Applied Data Science in Europe

Challenges for Academia in Keeping Up with a Highly Demanded Topic

T. Stadelmann, K. Stockinger, M. Braschler, M. Cieliebak, G. Baudinot, O. Dürr and A. Ruckstuhl

Institute of Applied Information Technology
Zurich University of Applied Sciences
Winterthur, Switzerland
[stdm, stog, bram, ciel, baug]@zhaw.ch

Institute of Data Analysis and Process Design
Zurich University of Applied Sciences
Winterthur, Switzerland
[dueo, rkst]@zhaw.ch

***Abstract*—Google Trends and other IT fever charts rate *Data Science* among the most rapidly emerging and promising fields that expand around computer science. Although Data Science draws on content from established fields like artificial intelligence, statistics, databases, visualization and many more, industry is demanding for trained data scientists that no one seems able to deliver. This is due to the pace at which the field has expanded and the corresponding lack of curricula; the unique skill set, which is inherently multi-disciplinary; and the translation work (from the US web economy to other ecosystems) necessary to realize the recognized world-wide potential of applying analytics to all sorts of data.**

In this contribution we draw from our experiences in establishing an inter-disciplinary Data Science lab in order to highlight the challenges and potential remedies for Data Science in Europe. We discuss our role as academia in the light of the potential societal/economic impact as well as the challenges in organizational leadership tied to such inter-disciplinary work.

***Index Terms*—Data Science, Multi-disciplinary, Curriculum development.**

I. Introduction

Data Science seems to be the new magic bullet that could solve many problems of today's enterprises and even societies. The term is often used in close relationship with Big Data and is ranked high on the agenda of CIOs across all industries. Definitions vary, but have in common that Data Science comprises a unique blend of skills from analytics, engineering and communication aiming at generating value from the data itself [1][2]. This makes Data Science inherently an applied science, with the goal of applying various tools and techniques to data in order to gain a *data product*, an exploitable insight derived from collected facts.

It is interesting to see that especially in industry Data Science is often considered as the "new kid on the block" even though data-intensive sciences such as high-energy physics, astrophysics or bioinformatics have been using certain data science concepts already for decades – for instance in the field of scientific data management [3] and eScience [4]. These communities are used to processing, analyzing and visualizing terabytes and even petabytes of data to enable domain scientists achieve scientific breakthroughs. Recent trends in social media, however, allowed end-users to produce large amounts of data that were originally only exploited by large web companies such as Amazon, eBay, Google or Facebook. This is a strong publicly visible sign of Data Science's potential impact on society and economy.

II. Perspectives towards Industry

For example, many companies realized that the changing user behaviors and the higher affinity to online media of younger generations have large impacts on their businesses. In order for established enterprises to attract these new generations, they must adapt their business models to the new needs in order to avoid being "taken over" by web companies. Why would customers go to physical shops if a majority of the products can be bought online on Amazon - that in turn even suggests articles that are bought by like-minded people? Why would future generations go to expensive financial advisors of established banks, when Google offers often better financial advice by analyzing search behavior using Google Trends?

Understanding the needs of the new online society is key for succeeding in today's business world, and Data Science is one approach towards data-driven decision making as opposed to using "gut feelings". We currently see a shift in industry that already happened in data-intensive science decades ago. Hence, industry should learn from science – a discipline that is used to observing, modeling, simulating and explaining nature and the respective natural laws: For instance, a recent success story has shown that data mining techniques based on neural networks that were originally designed for high-energy physics analysis at CERN, have helped analyzing and interpreting the user behavior of the German retailer Otto [5].

However, many industries are still struggling with so-called Big Data-demands, as they are not quite sure how the new trend fits into the current enterprise application landscape. Analytical questions, as for instance in marketing, have successfully been answered before using data warehousing and business intelligence technology with data mining methods. What is the additional benefit of Data Science? And how does it scale from web economy problems to applications in established European industries like industrial production, as suggested by the concept of "Industry 4.0" [6]? Data Science education and research are needed to find answers to these questions.

III. Perspectives towards Teaching and Research

Based on the demands from industry, the new job profile *data scientist* has emerged whose interpretation differs widely

between various industries. Our goal is to foster a curriculum as well as a research agenda for future data scientists that are at the frontier of both applied research and industrial needs. The skill set of these data scientists is multi-faceted as shown in Figure 1. What follows portrays our work-in-progress and is intended to start a discussion in Europe.

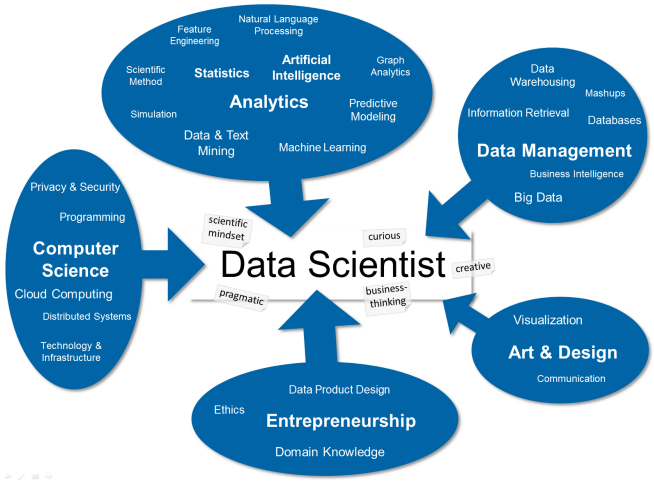

Fig. 1. Our interpretation of the Data Science skill set.

How do we ensure that we will be able to teach these diverse skills? Our approach at Zurich University of Applied Sciences is to address these challenges as part of the Datalab, a multi-disciplinary virtual organization spanning several institutes and departments for a collaboration between researchers and practitioners of various areas that altogether cover most of the depicted skill set. Data scientists require entrepreneurial experience, pursue different research interests from e.g. algorithms to visualization and come from backgrounds like computer science, statistics or law.

Our experiences from various projects in both research and industry have shown that truly multi-disciplinary teams are most suited for solving complex problems that require innovative "out-of-the-box-thinking". Knowing one specific discipline is often not enough. A broader perspective also helps in picking up necessary domain knowledge faster. Hence, one of the main aspects of our curriculum and applied research agenda is to foster multi-disciplinary approaches, which requires a broad education based on sound scientific methods in computer science, statistical learning etc. as well as best practices based on case studies from industry that pass the test of real life. The ideal data scientist is thus a scientist with entrepreneurial skills, who is used to asking the right business questions, understands the techniques and is familiar with the tools for solving them.

We believe that as a university of applied sciences we are ideally positioned to address these dual challenges as most of our research is already conducted within the confines of projects executed with industry partners. However, in order to make sure we are not trapped in local optima (i.e. country specific adaptations), another goal is to establish the basis for a European-wide Data Science curriculum and research agenda that can be generalized to a wider range and is accredited by the international Bologna system.

Our proposed curriculum and research agenda cover the following topics:

- Database and Cloud Computing technology for Big Data
- Data Mining, Statistics and Predictive Modeling
- Machine Learning and Graph Analytics
- Information Retrieval and Natural Language Processing
- Business Intelligence and Visual Analytics
- Data Warehousing and Decision Support
- Communication and Visualization of Results
- Privacy, Security and Ethics
- Entrepreneurship and Data Product Design

The goal of the data science curriculum foresees that students will work on coherent real-life projects and will apply data science to various case studies from science and industry to learn crafting business cases that have impact.

One of the challenges of inter-disciplinary education and research is that different university departments need to work more closely together to come up with joint agendas. What has been shown to work successful in large-scale sciences such as e.g. astrophysics or bioinformatics could be a role model for data science centers: departments, universities, labs and industry join necessary forces to cope with the high complexity of the subject. This might require re-thinking established scientific disciplines and organizational structures, being open for applied research topics that at first glance do not fit directly into the core competences of the respective departments. However, the best scientific results with positive impacts on society often happen at the crossroads between well-established fields, i.e. where inter-disciplinary teams tackle larger problems jointly rather than staying in the comfort zone of traditional lab sciences with narrow focuses.

IV. Conclusions

Data Science is a new paradigm that has the potential to revolutionize enterprises in a similar way as the introduction of computers and IT changed business processes in the first place. In order to make sure that this paradigm shift does not only happen in Silicon Valley but also in Europe, a streamlined curriculum and research agenda is required that adapts to the European market. These new challenges offer great opportunities but also responsibilities for universities and higher education in general in order to stay abreast of technological advances and to shape them for the benefit of society.

References

[1] http://radar.oreilly.com/2010/06/what-is-data-science.html
[2] http://online.liebertpub.com/doi/pdf/10.1089/big.2013.1508
[3] https://sdm.lbl.gov/sdmcenter/
[4] research.microsoft.com/en-us/collaboration/fourthparadigm/
[5] http://www.big-data-europe.com/frankfurt-ruckblick-2012/
[6] http://www.bmbf.de/de/9072.php