

Axe stratégique de recherche 2017–2020

Le Domaine des EPF lance l'Initiative for Data Science in Switzerland

Avec le lancement de l'*Initiative for Data Science in Switzerland*, le Domaine des EPF entend favoriser l'expansion de la science des données en développant l'enseignement et la recherche et en fournissant l'infrastructure nécessaire aux utilisateurs des différentes disciplines.

Au cours des dernières années, la science des données a fortement gagné en importance au niveau international, voyant notamment la majorité des institutions qui pratiquent une recherche et un enseignement de pointe investir des sommes considérables dans des centres et des programmes consacrés à cette discipline. L'*Initiative for Data Science* entend promouvoir la collaboration avec des institutions de premier plan, qui comprendra l'échange de bonnes pratiques dans le développement des programmes d'enseignement ainsi que la réalisation de projets de recherche conjoints destinés à développer ce champ d'investigation.

La science des données constitue un axe stratégique de recherche pour le Domaine des EPF durant la période 2017–2020. L'*Initiative for Data Science in Switzerland* aura pour objectif de garantir que le Domaine des EPF et la Suisse acquièrent l'expérience nécessaire dans cette discipline et restent compétitifs à l'échelle mondiale.

Swiss Data Science Center

Le *Swiss Data Science Center (SDSC)* sera mis sur pied dans le cadre de l'*Initiative for Data Science in Switzerland*. Il sera dédié à la fragmentation des paysages de données et des perspectives analytiques à laquelle on assiste actuellement (voir encadré). Dirigé conjointement par l'EPFL et l'ETH Zurich, le SDSC sera constitué d'une équipe multidisciplinaire répartie sur différents sites et réunissant des chercheurs en science des données ainsi que des experts en santé personnalisée et en médecine personnalisée, en sciences de la terre et de l'environnement, en sciences sociales et en humanités digitales ainsi qu'en économie. La recherche génomique en médecine personnalisée nécessite par exemple une analyse des données moléculaires et cliniques d'un genre nouveau, qui soit efficace et respecte la confidentialité. Pour le Big Data, il s'agit là d'un immense défi qui constitue actuellement un obstacle majeur.

La vision du SDSC est de dégager des connaissances exploitables à partir de différentes sources de données et de



Science des données – Découvrir des connaissances exploitables à partir de sources de données complexes.
(Illustration: iStock.com/derrrek)

réunir fournisseurs de données, experts en science des données et en informatique et autres spécialistes autour d'une plateforme analytique de pointe offrant des *Insights-as-a-Service*, tout en prenant en compte les aspects liés à la sécurité des données et à la confidentialité. Les synergies inédites qui seront générées grâce au SDSC entre les institutions du Domaine des EPF ainsi qu'entre le monde académique et l'industrie – aussi bien en science des données que dans d'autres champs d'investigation sélectionnés avec soin – permettront de favoriser des découvertes scientifiques ayant un impact majeur pour la société.

Afin d'accomplir cette mission, le centre aura pour objectif de:

- développer en son sein un réseau d'assistance en science des données pour encourager une collaboration étroite avec les groupes de recherche et promouvoir la coopération entre utilisateurs et scientifiques;
- offrir des services complets en science des données aux personnes actives dans la recherche et le développement en Suisse comme à l'étranger en leur proposant en particulier des solutions de logiciels hébergés pour des *Insights-as-a-Service*;

- créer une communauté pour favoriser le partage d'outils, de méthodes et de connaissances dans cette discipline.

Plutôt que d'élaborer une infrastructure distincte pour ses activités, le SDSC utilisera les infrastructures du Domaine des EPF pour diffuser ses services en ligne (et plus particulièrement les ressources du Centre suisse de calcul scientifique, CSCS), ainsi que celles de SWITCH et de fournisseurs de services informatiques en nuage (*cloud computing*).

Science des données

La quatrième révolution industrielle se caractérise par la convergence de systèmes connectés, dits intelligents (*smart systems*), et par des percées dans des secteurs tels que le séquençage génomique et les nanotechnologies. Alors que les données sont considérées comme le nouveau pétrole de la recherche, on ne leur accorde que peu de valeur lorsqu'elles sont à l'état brut. En revanche, si ces informations brutes pouvaient être extraites, affinées et acheminées jusqu'aux endroits où elles ont une influence sur les décisions, leur valeur serait susceptible d'augmenter de manière spectaculaire.

La science des données incarne la nouvelle approche qui permettra de réaliser cette vision. A mi-chemin entre plusieurs disciplines académiques, dont la gestion et l'ingénierie des données, la statistique, l'apprentissage automatique, l'algorithmique, l'optimisation et la visualisation, elle offre aux sciences sociales, à l'économie, à la médecine, aux sciences de l'environnement et à bien d'autres encore un nouvel outil pour comprendre et influencer des systèmes complexes concrets et faire face aux grands enjeux de notre époque.

Malheureusement, les données ainsi que les méthodes et systèmes analytiques disponibles actuellement sont extrêmement fragmentés. Pour des non-spécialistes, il est très difficile de les exploiter correctement, et l'interprétation des résultats est particulièrement ardue. En d'autres termes, il existe un fossé entre les personnes qui créent les données, celles qui développent l'analyse et les systèmes et celles qui pourraient potentiellement en extraire la valeur. Ce fossé se creuse d'autant plus que les données en question revêtent la plupart du temps un caractère sensible. Pour que la science des données puisse être exploitée au maximum de son potentiel, il est donc essentiel d'instaurer un climat de confiance grâce au respect des dispositions légales (*compliance*) et à la préservation de la confidentialité, ainsi qu'à la sécurité du stockage et du traitement des données.

L'*Initiative for Data Science in Switzerland* entend relever ces défis dans l'intérêt de la Suisse, de la communauté scientifique et du monde entier.

Master en science des données

L'essor rapide de la discipline suscite non seulement un besoin accru de spécialistes bénéficiant d'une formation de pointe doublée d'un profil d'expertise adapté aux nouvelles exigences, mais aussi une nécessité pour les chercheurs et praticiens issus de différentes disciplines de posséder des connaissances méthodologiques de base. Le nombre d'étudiants participant à des cours en rapport avec la science des données augmente rapidement. Pour répondre à cette demande, des cursus de master en science des données seront proposés à l'EPFL et à l'ETH Zurich, et une large palette de cours destinés à différents domaines d'application seront mis sur pied. Par la suite, ces cours pourront être intégrés à un programme de formation continue destiné aux professionnels en Suisse.

Le but de ces nouveaux cursus de master est de former des diplômés qui maîtrisent parfaitement les aspects les plus variés de la science des données, depuis les bases de mathématiques et d'analytique aux systèmes et logiciels, en passant par l'application pratique des différentes méthodes.

La partie principale des cursus de master sera constituée d'une formation complète en analytique des données, composée de cours sur l'apprentissage automatique, les statistiques, l'exploration de données, le langage naturel, le traitement d'images et de signaux et l'intelligence artificielle. A ce module de base viendront s'ajouter des cours portant sur la gestion des données, le *cloud computing* et les systèmes informatiques, la sécurité et la confidentialité ainsi que sur les modèles de programmation. Ces différents enseignements permettront aux diplômés d'exceller dans leur profession. L'EPFL et l'ETH Zurich prévoient de proposer ces nouveaux cursus de master à partir du semestre d'automne 2017.

Pour plus d'informations

www.datascience.ch

Editeur

Conseil des EPF
Fritz Schiesser
Président
Haldeliweg 15, CH-8092 Zurich
fritz.schiesser@ethrat.ch
www.cepf.ch

Renseignements

Swiss Data Science Center (SDSC)
Olivier Verscheure
Directeur exécutif
EPFL SDSC, INN 314, Station 14
CH-1015 Lausanne
olivier.verscheure@epfl.ch

Zurich et Berne, mai 2016