



SDSI2015

2nd Swiss Workshop on
Data Science

Program

12th June 2015



Preface



A warm welcome to the SDS|2015 conference on Data Science!

It has been more than a year ago that my personal expectations for SDS|2014 have been exceeded greatly: We from the ZHAW Datalab planned for a small event, expected at best 50 attendees (including us), and have been overwhelmed by 120 participants, great support from industry and very encouraging feedback from the community. That last item mattered most to us: From the start, the SDS conferences have been planned as community events. Participants should experience the emergence of a Swiss Data Science community and realizing themselves as part of it. Thus, ample time for discussions

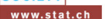
and networking are of immense importance to us as organizers. That's why we are very happy to strengthen this community aspect this year in several respects:

- Most of the talks are contributions from the community (academic and industry alike), and we are very grateful for that!
- Our sponsors enabled us to keep the registration fees as low as possible by funding most of the organizational expenses. Additionally, the Hasler Stiftung granted considerable deficit guarantees that will most likely cause SDS|2015 to have a balanced result financially – thank you all very much!
- Several individuals and companies have been involved in realizing the Interactive Track, the design goal of which is to bring new people into spontaneous close collaboration – community building on a small scale.

Anyway, looking ahead to SDS|2016, some things remain to be improved (maybe starting with registration options...). And so I want to close the preface with a call to the community: Consider, together with your organization, contributing your resources to SDS|2016. Have a location? Want to manage a track? Have another idea and the resources to implement it? Get in touch!

Enjoy the conference, and have great ideas & contacts!

Thilo Stadelmann
SDS|2015 General chair



HASLERSTIFTUNG

Content

Talks & Speakers.....	4
Interactive Track.....	18
Program	20
Sponsors & Supporters.....	19
Feedback.....	36
Notes	37
Impressum.....	42

The use of ever increasing datasets in macroeconomic forecasting

Abstract. KOF Swiss Economic Institute is probably best known for its forecasting activities. Besides by macroeconomic theory, these forecasts are largely driven by ever bigger datasets. Whereas traditionally microeconomic, i.e. firm- and household-specific, information is first aggregated to the macroeconomic, i.e. national, level data before macroeconomic relationships are estimated and used in producing such forecasts, there is a clear tendency to more and more analyse macroeconomic relationships using micro data more directly. This sometimes allows new insights that cannot be achieved by looking at solely macroeconomic data.

An example of this is given by the KOF Surprise Indicator. It is based on the semantic identification of survey answers of the KOF survey in the Swiss manufacturing industry. By using the firm-specific answers, it can estimate to what extent each firm has been surprised during the previous quarter. By subsequently aggregating this information, the KOF Surprise Indicator identifies the growth path of the Swiss economy and allows for accurate estimates of the current growth rate of the Gross Domestic Product (GDP).

Data limitations still make such examples in which microeconomic data are directly used in producing a macroeconomic indicator rare. Nevertheless, also the substantial increase in macroeconomic time series allows for the use of more sophisticated techniques to aggregate information. Since 1976, the KOF Swiss Economic Institute has been publishing a leading indicator, the KOF Economic Barometer, predicting how the Swiss economy will develop in the near future. Whereas in the previous versions of the KOF Barometer six to 25 variables were selected based on expert knowledge, the 2014 version contains over 200 variables selected out of a set of close to 4500. For that a relatively automated and robust selection procedure has been developed. This procedure is re-run every autumn allowing the Barometer to 'learn'. Subsequently, the selected variables are subsequently transformed into one leading indicator. The increased number of variables allows a – compared to its previous versions – relatively stable and robust KOF Barometer that hence no longer has to rely on filtering techniques to reduce the noise in the final indicator.

Biography. Jan-Egbert Sturm (Ph.D. University of Groningen, 1997) is Professor of Applied Macroeconomics as well as Director of the KOF Swiss Economic Institute at the ETH Zurich. He was researcher at the University of Groningen, The Netherlands, until 2001, and taught as Visiting Professor at the School of Business, Bond University, Gold



Coast, Australia, 2000 and 2005. As Head of the Department for Economic Forecasting and Financial Markets at the Ifo Institute for Economic Research, he was also Professor of Economics at the University of Munich (LMU) at the Center for Economic Studies (CES), 2001-2003. He held the Chair of Monetary Economics in Open Economies at the University of Konstanz, Germany, which was coupled with the position of Director of the Thurgau Institute of Economics (TWI) in Kreuzlingen, Switzerland, 2003-2005.

In his research, Jan-Egbert Sturm relies heavily on empirical methods and statistics, concentrating on monetary economics, macroeconomics as well as political economy and with a special interest in fields which are closely related to practical and current problems. His applied studies have focused on, for example, economic growth and central bank policy. He has published several books, contributed articles to various anthologies and internationally renowned journals. Jan-Egbert Sturm headed the Ifo research team at the Joint Analysis of the Six Leading German Economic Research Institute, 2001-2003.

Since 2001 he has been member of the CESifo Research Network and since 2003 Research Professor at the Ifo Institute. In 2005 he was appointed to the European Economic Advisory Group of the CESifo. In 2006 he was appointed member of the user advisory council of the Ifo Institute. At the beginning of 2007, he became President of the Centre for International Research on Economic Tendency Surveys (CIRET). Since 2013 he is editor of the European Journal of Political Economy. Currently he is member of several advisory commissions in Switzerland.

**The Emerging Discipline of Data Science:
Principles and Techniques for Data-Intensive Analysis**

Abstract. The Scientific Revolution (1550-1700) led to the increasing significance, potential, and risks of empiricism – 17th Century knowledge discovery - that in turn led over 400 years to the Scientific Method – a body of principles and techniques for *investigating phenomena, acquiring new knowledge, and correcting and integrating previous knowledge*.

The Computing Revolution (1940-1970) led to the increasing significance, potential, and risks of software – 20th Century knowledge work – including the software crisis (1968) that in turn led over 40 years to Software Engineering - a body of principles and techniques for *the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software*.

The Digital Revolution (1970-) with the emerging Digital Universe and Big Data Revolution (2000-) is leading to the significance, potential, and risks of data-intensive analysis – 21st Century knowledge discovery – that is leading to the need for Data Science – an emerging discipline currently in its infancy, analogous to the scientific method and software engineering in their revolutions. The importance of Data Science can be seen in the potential impact on the quality of lives of the US Government’s Precision Medicine Initiative for “Delivering the right treatments, at the right time, every time to the right person.”

This exploratory talk examines Data Science from data analysis to data-intensive analysis. Data analysis with roots in Babylonia (1700-1200 BCE) and India (1200 BCE) is applied in most human endeavors following well-established principles, e.g., statistics, and guidelines, e.g., the Cross-Industry Standard Process for Data Mining. The roots of data-intensive analysis are in Big Data (~2000) that, just emerging, is opening the door to profound change – to new ways of thinking, problem solving, and processing that in turn bring new opportunities and challenges. Since 2007, this *Fourth Paradigm [10]* of science is being applied to evidence/data-based analysis in most human endeavors.

The talk presents an emerging data-intensive analysis workflow that augments the previously dominant data analysis phase with an equally important and substantial data management phase and correspondingly augments the scope of Data Science. Through use cases we identify opportunities and challenges across the data-intensive analysis workflow and their requirements for *principles and techniques to measure and improve the correctness, completeness, and efficiency of data-intensive analysis*.

Biography. Dr. Brodie has over 40 years experience in research and industrial practice in databases, distributed systems, integration, artificial intelligence, and multi-disciplinary problem solving. He is concerned with the Big Picture aspects of information ecosystems including business, economic, social, application, and technical. Dr. Brodie is a



Research Scientist, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology; advises startups; serves on Advisory Boards of national and international research organizations; and is an adjunct professor at the National University of Ireland, Galway. For over 20 years he served as Chief Scientist of IT, Verizon, a Fortune 20 company, responsible for advanced technologies, architectures, and methodologies for Information Technology strategies and for guiding industrial scale deployments of emergent technologies. His current research and applied interests include Big Data, Data Science, and data curation at scale and a related start up Tamr.com. He has served on several National Academy of Science committees. Dr. Brodie holds a PhD in Databases from the University of Toronto and a

Doctor of Science (honoris causa) from the National University of Ireland.

Tackling Challenging Problems in Academia and Industry: An Interdisciplinary Approach

Abstract. The ZHAW Datalab was founded in early 2013 as one of the first Data Science laboratories in Europe. The main goals are to foster interdisciplinary teaching and applied research to tackle challenging problems both in academia and industry.

In this talk we shed light on the interdisciplinary approach of the ZHAW Datalab. First, we give an overview on our teaching activities and discuss how we educate aspiring Data Scientists of the future. Next, we highlight various lessons learned from three different applied research projects that we have been doing jointly with industry partners. In the first project we applied state-of-the-art data warehousing and machine learning techniques for discovering patterns in the user behavior of an e-commerce application. In the second project we used deep learning techniques for face recognition. Finally, in the third project we applied innovative techniques for text understanding such as sentiment analysis. The talk should inspire the audience of the workshop to engage in similar projects and, hence, to boost the innovation potential of an enterprise. to get personalized insights and digital services beyond terrifying Big Brother scenarios.

Biography. Dr. Kurt Stockinger is an Associate Professor (Dozent) of Computer Science and Director of Studies, Data Science at Zurich University of Applied Sciences (ZHAW). His research interests cover various aspects of Data Science, namely Big Data, data warehousing, business intelligence and advanced database technology. He is also on the Advisory Board of Callista Group AG. Before joining ZHAW in 2013, he was a Data Warehouse (DWH) and Business Intelligence (BI) Architect at Credit Suisse, Zurich where he worked on designing and implementing algorithms for a terabyte-scale enterprise data warehouse, data security as well as on several joint research projects between Credit Suisse and ETH Zurich. From 2004 to 2007 Kurt was a researcher at Lawrence Berkeley National Laboratory with focus on multi-dimensional indexing & query methods for large-scale scientific data as well as high-performance visual analytics. From 2000 to 2003 Kurt was heading the Replica Optimization Team of the EU Data Grid Project at CERN. In 2001 he was a visiting researcher at California Institute of Technology. Kurt has more than 60 publications in both scientific and industry journals and conferences. In 2008 he received an "R&D 100 Technology Award" (which is considered as the Oscar of innovations in the USA). He holds a Ph.D. in computer science from CERN / University of Vienna.

MIDATA.COOP – Citizen-Controlled Reuse of Personal Data

Abstract. First, by using free online services citizens increasingly lose control over their personal data and thus their digital identity. MIDATA.COOP organized as a Swiss cooperative empowers citizens and enables their **digital self determination** by creating a citizen-owned personal data platform on which citizens can securely store, access, control, manage and share their personal data. Each citizen should own and fully control her own 'data warehouse'. Citizens and society should be the primary beneficiaries of the growing economic and societal value of **the reuse** of aggregated personal data. The presentation contains both conceptual aspects of the MIDATA cooperative and a short demo of the MIDATA personal data platform prototype developed at ETH.

Biography. André Golliez studied computer science at ETH Zurich until 1986. He then worked as an IT manager in the engineering industry and the financial sector. In 1999 he and his partners founded the company itopia ag. Furthermore, he specialized in strategic IT consulting for public administrations. From 2004 to 2009 André Golliez was the president of the Swiss computer science society. The Open Government Data (OGD) – Study Switzerland was conducted under his direction. Also, he accompanied the running federal and cantonal OGD projects and co-authored the OGD Strategy Switzerland approved by the Federal Council in April 2014. Since January 2012, he is the president of the Association Opendata.ch (www.okfn.org), the Swiss Chapter of the Open Knowledge Foundation (www.okfn.org). Since February 2015 he is co-founder and boardmember of the MIDATA cooperative. In July 2015 André Golliez starts his own company Golliez Open Data Consulting.

Big Data Analytics - Evolving Toward Analytics 3.0

Abstract. Terabytes and petabytes of data pour into the organization from operational and transactional systems, from scanning and facilities management systems, from inbound and outbound customer contact points, from social/mobile media and the web.

The hopeful vision is to be able to harvest and harness every byte of relevant data and use it to make supremely informed decisions. By applying analytics to the seemingly unlimited flow of available data, data scientists can understand and address complex business issues in ways never before imagined – basing decisions on data-driven insight rather than intuition. Being data-driven pays off. Study after study confirms the obvious: Companies that invest in big data analytics develop deeper customer insights, discover new business leads, report higher win rates and are more profitable than their not-so-data-driven counterparts. You would think boards of directors and executives would be quick to embrace this winning culture, but it isn't happening to the extent you might imagine. At least not yet. Big data analytics is currently entering the operational areas where predictive analytics, such as data mining, were already well rooted and data-driven decisions were the norm – such as campaign management, churn and credit risk. However, the strongest performing companies will take a more holistic view of big data. They are not just investing in analytics technologies within discrete operational areas; they are creating a pervasive analytics culture – and that is still rare.

A first step into this direction can be the democratization of analytics or „Agile Business Intelligence“, where business users get much more freedom to gain insights out of (small and big) data. Technically this requires an In-Memory architecture that allows highly distributed computing, including analytical and statistical functionality.

Biography. Guido Oswald is Sr. Solution Architect at SAS and helps organizations to bring their Business Analytics infrastructure and processes to the next level. His focus areas are new technologies (like Hadoop, Docker, Cloud Foundry, SAS High Performance Analytics, etc.), Data Management and Agile Business Intelligence.

Cross-functional data analytics for Client Relationship Advisors for a global American Bank

Abstract. The talk will focus on introducing briefly the situation at the bank, detail their requirements, and discuss the solution. The main body of the talk will revolve around project lessons learned and how this American bank teams up with a small Swiss vendor to approach big data analytics of unstructured data in novel ways. It will present especially:

- Novel approaches in data visualization
- Lessons learned of implementing such a solution across the Atlantic
- Best practices in cloud vs on-premise deployment of big data analytics

Biography. Dr. Dorian Selz is a co-founder of Squirro. Before that he co-founded the Swiss search platform local.ch and was its CEO until December 2008. He was a partner & COO at Namics, Switzerland's largest E-Business consultancy and has a PHD in Information Systems from the University of St. Gallen.

50 shades of truth: visualisation and communication of uncertainty

Abstract. While the progress in data storage, data capacity, treatment and algorithms has been incredible in the recent few years the “last mile” between the output and the decision takers seems to be still lacking behind.

One of the reason is that it has not only to do with science and technology but also, possibly event mainly with human psychology, general affinity to math and sometimes exaggerated expectations from data analysis, statistics, etc.

We are planning to present a few aspects of things to consider in order to improve the quality of the information transfer in that last mile: especially considering the background of decision taking people, the attractiveness/sexiness of graphical representation of results and the personal affinity to statistical results.

We will also present concrete examples how these issues can be solved. Especially we will focus on two innovative ways of representing data analytics results that help understanding of complex information: the shaded radar plot and the bicolored pie chart.

Biography. Rodolphe Dewarrat studied mathematics and engineering sciences at the Universities of Fribourg (Switzerland) and Cambridge (United Kingdom). Since 2001 he has been working for various companies and start-ups as scientific developer and consultant. In 2012 Rodolphe Dewarrat has been launching his own consulting company, IMSD Sàrl, with offices in the so-called Watch Valley and in Zürich. IMSD is consulting company for data analysis issues or more general scientific communication issues in Switzerland and in Europe.

Privacy in the Big Data Era

Abstract. As we look at the dynamic change in our data-driven world, we see a global dialogue, polarized along competing but fundamental principles, unclear and pulled in variety of directions. Industry, government and civil society are uncertain on how to create a balanced privacy and data protection framework in the era of big data. As an ever larger amount of data is digitized and travels across organizational boundaries, there is a set of policy issues that will become increasingly important, including, but not limited to privacy, security, intellectual property and liability. Scigility addresses such legal and technological issues providing a unique Governance Framework for companies and public-sector organizations enabling them to cope with the values and the legal framework our society must apply to the new technologies and circumstances of our time.

Biography. Olivier Heuberger is a lawyer mainly active in the fields of ICT and intellectual property law, with a focus on privacy and data protection law, licensing, internet and software law. As Legal Counsel of Scigility, he advises companies, individuals and public-sector organizations in all fields of ICT-/IP and Contract Law. Before joining Scigility Olivier Heuberger studied Law and Economics at the Zurich University of Applied Science and made his master's degree in law at the University of Lucerne in 2009. He worked for a medium sized law firm in Zurich, as a Law Clerk at the District Court in Bülach as well as for a Big4-Consulting firm in Zurich. Olivier Heuberger made his bar exam in 2012 and is since then working as a lawyer in Zurich and as Legal Counsel for Scigility. He is currently writing his Ph.D. on IT-Law at the University of Lucerne.

Re-engineering IoT Legacy Analytics with Big Data

Abstract. In this project, we rescued a few legacy IoT solutions and made them faster by exploiting speed and performance of a big data platform based execution.

In our analytics solution park, a number of workflows are dedicated to Internet of Things applications, and particularly to the analysis of energy usage time series. One solution, in particular, predicts the amount of electrical energy usage for clusters of smart meter IDs in Ireland. The bottle neck of this solution, however, lies in the first ETL process, which takes a very long time to execute. This made the solution difficult to use in production, and challenging for re-training.

Relying on KNIME big data access and manipulation nodes, we transformed all ETL processes of our legacy solution into In-Database ETL processing nodes for Cloudera Impala, Apache Hive, and ParStream. The Database Manipulation nodes helped us to build the complex required SQL queries without a deep knowledge of the big data platform specific SQL dialect.

The (smaller) resulting data set was then pulled back into the KNIME analytics platform to proceed with the analytics and build the time series prediction model. The execution of this re-engineered ETL process allows for significant speed-ups and for more frequent model re-trainings!

During this talk we will guide the user step-by-step to connect to and run a series of ETL operations on any big data platform from the comfort of a visual analytics workbench, such as the KNIME Analytics Platform.

Biography. Dr Rosaria Silipo has been mining data, big and small, since her master degree in 1992. She kept mining data throughout all her doctoral program, her postdoctoral program, and most of her following job positions. She has many years of experience in data analytics, data visualization, data manipulation, reporting, business intelligence, training, and writing. In the last few years she has been using KNIME for all her projects, becoming an expert KNIME user and a KNIME certified trainer. She is also the author of more than 50 scientific publications and 3 books for data analysis practitioners.

Dr Bernd Wiswedel is one of the founders of KNIME and arguably the author of the first line of code. He has been with KNIME since the beginning and is CTO at KNIME.com AG, Zurich, Switzerland, where he coordinates the development activities of both the open source KNIME platform and the enterprise scale extensions. He holds a PhD with the University of Konstanz, Germany; his background is on data and information mining.

0 lines of code and in under 20 minutes: Building a customer churn application

Abstract. Handling customer churn is crucial to the success of every business and the key to sustainable customer management. However, as a busy data scientist, the churn problem is just one of many stacking up on your desk. Thanks to Azure Machine Learning, it is possible today to build predictive applications within hours rather than days or weeks, enabling you to keep up with all your projects and the business' expectations. Equipped with nothing but a web browser we will build a model that predicts customer churn and will operationalize it to make it ready for consumption by any line of business application within a few seconds, ready to deliver true business impact - all of this in less than 20 minutes and without requiring a single line of code.

Biography. Marc Schöni is a Technical Solution Professional for BI & Analytics at Microsoft in Switzerland with years of project experience in this space – and a hopeless data aficionado. Taking care of large and medium sized customers across the country, he supports them in strategic evaluations of analytics platforms and applications and delivers hands-on, proof-of-concept type of projects on a daily basis.

From microscopes to satellites: Interactive Image Analysis at Scale Using Spark

Abstract. Many important scientific questions ranging from the micro-scale mechanical behavior of foam structures to the process of viral infection in cells demand not only high spatial, but also temporal resolution. Detector improvements have made realizing many of these experiments possible, and have consequently produced a flood of rich image data. At the TOMCAT beamline of the Swiss Light Source peak acquisition rates reach 8GB/s [1] and frequently cumulate to 10s or 100s of terabytes per day. Parallel developments in satellite imaging have led to petabytes of multichannel images being collected every year providing high resolution images of the entire planet every few days. While visual inspection is invaluable, detailed quantitative analysis is essential for summarizing and comparing measurements and performing hypothesis tests. Even more important is the ability to detect outlier events and unexpected structures buried deep inside the voluminous data. Existing tools scale poorly beyond single computers and make this type of interactive exploration very difficult and time consuming. We have developed a scalable framework based on Apache Spark and the Resilient Distributed Datasets proposed in [2] for parallel, distributed, real-time image processing and quantitative analysis [3]. The distributed evaluation tool performs filtering, segmentation and shape analysis enabling data exploration and hypothesis testing over millions of structures with the time frame of an experiment. The tools have been tested with clusters containing thousands of machines and images containing more than 100 billion voxels. Furthermore we have expanded this tool using technologies like OpenLayers [4] and D3.js [5] to make the analysis reachable to even non-technical users. We show how these tools can be used to answer long-standing questions and see small genetically driven structural changes in bone, topology rearrangements in liquid foam, and track the course of infection in living cells. We finally demonstrate our future road map including real-time image processing using Spark Streaming and approximate analysis using BlinkDB.

Biography. Kevin Mader is the founder of 4Quant and a lecturer in the X-ray Microscopy Group within the Department for Information Technology and Electrical Engineering at ETH Zurich. His work focuses on turning big hairy 3D images into simple, robust, reproducible numbers without resorting to black boxes or magic. In particular, as part of several collaborations, he is currently working on automatically segmenting full animal zebrafish images, characterizing rheology in 3D flows, and measuring viral infection dynamics in cell lines.

Automated Data Curation at Scale

Abstract. Today, data comes from an ever increasing number of different data sources in various formats. Data scientists spend up to 80% of their time preparing data. Curating several hundred sources into a high-quality, unified data view is not even feasible, yet doing so automatically can boost both breadth and depth of insights acquired during data analysis. We present automated approaches to data de-noising and unification, and show how the improved data quality leads to substantially better results when using machine learning technics.

Biography. Tobias Widmer is Co-Founder and CTO at Wealthport, a data curation startup in Zürich, Switzerland. He earned a MSc from ETH Zürich in Computer Science. Prior to Wealthport, he was a software architect and project manager at Avaloq where he pioneered modern Java enterprise technologies and data architecture practices, primarily focusing on scalable cloud-based software development and deployment environments. Previously consulted various financial services companies in data management, process management, requirements engineering and quality management. Tobias worked with IBM Rational Research for the Eclipse open-source project and was part of the team building the first version of Jazz and Rational Team Concert.

Dr. Noah S. Bieler earned his PhD from ETH Zürich in Computational Chemistry. His research involved the production of large amounts of simulation data and their efficient storage, transformation, processing and analysis. Currently working as a Principal Data Scientist at Wealthport, a data curation startup located in Zürich, Switzerland, where he is in charge of designing efficient machine learning and data processing algorithms for automated data curation.

Distributed Training of Machine Learning Models: Classification and Regression on a Large Scale

Abstract. Training machine learning models is still very challenging as soon as the data becomes too large to fit onto a single computer. We study distributed algorithms for such tasks, in particular for classification, regression, sparse modeling, structured prediction, and other techniques.

Traditional distributed computing platforms such as Hadoop MapReduce are not suitable for these tasks, due to the high iteration and communication cost. We propose a communication-efficient algorithmic framework, CoCoA, that uses local computation in a primal-dual setting to dramatically reduce the amount of necessary communication. We provide a provable convergence speed for this class of algorithms, as well as experiments on real-world distributed datasets.

On the practical side, we're actively developing scalable and easy to use open source implementations of the proposed new methods, built upon Apache Spark. In this talk, we'll present some state of the art results of our software for classification, regression and structured prediction.

Dr. Martin Jaggi is a post-doctoral researcher in machine learning at ETH Zurich. Before that, he was a research fellow at the Simons Institute in Berkeley, US, working on the theory of big data analysis, and a postdoctoral researcher at École Polytechnique in Paris, France.

Biography. He has earned his PhD in Machine Learning and Optimization from ETH Zurich in 2011, and a MSc in Mathematics also from ETH Zurich, interrupted with several shorter stints in industry (Google, Netbreeze, Avaloq). He is broadly interested in methods for the analysis of large datasets, distributed training algorithms, open source software and machine learning applications for example in medicine, computer vision or text analysis. He is a co-founder of the startup SpinningBytes.com, and also the founder and co-organizer of the Zurich Machine Learning and Data Science Meetup, the largest technology meetup in Switzerland.

Mining the Voice of the Customer at Hilti

Abstract. The proposed talk will present lessons learned from applying natural language processing and machine learning techniques for automatically analyzing hundreds of thousands of unstructured and textual customer feedback documents at Hilti, a global manufacturer of fastening technologies. The presentation will include a live demo.

Biography. Oliver Müller is Assistant Professor at the Institute of Information Systems at the University of Liechtenstein. Oliver's research deals with the design of data-driven decision support systems, with a special interest in systems that analyze unstructured and/or textual data. Oliver holds a Ph.D. in Economics from the University of Münster.

Stefan Debortoli is a Research Assistant and Ph.D. Candidate at the Institute of Information Systems at the University of Liechtenstein. His doctoral studies focus on applying Big Data Analytics (especially text mining) as a new strategy of inquiry in Information Systems Research.

Karina Kisselite is a Data Scientist at Hilti, leveraging advance analytics techniques for various business needs, e.g. creating predictive analytic models for direct marketing campaigns or using process mining for understanding customer engagement.

Invitation. We invite you to a “data expedition”, where you will travel deep into unmap-ped territories of data. Together with people from different backgrounds, you will playfully explore your way through a full-scale data investigation and – hopefully - exit this quest with an innovative data product. Teams of data explorers help each other learn how to find data, overcome obstacles in using it, design a product and communicate it strate-gically.

Are you intrigued? Then you are the right person for this session! We will provide you with Twitter data on hot political discussion in Switzerland (e.g. national referendum next Sunday), and team you up with the right people. Data science experts will be there to guide you whenever you get stuck. You do not need to be a programmer or data analyst to participate – just bring your computer, your personal skills and a huge amount of cu-riosity and creativity!

If you did not yet register for the track, but planning to attend, please visit the registration desk.

Organisation. Mark Cieliebak and Oliver Dürr

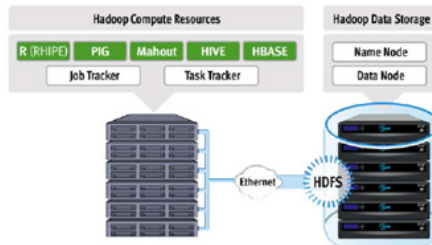
informatio

INFRASTRUCTURE LIFETIME SERVICES

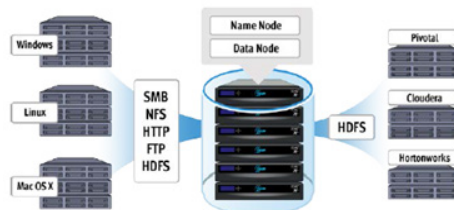
Informatio ist eine kleine Firma, die sich seit der Gründung im Jahre 2007 mit der Überwachung und dem Betrieb von Infrastrukturen befasst. Derzeit werden gut 8 PB an Kapazität in diversen Systemen überwacht.

Seit zwei Jahren unterstützt Informatio mit EMC Isilon das erste Scale-Out NAS System mit nativem Hadoop Filesystem Support. Isilon bietet bei Big Data Projekten enorme Vorteile.

Die Trennung von Speicheraufgaben von den Rechenaufgaben ermöglicht eine voneinander unabhängige Skalierung der Ressourcen. Die effiziente Speicherung der Daten erlaubt Auslastungen von über 80% der installierten Speichermenge welche mit Deduplikation noch erhöht werden kann.



Isilon erlaubt den Datenzugriff auf die gleichen Daten über mehrere NAS Protokolle. Quelldaten können von vorhandenen Systemen via Windows oder Unix/Linux Protokolle generiert werden. Via HDFS kann ohne notwendigen Datentransfer auf die Dateien zugegriffen werden. Aufwendige und lange andauernde Ladezyklen entfallen dadurch. Gleichzeitig wird der Zugriff sowohl via HDFS 1.0 und 2.0 sowie aus mehreren Distributionen unterstützt. Einmal gespeicherte Daten können so flexibel genutzt werden und Resultate schneller erreicht werden.



Die Sicherheit der Daten wird durch EMC Isilon verbessert. Die bestehende Namenode Verfügbarkeit wird stark erhöht, da dieser Service auf jedem Isilon Node läuft und alle Namenodes auf den gleichen Datenbestand zugreifen können. Als Enterprise Speichersystem stehen weitergehende Funktionalitäten für Backup und Recovery, Archiverung und Disaster Toleranz zur Verfügung.

Gerne beraten wir Sie bei Ihren Big Data Infrastruktur Projekten umfassend in Bezug auf Server, Netzwerk und Speicher.

Program

Time	Room 1: TN E0.46
09:30	Welcome reception
	Unified
10:00	<p>Jean-Mar</p> <p><i>Keynote on „The Use of Ever</i> Jan Egbert Sturm, Dire</p> <p><i>Tackling Challenging Problems in</i> Kurt Stockinger, Ass</p>
11:30	Short Break
	Executive Track (TN E0.54)
11:45	<p>„MIDATA.COOP – Citizen-Controlled Reuse of Personal Data“ André Golliez,</p>
12:10	<p>„Big Data Analytics - Evolving Toward Analytics 3.0“ Guido Oswald, SAS Institute</p>
12:35	Lunch Break
14:00	<p><i>Cross-functional data analytics for Client Relationship Advisors for a global American Bank“</i> Dorian Selz, Squirro</p>
14:25	<p>„50 Shades of Truth: Visualisation and Communication of Uncertainty“ Rodolphe Dewarrat and Vera Gärttling, IMSD</p>
14:50	<p>„Privacy in the Big Data Era“ Olivier Heuberger-Götsch, Scigility</p>
15:15	Coffee Break
	Unified
15:45	<p><i>Keynote on „The Emerging Discipline of Data</i> Michael L. Brodie, Research scientist and e</p> <p>„Mining th Oliver Müller and Stefan Debort</p>
17:10	Apéro & Open End

Room 2: TN E0.54	Seminar Room: TS O2.07	
Track (TN E0.46 + TN E0.54)		
<p><i>Welcoming address</i> Marc Piveteau, President of ZHAW</p> <p><i>Increasing Datasets in Macroeconomic Forecasting</i> Director of Swiss Economic Institute, ETH Zurich</p> <p><i>In Academia and Industry: An Interdisciplinary Approach</i> Associate Professor (Dozent), ZHAW Datalab</p>		
Technical Track (TN E0.46)		
<p><i>„Re-engineering IoT Legacy Analytics with Big Data“</i> Rosaria Silipo and Bernd Wiswedel, KNIME</p> <p><i>„0 Lines of Code and in under 20 Minutes: Building a Customer Churn Application“</i> Marc Schöni, Microsoft</p>	Interactive Track (TS O2.07)	
		Data Expedition
<p><i>„Interactive Scientific Image Analysis and Analytics using Spark“</i> Kevin Mader, ETH Zurich / PSI</p> <p><i>„Automated Data Curation at Scale“</i> Tobias Widmer and Bernhard Bicher, Wealthport</p> <p><i>„Distributed Machine Learning: Algorithms and Open Source Implementations“</i> Martin Jaggi, ETH Zurich & SpinningBytes AG</p>		Data Expedition
Track (TN E0.46 + TN E0.54)		
<p><i>Data Science - Principles and Techniques for Data-Intensive Analysis</i> Entrepreneurial advisor, MIT Massachusetts Institute of Technology</p> <p><i>The Voice of the Customer at Hilti</i> ...li, University of Liechtenstein and Karina Kisselste, Hilti</p>		



SAS ist mit 3,09 Milliarden US-Dollar Umsatz einer der grössten Softwarehersteller der Welt. Kunden an weltweit 75.000 Standorten setzen SAS Lösungen ein, um aus ihren vielfältigen Geschäftsdaten (Big Data) konkrete Informationen für strategische Entscheidungen zu gewinnen und so ihre Leistungsfähigkeit zu steigern (Analytics). Damit ist SAS der grösste Anbieter von Big-Data-Analytics-Software.

Mit SAS entwickeln Unternehmen Strategien und setzen diese um, messen den eigenen Erfolg, gestalten ihre Kunden- und Lieferantenbeziehungen profitabel, steuern in Echtzeit die gesamte Organisation und erfüllen regulatorische Vorgaben.

Firmensitz der 1976 gegründeten US-amerikanischen Muttergesellschaft ist Cary, North Carolina. SAS Schweiz beschäftigt derzeit rund 55 Mitarbeiter und hat Niederlassungen in Zürich und Genf. Weitere Informationen unter www.sas.com/ch.



Microsoft

Since its establishment in 1989, Microsoft Schweiz AG has developed from a small operation with three staff members into a mid-sized company with more than 600 employees, thanks to controlled and continuous growth. The headquarters of Microsoft Schweiz AG are located in Wallisellen near Zurich, and additional offices are located in Bern, Basel, Alpnach and Geneva. The executive management is currently represented by Petra Jenner, Country Manager of Microsoft Schweiz. In global sales rankings, the Swiss subsidiary occupies 13th place out of 168 Microsoft subsidiaries. In the ranking of “Sales per PC”, it even holds first place. This is first and foremost due to the conviction of many Swiss companies that the intelligent use of information and communication technologies leads to greater efficiency and competitive advantages.

Microsoft Schweiz AG has grappled with the changes in the working environment and its own work culture for years. As a result, in 2011 the headquarters of Microsoft Schweiz in Wallisellen were renovated on the basis of the latest findings. For the duration of the three-month renovation, Microsoft was involved in a special experiment: the office was completely closed and the employees shifted all of their activity to their home office or a location of their choosing. The result was not surprising, but worth noting: business activity was not negatively affected at all, but the social environment was. It was made clear that the office will continue to be of great significance in the future. Not as an individual workspace, but as a place for meeting and a platform for social interaction. The journey into the new working world is by no means finished – it has only just begun.

We are convinced that the centre of knowledge in Switzerland can only ensure and expand upon its success for the long-term if more modern technologies are used efficiently and intelligently. Management and work culture play a central role here. The more autonomy and responsibility assigned to the individual – even with regard to scheduling and spatial planning – the greater the benefit to the organisation. In order to raise awareness on this topic among the Swiss population, Microsoft launched the Home Office Day in 2010. Together with partners SBB, Swisscom and Witzig The Office Company, the Home Office Day takes place every year in May. It serves as a call to think about your own mobility and working behaviour. This should not only increase your own productivity and well-being, it should also make an active contribution to protecting the environment.



Confidently Anticipate and Drive Better Business Outcomes with SAP Predictive Analytics

SAP Predictive Analytics brings predictive insight to business users, analysts, data scientists, and others in your company. Unlock the potential of Big Data from virtually any source with the power of predictive modelling. By automating the building and management of sophisticated models

to deliver insight in real time, this software makes it easier to make better, more profitable decisions across the enterprise. No statistical degree is required to create powerful predictive models with SAP Predictive Analytics. Visit our booth and hear companies like Raiffeisen Group Switzerland or EMS Chemie run better with SAP Predictive Analytics.

OBJECTIVES

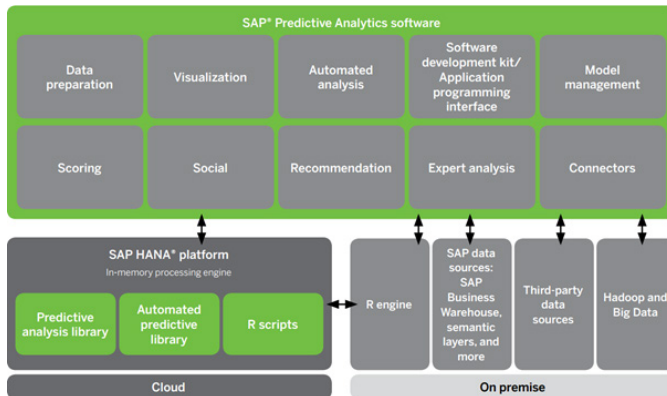
- Bring the power of predictive analytics to a broad spectrum of users
- Incorporate predictive analytics into line of-business applications and business processes
- Harness Big Data and the Internet of Things to predict and act in real time

SOLUTION HIGHLIGHTS

- Automated preparation of data
- Predictive modeling for users at all levels
- Advanced data visualization
- Predictive scoring, model management and social and recommendation functionality
- Access to high-performance algorithms
- 360-degree connectivity to Big Data and third-party data sources

BENEFITS

- Gain faster, more accurate predictive results with automated tools
- Improve decision making by embedding predictive results into business processes
- Uncover hidden insight with sophisticated modeling, unleashing the power of Big Data and the Internet of Things





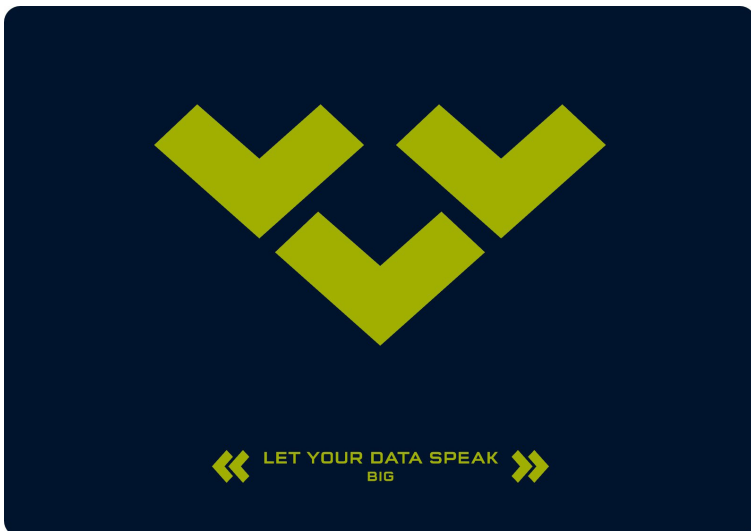
IT-Logix is your trustworthy Swiss Business Intelligence (BI), Data Warehouse and Big Data expert. We are recognized by SAP for Business Intelligence and by Microsoft (Gold Data Analytics) in Switzerland.

Our mission is to transform your structured and unstructured data into valuable business information. Your business users will be able to access, analyze and share information to make the decision process faster and more reliable each and every day. We know how to handle big volume, high variety and changing velocity to get maximum value from Big Data. In addition to requirements analysis and technical implementation, we will support you in visualizing the data according to the IBCS standards.

Our consulting services in BI strategy development, Business Intelligence Competence Center (BICC) organization set-up and agile project management complete our high quality end-to-end Business Intelligence services portfolio.

We are curious. We love to solve riddles. We challenge the status quo. We have common sense. We use proven methods and cutting-edge technologies to drive your business success.

This is why: We let your data speak.





Today's businesses are in a constant state of competitive upheaval. With the increasing momentum of the Internet of Everything (IoE) market transition, there will be 50 billion devices connected to the Internet by 2020—just five years from now. As billions of new people, processes, and things become connected, each connection will become a source of potentially powerful data to businesses and the public sector. Being successful in this environment requires that businesses continuously search for ways to deliver a competitive edge that can lower their costs, increase their revenues, and improve their speed. Organizations that can unlock the intelligence in this data can create new sources of competitive advantage, not just from more data but also from better access to better data.

Cisco has been working on making this vision of pervasive use of big data within enterprises a reality, and it is time to share the news. Our tools help you get the most from your data wherever it is with a flexible, high-performance, yet simplified data center infrastructure to let you see and manage your data, no matter how big it gets. Cisco also has the data virtualization, analytics, and application integration capabilities to help you find and use your data instantly, even at the very edge of the network. And we have the security to protect your data everywhere.

Cisco is the worldwide IT leader that enables amazing things to happen when you connect the previously unconnected. Cisco brings together people, process, data and things to transform how organizations meet next-generation demands. Our customers rely on the power of Cisco intelligent servers and networks to be competitive. Our collective strengths enable us to solve our customer's important business challenges. Cisco's true value isn't just in what we make- it's what we make possible.

www.cisco.com/go/bigdata



Our position

The ZHAW Data Science Laboratory, Datalab, is a virtual research group spanning several departments and institutes of one of Switzerland's biggest universities of applied sciences. Being very strong in research, it brings together all scientists engaged in the particular aspects of Data Science under one roof for collaborative research and industry projects. Founded in early 2013 as one of Europe's first groups dedicated to Data Science, Datalab currently comprises more than 60 researchers from as diverse areas as law, analytics, computer science and entrepreneurship. It is one of the leading Data Science research centers in Switzerland and beyond.

Our offer: R&D projects

Datalab offers the perfect team composition for each and every individual analytics project. From our pool of domain experts, we form interdisciplinary project teams with the right mix of methodical expertise. We are used to working together and crossing departmental borders. Possible project settings are third party- or directly funded R&D projects as well as student's thesis projects, feasibility studies and consulting. Typical project durations vary from a few weeks to several years. We adopt our processes, tools and technologies to our partner's needs.

Our service: Education & community building

Datalab designed one of the first dedicated Data Science curricula in Europe: The "Diploma of Advanced Studies in Data Science" for professional education. Our associates also commit considerable amounts of time to various Bachelor's and Master's degree programmes, thus sustaining the idea interchange between projects, students and industry.

We are committed to the Swiss Data Science community by means of several events and conferences we (co-) organize. One particular example is the "Swiss Workshop on Data Science" conference series that is very well attended and supported by industry and academia alike. We highly value this chance for idea exchange, trend spotting and passing on of lessons learned.

Get in touch

Find more information on the web (e.g., about individual team members and projects:), contact us directly via datalab@zhaw.ch, or follow us on Twitter: [@DataScienceCH](https://twitter.com/DataScienceCH).



References to geographic locations (zip codes, coordinates, addresses) are contained in 60 to 80% of all business data, such as customer, sales, POS and competitor data.

Crosswind helps companies to identify the business value of this location data by answering questions like “Where are my most valuable customers?”, “How can we optimize our point-of-sales network?”, or “Where are my competitors?”. Answers to these questions are of great value for marketing, product management, or logistics.

Our expertise in Location Intelligence – the combination of traditional Business Intelligence and geo-analytics – is translated into data models that fit, data warehouses with perfectly integrated geo data, custom reports and dashboards, and intelligent maps to gain deeper insights into the business. Crosswind – we unlock the Where.

<http://www.crosswind.ch>

CALLISTA GROUP

Callista Group AG – Transform Data to Business

Callista Group AG is your partner for design, development and implementation of Data Warehousing, BI, Analytics and CRM systems. We provide classic architectural approaches as well as sophisticated big data environments enabling the business to benefit from new technologies and methods.

Successfully managing customer relations requires a thorough understanding of internal business processes and expertise in building analytical systems. One of the most important aspects is the knowledge of customer communications and interactions through traditional channels and social networking (customer journey). In numerous projects the Callista experts have proven that such complex scenarios can be implemented successfully.

In-depth analytical observations have long been the basis for business decisions, and supplemental external data is often used as a data source in addition to the data warehouse. The incorporation of such analyses in processes and their automation and adaptation is still the “king discipline” of modern data warehouse environments. Especially decision automation based on advanced analytics including big data analytics can result in significant competitive advantages for your company.

Drivers such as „big data“, „advanced analytics“ or „operational BI“ are setting new standards for integration, performance and monitoring capabilities. These Drivers cause challenges and conflicting goals that need to be resolved.

Use our expertise, methods and solutions to transform your potential into real business and develop your customers.

<http://www.callista-group.com>



Who we are

Scigility is the leading expert for Big Data Solutions in Switzerland. We combine academic and practical knowhow to a unique service offering for our customers.

Our Goal

Our DevOps experts support customers in defining, building and operating integrated Big Data platforms. Based on Big Data platforms our Big Data Developers and Analysts build applications to create new insights to our customer's data. In addition to technical services our legal experts supports customers on how to use Big Data in accordance with Swiss and European law and how to create Big Data applications that includes data privacy, security, intellectual property and liability as fundamental design pattern in their architecture.

Research Collaboration

We closely work together with Research and Development Unites of our Partners, enhancing the productivity and create competitive advantage and optimized business performance of our clients.



What we do

We believe innovation requires an entrepreneurial mindset combined with the courage to push boundaries and break new ground.

Zühlke helps its clients to transform their vision from a smart idea into a resounding market success. Our tailor-made data analytics solutions support our customers in optimizing their processes, help to make better, data-based decisions and provide the basis for disruptive, new products and services. Our promise: as a strong solutions partner,

we contribute our in depth and broad industry expertise, and take responsibility for the success of our clients' projects.

We live for «empowering ideas».

Who we are

Zühlke is an independent service provider for product and software engineering, management consulting, and start-up financing. Zühlke provides added value as a result of the experience gained through more than 8000 successful international projects, as well as continued investment in business and technology knowledge and understanding. Founded in 1968, the Zühlke Group today has local teams in Austria, Germany, Serbia, Switzerland and the United Kingdom. In 2014, Zühlke generated CHF 119 million in revenue, employing more than 700 staff.



The main objective of the Swiss Association for Analytics is to spread the word about analytics in Switzerland. The goal is also to show added value of predictive analytics, data mining and machine learning to Swiss companies. Allowing analytics practitioners to exchange ideas and experiences is also our objective. The Swiss Association for Analytics provides latest updates, trends, news and original content to analytics practitioners.

For more information, visit www.swiss-analytics.com

swiss group for artificial intelligence
and cognitive science



Special Interest Group of the Swiss Informatics Society for Artificial Intelligence and Cognitive Science

The Special Interest Group for Artificial Intelligence and Cognitive Science (SGAICO) brings together researchers, practitioners and other parties interested in the subjects of artificial intelligence and cognitive science (AI/CO).

SGAICO pursues the goal of promoting intelligent technologies for innovation in our society. It provides a platform for exchange between industry and universities. You are cordially invited to join in our various activities and become a member of our Swiss network of experts as a platform for your ideas and initiatives.

- Discussion and dissemination of AI/CO-related knowledge
- Exchange on application problems occurring in Swiss enterprises and establishment of contacts between users and experts in Switzerland
- AI/CO methods and technologies in interdisciplinary contexts such as for example engineering, medicine, psychology and law
- Successful applications of AI/CO around the world and their relevance for Switzerland
- Topics and needs of AI/CO education in Switzerland
- SGAICO holds regular events on AI /CO-related topics and supports Swiss researchers in a wide range of activities such as community exchange, establishing contacts, or the organization of events.

SGAICO is a member of the European Coordinating Committee for Artificial Intelligence (ECCAI), which is coordinating the European national AI/CO societies.

More Information: <http://www.s-i.ch/en/special-interest-groups/sgaico>

Contact: sgaico@s-i.ch

President: Jana Koehler, Hochschule Luzern

Deputy Presidents: Jean-Daniel Dessimoz, HESSO-HEIG-VD, Marc-Oliver Gewaltig, EPFL, Thilo Stadelmann, ZHAW

SWISS STATISTICAL SOCIETY

www.stat.ch

There is no data science without statistics

“Statistics is the language of the future data economy”, wrote John Pullinger, the then-president of the Royal Statistical Society, in 2013. Today, two years later, the talk of a future data economy already seems outdated. But sure enough, statistics is more important than ever. There can be no data analysis without statistics be it of big or small data, of data stored in lakes or warehouses, of technical or personal data, in finance or life sciences. Any analysis will use statistical knowledge and methods, and rightly so!

But although being called the sexiest métier, statistics as a discipline tends to be buried in oblivion underneath all the buzz about hot new data trends. The Swiss Statistical Society (SSS) is the association of statisticians in Switzerland. It promotes application and development of statistics in Switzerland, represents the interest of professionals and contributes to the recognition of statistics as a scientific discipline in its own right. The Society organizes conferences, meetings, networking-events and courses. It is also committed to promoting young talents in the field, inter alia by means of the Lambert Award for Young Statisticians.

The Swiss Statistical Society was founded in 1988 and has since incorporated 500+ members from all over Switzerland. It is organized in the three sections Business & Industry; Official Statistics and Education and Research which represent the professional landscape. It is part of national and international networks and umbrella organizations and supports cooperation between institutions with similar goals and interests.

The Swiss Statistical Society is a member of the Swiss Academy of Sciences (SCNAT) and of the Swiss Academies of Humanities and Social Sciences (SAHS).

More information and membership application form at: www.stat.ch

Contact:
Swiss Statistical Society (SSS)
Administration
CH-3000 Bern
+41 44 350 03 41
sss@stat.ch
www.stat.ch

HASLERSTIFTUNG The Hasler Foundation was set up in 1948 by Gustav Hasler as the „Hasler Works Foundation“ and was originally a corporate foundation intended to ensure that the Hasler Works which had developed out of the Eidg. Telegraphenwerkstätte (Federal Telegraphy Works) would continue to operate in the spirit of Gustav Hasler even after his death (1952). It would also promote „Swiss telephony and telegraphy“.

Nowadays, the Hasler Foundation is an independent foundation, free of any corporate interests, which aims to promote education, research and innovation in the field of information and communications technology (ICT) for the well-being and benefit of Switzerland as an intellectual and industrial centre.

The Foundation provides its support by financing or co-financing selected educational, research and innovation projects and undertakings in the field of ICT. The annual funds distributed by the Foundation amount to several million Swiss francs.

The Foundation is politically and religiously neutral and free of any commercial interests. Although it is a non-profit-making organisation, it is not a charity, in that its support is always results-based, in keeping with the purpose of the Foundation.

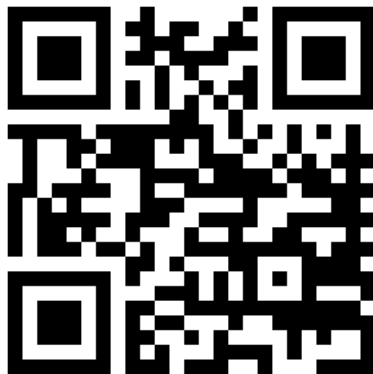
On our website, you can find out what and whom the Hasler Foundation supports, and how to make an application. There is also a brief introduction to the Foundation itself.

<http://www.haslerstiftung.ch/en>

Feedback

Please consider taking 2 minutes to fill out the following questionnaire in order to help us make your thoughts actionable:

www.zhaw.ch/datalab/feedback



Program Committee

Prof. Gerold Baudinot
Prof. Dr. Martin Braschler
Dr. Mark Cieliebak
Dr. Oliver Dürr
Melanie Imhof
Prof. Dr. Andreas Ruckstuhl
Dr. Thilo Stadelmann
Dr. Kurt Stockinger

Organization Committee

Gerold Baudinot, Martin Braschler, Mark Cieliebak, Oliver Dürr, Melanie Imhof, Bettina Bhend, Andreas Ruckstuhl, Thilo Stadelmann, and Kurt Stockinger.

SDS|2016

We hope you enjoyed SDS|2015 and got useful insights and contacts. Please use the feedback forms provided to help us optimize SDS|2016. You can also send your comments and ideas for participation to

datalab@zhaw.ch

Thank you!

Impressum

Datalab
c/o ZHAW School of Engineering
Frau Annette Zani
Postfach, 8401 Winterthur
Phone: +41 58 934 69 60
Fax: +41 58 935 69 60
Email: datalab@zhaw.ch



Check www.dlab.zhaw.ch after the event for documents regarding the individual presentations.