

## ENTRETIEN AVEC DES MEMBRES DU LABORATOIRE LIRIS

*Atila Baskurt, professeur à L'INSA de Lyon et directeur du LIRIS, Jean-François Boulicaut, professeur à l'INSA de Lyon et directeur adjoint du LIRIS, Mohand Saïd Hacid, professeur à L'Université Lyon 1 et directeur adjoint du LIRIS et Sylvie Servigne, maître de conférences au LIRIS, INSA de Lyon.*

Interview réalisée par Anne-Caroline Jambaud  
Novembre 2012

### Qu'est-ce que le laboratoire LIRIS ?

**Atila Baskurt** : Le LIRIS, Laboratoire d'InfoRmatique en Image et Systèmes d'information, est une Unité Mixte de Recherche CNRS qui regroupe 300 chercheurs dans 12 équipes de recherche et qui dépend, en plus de la tutelle nationale CNRS, de 4 tutelles régionales : INSA Lyon, Université Lyon 1, Université Lyon 2 et Ecole Centrale de Lyon. Créé en 2003 avec la fusion de plusieurs laboratoires ou équipes, le LIRIS se positionne comme un laboratoire de site sur le campus de la Doua (LyonTech), fédérant les compétences en Sciences et Technologies de l'Information (STI). Nous avons été très impliqués dans la création du laboratoire d'excellence (LabEx) Intelligences des Mondes Urbains (IMU) : aujourd'hui, 90% de nos chercheurs sont impliqués dans IMU. Nous sommes des traiteurs de l'information et intervenons dès que les données existent. Les données – scientifiques, d'enquête, d'observation, etc. - sont accompagnées d'une problématique qui nous vient soit des autres disciplines scientifiques, soit des praticiens. À partir de là, nous intervenons sur toute la chaîne qui permet de structurer ces données, de les gérer, de les indexer, de les annoter, d'y accéder efficacement, de les visualiser, etc. Quand nous avons une très grande masse de données, nous sommes capables de fouiller afin d'extraire de la connaissance. Nous nous intéressons aux données alpha numériques et également aux données de type images, vidéo, 3D, 3D+T, 4D, des flux d'information, etc.

Les aspects visualisation sont importants : il s'agit d'avoir une vue synthétique ou de visualiser les informations sous une forme compréhensible pour les experts ou les usagers qui vont prendre la décision finale. Nous améliorons sans cesse nos modèles, nos algorithmes de manière à mieux répondre aux besoins des experts dans une optique d'aide à la décision.

Par exemple, pour savoir quelle est la philosophie de déplacement en ville avec Vélo'v, nous fouillons dans les données Vélo'v afin de trouver, grâce à des modèles assez théoriques, des formes d'utilisation, des tendances : ce que font les gens, ce qu'ils ne font pas... Nous montrons cela aux experts afin de comprendre ensemble, et éventuellement nous complétons par d'autres tests sur les données.

*« La donnée, c'est l'or noir du 21e siècle »*

Nous ne sommes pas des prestataires de services. Nous travaillons sur des problématiques qui nécessitent des investigations théoriques et l'élaboration de nouveaux concepts et approches pour apporter des solutions appropriées et effectives.

**Les masses de données croissantes, le big data, posent de nouvelles problématiques. Travaillez-vous sur ces questions ?**

**Atila Baskurt** : Le fait de passer à l'échelle, d'avoir de plus en plus de données, constitue une rupture dans nos techniques. Un exemple : auparavant, nos méthodes traitaient des images 256 x 256 pixels ; aujourd'hui, on se retrouve face à des images en gigapixels (plusieurs milliards de pixels) générant des bases d'images de plusieurs téraoctets, voire pétaoctets pour les images astronomiques comme dans le projet de très grand télescope LSST (<http://lsst.in2p3.fr>) par exemple. Le stockage, la visualisation ainsi que l'accès aux données pertinentes deviennent des défis scientifiques. Il faut donc complètement revoir nos modèles de traitement, d'analyse et de visualisation de ces données. Il en est de même actuellement pour les données vidéo ou les données 3D qui peuvent comporter plusieurs millions voire milliards de mailles. Il faut nous adapter à cette nouvelle problématique. Ainsi, le LIRIS se positionne clairement sur le créneau de traitement multi-échelles capable d'explorer de très grandes quantités de données. C'est l'une des raisons pour laquelle notre partenariat avec le Grand

Lyon s'est accéléré ces dernières années, notamment pour le traitement de ces grandes masses de données.

**La plupart de grands groupes qui travaillent sur les problématiques data, notamment dans le champ des smart cities, sont bardés d'ingénieurs. Quel est l'apport distinctif des chercheurs dans ce domaine ?**

**Jean-François Boulicaut :** La plupart des chercheurs du laboratoire LIRIS travaillent sur des problèmes ouverts, à la résolution desquels les progrès de l'industrie informatique ne changent pas grand-chose. C'est précisément la différence entre une approche reposant sur l'ingénierie et une approche fondée sur une démarche scientifique. Nous travaillons uniquement sur des problèmes qui sont algorithmiquement difficiles, où le passage à l'échelle ne se résout pas par un processeur deux fois plus rapide qui crée des images deux fois plus grandes. Il y a une croissance exponentielle de la complexité, or nous avons des processeurs qui ne permettent d'améliorer que ce qui est non exponentiel. Cela veut dire que nous ne pouvons pas nous contenter d'attendre la prochaine génération de processeurs. Cela ne changera quasiment rien aux capacités de traitement. Nous travaillons sur des questions fondamentales du traitement de l'information et le passage à l'échelle est clairement une préoccupation des chercheurs. La réponse n'est pas toujours technique et ne dépasse souvent pas le cadre de l'ingénierie. Elle nécessite des concepts et outils qui reposent sur des fondements scientifiques. Nous utilisons le terme scientifique pour souligner que cette question est ouverte, elle n'est pas résolue. C'est cela qui nous intéresse : ce qui n'est pas résolu. Ce n'est pas parce qu'on rajouterait des barrettes de mémoire que le problème serait réglé. Il faut vraiment penser le traitement de l'information différemment.

**Mohand-Saïd Hacid :** Depuis le milieu des années 80, les vitesses des unités centrales ont été améliorées de 50 à 100% par an, alors que la latence de la mémoire n'a été améliorée que de 7% par an. Par conséquent, les traitements de données qui relèvent de la fonction des systèmes de gestion de bases de données ne tirent pas profit des possibilités de traitement. D'autre part, le stockage des

très grandes masses de données qui sont et seront produites, nécessiteront des architectures élaborées pour pouvoir accueillir les traitements de plus en plus « gourmands » en ressources.

**Un département du LIRIS est intitulé « Données – connaissances – services ». Pouvez-vous explicitez ces trois mots ?**

**Mohand-Saïd Hacid :** La donnée, c'est vraiment l'information brute captée avec un capteur ou qui a été générée par l'utilisateur lui-même. A partir de là, notre travail au laboratoire LIRIS consiste à transformer cette donnée pour générer de l'information intelligente ou intelligible. Certains font subir des traitements sophistiqués à des données pour générer des connaissances : ce sont les fouilleurs de données. Viennent ensuite les services : tout ce que nous greffons autour pour exploiter cette chaîne de transformation de données.

Les chercheurs qui étudient les services travaillent sur des verrous scientifiques liés à la découverte de nouveaux services, leur composition, leur tolérance aux pannes et leur sécurité ; ce sont des mécanismes fondamentaux pour les services. Chaque fois que nous produisons des données en masse, cela va très souvent se retrouver sous la forme de services. Par services, nous entendons tout l'environnement que nous offrons aux gens pour réaliser leur travail, exécuter des tâches, en accédant à des informations ou en sollicitant des composants qui vont les aider à réaliser ces services. Ce sont des fonctionnalités avancées qui sont disponibles pour des usages à la volée.

**On parle beaucoup de « cloud computing ». Quelles problématiques posent l'externalisation des données ?**

**Mohand-Saïd Hacid :** Nous ne sommes plus seulement dans l'usage classique, centralisé, qu'on a connu jusqu'aux années 2000 : nous ramenons les outils vers les utilisateurs sur leurs ordinateurs ; les données peuvent être localisées et/ou distribuées. Désormais, on branche son ordinateur et on accède à des services qui sont ailleurs ; les données sont elles aussi ailleurs. C'est la notion de « cloud » ; les ressources sont dispersées, elles sont accessibles à tous. Les informations et les traitements nous suivent quand nous

nous déplaçons, grâce aux Smartphones, tablettes, etc. La mobilité s'est installée dans le monde de l'information. Cela demande de nouveaux outils, de nouveaux modèles pour accéder à ces données réparties. Cela élargit la palette des services aux usagers et amène de nouvelles contraintes liées à la sécurité, la confidentialité et la qualité de services dont il faudrait tenir compte.

**La plupart des grands groupes (Thomson, Sony, Siemens, Cisco, etc.) se positionnent sur le marché de la donnée, signe que les enjeux sont forts. N'y a-t'il pas un risque de privatisation des données ?**

**Jean-François Boulicaut :** La donnée, c'est l'or noir du 21<sup>e</sup> siècle. Or actuellement, il y a effectivement des tentatives de privatisation de la donnée ; c'est préoccupant.

C'est comme si, avec le recul de plusieurs siècles d'industrialisation, nous analysons l'impact d'avoir privatisé la matière première. Nous savons tous à quel point la possession de la matière première peut être génératrice de conflits. Et bien là, nous sommes confrontés à la privatisation de la donnée, c'est-à-dire ce sur quoi vont se construire les nouveaux services de demain.

Nous ne savons pas encore exactement ce que nous allons pouvoir faire avec les données. Mais il faut préserver cette matière première car ensuite, ce ne sera pas la peine d'exécuter de nouveaux services sur des données qui n'existeront pas ou pour lesquelles il faudra payer cher pour y accéder. Une nouvelle économie pourrait se développer mais pas forcément dans le sens de l'intérêt citoyen.

**Avez-vous des exemples de tentatives de privatisation de la donnée ?**

**Atila Baskurt :** À Lyon Confluence, dans le cadre de l'expérimentation NEDO, ce sont des organismes ou entreprises japonais qui pourraient posséder, voire qui possèdent déjà, des données concernant les bâtiments intelligents, les réseaux de capteurs, etc. Cela nous alerte !

Nous serions très intéressés pour disposer de ces données, parce que nous pourrions les traiter, être vraiment originaux dans notre approche et répondre aux besoins des

praticiens. Le terrain de jeu peut être Lyon Confluence.

**Jean-François Boulicaut :** Ces données n'existent probablement pas encore. Si tel est le cas, la question se posera quand elles existeront. Mais cela pourrait être trop tard ! Car entre temps, les propriétaires des données auront été figés par les « consortium agreements ».

De la même façon, des représentants de grands groupes déclarent publiquement que les données publiques qu'ils peuvent récupérer ne sont plus vraiment publiques parce qu'ils les structurent, les nettoient ou les enrichissent. Ils sont en train de revendiquer que les données publiques qu'ils vont aspirer peuvent leur appartenir. Il est étonnant de voir à quel point les propriétaires de données, ceux qui collectent les données, mésestiment les possibilités d'exploitation qui pourront en être faites.

Je me souviens très bien qu'il y a 20 ans, au plus haut niveau de l'Etat français, des gens étaient prêts à signer la cession de toutes les images numériques des œuvres de nos musées. A l'époque, cela avait été arrêté au dernier moment. Je me souviens d'échanges avec notamment des gens qui représentaient le monde de la culture qui disaient : ce n'est pas grave, de toute façon les images numériques sont de qualité médiocre. Ils n'avaient pas pu, à l'époque, imaginer ce que pouvait être le numérique dans 10, 20, 40 ans. Or si nous cédon les droits, nous les cédon pour toujours.

Les praticiens, élus et décideurs ont tout intérêt à se mettre en rapport avec les chercheurs, qui travaillent sur le long terme, pour avoir une vision prospective des enjeux.

**Atila Baskurt :** D'où l'importance d'outils comme le LabEx IMU. C'est là où peuvent se rencontrer les chercheurs, les praticiens de la ville, les collectivités – Lyon ou Saint-Etienne -, les agences d'urbanisme, les architectes, les archéologues, les ingénieurs... pour pouvoir réfléchir sur les enjeux de demain et lancer des projets de recherche et de développement pour y répondre.

**Selon que la donnée est brute, enrichie ou transformée, la propriété ne serait pas la même ? Que dit le droit ?**

**Mohand-Saïd Hacid** : Il y a là un vide juridique qui contribue à freiner le partage des données, leur diffusion. Quelle est la source de la donnée ? Comment assurer sa traçabilité ? Ce sont des problèmes très compliqués sur lesquels travaillent les chercheurs, ils appellent cela « la provenance des données ». Mais lorsqu'elles ont subi des transformations, que des éléments ont été enlevés ou ajoutés, qu'en est-il à l'arrivée ? La traçabilité des données est un point clé.

Il y a toute une branche de la recherche qui s'intéresse à l'aspect légal de l'usage des données. C'est un mariage du droit et de l'informatique. Il y a des sujets de recherche très pointus sur ces aspects.

**On sait depuis longtemps qu'avoir l'information, c'est détenir le pouvoir...**

**Jean-François Boulicaut** : Pour prendre un exemple qui est sur la place publique, Google est le seul à savoir quelles sont les requêtes qui lui sont transmises ; il ne les partage pas. C'est là dessus que se construit l'économie de demain. Google est le seul à savoir qui fait quoi avec son moteur.

**Atila Baskurt** : Ce n'est pas seulement la donnée qui est de l'or noir, c'est aussi l'interface, l'usage qui en est fait et les services qui vont se développer.

**Vous insistez beaucoup sur la protection des données ; en quoi est-ce essentiel pour vous ?**

**Jean-François Boulicaut** : Ce qui nous intéresse c'est de valoriser les données : développer de nouveaux services, découvrir des connaissances, etc. À partir du moment où nous avons les données, nous voyons parfaitement ce que nous pouvons en faire. Mais le problème, c'est que nous ne les avons pas, ou que nous ne les aurons pas ! Il y a un verrou qui ne dépend pas de nous pour pouvoir bien travailler. Développement de nouveaux services, découverte de structures, maîtrise de la qualité des données... Il y a plein de sujets de recherche sur lesquels nous travaillons déjà et pour lesquels nous avons besoin de matières premières. Il ne s'agit pas seulement de données brutes, mais de gens qui ont des données brutes et qui ont des problèmes à régler, des questionnements, des intentions : améliorer la maîtrise de l'énergie

sur l'agglomération par exemple. Pour cela, il faut avoir les données. Nous insistons sur ce problème parce qu'il y a là un verrou qui échappe à la recherche.

**La première motivation de l'open data est le développement économique : « c'est bon pour le business ». Comment la recherche publique s'accommode t'elle de cette vision très libérale ?**

**Jean-François Boulicaut** : Quand un chercheur travaille à la compréhension de la virulence de certains virus, c'est indispensable pour l'humanité mais ça peut permettre à certains laboratoires pharmaceutiques de s'enrichir de façon éhontée, et cela peut aussi permettre à d'autres de faire des armes chimiques. Les trois aspects existent. Le chercheur ne va pas arrêter de travailler sur la virulence des virus sous prétexte qu'il peut y avoir un usage dévoyé.

Si l'open data génère des initiatives économiques créatrices d'emplois, il y a bien une retombée directe sur le bien être du citoyen. Tout cela est très complexe. Nous n'allons pas nous interdire le développement de services économiques sur les données sous prétexte qu'il pourrait y avoir des usages dévoyés.

**Qui produit de la donnée ?**

**Sylvie Servigne** : La question c'est à la fois qui ou quoi, et comment ? Parfois, on lance un système, de récolte de données et quand le système produit effectivement de la donnée, le donneur d'ordres ne se rappelle même plus qu'il a donné cet ordre. On peut mettre des capteurs partout. Les données arrivent, et on les utilise - ou pas. Mais à un moment, il y a bien un élément déclencheur. Ensuite, les systèmes fonctionnent et collectent. Des entités – organismes, entreprises, instituts – peuvent être producteurs de données, mais c'est aussi le cas de tout citoyen, de tout usager qui se connecte à un système électronique muni d'un système de transmission et de mémorisation.

**Les producteurs de données n'ont pas toujours conscience de produire de la donnée...**

**Sylvie Servigne** : Du moment qu'on met de la donnée sur un outil électronique, il faut être

conscient qu'avec Internet, avec les réseaux, cette donnée peut partir partout. C'est ce qui se passe avec Facebook.

**Jean-François Boulicaut :** Quand vous posez quelque chose sur le bureau de votre ordinateur, il n'y a que vous qui le voyez mais vous trouvez normal que l'administrateur système mette à jour vos logiciels sur votre poste, voire prenne le contrôle de votre poste. A partir du moment où votre système est géré à distance, la notion de propriété d'un document n'a aucun sens.

### **A vous écouter, nous serions donc dans une sorte de production auto générative de la donnée ?**

**Jean-François Boulicaut :** C'est un phénomène récent dans la collecte de données. Prenons le cas de *Voyager*, la sonde qui envoie des données depuis 30 ans alors qu'au départ cela ne devait durer que quelques années. Elle est train de sortir du système solaire. Les gens qui au départ collectaient les données ne sont certainement plus là, et la sonde continue d'envoyer des données. Ce qui est nouveau c'est que nous sommes désormais passés à de la collection automatique, sans effort, quasiment sans coût et exhaustive de données sur certains phénomènes. C'est ce qui se passe avec les capteurs qui vont être coulés dans le béton (pavés intelligents). Dans certains cas, la donnée continue d'être produite alors qu'on ne sait plus quelle était l'intention initiale.

Les gens qui mettent en place ou commandent ces dispositifs, sous estiment systématiquement les usages qui pourront être faits des données. Ils vont récolter par exemple les déplacements de tout le monde partout dans un bâtiment, parce qu'ils veulent faire des comptages à l'entrée d'un restaurant, mais ces données permettent bien d'autres usages. La CNIL, Commission Nationale de l'Informatique et des Libertés, ne dit pas autre chose : les gens collectent aujourd'hui des données avec des objectifs extrêmement restreints et ils ne se rendent pas compte que ces données vont persister, qu'elles existent. Ils sous estiment totalement l'exploitation qui peut en être faite. En bien comme en mal. Une fois que des capteurs sont posés, ils collectent, d'autant qu'il y a de plus en plus de réseaux de capteurs autonomes. Ils collectent, ils émettent. Et l'intention est perdue.

### **Qu'est-ce qu'une donnée utile ?**

**Mohand-Saïd Hacid :** Une donnée utile est une donnée qui peut aider à accomplir une tâche, à résoudre un problème, à comprendre un phénomène.

**Jean-François Boulicaut, Sylvie Servigne :** Il nous paraît fondamental de souligner que la raison pour laquelle des données sont collectées aujourd'hui, c'est parce qu'un usage a été perçu. Mais nous ne savons pas ce qu'il sera possible de faire sur ces données demain. En fait, les usages, les intérêts des données collectées vont apparaître au fur et à mesure du temps si ces données ont été conservées.

**Atila Baskurt :** D'un autre côté, nous aurons sans doute demain une manière si différente de capter l'information – avec des nano capteurs, etc. – que les données que nous sauvegardons aujourd'hui ne seront peut-être jamais utilisées, car la finesse de la résolution que nous aurons en imagerie, en réseau de capteurs aura évolué.

**Sylvie Servigne :** J'ajouterai que pour définir une donnée utile, il y a une temporalité, une actualité.

### **En quoi consiste le traitement des données ?**

**Jean-François Boulicaut :** La quantité de choses que je peux faire avec des données est phénoménale ! Quand on met en place des dispositifs, il y a les attentes, les prévisions par modèles et les faits – qui sont capturés par les données. A partir de ces trois flux on peut faire énormément de choses : on peut s'interroger sur le lien entre les attentes et ce qui se passe, entre ce qui est modélisé et ce qui se passe, entre les attentes et les modèles, etc. On peut également travailler sur chacun de ces flux. A partir des déplacements de toutes les personnes individuellement dans Lyon, on peut tirer beaucoup de choses très intéressantes que je ne pourrai même pas énumérer aujourd'hui. Mais cela suppose d'être capable de construire les bons résumés par rapport à un questionnement. La construction de bons résumés est particulièrement importante sur les masses de données.

### **Qu'est-ce qu'un bon résumé ?**

**Jean-François Boulicaut :** Un bon résumé est la vue synthétique d'un ensemble de données. Des gens prennent des centaines de mesures sur un phénomène et à la fin ils ramènent tout cela à une moyenne d'un débit dans un tuyau. C'est un gros résumé car on est passé de 300 valeurs à une. On pourrait imaginer des résumés un peu plus riches ! De plus, la plupart des phénomènes qui intéressent les gens aujourd'hui, sont des phénomènes dynamiques. Il s'agit de comprendre ce qui bouge, ce qui change au cours du temps. Comprendre les phénomènes dynamiques suppose d'avoir des résumés sur ce qui bouge.

**L'un des enjeux de l'open data est de générer des nouveaux services par des croisements des données. Cela pose-t-il des problèmes spécifiques ?**

**Jean-François Boulicaut :** De nouvelles questions apparaissent lorsque l'on juxtapose des collections de données. Par exemple on peut avoir pensé très bien contrôler le type de service qui pourrait être construit à partir de tel type de donnée de mobilité. Et puis on agrège les données de la météo et des données socio économiques, et on s'aperçoit qu'il peut y avoir des choses complètement nouvelles qui peuvent être faites. C'est bien cela qui rend extrêmement compliqué le problème de la protection de la vie privée ou de la protection économique. Toutes les sociétés concurrentes sont amenées à avoir des projets collaboratifs. Elles sont amenées à partager leurs données. Elles veulent bien, mais en ayant une vision de ce que les autres peuvent voir dedans. Et ça, c'est très compliqué. Si une information est retirée d'une base de données, il est possible de la reconstruire à partir de deux autres, par requête par exemple.

**Mohand-Saïd Hacid :** On peut contrôler l'accès individuellement aux données, mais quand il s'agit de les mélanger pour répondre à un besoin spécifique, elles peuvent contribuer à révéler des informations que l'on voulait protéger.

**Si l'on peut reconstituer une donnée privée ou sensible par déduction ou décroisement, comment ouvrir les données au public sans risquer d'attenter à la vie privée et au bien public ?**

**Sylvie Servigne :** C'est un problème très

difficile de recherche. La difficulté est de prouver que les informations privées concernant les citoyens ou sensibles sur le plan économique et industriel ne peuvent pas être retrouvées à partir du croisement de données dites non sensibles ou anonymisées.

**Jean-François Boulicaut :** Nous pouvons appliquer la même technique qu'en cryptologie : si l'on estime qu'il faut trois ans pour décrypter le code, il suffira de changer le code avant trois ans. Ici, si on estime qu'il faut au minimum 10 requêtes sur les bases de données pour reconstruire l'information, on peut considérer qu'elle ne sera pas reconstruite car il y a peu de chances que quelqu'un trouve une telle séquence. Il faut essayer de quantifier l'effort que demande la reconstruction. S'il est considéré comme gigantesque, le risque est faible.

**Dans quels grands programmes de recherche sont engagées les équipes de LIRIS ?**

**Atila Baskurt :** Le CNRS affiche clairement quelques débats prioritaires au niveau national, dont la sécurité informatique, et les bases de données. Nous avons deux projets qui ont été retenus récemment dont AMADOUER, programme d'analyse de masse de données de l'urbain et de l'environnement (lié à la complexité de la nature des données), et PETASKY, programme de gestion et d'exploration de grandes masses de données scientifiques issues d'observations astronomiques grand champ (lié à la masse des données).

**Sylvie Servigne :** Un des enjeux d'AMADOUER est d'engendrer une fertilisation de chercheurs du LIRIS et de chercheurs issus de différentes disciplines pour mieux comprendre les phénomènes urbains, environnementaux ou naturels. L'objectif est de croiser des masses données issues de différents milieux, des données pluridisciplinaires provenant des SPI (Sciences pour l'ingénieur) et des SHS (Sciences humaines et sociales).

**Mohand-Saïd Hacid :** PETASKY s'inscrit dans la participation du CNRS à un vaste programme qui consiste à observer l'univers (LSST, <http://lsst.in2p3.fr/>). Ils construisent un télescope géant au Chili par lequel ils vont observer le ciel pendant dix ans et acquérir

des images toutes les 17 secondes. Il faudra stocker toutes ces images. On atteindra des capacités jamais enregistrées : 140 petaoctets de données. Certains traitements sur ces données sont extrêmement coûteux. Certaines requêtes nécessiteront par exemple de faire coopérer, en l'état actuel de la technologie, 8000 machines en parallèle pour produire des résultats au bout de quelques heures. L'ensemble des données qui seront fournies permettront de servir un champ large de l'astrophysique : de l'étude du système solaire à la cosmologie.

On intervient pour prévoir les outils et les techniques qui permettront d'exploiter les données.