Data Science, one of the most significant developments in computing in the 21st century, is a discipline in the making, drawing principles, methods and tools from established fields like computer science, statistics, economics, and many more [1,2]. It is a means of accelerated (scientific) discovery and is transforming science, business, politics, and any domain with adequate data. Yet, data science is immature. What are its underlying principles and techniques (models, methods) that are applicable across different use cases and fields of application? What «science» underlies this emerging discipline (see [3, section 4.2])? We are launching an activity to develop a Reference Framework for Data Science and invite your participation. A more detailed description including an initial structure will be provided in early December as a starting point.

We will pursue these questions together in spring 2020 in three-stages:

1. A local *workshop* on Feb 27 to share and discuss preliminary thoughts and hypotheses of the participants; we expect participants predominantly from but not limited to ZHAW, ETH, and UZH;
2. *Collaborative advancement sessions*: to Mar 13, Dr. Michael Brodie of Harvard University, who is conducting research on the foundations of data science, volunteers to do interviews and potentially co-working with each participating research group to create a systematic overview of cross-field learnings as well as a refinement of the presented ideas (time investment per participant: ca. ½-3 days);
3. A future international *symposium* to solicit participation of the international community (time and location TBD, e.g. at Schloss Dagstuhl in Germany or co-located with a major data science conference; including proceedings)

Subjects of interest include, but are not limited to:

- Data science principles and techniques drawn from constituent disciplines of data science
- Engineering principles required to apply existing foundations, e.g., in statistics, in practice. See [4,5]
- Open research questions into the “more than the sum of its parts”-aspects of data science
- Cross-field lessons learned

To participate in the workshop (stage 1) and the subsequent advancement sessions (stage 2) each participant is asked to formulate and present their (preliminary) ideas on the subject in their area of expertise. We suggest references [1, 3, 4, 5] to put their contribution into context.

We invite contributors to present their ideas at the workshop (stage 1), to engage in collaborative advancement sessions (stage 2) and, to continue the development of a Reference Framework for Data Science as an activity in the international data science community (stage 3).

Important dates 2020

- *Abstract submission*: January 16 (registration happens by submitting an extended abstract (1-2 pages) containing one’s own hypotheses / findings / preliminary ideas on the subject to alho@zhaw.ch
- *Workshop* (stage 1): February 27 (13:00-19:00, ZHAW campus Winterthur; every participant is given a speaking slot of a few minutes to present own ideas, with ample time to discuss; apero provided)

Organizers

- Dr. Thilo Stadelmann, Prof. AI/ML, ZHAW School of Engineering
- Dr. Michael L. Brodie, visiting scholar, Harvard School of Engineering and Applied Sciences
- Dr. Robert Vorburger, head of Datalab, ZHAW School of Life Sciences and Facility Management

References (all available online)

From Braschler et al. (Eds.), «Applied Data Science – Lessons Learned for the Data-Driven Business», Berlin, Heidelberg: Springer, 2019.
[1] Brodie, «What is Data Science?».
[2] Braschler et al., «Data Science».
[3] Brodie, «On Developing Data Science»
[4] Jordan, M. Artificial Intelligence — The Revolution Hasn’t Happened Yet; UC Berkeley, April 18, 2018
[5] Jordan, M. I. (2019). Dr. AI or: How I Learned to Stop Worrying and Love Economics. Harvard Data Science Review.