et sur différents appareils. Elle comprend non seulement une présentation détaillée des statistiques et des contenus sur les circuits de diffusion du Bureau australien de statistique, mais aussi la présentation des mêmes statistiques et des mêmes données après massification au profit des tiers. Elle comprendra un nombre croissant d'applications générales ou spécifiques mises au point par nous-mêmes ou par d'autres intervenants.

V. Notre propre utilisation des données massifiées

18. L'application «données par région» (DbR, ou Data by Region) (www.abs.gov.au/databyregion) illustre l'utilisation que nous pouvons avoir de la massification des données.

19. Les utilisateurs souhaitent se renseigner sur l'endroit dans lequel ils vivent. Ils cherchent à comparer leur région avec les autres régions pour comprendre comment elle s'insère dans le cadre général de l'Australie. Depuis dix ans, le mécanisme national des profils régionaux établi par le Bureau australien de statistique est un des dispositifs qui permet d'obtenir cette information. Il regroupe les données régionales provenant d'ensembles de données multiples au sein d'une série chronologique unique et cohérente. Autrefois, nous procédions manuellement à la compilation et à la validation de milliers de pages Web régionales en vue d'établir les profils régionaux. Certes, cela évitait aux utilisateurs de devoir trouver et explorer eux-mêmes chaque ensemble de données, mais ce processus demandait des mois de travail et des ressources considérables pour une efficacité limitée. Une fois qu'elles étaient publiées, les données issues du profil national des régions n'étaient pas actualisées avant la publication annuelle suivante, alors que de nouvelles données sur les régions étaient régulièrement mises à la disposition du public au cours de l'année.

20. En juillet 2014, le Bureau australien de statistique a mis au point l'application DbR, qui utilise les trois couches pour apporter à nos utilisateurs un service innovant:

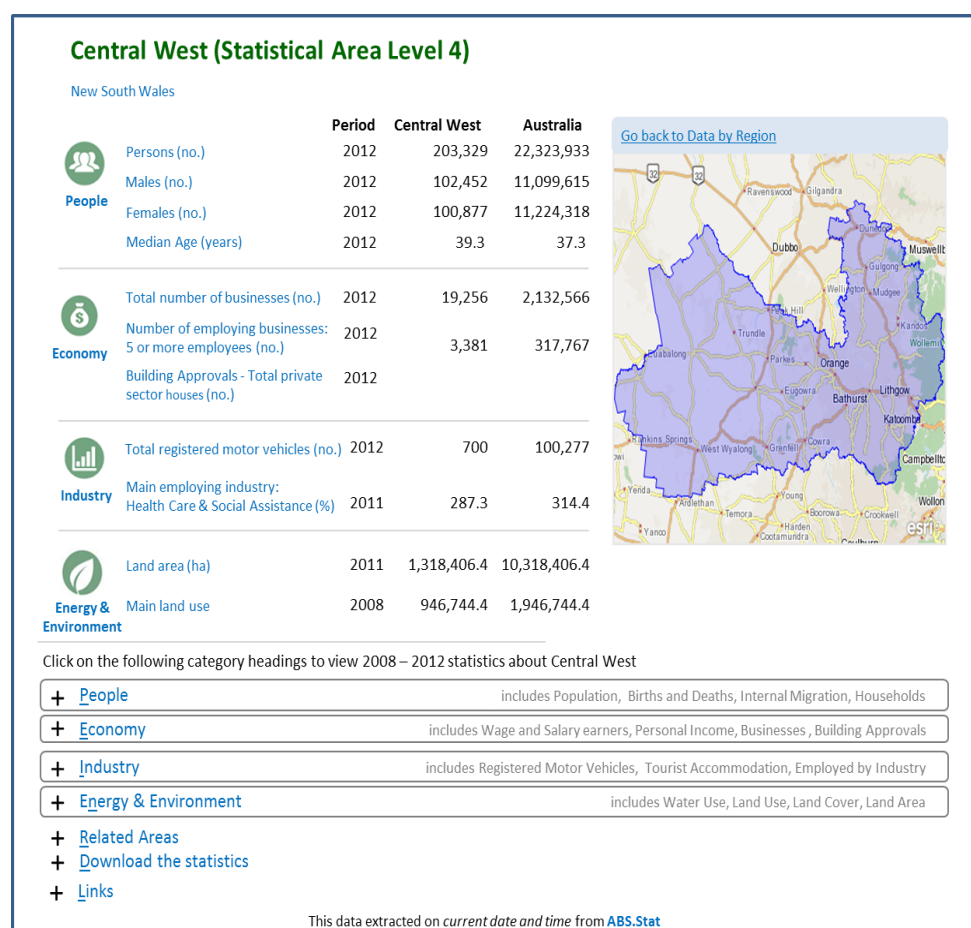
- Une couche contenu («Content Layer» dans la figure) fournie au moyen d'ensembles de données dans ABS.Stat (basé sur le produit OECD.Stat utilisé par un certain nombre d'organisations de statistiques officielles);

- Une couche fourniture fournie au moyen des services Web de ABS.Stat et un service Web spécialement élaboré par le Bureau australien de statistique; et
- Une couche présentation fournie au moyen de l'application DbR.

21. L'application DbR offre un point d'entrée unique pour l'accès aux statistiques par région du Bureau australien de statistique. Elle comprend un système de navigation intuitif basé sur un outil cartographique et propose les nouvelles statistiques à mesure qu'elles sont publiées. La principale révolution apportée par l'application DbR se trouve «sous le capot», c'est-à-dire dans la façon dont l'application appelle l'information et les données depuis un dépôt centralisé unique. Lorsque l'utilisateur sélectionne la région qu'il souhaite visualiser, les données sont extraites dynamiquement du dépôt. Nous n'avons plus besoin de construire manuellement les ensembles de données pour chaque région. Cette approche consiste effectivement à massifier nos propres données pour qu'elles puissent être consultées en détail dans nos propres applications. Elle nous aidera à élaborer une méthode viable et souple de diffusion de notre contenu.

Figure 3

Données sur le Centre-Ouest de la Nouvelle-Galles du Sud (application DbR)



VI. Autres utilisations des données que nous massifions

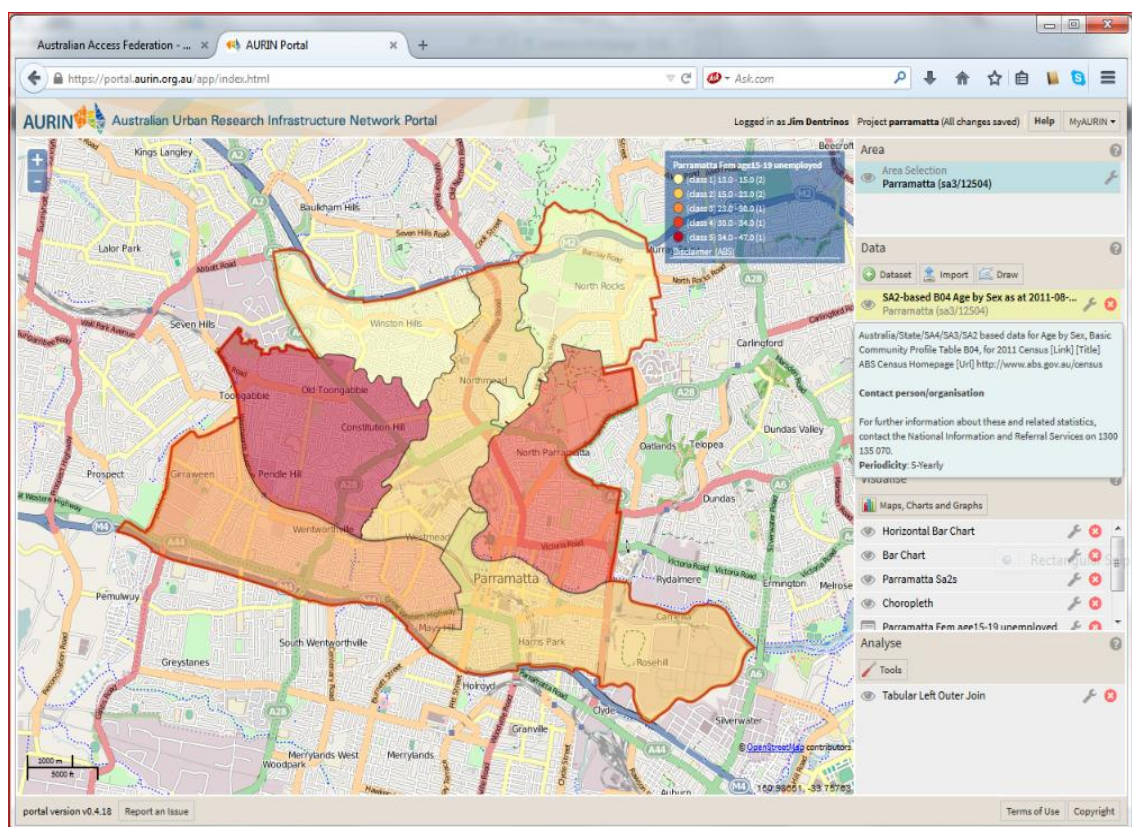
22. Les nouveaux services Web du Bureau australien de statistique permettent également à des utilisateurs externes d'accéder en temps réel à des données stockées en lieu sûr sur ses serveurs, de les intégrer à d'autres ensembles de données et de présenter les résultats dans leurs propres systèmes. On peut citer, par exemple, le Réseau australien d'infrastructures de recherches urbanistiques (le réseau AURIN), qui est utilisé dans toutes les universités d'Australie et au-delà (voir <http://aurin.org.au/>).

23. L'objectif du réseau AURIN est de promouvoir les capacités de recherche et de décision des agglomérations en développant une infrastructure informatique capable de piloter et d'anticiper l'innovation, les nouvelles méthodes et les nouvelles pratiques. À la fin du premier semestre de 2013, le réseau AURIN et le Bureau australien de statistique ont engagé un projet conjoint destiné à rendre les ensembles de données du Bureau accessibles en ligne et en temps réel au réseau au moyen de solutions machine à machine.

24. Le projet a permis de faire la démonstration de l'accès machine à machine aux résultats du recensement de 2011 par le biais d'un modèle de portail de données fédéral qui a diffusé des données dans toute l'Australie. Le réseau AURIN a utilisé les connaissances et les outils de ses membres pour présenter les données grâce à une interface utilisateur perfectionnée que le Bureau australien de statistique n'aurait jamais pu concevoir. Par essence, le Bureau a massifié ses données à l'intention du réseau AURIN, afin de produire un nouveau service à valeur ajoutée (le portail AURIN).

Figure 4

Capture d'écran de la partie du portail AURIN qui utilise les données du Bureau australien de statistiques



25. La massification des données par le Bureau australien de statistique a procuré au réseau AURIN et au Bureau lui-même une série d'avantages, en particulier:

- Les données présentées par le réseau AURIN sont toujours les données les plus récentes du Bureau australien de statistique;
- Les services sont réutilisables. Le réseau AURIN est un portail consacré à l'infrastructure urbaine. Nos services Web pourraient tout aussi aisément répondre aux besoins d'un portail spécialisé dans les questions environnementales ou économiques;
- Les utilisateurs du réseau AURIN ont réalisé des gains de productivité;
- Le Bureau australien de statistique n'a pas eu besoin de développer la couche présentation;
- De nombreux développeurs ont fourni des techniques d'analyse (existantes et nouvelles) destinées à être utilisées dans le portail;
- Il est devenu encore plus nécessaire de produire des données et des métadonnées bien décrites;
- Nous avons pu expérimenter des aspects de notre future stratégie de diffusion;
- Nous avons enrichi notre expérience du travail avec les services d'information intermédiaires, un secteur en pleine croissance.

VII. Massification des microdonnées

26. Enfin, nous avons également progressé dans le domaine de la production dynamique d'agrégats à partir d'unités de données confidentielles. Nous avons investi dans la mise au point d'une suite de produits permettant aux chercheurs et aux décideurs externes d'accéder aux données sur les sous-groupes de population sans compromettre le respect de la vie privée et la confidentialité. Ces produits permettent aux utilisateurs d'obtenir des tableaux de microdonnées et de les analyser avant d'élaborer des résultats rendus confidentiels de façon dynamique.

27. Parmi les exemples de ces produits du Web, on peut citer TableBuilder (voir l'enquête Survey TableBuilder). Ce produit incorpore à la couche fourniture une méthode des perturbations destinée à protéger automatiquement les tableaux de dénombrement et les principaux résumés statistiques tels que les écarts, les totaux, les moyennes, les médianes et les quintiles. Cette méthode des perturbations réduit les risques de divulgation engendrés par les interrogations de tableaux similaires, les interrogations répétées des mêmes tableaux et les interrogations répétées d'une même cellule au sein de tableaux différents. TableBuilder intègre d'autres algorithmes de confidentialité et des protections qui s'activent instantanément tout en réduisant au minimum les pertes d'utilité, tout en garantissant que le produit obtenu ne permette pas d'identifier un répondant.

28. Pour l'heure, le Bureau australien de statistique produit la couche présentation pour les services de confidentialité, et les techniques d'analyse se limitent à celles que nous sommes en mesure de fournir. À l'avenir, toutefois, nous prévoyons d'assurer la confidentialité instantanée en tant que service destiné à permettre aux autres utilisateurs de générer des données confidentialisées au sein de leur propre système. Associés aux démarches tendant à proposer des solutions hors système d'accès aux microdonnées, ces services de confidentialité instantanée donneront à nos utilisateurs, en particulier aux chercheurs, davantage de possibilités d'interagir avec nos ensembles de données les plus détaillées et de les réutiliser.

VIII. Les perspectives d'avenir

29. Le Bureau australien de statistique est conscient du fait que l'utilisateur n'est plus uniquement un «lecteur» passif du Web, mais un véritable acteur de cet environnement. Il joue et aspire à jouer un rôle croissant dans la diffusion, la mise en commun, l'examen, la promotion et la réutilisation de l'information. La massification de cette information ouvre un vaste champ de possibilités de connexion à l'économie numérique. En accédant à une masse importante de données et de métadonnées, les utilisateurs pourront extraire les contenus du Bureau australien de statistique pour les employer dans leurs propres flux de données, que ce soit sur les nouveaux supports, dans leurs propres systèmes ou dans le cadre de toute autre future solution novatrice. Le Bureau devra toujours produire des contenus pour répondre aux besoins de ses utilisateurs, mais nous anticipons une hausse constante de la demande de données «massifiées».

30. Nous sommes des spécialistes de la collecte, du traitement et de l'analyse des statistiques. En revanche, nous ne sommes pas toujours des spécialistes de la présentation de ces données, et nous ne disposons pas des moyens nous permettant de les présenter dans tous les formats et dans tous les styles innovants auxquels nos utilisateurs aspirent. Nous devons par conséquent confier cette tâche à d'autres.

31. Un certain nombre de difficultés se posent:

- Le développement de la réutilisation, de l'agrégation et de la diffusion des données requiert des techniques perfectionnées pour garder la trace de l'origine et des auteurs des données et protéger notre nom;
- Notre façon de travailler et notre infrastructure doit être adaptée en vue de garantir un niveau cohérent de qualité et de performance;
- Ces nouvelles méthodes offrent des possibilités et engendrent des difficultés, et nous devons veiller à ce qu'elles répondent à nos objectifs stratégiques et à notre modèle opérationnel tout en protégeant la viabilité de notre organisation.

32. À travers des mesures de diverses envergures, le Bureau de statistique est engagé dans cette transformation dont il perçoit tout l'intérêt, et il constate que beaucoup d'autres services nationaux de statistique et d'organisations internationales ont engagé des transformations du même ordre. La massification des données est une composante importante de la modernisation et de l'optimisation de leur utilisation.

Bibliographie

Branson, M 2014, *Measuring Interest Group Expectations and Trust*, paper presented to 2014 IAOS Conference, Da Nang, Vietnam, 8-10 October 2014.

Borowik, J., Branson, M. and Watson, D. 2013, *Interactive Content and Dynamic Publishing – A vital part of an NSO'S output and communications strategy*, paper presented at IASE, Macau 2013.

Montgomery, T 2014, *Dynamic Publishing – how far can we go?*, paper presented at 2014 IMAODBC Conference, Evora, Portugal, September 2014.
