



Avalanche de données

L'avancée de la *eScience* n'est pas sans effets sur la pratique scientifique traditionnelle

Yuri Vasconcelos

PUBLIÉ EN NOVEMBRE 2014

L'époque est déjà loin où la difficulté des scientifiques était d'obtenir des données pour faire avancer leurs recherches. Dans nombre de domaines de connaissance, les récents progrès en technologie de l'information avec la démocratisation de l'informatique, l'augmentation des réseaux informatiques et la multiplication des sources d'information ont eu comme effet direct une production intensive de données. Cela concerne des champs aussi distincts que l'astronomie (régulièrement inondée par des milliers d'images et d'informations sur des corps célestes captées par des télescopes puissants), la biologie moléculaire (grâce à l'apparition de machines de séquençage génétique hautement performantes) ou l'écologie (favorisée par une série de technologies et de capteurs capables de documenter avec précision les transformations vécues par les différents biomes). Tout cela place les chercheurs devant un nouveau problème : comme traiter, élaborer et visualiser l'avalanche de données obtenues par les moyens les

plus divers ? Pour tenter de répondre à ce dilemme, une nouvelle branche de la science acquiert chaque fois plus d'importance : l'*eScience*, qui recourt à des modèles mathématiques et à des instruments informatiques pour analyser des informations et accélérer la recherche dans d'autres domaines de la connaissance.

Lors du *Microsoft eScience Workshop 2014* qui s'est tenu du 20 au 22 octobre à Guarujá (littoral de l'État de São Paulo), Carlos Henrique de Brito Cruz, directeur scientifique de la FAPESP, a déclaré : « L'idée de connecter la pratique scientifique traditionnelle avec l'accès, l'utilisation et le traitement de grandes quantités de données va modifier la manière dont est réalisée la science et augmenter sa potentialité. La FAPESP est à l'avant-garde de ce processus ; l'an dernier, en fin d'année, elle a lancé le programme *eScience* ». L'objectif du programme est d'organiser ou d'intégrer des groupes engagés dans la recherche sur les algorithmes, la simulation informatique et l'infrastructure de données avec des équipes de scien-

tifiques qui interviennent dans d'autres domaines de la connaissance comme la biologie, les sciences sociales, la médecine et les sciences humaines.

DÉFI MONDIAL

D'après Brito, « l'une des principales barrières sera peut-être la difficulté de communication entre les scientifiques des équipes nécessaires pour faire une science de cette manière, fortement basée sur des données ou des grandes quantités de données. Cela requiert une communication très effective entre des chercheurs en informatique et des scientifiques d'autres domaines. C'est un défi au Brésil et partout ailleurs ». le directeur scientifique de la FAPESP était un des participants de la table ronde L'importance stratégique de l'eScience, qui a également compté sur la présence des scientifiques Jason Rhody, directeur senior de l'Office of Digital Humanities de National Endowment for the Humanities, et Chris Mentzel, directeur de Gordon and Betty Moore Foundation, deux organisations nord-américaines qui possèdent des programmes de soutien à la science.

Mentzel indique qu'« actuellement, tous les champs de la recherche sont tou-

chés par l'échelle moderne de production de données ». Il souligne l'importance des *data scientists* (scientifiques des données), ces professionnels qui se penchent sur le grand volume de données générées par les recherches pour produire de nouvelles connaissances : « Ce sont des chercheurs qui travaillent entre les disciplines. Ils construisent des ponts ». À la Gordon and Betty Moore Foundation, Mentzel dirige un programme de 60 millions de dollars US destiné à encourager les initiatives dans le domaine de l'eScience. Pour Rhody, les scientifiques sont confrontés à un changement de paradigme : « Nous sommes en train de passer d'une culture de pénurie à une culture d'abondance de données ».

Créé en 1999 par John Taylor, le directeur du Cabinet de Sciences et Technologie du Royaume Uni, l'eScience ou e-Science est aussi appelée « science orientée par des données » (*data-intensive computing*) ou encore « informatique fortement orientée par des données » (*data-intensive computing*). Certains pays comme l'Angleterre et les États-Unis disposent déjà de programmes soutenus par le gouvernement pour le développement de ce nouveau domaine scientifique. Au Brésil, il faut signaler



le Centre de Recherche en eScience de l'Université de São Paulo (USP), institutionnalisé en 2012. Coordonné par le professeur Roberto Marcondes Cesar Junior, membre de l'Institut des Mathématiques et Statistiques (IME) et coordonnateur adjoint du département des sciences exactes et ingénieries de la Direction Scientifique de la FAPESP, le centre regroupe 20 chercheurs.

Le *Microsoft eScience Workshop 2014* a été réalisé en même temps que la 10^e Conférence internationale IEEE en eScience, organisée par Computer Society de l'Institute of Electrical and Electronics Engineers (IEEE) – une entité fondée aux États-Unis par des ingénieurs électriciens et électroniciens. Pendant l'événement, des chercheurs boursiers de l'Institut Virtuel de Recherches FAPESP-Microsoft Research ont présenté des travaux qui relient les applications de la science informatique aux défis de la science de base dans des domaines en relation avec les changements climatiques et l'environnement. L'un de ces travaux explore des solutions novatrices pour la surveillance de plantes dans les tropiques en associant la recherche informatique et la phénologie. La phénologie est une des branches les plus anciennes de la science ; elle est un domaine de l'écologie qui étudie les phénomènes cycliques de plantes (l'apparition de feuilles, boutons, fleurs et fruits) et leur lien avec les conditions environnementales.

Coordonné par le professeur Leonor Patrícia Morellato du Laboratoire de Phénologie de l'Institut de Biosciences de l'Université d'État Paulista (Unesp)

Connecter les applications de l'informatique à la phénologie, qui étudie les phénomènes cycliques des plantes



à Rio Claro (État de São Paulo), le projet prévoit la combinaison de technologies pour contrôler les altérations survenues au cours du temps sur des végétations natives du *cerrado* [note de trad. : région de savane d'Am. du Sud], de la forêt atlantique, des *campos rupestres* [écosystèmes néotropicaux d'altitude] et même de la *caatinga* [forêt épineuse]. La zone centrale de la recherche se trouve à Itirapina, dans la province de São Paulo. Morellato explique : « En plus de l'observation directe des plantes au sol, nous avons installé un appareil photo en haut d'une tour de 18 mètres pour photographier chaque jour la végétation et une saison météorologique. Nous allons aussi disposer d'un drone équipé d'un capteur hyperspectral et d'un appareil photo pour additionner une échelle spatiale à la collecte de données ». D'une grande résolution spatiale, les capteurs hyperspectraux peuvent fournir des détails sur les caractéristiques physico-chimiques et les réponses physiologiques des plantes observées. Pour la chercheuse, la phénologie est un des meilleurs outils pour comprendre les effets des changements climatiques sur les plantes : « Cela est déjà établi dans des régions tempérées, où le déclencheur de la phénologie est la température environnementale et la durée de la journée. Mais on sait peu de choses sur ce qui se passe dans les végétations tropicales. Avec les données des appareils photos et du capteur hyperspectral, nous cherchons à savoir quels sont les déclencheurs de la phénologie dans les tropiques, c'est-à-dire ce qui fait apparaître à des moments donnés les fleurs, les fruits et les feuilles sur les plantes ».

ANALYSER DES IMAGES

De l'avis de Morellato, la recherche serait impossible sans l'aide des chercheurs et des ressources informatiques : « Le volume de données que nous allons recueillir est gigantesque. À lui seul, un appareil photo numérique prend 60 photos par jour. 11 appareils surveillent 6 types de végétation et nous devons observer l'évolution pendant au moins une saison de croissance pour pouvoir ensuite établir un lien avec le climat. Puis il faut traiter et analyser toutes les images, ce qui serait impossible avec un simple formulaire électronique. Il faut de l'aide pour travailler avec ce *big data*. C'est pour cette raison qu'une étudiante

Des capteurs sans fil installés dans des forêts fournissent de nombreuses données sur les processus naturels

de master a créé une banque de données spéciale pour le projet et qu'un post-doctorant a élaboré un logiciel pour la visualisation et l'organisation d'images ».

La recherche de Morellato compte sur la participation du scientifique Ricardo Silva Torres, directeur de l'Institut d'Informatique de l'Université d'État de Campinas (Unicamp), qui mène également un projet dans le cadre de l'accord FAPESP-Microsoft Research. L'objectif de son étude est de développer de nouvelles techniques analytiques et des outils informatiques pour le traitement d'images de télédétection afin d'analyser la dynamique de certains biomes à des échelles régionales et continentales. Réalisé en partenariat avec le professeur Marina Hirota du Département de physique de l'Université de Santa Catarina (UFSC), le travail cible les biomes tropicaux sud-américains.

Un autre travail présenté pendant la rencontre de Guarujá est celui de l'écologue de l'Unicamp Rafael Silva Oliveira, développé en partenariat avec les chercheurs Antonio Alfredo Ferreira Loureiro, du Département d'informatique de l'Université Fédérale de Minas Gerais (UFMG), et Stephen Burgess, de l'University of Western Australia : « Notre étude a comme objectif l'étude de la dynamique de l'eau et du carbone dans des forêts de nuages, des pâturages et dans la zone de transition entre eux. [...] Nous voulons comprendre comment des éléments-clés, comme l'absorption et le stock de carbone, la transpiration

des arbres et la captation d'eau du nuage par la végétation, sont affectés par des changements d'utilisation de la terre et par des variations du climat ».

Les études sur le terrain ont lieu dans la région montagneuse da Mantiqueira (région de Campos do Jordão, dans la province de l'État de São Paulo). D'après Oliveira, un réseau de capteurs sans fil est installé sur place pour contrôler, sur trois couches de l'écosystème (atmosphère, végétation et sol), des paramètres microclimatiques de métabolisme de la végétation et de dynamique de l'eau dans le sol : « Ces données pourront améliorer la prédiction des impacts environnementaux provoqués par des changements d'utilisation de la terre et, en même temps, permettront le développement de modèles hydrologiques et de circulation biosphère-atmosphère avec une meilleure capacité prédictive ».

Projets

1. Towards an understanding of tipping points within tropical South American biomes (2013/50169-1); **Modalité** Programme de Recherche 'Partenariat pour l'Innovation Technologique' (Pite) et Accord FAPESP-Microsoft; **Chercheur responsable** Ricardo da Silva Torres (Unicamp); **Investissement** 384 838,38 *reais* (FAPESP).
2. Combining new technologies to monitor phenology from leaves to ecosystems (2013/50155-0); **Modalité** Programme FAPESP de Recherche sur les changements climatiques mondiaux - Programme de Recherche 'Partenariat pour l'Innovation Technologique' (Pite) et Accord FAPESP-Microsoft; **Chercheuse responsable** Leonor Patrícia Cerdeira Morellato (Unesp); **Investissement** 1 115 752,48 *reais* et 535 902,72 dollars US (FAPESP).
3. Interactions entre sol-végétation-atmosphère dans un paysage tropical en mutation (2011/52072-0); **Modalité** Programme de Recherche 'Partenariat pour l'Innovation Technologique' (Pite) et Accord FAPESP-Microsoft; **Chercheur responsable** Rafael Silva Oliveira (Unicamp); **Investissement** 644 800,74 *reais* et 663 429,82 dollars US (FAPESP).

