

## Chapter 2

# Big Data Panel at SIGDSS Pre-ICIS Conference 2013: A Swiss-Army Knife? The Profile of a Data Scientist

Barbara Dinter, David Douglas, Roger H.L. Chiang, Francesco Mari, Sudha Ram, and Detlef Schoder

**Abstract** The purpose of the big data panel was to provide a forum for exchange of ideas on curricula content in the area of data science and big data. The panelists were from a broad range of academic institutions designed to provide different perspectives. Industry perspectives are vital as they will be the ones employing the graduates of these programs. Thus, the panel included an industry expert from a company that is a leader in data science and big data. Although there was agreement on selected skills as being foundational, it was recognized that a curriculum would not provide all the skills a data scientist would need for many big data projects—thus the team approach to projects.

**Keywords** Big data • Data science • Data scientist • Curricula • Careers

---

*Panel Coordinators:* Barbara Dinter and David Douglas

*Panelists:* Roger H.L. Chiang, Francesco Mari, Sudha Ram, and Detlef Schoder

B. Dinter (✉)

Chemnitz University of Technology, Thuringer Weg 7, 09126 Chemnitz, Germany

e-mail: [barbara.dinter@wirtschaft.tu-chemnitz.de](mailto:barbara.dinter@wirtschaft.tu-chemnitz.de)

D. Douglas

Information Systems, University of Arkansas, Business Building 204B,

72701 Fayetteville, AR, USA

e-mail: [ddouglas@walton.uark.edu](mailto:ddouglas@walton.uark.edu)

R.H.L. Chiang

University of Cincinnati, Cincinnati, OH, USA

F. Mari

SAP, Milan, Italy

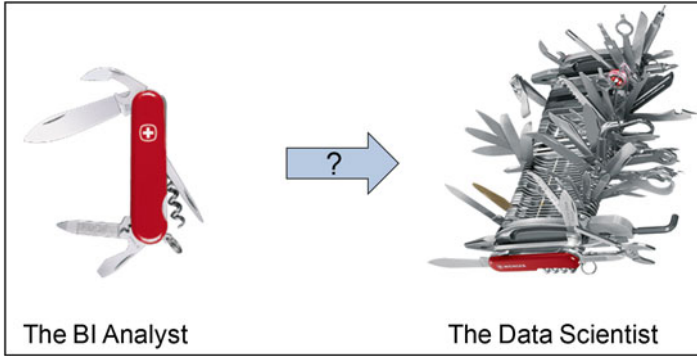
e-mail: [francesco.mari@sap.com](mailto:francesco.mari@sap.com)

S. Ram

University of Arizona, Tucson, AZ, USA

D. Schoder

University of Cologne, Cologne, Germany



**Fig. 2.1** The analogy of the Swiss-army knife (Pictures taken from [www.wengerna.com](http://www.wengerna.com))

Big data as a recent hot topic not only affects the information systems (IS) landscape in industry, it also has significant impact on academia. The need for so-called data scientists has recently attracted widespread attention, in particular initiated by Davenport et al. (2012). Well-known is the Data Science Venn Diagram (Conway 2010) which was a first attempt to illustrate the different skill areas a data scientist should cover. Although differing slightly in detail, there seems to be a consensus that the following skill areas are needed for those profiles (Chen et al. 2012; Chiang et al. 2012; Conway 2010; Davenport et al. 2012; Laney and Kart 2012; Patil 2011): analytical skills, IT and programming skills, business and domain knowledge, and interpersonal skills (such as communication and curiosity).

The panel discussion on “A Swiss-Army Knife? The Profile of a Data Scientist” was aimed at sharing experiences in big data education and identifying best practices for pedagogy to produce well prepared students for a career in industry. The panel title was motivated by drawing an analogy between the capabilities of a data scientist and the functionalities of a Swiss Army knife. While “traditional” analysts need specific but limited skills, such as business intelligence, data warehousing, and data management, the profile of a data scientists seems to be much broader. Based on the analogy, the latter corresponds to the so-called Wenger giant knife as illustrated in Fig. 2.1.

The panelists come from both, academia having a background in big data education and industry, working with data scientists. The panel has been structured by three topic areas and one corresponding “guiding” question. For each of them the panelists were asked in advance to prepare a short statement:

- The profile of a data scientist: What skills does a data scientist need?
- Data science curriculum: Do we need dedicated data science programs? If so, what facilities/disciplines should be in charge of the program?
- Data science career in industry: How students can be motivated and enthused for a data science career?

After a short presentation of the purpose of the panel and an introduction of the panel members each panelist presented his/her point of view for the aforementioned topics. The profile of a data scientist was described similar to the aforementioned data science Venn diagram. Required knowledge and skills focused on several areas including analytical, IT, business, domain, and communications. The analytical knowledge and skills included data mining, statistics, text mining, optimization, sentiment analysis, network and graph analysis, econometrics, and predictive modeling. The IT knowledge and skills included relational database and data warehousing, ETL, OLAP, dashboards and visualization, Hadoop and MapReduce environments, cloud computing, and all types of data management.

Another approach to characterize the required skills of a data scientist was guided by the Hal Varian's definition (Varian 2014) of data science which includes ability to take data and to be able to understand it, to process it, extract value from it, visualize it, and communicate it. Based on this rather process-oriented view specific data science skills can be identified in terms of technologies, techniques, and problem formulation. The technologies include programming languages such as R, Java and Python as well as the Hadoop and MapReduce environments and visualization tools such as D3 and GEPHI. Techniques include mathematical modeling, graph and network theory, machine learning, algorithms, and data structures. Furthermore, data scientists must be able to do problem formulation, have computational thinking and curiosity, ask lots of questions and have the ability to tell a story with the data. Indeed there was quite a bit of consensus from the panelists on the skills needed from a data scientist.

Francesco Mari, representing the practitioner's perspective, noted that only very few data scientists might have such exceptional capabilities. He argued that instead hybrid roles are needed and that it is critical to build teams to cover all required skills.

The panelists emphasized that big data education requires joint efforts and collaboration between academia and industry. The obvious "learning by doing" approach was noted as being more relevant than ever. Francesco Mari raised a point which makes curriculum design even more challenging. "The knowledge of the field is in the data" – i.e., the data scientists along with their domain knowledge are able to have the data tell compelling stories. Even in the same industry and/or the same problem domain, "the data can tell different stories and the corresponding impact of the stories can be huge." The data scientist knowhow of storytelling from the data is more difficult to teach and transfer than traditional IT topics such as programming, etc. Such a variety of cases makes data scientist education even more challenging as not only tool and technology knowledge is required. There is also a need for concrete projects and real data in teaching.

Finally, the panel members presented their assessment of data scientist career paths. Although there is definitely no shortage of careers, often appropriate career paths in organizations do not exist. Careers as data scientists are available in practically every industry including retail, manufacturing, health care, and services. Excellent data scientists skills, e.g. in programming, are not rewarded because their

work is mostly in the background and not visible and thus their value not recognized. Francesco Mari suggested that one potential improvement may be to link them with the R&D department.

After the panelists' presentation, the floor was opened for questions from the audience. The issue around roles of data scientists was raised and the discussion focused on which roles in a data science/big data team make sense. This question is hard to answer, as it always depends from the concrete context and project. Experience from practice shows that in most cases up to three team members can provide all the skills that are required in big data projects. In addition, people with a background in other disciplines (e.g. physics) can enrich such teams.

In conclusion, there is a lot of potential for technical MIS programs to develop inter-disciplinary and innovative curricula in partnership with industry for meeting the burgeoning demand for data scientists.

## Biography

**Barbara Dinter** is Professor and Chair of Business Information Systems at Chemnitz University of Technology, Germany. She holds a Ph.D. from the Technische Universität München, Germany, where she previously earned a master's degree in computer science. Barbara Dinter worked for several years at University of St. Gallen, Switzerland as a Post-Doc and project manager. In her role as an IT consultant, she worked with a variety of organizations. Her research interests include business intelligence and analytics, big data, data driven innovation, and information management. She has published in renowned journals such as *Decision Support Systems*, *Journal of Database Management*, and *Journal of Decision Systems*, and on conferences such as ICIS, ECIS, and HICSS.

**David Douglas** is a University Professor, Co-Director of the Institute for Advanced Data Analytics and Director of Enterprise Systems at the University of Arkansas. He holds a Ph.D. in Industrial Engineering from the University of Arkansas. He teaches a wide variety of information systems subjects with emphasis on enterprise systems and global IT, as well as business intelligence/knowledge management focusing on data mining and data warehouses. He has taught in several countries and presents workshops world-wide. He has received a number of honors and awards including International Educator of the Year (IACIS), Majdi Najm Outstanding Service Award (SAP University Alliances), IBM International Professor of the month and NSF co-principal investigator for Enterprise Computing Community. His research interests include enterprise systems, business intelligence and data analytics. His publications have appeared in various journals including *Communications of the ACM*, *Decision Sciences Journal of Innovative Education*, *the Journal of Computer Information Systems*, *the Journal of Organizational and End User Computing*, *Information and Management*, and *the Journal of Management Information Systems*, as well as international, national and regional *Proceedings* of various Conferences.

## References

- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems*, 3(3), 12:1–12:13.
- Conway, D. (2010). The data science Venn diagram. [Online]. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How ‘Big Data’ is different. *MIT Sloan Management Review*, 54(1), 22–24.
- Laney, D., & Kart, L. (2012). Emerging role of the data scientist and the art of data science. Gartner Group. White paper.
- Patil, D. (2011). *Building data science teams: The skills, tools, and perspectives behind great data science groups*. Cambridge: O’Reilly Radar.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.

Reshaping Society through Analytics, Collaboration,  
and Decision Support

Role of Business Intelligence and Social Media

Iyer, L.S.; Power, D.J. (Eds.)

2015, IX, 269 p. 58 illus., 43 illus. in color., Softcover

ISBN: 978-3-319-11574-0