

Delegates Summit: Best Practice and Definitions of Data Sciences – Beyond Statistics September 25, 2017

The Seventh Symposium on
Advanced Computation and Information in Natural and Applied Sciences
The International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017)
September 25 – 30, 2017, Thessaloniki, Greece



Dr. rer. nat. Claus-Peter Rückemann^{1,2,3}



¹ Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany

² Leibniz Universität Hannover, Hannover, Germany

³ North-German Supercomputing Alliance (HLRN), Germany

ruckema(at)uni-muenster.de



Delegates Summit: Best Practice & Definitions of Data Sciences . . .

Delegates and Contributors

- *Claus-Peter Rückemann* (Moderator),
Westfälische Wilhelms-Universität Münster (WWU) /
Knowledge in Motion, DIMF / Leibniz Universität Hannover /
North-German Supercomputing Alliance (HLRN), Germany
- *Oleg O. Iakushkin*,
Department of Computer Modelling and Multiprocessor Systems
at the Faculty of Applied Mathematics and Control Processes,
Saint-Petersburg State University, Russia
- *Lutz Schubert*,
IOMI, University of Ulm, Germany
- *Friedrich Hülsmann*,
Knowledge in Motion, DIMF, Germany
- *Birgit Gersbeck-Schierholz*,
Knowledge in Motion, DIMF, Germany
- *Olaf Lau*,
Knowledge in Motion, DIMF, Germany

The International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017),
The Seventh Symp. on Advanced Computation and Information in Natural and Applied Sciences,
CFP: <https://research.cs.wisc.edu/dbworld/messages/2017-05/1493741666.html>
Program: http://icnaam.org/sites/default/files/Preliminary%20Program%20of%20ICNAAM%202017_ver_3.pdf

Recall: Last Years' Post-Summit Results

In 80 Words Around The World.

Knowledge and Computing Definitions (Delegates and other contributors)

- “Knowledge is created from a subjective combination of different attainments as there are intuition, experience, information, education, decision, power of persuasion and so on, which are selected, compared and balanced against each other, which are transformed, interpreted, and used in reasoning, also to infer further knowledge. Therefore, not all the knowledge can be explicitly formalised. Knowledge and content are multi- and inter-disciplinary long-term targets and values. In practice, powerful and secure information technology can support knowledge-based works and values.”
- “Computing means methodologies, technological means, and devices applicable for universal automatic manipulation and processing of data and information. Computing is a practical tool and has well defined purposes and goals.”

Citation: Rückemann, C.-P., Skurowski, P., Staniszewski, M., Hülsmann, F., and Gersbeck-Schierholz, B. (2015): *Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing*; Sep. 23, 2015, The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 13th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sep. 23–29, 2015, Rhodes, Greece. URL: http://www.user.uni-hannover.de/cpr/x/publ/2015/delegatessummit2015/rueckemann_icnaam2015_summit_summary.pdf

Delegates and contributors: Claus-Peter Rückemann, Friedrich Hülsmann, Birgit Gersbeck-Schierholz, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Przemysław Skurowski, Michał Staniszewski, Silesian University of Technology, Gliwice, Poland; International EULISP post-graduate participants, ISSC, European Legal Informatics Study Programme, Leibniz Universität Hannover, Germany

Recall: Last Years' Post-Summit Results

In 80 Words Around The World.

Data-centric and Big Data Definitions (Delegates and other contributors)

- “The term data-centric refers to a focus, in which data is most relevant in context with a purpose. Data structuring, data shaping, and long-term aspects are important concerns. Data-centricity concentrates on data-based content and is beneficial for information and knowledge and for emphasizing their value. Technical implementations need to consider distributed data, non-distributed data, and data locality and enable advanced data handling and analysis. Implementations should support separating data from technical implementations as far as possible.”
- “The term Big Data refers to data of size and/or complexity at the upper limit of what is currently feasible to be handled with storage and computing installations. Big Data can be structured and unstructured. Data use with associated application scenarios can be categorised by volume, velocity, variability, vitality, veracity, value, etc. Driving forces in context with Big Data are advanced data analysis and insight. Disciplines have to define their ‘currency’ when advancing from Big Data to Value Data.”

Citation: Rückemann, C.-P., Kovacheva, Z., Schubert, L., Lishchuk, I., Gersbeck-Schierholz, B., and Hülsmann, F. (2016): *Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering*; Sep. 19, 2016, The Sixth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 14th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sep. 19–25, 2016, Rhodes, Greece.
URL: http://www.user.uni-hannover.de/cpr/x/publ/2016/delegatessummit2016/rueckemann_icnaam2016_summit_summary.pdf

Delegates and contributors: Claus-Peter Rückemann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Zlatinka Kovacheva, Middle East College, Department of Mathematics and Applied Sciences, Muscat, Oman; Lutz Schubert, University of Ulm, Germany; Iryna Lishchuk, Leibniz Universität Hannover, Institut für Rechtsinformatik, Germany; Birgit Gersbeck-Schierholz, Friedrich Hülsmann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany

Best Practice and Definitions

In 80 Words Around The World.

Statements on Data-Sciences (1/3) (Delegates and other contributors)

- **“Data Science is directly related to the processing and cross-correlation of individual sources of information.
An example of such data are images and video sequences received by users of social networks. These data allow data scientists to analyze the selected paths, range of interests and the emotional response of users at the locations of interest.
Such research is directly related to the tasks of clustering and high-performance data processing which we are investigated at Saint Petersburg State University.”**

Oleg O. Yakushkin, Department of Computer Modelling and Multiprocessor Systems at the Faculty of Applied Mathematics and Control Processes, Saint-Petersburg State University, Russia

Contact: o.yakushkin@spbu.ru,

Universitetskii prospekt 35, Petergof, Saint Petersburg, Russia 198504

Best Practice and Definitions

In 80 Words Around The World.

Statements on Data-Sciences (2/3) (Delegates and other contributors)

- **“Data Science: Do we still understand data? Big data and artificial intelligence all boil down to statistical analysis with no understanding of the meaning of data - how reliable is the information we gain from this and what is needed to make data and derived information explicable?”**

Lutz Schubert, IOMI, University of Ulm, Germany.

Contact: lutz.schubert@uni-ulm.de

Best Practice and Definitions

In 80 Words Around The World.

Statements on Data-Sciences (3/3) (Delegates and other contributors)

- **“Qualified Data, especially for an enterprise, represents frozen knowledge or in other words frozen value. The ability to manage these data is what we call data science. Hence, data science is by definition secondary to data. The essence of Data Science is to give qualified access to relevant data to the owners and users. Hardware and software and their implementation represent the tertiary level of qualified and high level data.”**

Data results from action!

Examples are from insurance companies to research data focussed disciplines.

Claus-Peter Rückemann, Friedrich Hülsmann, Birgit Gersbeck-Schierholz, Olaf Lau, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany.

Contact: ruckema@uni-muenster.de

Best Practice and Definitions

In 80 Words Around The World.

Case: Natural sciences & research Source: Hülsmann, Rückemann, Gersbeck-Schierholz (KiM, DIMF)

- Focussed on research insight
- On-purpose components
- Local workplaces (e.g., institute)
- Research data management, e.g., in most cases only archive
- Data exchange possible, long-term interest maybe limited
- Rarely common standards with respect to data structures and documentation
- Data used for new insight
- Single data-sets can grow up to >1 TB
- Long-term storage / archiving
- Data Science: Knowledge focus, file type, data centricity, small percentage of statistics, specialised algorithms, different requirements regarding long-term availability research purpose, insight / research

Best Practice and Definitions

In 80 Words Around The World.

Case: Insurance and business Source: Lau (KiM, DIMF)

- Focussed on preset purpose
- Ancillary conditions implemented in the system
- Distributed workplaces (e.g., different sites)
- Databases (e.g., using SAP)
- Limited data exchange
- Weather data integration with case files (e.g., images from national weather service)
- Data used for decision making
- Case files per case up to >100 MB
- Mid-term storage / archiving (esp., according with European regulations, code of conduct)
- Data Science: Documentation, database type, statistics, long-term availability, business purpose, legal and financial aspects

Best Practice and Definitions

In 80 Words Around The World.

Statements on Data Science / Data Sciences (Delegates and other contributors)

- **How should Data Science / Data Sciences be defined?**
- **Which Best Practice for Data Science / Data Sciences can be summarised?**

Conclusions, Discussion, Networking

Data Sciences (Delegates and other contributors)

- Data results from action.
- It is not reasonable to apply statistics without sufficient understanding of the data and originating context.
- Deriving explicable information requires qualified data and understanding.
- Data Science is directly related to processing of data, cross-correlation of individual sources of information, and data analysis.
- Data, especially Qualified Data, can represent knowledge, respective value.
- Implementation of methods and algorithms can be seen consecutive to the creation of data.
- Implementation of hardware and software resources and services can be seen consecutive to methods and algorithms.
- Methods and algorithms experienced certain development over the last decades, nevertheless, a common basic understanding is still not widely present.
- A ternary view of “Data-Algorithm-Implementation” can suitable to characterise an individual scenario, e.g., via ternary diagrams. More complex views can be created when required with advancing constellations.

Networking and Outlook

Thank you for your attention!

**Wish you an inspiring conference
and a pleasant stay in Thessaloniki!**

**Looking forward to seeing you again next year for the
Symposium on Advanced Computation and Information!**

Post-Summit Results

In 80 Words Around The World.

Data Science Definition (Delegates and other contributors)

- “Qualified Data, especially for an enterprise, represents frozen knowledge or in other words frozen value. The abilities to understand and manage these data is what we call data science.

Data results from action, hence, data science can be defined secondary to data. The essence of Data Science is to give qualified access to relevant data to owners and users.

Hardware and software and their implementation represent the tertiary level of qualified and high level data.”

Citation: Rückemann, C.-P., Iakushkin, O. O., Gersbeck-Schierholz, B., Hülsmann, F., Schubert, L., and Lau, O. (2017): *Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data Sciences – Beyond Statistics*; Sep. 25, 2017, *The Seventh Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 15th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM)*, Sep. 25–30, 2017, Thessaloniki, Greece. URL: http://www.user.uni-hannover.de/cpr/x/publ/2017/delegatessummit2017/rueckemann_icnaam2017_summit_summary.pdf

Delegates and contributors: Claus-Peter Rückemann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Oleg O. Iakushkin, Department of Computer Modelling and Multiprocessor Systems at the Faculty of Applied Mathematics and Control Processes, Saint-Petersburg State University, Russia; Birgit Gersbeck-Schierholz, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Friedrich Hülsmann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Lutz Schubert, IOMI, University of Ulm, Germany; Olaf Lau, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany.