

## Universités : nécessité d'une politique sur les données de la recherche en discussion à la CPU

Paris - Publié le vendredi 26 janvier 2018 à 18 h 31 - Actualité n° 111403

« Il faut donner aux chercheurs les moyens d'accéder aux données de la recherche, leur disponibilité constitue un enjeu majeur pour faire de nouvelles recherches et rester compétitifs », indique [Paul Indelicato](#), conseiller recherche de la CPU, lors d'une journée thématique sur les données de la recherche organisée à Paris par le réseau des vice-présidents « recherche et valorisation » de la CPU, le 16/01/2018.

« Il y a une absolue nécessité qu'il y ait une politique des établissements sur la PI des données », affirme [Marin Dacos](#), conseiller scientifique pour la science ouverte auprès du DGR, présent lors du colloque. Selon lui, « la clé de la souveraineté scientifique c'est la PI, et nous pourrions reproduire l'erreur que nous avons commise dans les années 70 à 80 en cédant peu à peu la PI de nos publications à quelques acteurs. Ce n'est pas grave de céder le travail de diffusion à un acteur privé, mais le céder à titre exclusif c'est autre chose ». Il rappelle à ce titre que, contrairement aux publications, « les données sont la propriété de l'employeur ou du financeur ».

Pour Paul-Antoine Hervieux, VP délégué aux données de la recherche et à la science ouverte à l'Université de Strasbourg, « les données de la recherche constituent un patrimoine pour les universités, qui nécessitent de la création d'écosystèmes qui peuvent être assez complexes, pour les gérer au mieux ». Selon lui, ces données « certifiées par les chercheurs » sont de « fantastiques sources d'interdisciplinarité ». « Certains organismes de recherche thématiques, comme l'[Inra](#) et l'[Ifremer](#) ont déjà une stratégie sur les données, mais quid des universités qui sont pluridisciplinaires et dont les chercheurs produisent des données multiformes ? Et comment gérer cela au niveau d'une vice-présidence recherche ? », s'interroge-t-il.

Miser sur l'interdisciplinarité, importance de la formation, échanges avec les chercheurs... des porteurs présentent trois structures de gestion des données développées au niveau de leur territoire — le Center for Data Science de l'Université Paris Saclay, le Data Institute de la Comue UGA et le SIS (Scientific IT Services) de l'ETH Zürich — que News Tank retrace à travers plusieurs thématiques identifiées par les porteurs.

« En ce qui concerne le financement, au SIS, nous travaillons tous pour de l'argent, et non pas pour des publications ou des collaborations scientifiques, donc nous sommes financés par les chercheurs qui sont nos clients », souligne notamment Balazs Laurenczy de l'ETH.

---

### Enjeux économiques et de propriété intellectuelle vus par Marin Dacos

Selon Marin Dacos, conseiller scientifique pour la science ouverte auprès du DGRI, « les données représentent un énorme marché potentiel. »

« Elsevier a annoncé en 2017 la version beta de 'data search', qui ambitionne d'être le moteur de recherche des données. Amazon a annoncé un programme dédié aux données scientifiques parce que ce sont des marchés potentiellement élevés. »

Face aux multiples enjeux liés à ces données de la recherche, « la France s'est dotée d'un administrateur général des données [en 2014], qui a pour mission de coordonner l'action de l'État et de l'administration dans le domaine de l'inventaire, de la gouvernance, de la production, de la circulation et de l'exploitation des données par les administrations dont les universités », rappelle Marin Dacos.

« Cet administrateur constitue un point de repère, une ressource. Les chercheurs peuvent le saisir pour lui demander des avis notamment s'ils se sentent démunis face à la complexité de mise en place d'un plan de gestion des données pour tel ou tel projet. »

Par ailleurs, l'administrateur coordonne le plan d'action 2018-2020 de la France dans le cadre de l'OGP (partenariat pour un gouvernement ouvert), dont le but est d'augmenter la transparence de l'action publique. La feuille de route en 11 points de l'engagement 14 de ce plan, « construire un écosystème de la science ouverte », a été dévoilée par Alain Beretz, DGRI, le 22/01/2018.

### Trois projets universitaires de gestion de la donnée

À l'occasion de cette journée, trois exemples d'organisation locale transdisciplinaire sont présentés :

- le Data Institute Grenoble Alpes/UMS Gricad ;
- le Center for Data Science de Paris-Saclay ;
- et le Scientific IT services de l'ETH Zürich.

Plusieurs thématiques sont identifiées par les porteurs de projets.

S'adapter à deux grands publics de chercheurs

Pour Balazs Laurenczy (ETH), on peut distinguer deux groupes de communautés de chercheurs : « les experts informatiques et les utilisateurs de l'informatique ».

- « Pour les premiers, nous devons les aider à trouver plus rapidement une solution :
  - en explorant et en triant les technologies et les programmes ;
  - en mettant en place pour eux une infrastructure de calcul ;
  - ou en les aidant à écrire ou à améliorer des logiciels ».

« Pour eux, le but final est d'accélérer la recherche et de les débarrasser pour qu'ils puissent vraiment se concentrer sur la recherche ».

- « De l'autre côté, nous avons des chercheurs qui ont juste besoin d'une solution qui marche et souvent que l'on fabrique de toute pièce, de A à Z pour eux :
  - en explorant les technologies et en construisant des interfaces graphiques faciles et pratiques ;
  - parfois, nous faisons de la co-analyse, qui consiste à analyser leurs données ensemble et en les aidant à maîtriser les outils d'analyse ».

« Ici, nous rendons possible la recherche et l'analyse de données qui ne serait autrement pas accessible à ces chercheurs. »

### Avant tout : échanger avec les chercheurs

Pour Balazs Laurenczy, « la première chose à faire est de discuter avec les chercheurs, de comprendre ce qu'ils font et leur besoins informatiques ». Une première étape cruciale selon lui car :

- les différentes disciplines ont des manières de penser différentes ;
- certains chercheurs font des expériences assez appliquées, d'autres de la recherche plus théorique ;
- les niveaux de connaissances informatiques sont très variés entre les chercheurs ;
- les différentes communautés de chercheurs ont chacune leur propres défis.

« Il faut vraiment partir du besoin des utilisateurs », souligne également Violaine Louvet, ingénieur de recherche calcul scientifique au CNRS et directrice de l'UMS Gricad, citant l'exemple d'une collaboration avec le CHU Grenoble Alpes pour la mise en place d'un entrepôt de données.

### Miser sur l'interdisciplinarité

« Pour avoir un centre qui marche, l'interdisciplinarité est également très importante », selon Sarah Cohen-Boulakia (CDS Paris-Saclay) pour qui trois composantes doivent être associées dès le début du projet :

- des spécialistes de data science ;
- des domain scientists ;
- et des experts en ingénierie.

« Les outils pour aider à la gestion et les aspects software [doivent être pensés] dès le début du projet », ajoute-t-elle, précisant que le Center for Data Science de Paris-Saclay ne finance que les projets qui contiennent au moins des data scientists et des domain scientists.

### L'importance de la formation

Violaine Louvet met également l'accent sur l'importance de l'animation et de la formation. « Calcul intensif, traitement de données, stockage, calcul à la demande, diffusion et valorisation de données : il faut accompagner les communautés dans leurs nouveaux besoins croissants », explique-t-elle.

« Et il y a aussi tout un conseil d'accompagnement à adapter aux communautés "neuves" dans le domaine, qui n'ont pas encore la pratique de ces infrastructures du numérique », ajoute-t-elle, avant de préciser que l'accompagnement peut être très différent d'un laboratoire à l'autre.

Du côté du SIS, une grande partie du travail concerne aussi la formation avec :

- de réguliers « code clinic », qui consistent à aider les chercheurs à créer leur logiciel ou à rendre leur code meilleur ;
- divers cours en lien avec l'infrastructure HPC, le développement de logiciel ou l'analyse de données ;
- et de la consultation dans le domaine de la gestion de données.

## Une ERC née du CDS

« Nous n'avons pas encore listé toutes les ANR qui ont été reçues et qui sont nées aussi de collaborations dans le CDS, et on a une ERC qui est née clairement dans le CDS », note Sarah Cohen-Boulakia du Center for Data Science de Paris-Saclay.

## SIS de l'ETH Zürich : le modèle payant

Le service informatique de l'ETH Zürich est composé de neuf sections totalisant 260 employés. Le SIS est l'une de ces sections et compte environ 30 experts dans différents domaines du calcul scientifique, la plupart avec une thèse.

L'organisation s'articule en trois sous-groupes :

- High performance computing.
- Scientific software.
- Data analysis and management.

« En ce qui concerne le financement, au SIS, nous travaillons tous pour de l'argent, et non pas pour des publications ou des collaborations scientifiques, donc nous sommes financés par les chercheurs qui sont donc nos clients », précise Balazs Laurenczy.

Les clusters sont cofinancés par le service technologies de l'information, et les utilisateurs grâce au modèle du « Shareholder ». Concrètement, des professeurs ou des instituts ou même des départements peuvent acheter des « shares », donc des parts du cluster et ils reçoivent en échange une certaine priorité dans la file d'attente du cluster ; cela leur garantit que — en moyenne — ils auront le quota de ressources qu'ils ont payées : temps de calcul, mémoire ou stockage. Le SIS établit des SLA (service licence agreement) qui définissent les termes de l'abonnement.

« L'avantage de ce modèle est que tous les chercheurs de l'ETH peuvent gratuitement accéder au cluster, grâce à la part financée par le service informatique de l'ETH, et qu'il est possible d'utiliser plus que les ressources que l'on a payées si la file d'attente est vide. »

En ce qui concerne la partie service, qui paie le reste des coûts qui ne sont pas liés aux deux clusters, le SIS a un modèle d'abonnements qui peuvent être faits pour un professeur, un institut ou même un département.

Plus rarement, le SIS peut aussi proposer des contrats individuels pour une tâche donnée, par exemple à la journée pour les formations. Enfin, il participe aussi à des projets nationaux suisses ou internationaux, qui soit le financent ou ont un bénéfice pour le SIS.

## Conférence des Présidents d'Université



Statut : association Loi de 1901

Composition : une centaine de membres votant (présidents d'université, directeurs d'écoles normales supérieures, d'INP, d'INSA, administrateurs généraux) et des membres associés.

Bureau

Gilles Roussel président

Fabienne Blaise, vice-présidente

Khaled Bouabdallah, vice-président

Missions :

- Force de proposition et de négociation auprès des pouvoirs publics, des différents réseaux de l'enseignement supérieur et de la recherche, des partenaires économiques et sociaux et des institutions nationales et internationales.
- Soutien aux présidents dans leurs nouvelles missions et de promotion de l'Université française et de ses valeurs en France et à l'étranger.

Moyens :

- une équipe permanente (27 personnes)
- des conseillers et consultants (7 personnes)
- s'appuie sur l'Amue, (Agence de mutualisation des universités et établissements) qui contribue à l'élaboration d'une offre logicielle performante et à la formation des personnels de l'enseignement supérieur.

---

Conférence des Présidents d'Université

103 boulevard Saint-Michel

75005 Paris - FRANCE



Fiche n° 1765, créée le 05/05/14 à 12:19 - MàJ le 13/10/14 à 13:00

---

© News Tank 2018 - Code de la propriété intellectuelle : « La contrefaçon (...) est punie de trois ans d'emprisonnement et de 300 000 euros d'amende. Est (...) un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une oeuvre de l'esprit en violation des droits de l'auteur. »