

**Zeitschrift:** bulletin.ch / Electrosuisse  
**Herausgeber:** Electrosuisse  
**Band:** 105 (2014)  
**Heft:** 8

**Artikel:** Les défis à relever dans le domaine des données massives  
**Autor:** Hengsberger, Cynthia  
**DOI:** <https://doi.org/10.5169/seals-856280>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 14.04.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# Les défis à relever dans le domaine des données massives

## Retour sur l'IC Research Day de l'EPFL

Il y avait foule le 12 juin au SwissTech Convention Center de l'EPFL pour l'édition 2014 de l'IC Research Day. Cet événement désormais traditionnel était consacré cette année au thème « Challenges in Big Data », soit les défis à relever notamment dans les domaines du stockage, de la transformation, de l'analyse et de la sécurité de grandes quantités de données informatiques. Un sujet actuel, en plein essor, pour une journée riche en informations.

### Cynthia Hengsberger

Organisé une fois l'an par la Faculté Informatique et Communications (IC) de l'EPFL, l'IC Research Day rencontre toujours un aussi franc succès! Pour cette 18<sup>e</sup> édition, il a investi le tout nouveau SwissTech Convention Center inauguré en avril dernier. Contrairement aux années précédentes, les organisateurs n'ont ainsi pas été contraints à refuser du monde par manque de place. Ce sont donc plus de 700 participants qui ont pu suivre, le 12 juin, la matinée de conférences consacrée au thème « Challenges in Big Data » et se renseigner sur les projets en cours à la Faculté IC lors de l'après-midi.

L'événement a également offert une excellente occasion de découvrir le dernier-né des bâtiments de l'EPFL, dont l'élégance et le côté novateur de l'architecture séduisent instantanément. Afin de limiter sa consommation énergétique, le SwissTech Convention Center fait en outre appel à des technologies innovantes telles que la climatisation par récupération de froid dans le sol par le biais des piliers de soutien ou la génération d'électricité grâce aux panneaux photovoltaïques orientables en cellules à colorants. Ces derniers présentent d'ailleurs l'avantage d'allier l'utile à l'agréable, puisque les cellules de Grätzel semi-transparentes de différentes couleurs créent une atmosphère chaleureuse bienvenue à l'intérieur d'un si grand bâtiment. Mais revenons au thème du jour...

### Le Big Data et ses challenges

Nous sommes récemment entrés dans l'ère de la révolution digitale. En effet, si

le nombre d'ordinateurs personnels vendus par année de par le monde ne cesse de croître, les chiffres concernant les tablettes, et surtout les smartphones, explosent. Si bien que l'on prévoit pour 2020 plus de 50 milliards d'objets connectés!

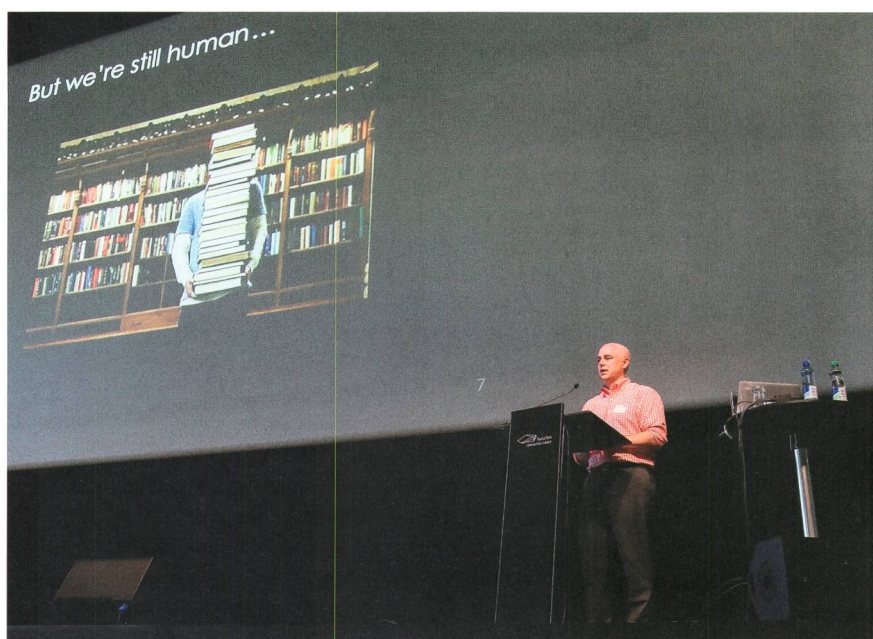
Rien d'étonnant donc dans le fait que le volume des données croisse de façon démesurée... Nous en générons tous, tout comme divers services (médicaux ou financiers, par exemple). C'est aussi le cas des objets et des nouvelles applica-

tions avec les réseaux de capteurs intelligents, les systèmes embarqués ou encore les futurs smart grids. En outre, l'éducation avec les tout nouveaux MOOCs (massive open online courses), la numérisation d'archives et les sciences (les observations, mesures, simulations, etc.) produisent également des volumes considérables de données.

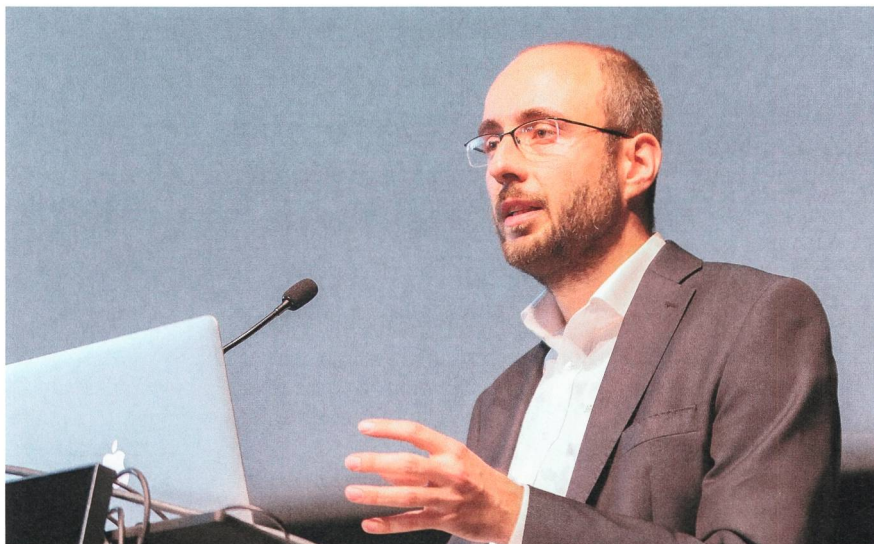
La révolution digitale offre certes une quantité d'avantages et de nouvelles possibilités, mais elle soulève également quelques défis de taille à relever... En effet, comment stocker ce déluge de données? Comment transformer ces dernières, les sécuriser, en tirer efficacement de l'information? Autant de questions auxquelles il devient urgent de répondre.

### Extraire efficacement de l'information

L'Homme a acquis au cours du temps des connaissances considérables. Mais lors de la réalisation d'un projet, il peut se révéler très compliqué de réunir toute l'information nécessaire. Impossible en effet de lire des milliers de livres ou



Christopher Re, Professeur assistant à l'Université de Stanford, Californie, développe un outil de recherche performant pour extraire efficacement les informations convoitées contenues dans des quantités phénoménales de publications.



Dans un exposé des plus intéressants, Martin Rehak, Cisco, Prague, a présenté une technique permettant de détecter des malwares au sein de données massives tout en préservant la confidentialité des données.

articles pour en relever les données pertinentes. Comment extraire de manière ciblée de l'information dans des publications? Faire une recherche avec des mots clés une fois les contenus digitalisés aide certainement, mais il n'en reste pas moins quelques obstacles à surmonter. En effet, certaines informations sont données sous la forme de graphiques ou de tableaux, ou sont difficilement interprétables une fois sorties de leur contexte.

Christopher Re, Professeur assistant à l'Université de Stanford, Californie, a présenté DeepDive, un système de données pour la science macroscopique, soit la science qui requiert de retirer de l'information d'une grande quantité de sources différentes. Ce système statistique d'extraction et d'intégration de données a fait preuve d'excellents résultats. Basé sur des techniques classiques de gestion de données complétées par des procédés de calculs statistiques particulièrement efficaces, il a même surpassé des volontaires, en quantité d'informations extraites mais aussi en précision, lors de recherches dans le domaine de la paléobiologie.

### Question de sécurité

Cerner un malware dans une quantité phénoménale de données revient à chercher une aiguille dans une botte de foin. Une tâche d'autant plus compliquée que les cybercriminels développent des stratégies de plus en plus perfectionnées. Alors, comment procéder? Martin Rehak, Principal Engineer chez Cisco, Prague, utilise une technique similaire à

celle des chercheurs d'or: éliminer de façon automatique les données sûres par pas successifs, pour ne garder qu'un volume réduit de données suspectes. Ces dernières sont alors examinées plus en détail.

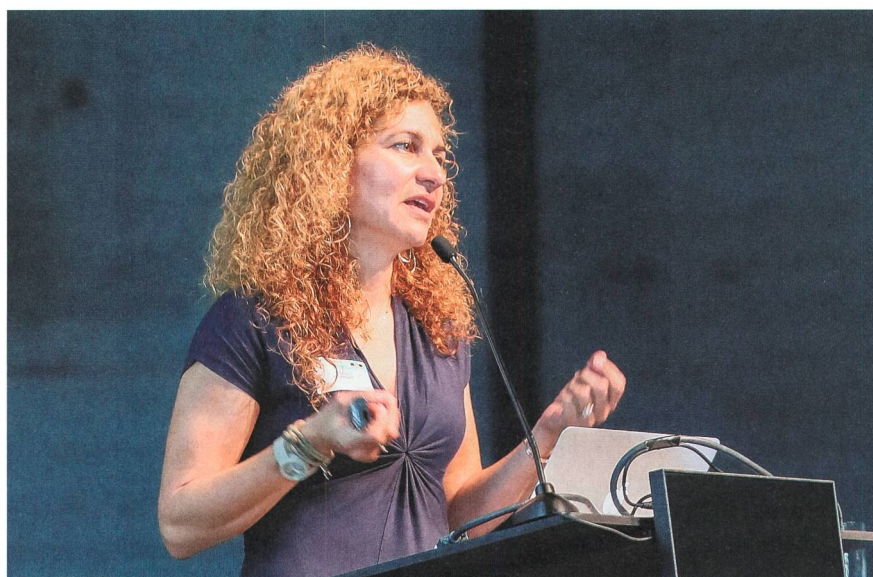
Pour ce faire, diverses approches et techniques (apprentissage automatique, calcul distribué, analytique des données massives, etc.) sont utilisées successivement. Les flux de données sont dans un premier temps analysés « en vol » (sans étape de sauvegarde préliminaire) par le biais de techniques d'apprentissage automatique très efficaces destinées à détecter des anomalies dans les flux tout en respectant la confidentialité des données. Celles qui n'ont pas pu être éliminées

(car pouvant représenter un danger potentiel) sont ensuite transformées, stockées, puis analysées plus intensément. Ce procédé permet de réduire le nombre de données à analyser d'un facteur  $10^6$  (1000 à 5000 incidents à vérifier pour 5 milliards de requêtes quotidiennes).

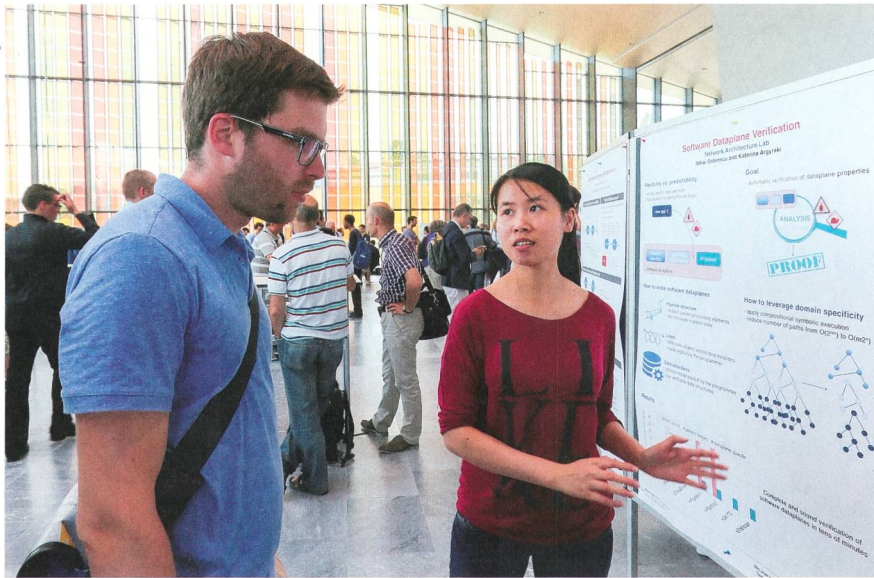
### Traitement des données in-situ

Le volume de données produites annuellement en 2020 devrait être multiplié par un facteur 50 par rapport aux chiffres de 2010. Dans certains domaines de recherche scientifique, il n'est déjà plus possible de toutes les enregistrer, qu'elles proviennent d'observations, de mesures ou de simulations, même temporairement. Par exemple, le LHC (Large Hadron Collider, le plus puissant accélérateur de particules au monde) du CERN génère 20 millions d'événements bruts par seconde, chacun d'un volume de 1 à 2 MB. Ces données ne peuvent bien évidemment pas être toutes enregistrées, c'est pourquoi des filtres sont utilisés successivement en vol afin d'en réduire le nombre effectivement stocké à environ 300 par seconde.

Cette problématique devient de plus en plus présente également dans d'autres domaines. Les neurosciences en sont un exemple avec la simulation du comportement des 86 milliards de neurones d'un cerveau humain pour le Human Brain Project, projet phare européen dont l'objectif consiste à comprendre le fonctionnement du cerveau humain, voire même la cognition. La Faculté IC



Anastasia Ailamaki, Professeure à la Faculté IC de l'EPFL, a expliqué comment trier des données massives à l'état brut afin de ne stocker que les plus intéressantes.



Les participants ont pu se renseigner sur les projets en cours à la Faculté IC, présentés sous la forme de près d'une centaine de posters et démonstrations.

de l'EPFL y participe d'ailleurs activement en développant les outils informatiques indispensables à son succès.

Dans un tel projet, il est également essentiel de pouvoir explorer les données à l'état brut de manière efficace au fur et à mesure qu'elles sont générées et de pro-

céder à une sélection pour ne conserver que les données réellement intéressantes. Comme l'a expliqué Anastasia Ailamaki, Professeure à la Faculté IC de l'EPFL, ceci peut être réalisé grâce à la virtualisation des données, c'est-à-dire en traitant les données à l'état brut (quel que soit

leur format) et en les analysant grâce à des requêtes, et ce, sans devoir au préalable intégrer ces données brutes dans une base de données (projet In-Situ).

### Les projets en cours à la Faculté IC

L'après-midi a, quant à lui, été consacré à la présentation sous forme de posters de près d'une centaine de projets de recherche. Le public a ainsi pu se renseigner sur des sujets aussi variés que la simulation de la propagation de la peste au Moyen Âge, la détection des émotions ou encore les solutions en vue d'améliorer le contrôle et l'efficacité des réseaux électriques de distribution.

Vous regrettez d'avoir manqué cet événement? Rien n'est perdu! Le 22 novembre 2014, la Faculté IC proposera une journée « Portes ouvertes » à l'occasion des 30 ans de la section Informatique et des 20 ans de la section Systèmes de communication (informations sous : [ic.epfl.ch/portes-ouvertes](http://ic.epfl.ch/portes-ouvertes)). N'hésitez donc pas à venir y découvrir par vous-même les nombreux et fascinants travaux de recherche lors de cet événement grand public!

Anzeige

## «Der Kontrollprofi»

### EurotestEASI MI 3100B SE

Neu



**Einfach – Selbsterklärend – menügeführt**  
**NIV und NIN Installationsprüfungen**  
**FI Prüfungen Typ B und B+**  
**Automatische Testabläufe, Messdatenspeicher**  
**Schnittstellen RS232 und USB**  
**Software EuroLinkPRO inkl.**

**ELKO**  
SYSTEME AG

Messgeräte • Systeme • Anlagen  
Zur Kontrolle und Optimierung des Verbrauches elektrischer Energie  
Brüelstrasse 47 CH-4312 Magden Telefon 061-845 91 45 Telefax 061-845 91 40  
E-Mail: [elko@elko.ch](mailto:elko@elko.ch) Internet: [www.elko.ch](http://www.elko.ch)



### LANZ – die sichere Kabelführung für Metro-, Bahn- und Strassentunnel zu international konkurrenzfähigen Preisen:

LANZ Produkte für den Tunnelbau sind **3-fach geprüft**

1. auf Erdbebensicherheit SIA 261 Eurocode 8 (EMPA)
2. auf Schocksicherheit 1 bar Basisschutz (ACS Spiez)
3. auf Funktionserhalt im Brandfall 90 Minuten (Erwitte)

Für die Kabelführung in Tunnel 3-fach geprüft sind:

- die LANZ G-Kanäle für kleine und mittlere Kabelmengen. Schraubenlos montierbar. Stahl PE-beschichtet und Stahl A4
- die LANZ Weitspann-Multibahnen (Kabelleiter nach IEC 61537). Für grosse Kabelmengen, hohe Belastung und weite Stützabstände. Stahl tauchfeuerverzinkt und Stahl rostfrei A4 WN 1.4571 und 1.4539
- die LANZ MULTIFIX C-Profilischiene mit eingerollter 5-mm-Verzahnung zur Befestigung u. a. von Rohren, Leuchten, Schildern.

Für die Stromversorgung in Tunnel 3-fach geprüft sind:

- die LANZ HE Stromschienen/Schieneverteiler IP 68 400–6000 A. 4-, 5- und 6-Leiter Alu und CU. Korrosionsfest giessharzvergossen.

Risiken vermeiden. Sicherheit erhöhen. LANZ montieren.

Rufen Sie LANZ an für Referenzen, Beratung, Muster und Offerten:  
**lanz oensingen ag CH-4702 Oensingen Tel. 062 388 21 21**



**lanz oensingen ag**  
CH-4702 Oensingen Südringstrasse 2  
Telefon 062 388 21 21 Fax 062 388 24 24  
[www.lanz-oens.com](http://www.lanz-oens.com) [info@lanz-oens.com](mailto:info@lanz-oens.com)