INFOCOMP / IMMM / DataSys 2018 International Expert Panel:

Panel on Information Processing and Ownership

General Data Protection on the Rise: Who Will be the Owner of Data and Who Will be in Control?

July 24, 2018, Barcelona, Spain

The Eighth Int. Conference on Advanced Communications and Computation (INFOCOMP 2018)
The Eighth Int. Conference on Advances in Information Mining and Management (IMMM 2018)

INFOCOMP / IMMM / DataSys
June 22–26, 2018 - Barcelona, Spain

INFOCOMP Expert Panel: . . . How Will Data Science Benefit to . . .

Panelists

- *Claus-Peter Rückemann* (Moderator),
  Westfälische Wilhelms-Universität Münster (WWU) /
  Leibniz Universität Hannover / HLRN, Germany
- *Iryna Lishchuk*,
  Leibniz Universität Hannover, Germany
- *Athanasios Tsitsipas*,
  Inst. of Inform. Resource Management, University of Ulm, Germany
- *Zlatinka Kovacheva*,
  Middle East College, Oman
- *Valery Covachev*,
  Inst. of Math. and Inform., Bulgarian Academy of Sciences, Bulgaria
- *Nikola Vlahović*,
  Faculty of Economics and Business University of Zagreb, Croatia

INFOCOMP 2018: `http://www.iaria.org/conferences2018/INFOCOMP18.html`
Program: `http://www.iaria.org/conferences2018/ProgramINFOCOMP18.html`

INFOCOMP Expert Panel: . . . How Will Data Science Benefit to . . .

Pre-Discussion-Wrapup / Panel Statements:

- Data cannot be protected. Other things could be protected.
- The data owner should be in control.
- Ownership of data is the central core of protecting rights and possessions of a person.
  How to strengthen the ownership?
- Behaviour and results of analytics may affect the privacy of an individual.
  How to control?
- The acceptance and combination and often even more the integration of practices and realisations is the core problem.

INFOCOMP Expert Panel: Post-Panel-Discussion Summary

Post-Panel-Discussion Summary (2018-07-24):

- "Data" itself cannot be protected. Other things should be protected!
- The widely promoted "data"-based concept is much too flat. The classical knowledge concept is a superior and much more suitable concept to address and handle issues in the overall context.
- The "owner" is the relation of "knowledge/data" to existence.
- The person who has (legal) default "entitlements" owns "data". "Ownership" is a central core of protecting rights and possessions of a (natural) person.
- The "data" owner (person with legal entitlements) should be in control.
- Applying algorithms, machine learning, artificial intelligence, analysis, automation, machinery ... can harm owners' interests.
- Behaviour and results of analytics may affect the privacy of an individual.
- The uncritical acceptance of combination and often even more of integration of practices and realisations are part of the core problem.
- Third parties (commercial and non-commercial), which are "using data" claim rights on "data" because of investments in their businesses.
- "Data gathering" needs to be controllable. Results need to be better controlled.
- "Providing data" for any use should always be by explicit, individual decision and consent of the owner.

INFOCOMP Expert Panel: Table of Presentations, Attached

Panelist Presentations: (presentation sort order, following pages)

- **The Endangered Species and the Beneficiaries of Data Protection – Cui Bono?** (*Rückemann*)
- **Privacy Implications of Big Data** (*Lishchuk*)
- **AI, Big Data and Privacy: A relentless war** (*Tsitsipas*)
- **Big Data Analytics: Statistical, Mathematical, Computational and Legal view** (*Kovacheva/Covachev*)
- **"Bringing Data Owners in Control. Economic and Social Implications of Data Protection."** (*Vlahović*)

INFOCOMP / IMMM / DataSys 2018 International Expert Panel:
General Data Protection on the Rise:
Who Will be the Owner of Data and Who Will be in Control?

The Endangered Species and the Beneficiaries of Data Protection – Cui Bono?

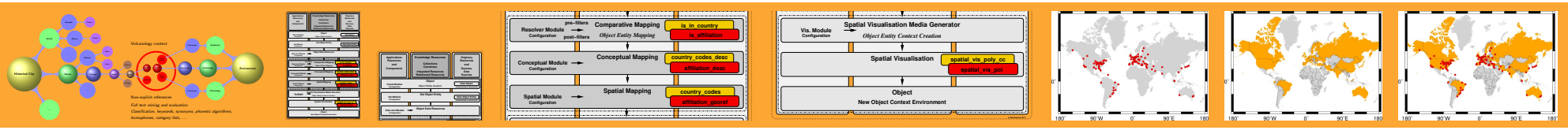
The Eighth Int. Conference on Advanced Communications and Computation (INFOCOMP 2018)
The Eighth Int. Conference on Advances in Information Mining and Management (IMMM 2018)

July 24, 2018, Barcelona, Spain

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



Last things first: Data protection, . . .

- Data protection based on **national law / national regulations**, e.g., Bundesdatenschutzgesetz (BDSG).
- **Data Protection Directive 95/46/EC**.
- **General Data Protection Regulation** (GDPR) (EU) 2016/679.
  Adopted April 14, 2016, two-year transition period, enforceable on May 25, 2018.
  Superseding the Data Protection Directive 95/46/EC.
- . . .

- Why are the terms "data" and "personal" data kept volatile and not defined in context of the data protection regulation?
- Can legal or technical aspects change the meaning of terms?
- What is the ultimate purpose of data protection?
- Why is there no clearly defined general force to opt-in?
- What is the relation of data to existence?
- Who is the "owner" of data? If not, can data exist on its own? . . .

Present situation

- How is data protected?
- Who is the “owner” of data?
- Who is in control of hardware capabilities?
- Who is in control of user space? . . .

Extrapolation

- How will data be protected?
- Who will be the “owner” of data?
- Who will be in control of hardware capabilities?
- Who will be in control of user space? . . .

What and why?

- **How can data be protected without an understanding and definition of data and the purpose?**
- **Ownership/entitlements of data are neglected/ignored!?**
- **Why are problems addressed in alternative ways?**
- **Cui Bono?**

Up-to-date example of buying goods to get ownership?: Automobiles

1. Automobile manufacturers **integrate data collection and surveillance in products**.
2. Legislative authorities **make features mandatory**.
3. Third parties, e.g., insurance companies, **claim data** (collected by the "owners of products") from manufacturers.
4. Tax authorities **claim taxes** on data.
5. Intelligence **claim access to data**.

⇒ Fights for data value: Politics vs. industry, insurance companies vs. manufacturers
⇒ Is there pressure on data owners / entitlement rights? Where is the position/lobby of the "owner"?
⇒ Is it coincidence that data protection and data tax claims meet at the same time?

Which of the following terms does not fit in the group?

- Wildlife protection.
- Animal protection.
- Data protection.
- Health protection.

⇒ Why? Because of its relation to existence.

Existence and cause

- **Existence is the ability to (physically) interact with the world/universe.**
  exsistere (lat.) = to appear, to arise, to become, to be,
  meaning "to stand out".
- **The four causes of existence or change in nature** (Aristotle: Metaphysics):
  - the **material cause**,
  - the **formal cause**,
  - the **efficient cause**,
  - the **final cause**.
- But while the efficient cause is sought in the case of genesis and destruction, the final cause is sought in the case of being also.
- Example: The moving cause of a house is the art or the builder, the final cause is the function it fulfils, the matter is earth and stones, and the form is the definition.
- **To exist is to have a specific relation to existence.** (after Bertrand Russell: The Principles of Mathematics, New York, W. W. Norton & Company, 1903, second edition 1937, pages 449-450)
- **The reference person / "owner" is the relation of knowledge and data to existence.**

Conclusions: First "ownership" protection then "data" protection

- **Data is the gold of the information age. "Ownership" matters.**
- **Endangered species is the owner.**
- **There are too many beneficiaries of data protection who are not the "owner".**
- **Data "owners" experience massive loss of control on their data.**
- Data protection is dominated by conflicting interests. Features, financial interests, technical requirements are no reasons not to protect ownership. Are there (business, political, ...) coincidences?
- Data cannot be protected on itself. Other things could be protected:
- The **most important concept is** that of **the owner / entitlement**. How to strengthen the ownership? Giving full rights to the "owner".
- **Claim data ownership, control, and protection.** "Ownership" of data is the core of protecting rights and properties of a person.
- **Claim universal opt-in. Foster auditing for "ownership".**
- **Scrutinise whenever any data related to your activities is claimed other than by yourself.**

HarmonicSS

GDPR on the Rise: Who Will be the Owner of Data and Who Will be in Control?

Privacy Implications of Big Data

Iryna Lishchuk, LL.M.

Institut für Rechtsinformatik, Leibniz

Universität Hannover

THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION HORIZON 2020

INFOCOMP, July 22- 26, 2018

Leibniz Universität Hannover

juristische | fakultät

Status

*"Knowledge is the engine of our economy. And data is its fuel"**

- Broad concept of personal data

  → All is information and all information is personal data

- From informational self-determination to default entitlements

  → From individual´ control to data driven economy

- Who is data owner?

  → He who has default entitlements owns the data

* *N. Kroes, The economic and social benefits of big data, Speech given on 23 May 2013 at Webcast Conference on BigData, Brussels*

Vision

→ What is the power of consent if the data uses are not known?

→ Where is the line between regulated and non-regulated actors in Big Data environment?

→ Where is the line between personal and non-personal data?

→ What if the learning algorithm is the sole entity to assess new data?

→ What is the Consent in Big Data environment?

AI, Big Data and Privacy: A relentless war

INFOCOMP 2018

ATHANASIOS TSITSIPAS
SENIOR RESEARCHER

IOMI, UNIVERSITY OF ULM

- AI & big data perform essentially **statistical analysis**. High risk of **making assumptions** that **do not hold true**.
- Because the system just analyses these relationships, **we don't even know what it learned.**
- Trend development prediction is short-lived and like all the other cases, it is **not clear what was actually learned**.
- **Lots of private data is implicit and hidden** among all the other data we freely give away

- **Data gathering** needs to be **controllable**: for the sake of performance and privacy
- The **relationship** between data provisioning and data mining needs to be **better understood**: which data could violate my privacy *indirectly*?
- **Results** need to be better **controlled**: not only for privacy reasons, but also because of potential misinterpretation
- **Results** need to be better **understood**: what is the implication of the relationship? Is the relationship coincidental or based on cause and effect? Etc.

Big Data Analytics: Statistical, Mathematical, Computational and Legal view

Zlatinka Kovacheva – Middle East College, Oman
Valery Covachev – Bulgarian Academy of Sciences

Statistical point of view: How to get usable information out of datasets that are too huge and complex for many of the traditional methods to handle?

Mathematical point of view: How to formalize the nature of Big Data and apply fuzzy logic, neural networks and other mathematical tools?

Computational point of view: How to solve the problems of data storage and management, communication and computation?

Legal and ethical point of view: How to ensure the privacy and confidentiality of the data?

Some unique characteristics of big date which have to be considered from different points of view:

- Big Data is often nourished by dynamic sources (intense networks of customers, clients and companies) and there is no automatic flow of data that is always available for analysis. There is almost voluntary generation of data;
- The data incorporation in tasks like fitting a suitable statistical model or making a prediction with a required level of confidence is challenging;
- The spontaneous nature of such real- time pro-active data generation can help us to capture complex, dynamic phenomena and enable data-driven decision making process;
- A big data generating mechanism may provide the desired statistical power, but the same may also be the source of limitations.
- Big data have a big potential of being used in unintended manner – e.g. phone records, social networking patterns, etc. The unintended usage of data leads to genuine concerns about individual privacy, data confidentiality and ethics;
- If the data generation costs are low, the people might generate data on as many samples and as many variables as possible;
- The high variety and high dimensionality of big data increase the number of sources of unstructured data such as text, maps, images, audio, video, news, signals and so on.

* Statistical challenges

From a statistical point of view, the large data could arise in the following cases - either huge numbers of predictors, huge numbers of sample size, or both. New statistical methods like Symbolic Data Analysis and Approximate Stream Regression have been developed. Adopting practices such as dynamic pricing overbooking needs statistical demand forecasting and optimization techniques. The use of statistical techniques on large data bases for business decision making is fast catching up in emerging economics.

The methods available now can broadly be categorized into **three groups**:

- **Divide and conquer method**: First, the original big dataset is divided into K small blocks that are manageable to the current computing facility unit. Then, the intended statistical analysis is performed on each small block. Finally, an appropriate strategy will be used to combine the results from these K blocks. As a result, the computation for the divide and conquer method can easily be done in parallel.
- **Fine to coarse method**: In order to make intended algorithms for the big dataset scalable, statisticians introduced a simple solution: rounding parameters. Hence the continuous real numbers of data are simply rounded from higher decimal places to lower decimal places. A substantial number of observations are degenerated to be identical. This idea was successfully applied to the functional data analysis using smoothing spline ANOVA models.
- **Sampling method**: Another more effective and more general solution for the big data problem is the sampling method. This means that we take a subsample from the original dataset with respect to a carefully designed probability distribution, and use this sample as a surrogate for the original dataset to do model estimation, prediction as well as statistical inference. The most important component for this method is the design of probability distribution for taking the sample.

Overall, the main advantage of the sampling method is its general application to various model settings. Moreover, it will automatically give rise to a random sketch of the full data as a byproduct, which is useful for the purpose of data visualization. However, the nontrivial part of using sampling method is the construction of sampling probability distribution, which plays a crucial role in sampling methods.

* Mathematical Aspects of Big Data

- Zh. Sun and P. Wang started to formulate some **mathematical foundations** of Big Data and the **fuzzy logic approach to Big Data**. For this purpose, the concept of the discrete mathematics was used to create a theory for big data. The concept of linguistic variables in fuzzy set, and logic theory allow to deal with a higher level of platform for the benefit of a big data theory. The word “big" was recognized as a mathematical operator. Furthermore, the semantic meaning of cardinality of big data was examined. The characteristics of an important concept of ”infinity" was closely associated with the big data. The relativity of big data is based on the operations of fuzzy subsets, recognizing that the linguistic variable “big" is appropriately being modeled as a mathematical linguistic variable suitable for mathematical evaluation.

- The other widely used tool for Big Data Analytics is **Neural Networks**. The layers size (neurons) of the artificial neural network needs to be increased to accommodate the increased dimensions of the input dataset. After certain point, the network size becomes so huge that it becomes almost infeasible to be implemented efficiently because of the increased complexity induced due to the exponential growth of the interconnections among the nodes (neurons) in the network. This phenomenon is generally phrased as the "the curse of dimensionality" in the field of machine learning. Therefore, there is a need to come out with an algorithm to process large dataset efficiently keeping the neural network size considerably small by optimizing the numbers of neurons and the interconnection between them. The future work will be on optimizing the neural networks.

* Computational solutions

- For computer engineers, a straightforward way to reduce computing time is to resort to more powerful computing facilities. Great efforts have been made to solve the problem of big data by **designing supercomputers**. Many supercomputers have been built rapidly in the past decade, such as Tianhe-2, Bluewater and Blue Gene. The speed and storage of supercomputers can be hundreds or even thousands of times faster and larger compared to that of a general-purpose PC. However, the main problem with supercomputers is that they consume enormous energy and are not accessible to ordinary users. Thus, although supercomputers can easily deal with large amounts of data very efficiently, they are still not a panacea.
- Instead, **cloud computing** can partially address this problem and make computing facilities accessible to ordinary users. Nonetheless, the major bottleneck encountered by cloud computing is the inefficiency of transferring data due to the precious low-bandwidth internet uplinks, not to mention the problems of privacy and security concerns during the transfer process.
- Another relatively new computational facility proposed is **the graphic processing unit** (GPU), which is powerful on parallel computing. However, a recently conducted comparison found that even high-end GPUs are sometimes outperformed by general-purpose multi-core processors, mainly due to the huge data transferring time. In brief, none of the supercomputer, the cloud computing, GPUs solves the big data problem efficiently at this point. Efficient statistical solutions are required, which makes big data problem manageable on general-purpose PCs.

* Legal and ethics aspects of Big Data

Central to the success of the Big Data projects will be four critical factors:

- Understanding the legal framework for Big Data and how it applies to the organization concerned;
- Effectively bringing together the organization's IT and legal functions in the Big Data project;
- A clear understanding of the organization's objectives for its Big Data operations;
- A structured approach to the strategy, policy and process aspects of Big Data governance.

The **key issues** have to be considered:

- Big data and **intellectual property** rights;
- **Data protection**.

General Data Protection Regulation ((EU 2016/679) (GDPR) (NIS Regulations) came into effect on 25 May 2018 pursuant to the Cyber security Directive (also known as the Network and Information Security or NIS Directive).

Any organizational conversation about **big data ethics** should relate to four basic principles that can lead to the establishment of big data norms:

- **Privacy;**
- **Confidentiality;**
- **Transparency;**
- **Identity.**

THANK YOU!

The theme is open for discussion

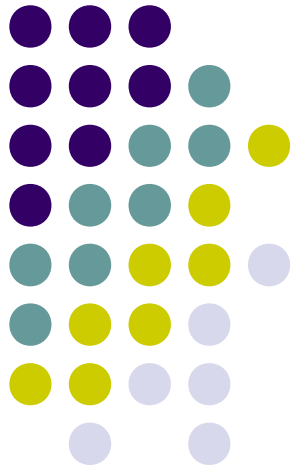
IARIA: DataSys 2018
INFOCOMP 2018 International Expert Panel,
Barcelona, July 24, 2018

General Data Protection on the Rise:

Who will be the Owner of Data? Who will be in Control?

asst. prof. Nikola Vlahovic, PhD

University of Zagreb,
Faculty of Economics & Business,
Zagreb, Croatia.



General Data Protection

- Who will be the Owner of Data?
  - Natural person remains the owner
  - Data collector, data processor – well defined consent from data owner
  - Data pseudonymization and full anonymity of data does not fall under GDPR in EU
  - High quality implementation of ISO/EIC 27000 standard may alleviate these processes

General Data Protection

- Who will be in Control?
  - Intention is tha the data Owners be In control
    - Some studies show high interest, involvement and perception of security risks by data owners especially individual people
    - Long-term involvement?
  - Impact on Businesses, Infrastructure, Public services & Governance
    - Digital economy?
    - E-government
    - Smart Cities?
    - Academic Research?

General Data Protection

- Economic & Social Implications
  - Favourable Business Environment?
  - Restrained Smart City Developement?
  - Increased Privacy Protection or Increased Social Alienation?

IARIA: DataSys 2018
INFOCOMP 2018 International Expert Panel,
Barcelona, July 24, 2018

General Data Protection on the Rise:

Who will be the Owner of Data? Who will be in Control?

asst. prof. Nikola Vlahovic, PhD

University of Zagreb,
Faculty of Economics & Business,
Zagreb, Croatia.