November/December 2017, No 3

Swiss Alliance for
Data-Intensive Services

Newsletter

Welcome to the third member newsletter. In this issue we want to introduce a new format: the member interview. The aim is to highlight different people, who are active within our alliance, and their diverse backgrounds. You'll find the interview right at the beginning of this newsletter. Furthermore, the expert group "Data Integration, Governance and Sharing" is introducing itself.
Please also note that the General Assembly of our alliance will take place on November 16 from 4.30pm at Berner Fachhochschule. Every representative is invited to bring up to two guests to the assembly. If you haven't already, please let the office know about your attendance.

Happy reading!
The Editors

Member Interview

Thorsten Busch, René Pfitzner

In this interview we are talking about data from a social science perspective, surveillance and data bias with Thorsten Busch.

Thorsten, who are you and what are you doing?
I am a business ethicist who specializes in all sorts of digital issues, e.g. workplace surveillance, social media, and digital games. I am especially interested in how private companies regulate public spaces, e.g. when it comes to challenges such as trolling and privacy. I studied political science, economics, and management in Oldenburg, did my Ph.D. in St. Gallen, was a visiting scholar at Harvard's Berkman Center for Internet and Society, and then worked as a postdoc in Montréal for a couple of years. Now I am back in St. Gallen and very much looking forward to seeing what interesting developments in Switzerland I have missed in recent years.

Why are you a member of Data+Service and what is your role in the Alliance?
Within the network, I am one of several people running our "Data Ethics" expert group. Within the expert group, we exchange ideas on various ethical questions from an academic but also, and especially, from a business perspective. In fact, we understand that many institutions struggle with the question which use cases that deal with large amounts of (customer) data are ethically legitimate, and which ones are not. For example, there are questions of confidentiality regarding customer data, ethical issues with new (data-driven) business models and the trade-off between the need to generate profits vs. the right of the customer to decide on the use of her personal data.

Where does your interest in "data topics" come from?
Teenage trauma, I guess. ;) I have been interested in technology from an early age on and maintained that interest. I wrote my Master's thesis about open source software, researched questions concerning social media governance during my Ph.D., and investigated digital games during my years as postdoc. Currently, at the University of St. Gallen, I am part of an interdisciplinary project that investigates the question how companies in Switzerland collect and evaluate information about their employees. There are a lot of open questions in this

fast-moving field, which constantly keeps technology and data topics exciting.

Your academic background is in the social sciences. How do you experience the communication with your more "technical" peers, and where do we still have to break down barriers?
That is a hard question. It is true that cross-disciplinary communication is complicated, as we all speak highly specialized "foreign languages". And if you try to bring together not only academics, but also managers, politicians and other groups, it gets even harder. In an ideal world, we would already start in high school, but during Bachelor studies at the very latest, to actively teach interdisciplinary thinking and argumentation. Actually, it's not that hard: Much can be achieved already if we start to actively listen to each other – although a bit of basic knowledge in how communication works would certainly help, as well. One particular issue when interacting with more quantitatively-minded people is that it is often hard to convince them based on qualitative arguments. On the other side, for more qualitatively thinking people like me, it can be challenging to understand technical jargon. In this case, it need attempts from both sides to be more transparent, clear and patient in their communication.

In your opinion: What are the biggest societal challenges in the field of "data"?
I think we have a real surveillance problem, which most likely will get even worse in the next few years. I often travel to the U.S., and the extent of surveillance there is horrific. Especially worrying is the amount of (well-documented) public-private partnerships in the surveillance industry. For instance, law enforcement agencies in democratic countries can subpoena user information from social networks for dubious reasons, and without the social networks even being allowed to talk about it to the public. Moreover, private companies from democratic countries frequently sell surveillance technologies to authoritarian regimes. These kinds of things may present enormous opportunities for economic growth. However, from a democratic perspective, they are very problematic and need to be discussed openly and transparently. This will be a hot debate for years to come, and it is difficult to find the right balance between hysteria and ignorance.

What are the most exciting developments in this field?
For me, it is particularly interesting to observe that many companies, after years of ignorance, have begun recognizing that data is not morally neutral. In recent years, there have been many scandals, e.g. with the face recognition algorithms of Google and HP, which were unable to recognize dark-skinned people. Now, not only nagging internet researchers like me complain about bias in AI and machine learning, but companies like Google themselves address the fact that we often have biased data sets. A lot has happened on this particular subject in a relatively short amount of time, and there is an acute sense of awareness for this kind of discrimination now, which makes me hopeful for the future.

Thorsten Busch is a Senior Research Fellow at the Institute for Business Ethics at University of St. Gallen. He is one of the organizers of the expert group "Data Ethics".

Expert group: Data integration, governance and sharing

CHRISTIAN SPINDLER

Imagine your organization had an efficient, compliant and trustable way to integrate data from various systems, securely exchange the data both internally and externally, and extract value. We call such data "Trusted Data" and formed an expert group around this vision, working and helping on questions of integration, governance and sharing organizational data. Having kicked-off in spring this year, we have identified three focus topics along the value chain of data processing in organizations: A thought-through software architecture is a key success factor in any data integration task. It allows for standardizing data formats, build-

ing flexible interfaces, and secures highest data quality for subsequent business processes. It enables building up modular data-driven systems that can rely on the defined quality standards to generate insights from the combination of formerly separated information.

Once a target architecture is in place, data feeds need to be combined from the various sources of an organization. This can turn out as a cumbersome task, as data is most often drawn from different source frameworks (data bases, plain tables, enterprise management systems, shared drives), is strongly diverse in quality (e.g. missing and/or wrong data, inappropriate granularity, redundant naming for same entity) and comes in both structured form (e.g. numeric values such as prices, timestamps or categorical text such as country names) and unstructured form (e.g. plain text from emails or documents, images, video and audio data e.g. from marketing or monitoring activities). Vice versa, data with similar naming can have totally different meaning in disparate systems.

Machine learning promises to automate at least parts of the integration and to greatly reduce the time and effort spent on data integration. Machine learning works by identifying patterns across data sources that can be exploited to learn if two data streams from distinct sources actually describe the same matter. We call this topic semi-automatic data integration.

Finally, organizations strive to govern and protect their valuable data to generate business models that possibly return the investment on the conducted data integration effort. The expert group is particularly interested in the questions and opportunities resulting from the upcoming regulatory revision of data protection in the EU and in Switzerland. As the vast majority of businesses are affected by the revision, we seek for answers on a general (strategic) and operational level, solving immediate needs of the expert group participants. If you would like to join this group, please feel free to contact Christian Spindler (christian.spindler@data-service-alliance.ch) or the Data+Service office (info.office@data-service-alliance.ch).

Call for Innovation Bootcamps

In a current call for innovation bootcamps, our industrial member PwC is seeking experts on the following topics (independent of each other): (1) Best practice in application of derivative-free optimisation methods, (2) Adversial Attacks on DL-Models and ensemble adversial training, (3) Best practice in generation of data when applying GANs, (4) Deep Reinforcement Learning. Each expert is asked to hold an informal half-day workshop (hands-on) for parts of PwC's analytics team; the workshops will be fully paid. Please inform the office at info.office@data-service-alliance.ch if you want to offer such a workshop until Nov 16.

Forthcoming

16.11.2017 - General Assembly
Data+Service General Assembly in Bern.

21.11.2017 - BPM Symposium
Experts from industry and academia present successful cases of digital projects.

22.11.2017 - IFZ FinTech Colloquium
FinTech Colloquium organized by Lucerne University of Applied Sciences.

23.11.2017 - Big Data Leadership
On ethics, privacy and leadership in a world of "big data", organized by the Lassalle Institute.

07.06.2018 - SDS 2018
Mark the date for the 5. Swiss Conference on Data Science.

More information about all these events can be found at:
www.data-service-alliance.ch

Impressum

Editors:
Alexander Grimm, René Pfitzner

Contact:
editors@data-service-alliance.ch

Editorial Deadline:
08.01.2017

Official Member Newsletter of the *Swiss Alliance for Data-Intensive Services*.

Zurich, November 10, 2017