



DATA SCIENCE AND HEALTHCARE

**Pascal Verdonck
Marc Van Hulle (et al)**



Koninklijke Vlaamse Academie van België
voor Wetenschappen en Kunsten, 2018
Standpunten 48 b

Data Science and Healthcare



KVAB Press

Published
by
the Royal
Flemish Academy
of Belgium
for
Science
and the Arts

Standpunten nr. 48 b



KVAB Press

Hertogsstraat 1
1000 Brussel
Tel. 02 550 23 23
www.kvab.be
info@kvab.be



Data Science and Healthcare

Pascal Verdonck
Marc Van Hulle
Bart De Moor
Erik Mannens
Rudy Mattheus
Geert Molenberghs
Femke Ongenae
Marc Peters
Bart Preneel
Frank Robben

Partial reproduction is permitted provided the source is mentioned.

Suggested citation: Pascal Verdonck, Marc Van Hulle (et al), *Data Science and Healthcare*, KVAB Standpunt 48 b, 2018 (Original text: Dutch).

© Copyright 2018 KVAB
D/2018/0455/01
ISBN 978 90 6569 918 11

Translation Dutch manuscript: Sandra McElroy

Photo and cover design: Anne-Mie Van Kerckhoven

Data Science and Healthcare

CONTENTS

Summary	2
Preface	3
Introduction and framework of this report	3
1. Big Data and data science: observations and definitions	5
2. Impact on professional training and jobs	7
3. Impact on medical/clinical research	8
4. Impact on the actors in healthcare	11
5. European legislation	13
6. Belgian legislation and approach to healthcare	15
7. Quality assurance of healthcare data	17
8. Privacy of the patient	20
9. Viewpoint: are informed consent and privacy still realistic?	22
10. Recommendations	23
11. Conclusion	24
References	26
Appendix	28

Summary

The Big Data ecosystem consists of five components: (1) data creation, (2) data collection and management, (3) analysis and information extraction, (4) hypothesis and experiment and (5) decision-making and action. We propose to use data science in which the systematic use of data through applied analytical disciplines (statistical, contextual, quantitative, predictive and cognitive models) leads to data-based decisions. For each segment of this ecosystem, there is a need for a ced professional education and job classification: the data engineer, data scientist and the data strategist.

The role of data science in healthcare is threefold (triple AIM): an increase in patient experience, quality and perception, better public health and cost savings. The EU General Data Protection Regulation (GDPR) reconciles two objectives: better protection of personal data for individuals, and more opportunities for business in the digital single market through simplification of regulation. The implementation of this for the individual Belgian patient - with access to his health data through a consolidated platform - must be realised by May 25, 2018.

Availability, accuracy, reliability and security are essential conditions for the added value of data science in healthcare. Data must be available anonymously for research purposes, in such a way that the identity of the patient is protected. The latter will be increasingly under pressure due to technological developments. Currently there is no legislation available to make adequately protected patient data available to parties, other than traditional healthcare providers, who may benefit from it (for research, product development ...) without the patient having to give his/her consent for a similar purpose of use. This should take into account European regulations that assign an important role to the data controller.

Preface

Position paper

The Academy's *Standpunten* series (Position Papers) contributes to a scientifically validated debate on current social and artistic topics. The authors, members and workgroups of the Academy write under their own name, independently and in full intellectual freedom. The quality of the published studies is guaranteed by the approval of one or several of the Academy's classes. This position paper was approved for publication by the meeting of the Class of Technical Sciences on 8 March 2017.

Introduction and context of this *Standpunt*

Health technology is one of the six essential building blocks that the World Health Organization (WHO) regards as essential for the achievement of a stable and sustainable global healthcare system. The other five are: financing, health workers, information, service and leadership/management. If one (or more) of these six facets is absent or cannot be developed sufficiently, then healthcare cannot function on the level that is required to improve the health of people and nations in a sustainable manner (WHO, 2010).

In the near future technology will have an even greater impact on the preventive, diagnostic and therapeutic possibilities of medicine and healthcare. Current-day medicine is increasingly evidence-based. Genetic and clinical parameters are becoming more identifiable and measurable, as a result of which we are confronted with a plethora of medical data. That offers enormous opportunities, provided that the data can be modelled and clustered with advanced numerical techniques. They can be used in the various phases of the health cycle (prevention, early risk, acute, chronic): by the patient (e.g. personalised websites), the care provider (e.g. decision support systems for assisting diagnosis, automatic pilots, genetic data analyses) and finally by the government as well (e.g. a more efficient health care system due to the modelling and monitoring of data).

Apart from the biomedical and clinical research communities, the pharmaceutical (and other) industries (mostly commercial) can also extract benefits from the re-use of patient data. But because scientific research is anything but a linear process, it is difficult to estimate in advance the potential of these data for a particular research question.

There is a clear consensus that patients, care providers, academics, and the healthcare and pharmaceutical industries can benefit from a better use of health-related data (*Electronic Health Record*, EHR): a more intensive study of such data will not only help us to better understand diseases, make better diagnoses

and evaluate existing therapies – and develop new ones – it will also create the transition to a modern healthcare system that strives for the best care for the patient. Big Data enables the researcher to process unprecedented quantities of information and discover unexpected patterns. There is a strong belief promoted that the challenges brought by chronic disorders such as cardiovascular diseases, as well as incurable diseases like dementia and cancer, can hereby be met head-on. For the care providers and health insurers, there is the possibility to stem the ever-increasing costs of health care. The EHR of the individual may offer new insights, such as a better understanding of the effects of medical interventions and the efficacy of the care trajectories. The EHR may also help when quantifying the value of certain medical technologies.

1. Big Data and data science: observations and definitions

Big Data is not the same as 'a lot of data and their metadata' (data about data) (*volume*). We are talking here about data sets in terms of petabytes (10th to the 15th power) or even exabytes, zettabytes and brontobytes. By way of comparison: an average hospital now creates 750 terabytes of information every year. The European BioInformatics Institute possesses 20 petabyte of data about genes, proteins and small molecules. Moreover, the total electronic data volume doubles every two to three years. It is also characterised by a very great variety of data as concerns the following:

- type: structured versus non-structured, in diverse formats. The data are heterogeneous;
- area of application: personal versus business-related, government. The data are extremely diverse;
- sources from which data can be mined and combined or linked. That goes together with the far-reaching automation of data gathering;
- granularity. This is a consequence of the 'datafication' of all aspects of everyday life that were never or rarely quantified in the past: locations, friendships (social media), academic competencies, consumer behaviour, surfing behaviour, physical activity ...

Big Data is supply driven: data are generated without much accountability by e-mails, personal images and videos, online transactions (purchases, payments, etc.), online search terms, streaming data, messaging, social interactions, knowledge networks, medical records, sensors (chemical, biological, electronic, ...), interconnections in medical and civil equipment in care institutes and the home environment ...

Big Data is also technology driven: consider the automation of data gathering, the distributed storage of data (*in the cloud*), the centralised and distributed data management platforms, the methods for analysing and visualising data, even in real time.

Big Data is therefore more than just data: analysis is also an integral part of it. The Big Data ecosystem includes the use of advanced heuristics, statistical procedures, neural networks, machine learning algorithms, artificial intelligence techniques, ontology-based search strategies, inductive reasoning algorithms, pattern recognition, forecasting algorithms, etc. The intention is to discover possible hidden but significant characteristics, connections and patterns. The major challenge for Big Data is that the analysis should lead to authentic knowledge (veracity) that provides a demonstrable added value (*value*) on time (*velocity*) and in answer to a given research question.

Big Data is a comprehensive and tiered concept that relies on a versatile and self-renewing technology platform for mass data bundling in a (virtual) data pool, coupled to very specialist algorithms, techniques and software. The aim is to gain insight into the data and extract new knowledge that can be used in a timely fashion (*time to action*).

The Big Data ecosystem consists of five components: (1) data creation, (2) data gathering and management, (3) analysis and information extraction, (4) hypothesis and experiment, and (5) decision-making and action. As such, Big Data is undeniably a process: '*Big Data is the capacity to search, aggregate and cross-reference large data sets.*' (Boyd & Crawford). Taking all this into account, it is perhaps better to talk about data science in which the systematic use of data via applied analytical disciplines (statistical, contextual, quantitative, predictive and cognitive models) leads to data-based decisions for production, management, research, education, and healthcare ...

A recent report by the European Commission defines data science as follows: '*Big Data in Health refers to large routinely or automatically collected datasets, which are electronically captured and stored. It is reusable in the sense of multipurpose data and comprises the fusion and connection of existing databases for the purpose of improving health and health system performance. It does not refer to data collected for a specific study.*'

In healthcare data science will quickly lead to results:

- increased effectiveness and a rise in the quality of treatments by, for example: earlier intervention in vascular diseases, a reduced risk of side effects from drugs, fewer medical errors, the fusion of networks, such as social and disease networks;
- increased possibilities for disease prevention by identifying risk factors in the population, in a subpopulation and at the individual level, and by improving the effectiveness of interventions to help people develop a healthier lifestyle;
- improved patient safety due to the possibility of taking better-informed medical decisions, based on information delivered directly to patients;
- a greater ability to predict results;
- better distribution of knowledge;
- reduced inefficiency and wastage, and improved cost management.

International example projects are listed in the appendix to this text.

2. Impact on professional training and jobs

For each segment in the Big Data ecosystem, appropriate professional training and a job classification is required. For example, the data engineer, the data scientist and the data strategist:

- The data engineer breaks open the existing data silos. He automates the uploading and the (virtual) aggregation of data, so that it remains up-to-date, whilst also dealing with missing data standards – even in specific domains such as healthcare. He is a data manager. He is familiar with the internal and external data landscape and works closely with the CIO, *Chief Information Officer*.
- The data scientist is very different from the traditional quantitative data analyst. She is an explorer and her strength lies in associative thinking. She is familiar with the sector. She gets to work with the mathematical tools, loads the data from one algorithm to the next and writes the required code. She experiments with prototypes, builds descriptive or predictive models and creates systems for a continuous dialogue with the data, rather than traditional ad-hoc analyses. She visualises the results with a view to effective communication with the other business functions. She provides the end-users with really intuitive interfaces with speaking graphic possibilities. She provides insight into applied data treatment and processing.
- The data strategist forms the bridge between business decisions in the business units and the technical/scientific Big Data disciplines. He develops the Big data strategy, selects the opportunities with the greatest impact and is responsible for the implementation. His most important collaborator is undoubtedly the person responsible for the marketing channels (personalised e-mails). The data strategist is familiar with national and international legislation. He is also aware of the best practices and uses them. He safeguards the global social support for his Big Data objectives.

There is a shortage of specialists with these profiles. Finding the best people to fill these positions as soon as possible also helps in the pursuit and implementation of patient-oriented health technology.

3. Impact on medical/clinical research

Big Data is not a hype. It is a new way of getting to know the world and thus managing the world. It is a paradigm for knowledge and decision-making, and for the rationalisation and management of behaviour. Data maximisation offers opportunities for relevant, usable inventions, patterns or relationships that are often unexpected and cannot be retrieved in any other way. The value of data explodes when it is linked to other data. Data integration is therefore a significant value creator. It promotes a shift in scientific research methods: from the ubiquitous hypothesis-driven research to data-driven research that relativises hypothesis- and model-formation and the usefulness of the experiment. Correlation ousts the causal (original) link in the search for explanations and clarification. The figures speak for themselves. No more sampling to find an answer to the question or devoting energy to purified data. All you need to do is collect data with 'n = all' to be able to answer the question. The claims about data-driven management, decision-making, health care and an equally data-driven government all fit within the margins of this paradigm.

And yet a fair bit of research and development is still required to raise the level of the functionality of the analysis. Data filtering and compression, self-documented data containing the necessary and sufficient metadata, information extraction from text, speech, video, etc., are therefore needed. So too are data cleaning, data integration, an automated design of databases, data querying and mining ...

Benefits and opportunities

There is a clear advantage to making health-related data available for research ends: patient data, but also medical images, biobanks, test results, clinical trials. A more intensive study of clinical data will not only help us to understand diseases better, make better diagnoses, evaluate existing therapies and develop new ones, but also enable the transition to a modern healthcare system aimed at providing the best possible care for the patient. Consider also evidence-based medicine (EBM): the explicit, insightful and conscientious use of the best available evidence, such as double-blind and randomised clinical trials, in the choice of treatment.

The interest in the healthcare and pharmaceutical industry for Big Data is well documented. These sectors are already dealing with huge quantities of data from doctors and healthcare institutions. According to Richard Bergström, director-general of EFPIA (European Federation of Pharmaceutical Industries and Associations), the importance of broad consent as a means of making medical data available for re-use cannot be overestimated (Bergström, 2015). This would make the enormous quantity of data that has already been collected available for innovation. Effective legislation would therefore also have to contain a proviso that makes medical data – once adequately protected – available for research or

health-related ends, without the need to ask for consent each time. After all, it is unrealistic for the legislator to estimate the extent of such consent in advance. Legislators must also recognise that research is not a linear process and that it is rarely possible to estimate the potential of patient data in advance.

The development of drugs is an example of this. Not every genetic or lifestyle characteristic in the population is tested when a drug is allowed on the market. Thanks to the systematic analysis of the use of the drug, conflicts can be dealt with in time so as to spot unexpected side effects, prevent sequels and analyse the spectrum of outcomes. If specific information is available, it is appropriate to use this when treating the individual patient. That's where 'personalised medicine' comes in, with the aim of matching individual genetic and clinical characteristics with the best available treatment.

Biological pharmaceutical companies are looking at genetic/protein paths in the body. Finding out where best to intervene for the remediation or management of a disorder is like looking for a needle in a haystack. Big Data on the basis of genetic data from thousands of individuals, ideally coupled to the corresponding family and clinical data, is therefore a valuable tool.

The anonymisation of data – a given individual cannot be linked to his data – has to meet strict criteria for use in Big Data and must be externally certified. It will be extremely important to strictly monitor the distinction with regulated healthcare. It is after all crucial that all data remain accessible to doctors and care personnel for use in their profession and in their relationship with the patient.

It is clear that the availability of Big Data is an enormous treasure trove for the data analyst and the statistician, and that it also solves a number of traditional problems within empirical research. One example is that significance usually ceases to be an issue for the studied effects and relationships. Even when different (or many) effects are being studied, multiple comparisons are far less of a problem. In many cases the most conservative corrections for multiple tests will still deliver sufficient significant results.

Concerns

However, there are still a number of concerns. A significant effect or difference is not automatically clinically or epidemiologically relevant. Take, for example, a blood pressure study. A difference in blood pressure between two groups may well be significant if the magnitude of that difference is 0.1 mmHg, but the importance of that difference is still doubtful. Although Big Data provides a sufficient statistical capability and thus also significance, we must always be on the alert for bias. That is closely correlated to the set-up of the study. In terms of design, Big Data are often comparable with surveys and/or observational studies.

If we look at it from the standpoint of a survey, we have to check whether the data, however wide in scope, are representative for the population. If not, it is best to use the appropriate (weighting) techniques for correction purposes. These are available, but we should not make the mistake of thinking that they are now superfluous.

From an epidemiological perspective, there is another major problem. Even when the available data include the whole population, there can be distorting effects, such as confounding and effect modification. Breslow and Day (1987) showed that when, for example, a disorder has a different natural prevalence in two groups (e.g. men and women) and the same goes for a risk factor, corrections must be made for gender in order to obtain a pure estimate of risk, even when we know the whole population. That is counterintuitive. As a result there is a real danger that it will be forgotten in the context of study data that are selected from a really large data flow.

These observations suggest that, in order to draw meaningful conclusions on the basis of Big Data, we should (continue to) make use of good experimental and epidemiological design.

We can't escape the fact that the unprecedented availability of data offers possibilities that were previously unthinkable, such as personalised medicines and, closely associated with that, dynamic treatment allocation (Zhang et al. 2012). If we want to find out what the optimal treatment scenario is for a given patient, we cannot do it without a mass of data. In other words: Big Data has helped give rise to the development of new disciplines within statistics, such as dynamic treatment allocation. This concerns mathematically sophisticated methods that are also particularly relevant for patient, practitioner and care organiser.

There are of course concerns about the use of algorithms (based on Big Data). There is no more a gold standard here than with the verdict of the practitioner or a group of practitioners. There will be a silver standard at best, and then only in certain cases. Which means that we need to continue taking into account both false positives and false negatives. It will also be appropriate, therefore, to properly support the use of such methods with insights and methods from diagnostics.

An additional concern is what happens with the huge amounts of health-related data that are amassed outside healthcare itself, in neighbouring sectors and even in more peripheral applications. Insurance companies and banks may use them to increase or reduce premiums, or to refuse products. That will give rise to various problems, in addition to privacy concerns. A profile that is created using Big Data remains a prediction that may differ from reality to a significant extent, both in terms of the current and future state of health. Moreover, excessive individual

profiling may undermine the solidarity principle on which the insurance concept is based. The individual and collective must therefore be carefully and continuously weighed up against each other.

4. Impact on the actors in healthcare

There are numerous initiatives concerning data and people demanding care. Some developments are fast, others slow. Some are closely related to the patient, others are conducted more in the medical sector. One of the problems is that there is currently no common strategy from the perspective of the patient. Patients have high expectations of the digitisation of care, but there's still a lot to be done and the data still has to be positioned and used in the right context.

The digital transformation is about the increasing application of digital models and processes in all aspects of an organisation. The aim is to radically improve the value and performance of the organisation. The added value is the relationship between the clinical result (patient outcome) and the system costs.

Due to an ageing population our society has an increasing number of chronic patients with various disorders. In order to provide the best possible care and supervision it is essential that all care providers (doctors, pharmacists, nurses, carers,...) of the patients communicate quickly and efficiently with each other and always have an understanding of the most recent medical information relevant for their care tasks. Communication with the patient is also essential.

ICT offers the possibility to measure all sorts of things. The danger is that, in the process, care is compromised rather than supported if we don't upscale fast enough to obtain the buy-in of the three actors: the patient (person with the care need), the care provider and whoever pays the bill (the payer).

Digitisation is also a leveraging tool in the evolution towards paper-free care for everyone: patients, care providers, health insurance companies and the government. It is vital that care providers only have to register medical details once (only-once principle). This will reduce the amount of administration and allow them to spend more time with their patients. Patients and health insurance companies (payers) will also have less paperwork, enabling them to focus on other tasks.

The changes that the new technologies cause are disruptive. They offer not only extra possibilities, but also introduce a different way of working for people, organisations and society. The role of data science in health care is threefold (triple AIM):

- to increase the patient experience, quality of care and patient perception;
- to improve the health of the population;
- to reduce costs.

Putting the patient at the centre

Patients are becoming more empowered and are increasingly involved in the use of their data. In a modern care system they play a central role as co-producer of their own health, a task for which they are also best equipped. However, with the use of various data silos in healthcare, data exchange has not become any simpler. This has to be resolved with appropriate haste.

The person demanding care must assume a central position in the care system: he/she is the ultimate user. He/she can exercise influence in various ways and via various channels:

- as a consumer: he only really buys care products that are not covered by the insurance package, often as a means of prevention, such as health and fitness apps;
- via the care provider (individual): the patient can convey her request or need for care innovation to the care provider. He/she indicates, for example, a desire for video calls. If a great many patients make this request, care providers will become sensitive to it;
- via the care provider (joint): a care provider is the ideal person to organise the mass demands of patients. For example, a doctor can involve all his diabetic patients in the choice, implementation and evaluation of e-health applications in the area of diabetes care;
- via the health insurance company: the relationship between patient and health insurance company is primarily seen as a necessary transaction. The lack of trust on the part of the insured person is relatively high and the patient and health insurer rarely see each other as bed mates in the area of care innovation. Digital administrative simplification and the benchmarking of data are possible new roles for insurance companies.

Many care institutions are not used to innovating, whereas in fact the information society demands it. They would do well to invest in three main areas of their value proposition: patient experience, operational processes and business models. If the digital transformation is to be upscaled, the lifestyle and needs of the person requiring care must be taken as a starting point in the organisation of care.

These elements become more important when you make care providers responsible for improving their patients' clinical results. We have to realise that medical records alone are not enough to take 'self-management – empowerment' to a higher level and to improve care provision.

The person requiring care wants a digitally supported experience. That starts with the use of digital tools and the Electronic Patient Record (EPR). Those requiring care point the way with the use of apps and their demand for access to their medical records. The EPR is a strategic platform in the hospital and the new network forms. It should increase efficiency, facilitate the interaction between data, speed up innovation and put the patient at the centre of the data strategy so as to improve clinical results and keep costs under control.

'Self-management – empowerment' will allow the person requiring care to play an active role in his own treatment and create greater value. That can range from medication schedules, lifestyle and prevention to the right care at the right moment by the right care provider. In order to increase patient involvement and satisfaction, care organisations must make efforts to bridge the gap between supply and demand as regards digital tools and strategies. Those requiring care are willing to monitor their health with digital tools and to share these data with the care professional. This provides care actors with the opportunity to share data more transparently with the patient.

The future hospital is a network of care function components that relies on integrated information flows. It strives for the best resolution to the demand for care, with the best value and the highest clinical result at an acceptable cost.

5. European legislation

Data Protection Regulation is an important factor in monitoring the use of Big Data. On the one hand the EU has to offer the citizen protection and on the other also ensure that stakeholders have enhanced access to the data.

Citizens must have control over their own personal data: it is important to enable people to decide for themselves (*empower*) what risks they want to take by making personal data available. Nowhere are the benefits and risks more sharply defined than in the case of medical data. Citizens have to give their express consent for data to be collected about them (*opt-in*). Moreover, they have to be able to concur with the purpose for which their data are being used. As a further control, data controllers must be able to prove that this consent has in fact been given.

It has taken twenty years for the EU to update its 1995 (95/46/EC) regulation. This was drawn up at a time when most people had no access to the internet, mobile phones or social media. The European Parliament has repeatedly urged the European Council to update the EU regulation concerning data protection. Since 2012 the European Commission has been working on a new draft text and on 24 June 2014 talks between Council, Commission and Parliament officially started.

According to Jan Albrecht, reporter in the Parliament on regulations concerning data protection, the Council and the Parliament did not share the same opinions about the rights of the citizen and the responsibilities of the *data controllers*, but there was a general consensus on the fundamentals of the new regulation: one set of rules, valid for the whole EU, giving back control to citizens of their data, the same rules for companies inside and outside the EU and an effective simplification (*one-stop shop*) of life for citizens and companies. Finally, the new regulation had to be technologically 'neutral' and should therefore not close the door on future innovations.

On 15 June 2015 the European Council reached a political agreement on the basis of the negotiations with the European Parliament. The aim: a general agreement for a new EU regulation on data protection, adapted to the digital age (*Data Protection: Council agrees on a general approach*, 2015). The new regulation reconciles two objectives: better protection of personal data for individuals and more opportunities for business in the digital single market by simplifying the legislation.

Better rights on data protection give those concerned more control over their personal data:

1. easier access to their data;
2. clear understandable information about what happens with the data;
3. the right to remove personal data and be 'forgotten';
4. the right to portability, in order to effect a simple transfer of personal data.
5. establishing limits on 'profiling' or the automated processing of personal data with the aim of evaluating personal characteristics.

Citizens have the right to submit a complaint about the improper use of their data and in this event to demand remediation and compensation. Data controllers must implement the necessary security measures and promptly inform the supervising authority about any breaches to personal data and about who has suffered harm as a result. Finally, they must be able to give guarantees when personal data is transferred outside the EU.

There are more opportunities for business because the rules of engagement within the EU are the same for everyone.

On 4 May 2016 the official texts of the regulation (2016/679) were published in the *EU Official Journal* (Document 32016R0679). The regulation came into force on 24 May 2016, but is only valid from 25 May 2018 (*Protection of personal data, Regulation of 4 May 2016*).

6. Belgian legislation and approach to healthcare

We are rapidly moving towards eHealth. Here we give a brief overview of Roadmap 2.0, with a few points of focus for the various care actors.

The care actors

Every general practitioner manages an electronic patient record (EPR) for each patient and publishes and updates a SumEhR for each patient in the secure safe (Vitalink, Intermed or BruSafe). The general practitioner has access to all relevant, published medical information about his/her patients via his EPR.

Every hospital, psychiatric institution and laboratory makes certain documents electronically available with reference in the hub-metahub system and can consult relevant data from the secure safes. Every hospital has an integrated, multidisciplinary electronic patient record (EPR).

An EPR is defined for all other kinds of care providers. They too can consult and update certain information from their EPR in the secure safe. Medicines and medical services are prescribed electronically. Pharmacists publish information about the medicines administered in the shared pharmaceutical record (GFD), which feeds into the medication schedule. The patient's medication schedule is also in the secure safe and is shared by doctors, pharmacists, homecare nurses and hospital staff, among others.

An effort is made to create and publish as much medical information as possible in a structured and semantically interoperable way.

All care providers can communicate with each other via the ehealth box; there are a number of electronic standard forms for this purpose. The care providers can practice telemedicine using *mobile health* applications that are officially registered. This registration depends on a number of controls in the area of data protection, interoperability, an EU label for medical appliances and *evidence-based medicine* (EBM). The registers are optimised and standardised, and registration is automatic where possible from the EMD/EPR.

Implants and medicines are tracked according to international standards. All data is exchanged electronically between care providers and health insurance companies. The registers are optimised and standardised, and registration is automatic where possible from the EMD/EPR.

The care providers receive incentives for the use and meaningful application of eHealth; financial incentives can have both a federal part and a part for the various Communities and Regions.

Each care provider is also trained in eHealth, by means of the basic training package and with top-up training. Each care provider has a one-stop shop that provides all administrative information on behalf of RIZIV (the Belgian National Institute for Health and Invalidity Insurance), the FPS Public Health and the regions (only-once principle).

The patient

The patient has access to the information that is available about himself in the secure safes and via the hub-metahub system; filters may be defined for this (this is still under discussion). Research is underway to find out if it is feasible to provide a consolidation platform on which all the information about the patient is added as well as the analysis and translation tools for the patient, so that he/she can better understand the file. This last aspect contributes to his '*health literacy*'. The patient can add information himself, via the consolidation platform, in the secure safe, via a hub or in a secure cloud.

All the information from the hubs, the safes, the consolidation platform and potentially the secure cloud forms the PHR (*Personal Health Record*) of the patient. Other relevant information is also available via the consolidation platform from the health insurance funds, the National Companies Register of Social Security and other relevant sources, such as living wills concerning organ donation or euthanasia.

The patient has access via various channels to his/her PHR, e.g. via a smartphone app. The patient is hereby informed and brought up to date about his/her actual situation and can play a crucial role in his/her treatment. In theory the patient no longer receives anything on paper from the doctor (unless requested); the certificate outlining the provisions delivered are sent by the doctor to the health insurance company, the drug prescription is available in the medication scheme, the proof of work incapacity is sent electronically to the employer and the patient receives the proof of receipt in his electronic mailbox. All of the above requires the patient to give his/her informed consent in advance.

The platform aims for the effective organisation of the mutual electronic service provision and information exchange between all actors in health care, with the necessary guarantees concerning information security, the protection of personal privacy and professional confidentiality. This should lead to:

- optimisation of the quality and continuity of health care provision;
- optimisation of the confidence of the patient;
- simplification of the administrative formalities for all actors in healthcare;
- effective support for the healthcare policy.

The legislation should make provision for making patient data available (when adequately protected) to the parties mentioned, without the patient having to give his/her permission each time for a study. The legislation should also take into account public interest. The government standpoint in this regard states that:

- patients are owners of the data and allow certain people access to (parts of) their medical records, within a restricted time limit. So more of an opt-in than an opt-out: you have to actively give the right, it is not awarded by default, except to your general practitioner who manages the Global Medical Record (GMR) (+ basic versus advanced access to data).
- A patient must have the right to be 'forgotten' with regard to data.
- A patient has the right at all times to withdraw access rights that he or she has given.
- A patient always has the right to consult all medical data related to him that have been stored somewhere and to be informed of the existence of these data.
- Health data must never be sold to third parties.

7. Quality assurance of health data

The combination in healthcare of data from several sources can ensure that use is made of the existing synergies between data to support clinical decisions. An effective analysis of these integrated data can also result in completely new approaches to the treatment of diseases. The combination and analysis of multimodal data brings with it various challenges that can be managed with Big Data technologies. The current definitions of Big Data place the emphasis on the aspects of volume, variety, veracity and velocity. These are the 4 Vs of Big Data. Within the medical domain, with the associated data sensitivity, the following aspects are important: availability, veracity (i.e. quality, validity and correctness) and reliability.

Availability

One condition for the effective (re)use of different sorts of clinical data to support decision-making, patient follow-up and clinical research is that the data are FAIR: findable, accessible, interoperable and reusable.

Major barriers restrict access to and exchange of medical data between institutions and even between departments in the same institution. Research, clinical activities, hospital services, education and administrative services all have their own data silos. In many organisations each silo maintains its own (sometimes duplicated) data in the self-chosen structure. This heterogeneity obviously forms a barrier to

combining and analysing data in the various silos, and thus acquiring insights. There are two ways to solve this problem. One is top-down: the government introduces Big Data initiatives to enable hospitals, general practitioners, etc., to share their data with each other. The aim of the bottom-up approach is to make patients the owners of their data and to make the data patient-oriented. With this approach patients should have access to their own data and be able to decide with whom it is shared and for what ends it can be used. Examples include initiatives like PatientsLikeMe and openhumans. The social network PatientsLikeMe allows patients with the same disorder to interact with each other, builds up a database with personal data that can be used for analysis and offers a platform for linking patients to clinical studies. Openhumans goes further and requires that all "human" data have to be made public for research. By linking individual who are open to sharing research data about themselves with researchers who are interested in the use of those data, these data can be used again and again and the lessons can be built upon.

Even when the data are shared and thus available and accessible, they are still not interoperable and reusable. The data in healthcare are often fragmented or were generated by heterogeneous sources, impossibly incompatible formats that consist of both structured and unstructured data. Due to the lack of cross-border coordination and technology integration in healthcare there is a need for open standards to enable interoperability and compatibility between the various components in the Big Data value chain. An increasing concern is the lack of industrywide standards for capturing patient-generated health data (PGHD) and for the interoperability of medical devices, such as heart rate monitors. Although many developers already use the consolidated CDA standard, there are still many devices, such as Fitbit, that have their own format. This makes interoperability difficult because the patient often owns several devices. Standardisation organisations like HL7 are working on this challenge. They are currently focusing on standard methods for capturing PGHD, which they want to make interoperable with existing standards for structured documents, such as CDA. It is therefore important that existing health standards and terminologies are used as much as possible in IT. However, it is likely that as a result of the demands and wishes of the various stakeholders in the Big Data chain (patients, suppliers, EMD sellers, application developers, etc.) new standards will continue to be developed. Given that healthcare recommendations, standards and policy are evolving constantly, flexibility should be built into the new IT (Big Data) technologies, to deal with these continuous changes.

The shared data must be integrated in a *Big Data Repository* or *Warehouse*. That occurs in three steps: extraction, transformation and loading (ETL). Data flows must be analysed; a decision must be made as to which data to integrate, aggregate, clean up and leave out in order to guarantee qualitative data; the way in which different pieces of information can be related to each other must be looked

at; it must be decided which rules can be released on the data; an investigation must look at which cases require how much response time; analytical queries and algorithms which draw the necessary conclusions must be made; data governance must be examined so that it complies with legal requirements; the infrastructure must be set up to guarantee scalability, low latency and performances; and there must be an examination of how the data will be made available to different parties. Semantic data integration, whereby the meaning of the data also becomes clear across data silo boundaries, lies within reach thanks to Semantic Web technologies. These enable a context-sensitive interpretation of data from heterogeneous data sources. You can also obtain a graph-based representation of the data, showing the relationships and links between various data points. Ontologies are used for this purpose. An ontology models all concepts and their associated properties and relationships within a certain domain. There are already various (standardised) ontologies available for healthcare. In the Semantic Web data are then modelled using an RDF (*Resource Description Framework*). RDF presents sources (resources, data) and their relationships as triples in the form of subject-object-predicate (SOP). Such standardised ontology models can easily be requested by using the RDF query language SPARQL, which matches the query pattern to the underlying graph. The mapping of the heterogeneous datasets on standardised ontologies facilitates the sharing of such sets and the correct reuse of the data across various applications.

Veracity and reliability

Data about health care are collected in a broad context. As a result their quality and validity vary greatly. For instance, sensors that the patient uses at home can collect data about his/her health, for example heart rate. But the various sensors differ greatly in the accuracy with which they collect data. Given that the patient is not continuously being checked by medical personnel, the context in which these data were collected is not clear either. A poorly functioning device or network connection, or the forgetfulness of the patient, can mean that data are missing or incomplete. At the other end of the spectrum are the data that are collected in hospitals and laboratories. Although these are more reliable and more accurate, it is still often difficult to uncover the circumstances in which the measurements or samples were taken. There is a great desire for reliable and reproducible results, especially in medical and pharmaceutical research where data collection is very difficult and/or expensive. The bringing together of data and results with differing levels of accuracy, validity, quality and reliability is a very challenging problem. The first step in making heterogeneous datasets operable for analysis and conclusions is to offer researchers tools for the easy mapping of sets on each other or linking of them to each other. Despite the existence of a great number of tools, it is still difficult to make data from different sources and with different formats interoperable using semantics (e.g. linked open data (LOD) cloud). There are still no mapping formalisms indicating how such heterogeneous data can be mapped

on an RDF model, in an integrated and interoperable way. The recently developed RDF *Mapping Language* (RML) fills this gap. Thanks to RML, mapping rules can be simply defined from a certain dataset on a semantic model, in a source-agnostic and expandable way. This results in greater integrity within datasets and a more enhanced link between heterogeneous data sources.

A second step for reliably making data available is the accurate establishment of the provenance of the data. This gives some insight into the origins: not only how they were collected and under what circumstances, but also how they were processed and transformed. That is important not only for the reproducibility of the analyses, but also for estimating the reliability of the data. Given that the complexity of organisations and operations is increasing and new methods of analysis are very quickly becoming available, it is essential to establish the provenance of the data. The provenance can significantly affect the conclusion of the analysis. This is why comprehensibility and reliability must be basic requirements for all applications of data analysis within healthcare.

Algorithms, techniques and models must be developed, not only to model provenance, but also to extract this automatically and to annotate the data during the entire dataflow. The provenance of data can also be constructed on the basis of heuristics if this is not available. Here too Semantic Web technologies can play a role. The workgroup World Wide Web Consortium (W3C) Provenance has brought out a number of models and standards: these 3C Prov can be used to establish data provenance. The models are also modelled as RDF, whereby they can be easily integrated with other RDF models which, for instance, model the medical knowledge. Thanks to the generic core specification of W3C Prov nearly every use case concerning provenance can be modelled in an interoperable way. By using the W3C Prov model the data provenance of various ehealth applications and use cases can easily be made interoperable, whereby previously unforeseen links between applications and data can be exposed. However, it needs to be investigated which domain specific extensions can be used to expand the W3C Prov model in order to capture the provenance and context of all the health data collected.

8. Privacy of the patient

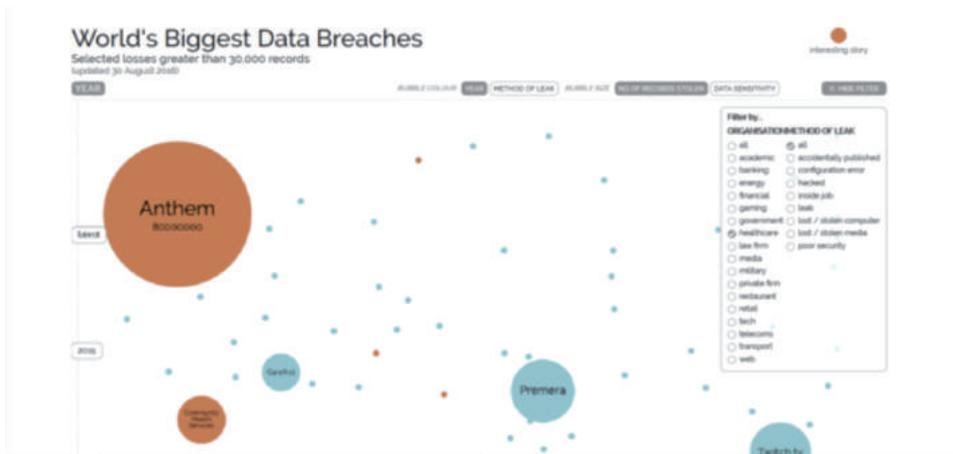
MEP Jill Evans states that data management and data security are challenges that we need to tackle (Mackay, 2015). The confidentiality of patient data must always be guaranteed and patients must feel secure in the knowledge that their medical information is not wrongly shared. There is room for sharing medical data and results of clinical studies if they are relevant for policy decisions, and on condition that the privacy of the individual is guaranteed and that, where necessary, personal

data are anonymised. According to Nicola Bedlington, secretary general of the European Patients Forum, EHR developers must not only take into account the needs of the users but must also mask sensitive data or give patients control over who can consult the data. Individual consent must be given for new applications of existing data, and the modalities to provide for this must be investigated, as must the role of data sources over which the patient has control.

According to Katrín Fjeldsted, chairwoman of the Standing Committee of European Doctors (CPME), personal information must always be used in an ethically responsible and secure manner, given that it forms the basis of the relationship of trust between patient and doctor (Mackay, 2015). Paolo Casali, the public policy chair of the European Society for Medical Oncology (ESMO) advocates the concept of one-time consent: this offers the patient the possibility of giving fully informed and revocable consent. Patients would thus be able to “donate” their personal data for research purposes, with strict conditions for the use thereof. Population-based data can inform health policy and play an important role in medical breakthroughs. However, this is only possible if the data are complete. An important challenge is therefore to translate the legislative framework (the law on the protection of privacy and GDPR) into workable technical solutions.

Problems and solutions

- Mass data leaks are a reality, also in the health sector (<http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-reaches-hacks>). It is not at all clear whether we have all the technological and organisational solutions to reduce the number of incidents. The harm to the patients involved and to society is very difficult to estimate. The possible risks must be properly weighed up against the possible benefits.
- The anonymisation and ‘pseudonymisation’ of data has so far played an important role in protecting medical data for research. A number of studies have shown that anonymisation is a legal fiction. Even for a very small number of data it is possible to uncover the identity of a person using open data sources. See for example http://ec.europa.eu/justice/data-protection/article-29/documentation/opinionrecommendation/files/2014/wp216_en.pdf
- Control in the hands of the patient: the processing of information is so complex (e.g. techniques for machine learning) that it is very difficult to understand how information can be used and what the possible results may be. In addition the value of Big Data lies precisely in the fact that information from many sources can be combined for a large number of objectives (which are not always determined in advance) As a consequence it is not always clear how this can be reconciled with the basic rights of consent and purpose.
- There is a continuum of solutions between collecting data in the cloud for analysis and the local storage of data with local analysis.



- In recent years serious progress has been made in the area of cryptographic techniques, as a result of which we can carry out joint calculations on data that remain stored locally and protected (*multiparty computation*). The techniques are still two to three orders of magnitude slower than a solution whereby all data is put in the cloud, and there is usually a large overhead in communication (gigabytes or even terabytes) associated with it.
- A second breakthrough is *Fully Homomorphic Encryption* (Gentry, 2009). This technique enables data to be stored in encrypted code in the cloud while still allowing calculations to be performed on it. In practice this only works for very simple calculations (simple statistical parameters), because the overhead and complexity of the calculations evolves very rapidly.
- There is also potential for intermediate solutions, whereby data are randomised to protect individual data (including techniques such as differential privacy).

To date these methods have only been used on small datasets as proof of concept. Much more research is needed to find out which solutions can be scaled up to realistic applications.

9. Viewpoint: are informed consent and privacy still realistic?

In closely controlled circumstances, such as clinical trials or social science research, researchers have time to explain – to the participant, the patient – what will happen to the data collected and how it will be processed. Individuals know what they are agreeing to and can therefore give their informed consent. But once collected, data can have a life of its own. How can people who have given their consent oppose the secondary, perhaps unforeseen use of their data?

Technological advances, such as Big Data, may give rise to applications that are as yet unknown.

The second point is the question about whether privacy and anonymity in the digital era are still realistic. There is no real way to make the future privacy-proof. An MIT study (de Montjoye et al., 2015) convincingly demonstrated how difficult it is to guarantee anonymity, even when personal data have been deleted: by identifying patterns in credit card statements the identity could be uncovered in 90% of cases.

According to Colin Mackay, former director of Communications at EFPIA and now the driving force behind a communication agency for healthcare in Brussels, legislators must take a different approach (Mackay, 2016). EU citizens must be asked whether they give consent to their data being collected (*opt in*), instead of not giving consent (*opt out*): “Metadata is here to stay, perhaps data privacy is not.”

If policymakers and businesses take full advantage of the potential of Big Data, especially to control the ever-increasing cost of healthcare, the privacy problem can be solved in previously unforeseen ways. Anonymised data must be made available for the purposes of research in such a way that the identity of the person behind the data cannot be uncovered. However, in the Big Data Era this is a difficult area: in five years’ time privacy might well be a completely empty concept for Big Data – with players such as Google+, Amazon, Facebook, etc., who link everything to everything else – and because of increasingly advanced techniques for machine learning. This is why <http://healthdata.be> is only available in aggregated form: there are X number of diabetes type 2 patients in Belgium who on average have this and that characteristic (age, gender...) and symptoms

10. Recommendations

Belgian law already stipulates a number of basic rules for access by care providers to medical records. As this report indicates, there are also other parties and new care objectives that justify the sharing of medical data. Moreover, there is new European legislation on data use and the role of the data controller.

We propose the following recommendations:

1. the distinction between a proliferation of data and regulated healthcare must be monitored, so that data remains accessible to care providers and so that their relationship with the patient is not harmed;

2. medical data can be shared, on condition that the privacy of the individual is guaranteed and that where necessary the personal data are anonymised or pseudonymised. Data controllers make sure that this is adhered to;
3. every care institution uses its own data format. This heterogeneity forms a barrier to combining data from data silos. As a consequence there is a need for cross-border coordination and technology integration in healthcare and a regulated use of open standards. This will enable interoperability and compatibility between the various components in the Big Data value chain;
4. *data provenance* provides insight into the origins of the data: not only how it was collected and under what circumstances, but also how it was processed and transformed. It is vital for the researcher to be able to accurately estimate the reliability of the data and the reproducibility of the analyses. *Data provenance* must therefore be a part of medical data gathering and the data controllers can monitor this;
5. policy must monitor the development in data standards, new Big Data technologies and new health care recommendations (as well as the results of data science) and be open to new developments;
6. the patient must give his/her express consent (*opt in*) for the purpose for which his/her data are to be used (and any possible risk for the patient). Each time there is a new purpose, he/she need not give his/her consent again (*one-time consent*). The patient can always revoke consent (time stamp acts as evidence), so that new requests can be rejected. The data controller monitors both aspects;
7. care institutions are not always used to innovating, whereas our information society demands it. As a result, a mechanism can be created to register the patient's demand or need for care innovation (e.g. after the use of apps and in relation to accessing medical records) at the care institution. The care provider is best placed to do this;
8. digital administrative simplification and the benchmarking of data are possible new roles for the insurance companies.

11. Conclusion

1. The Big Data ecosystem process consists of five components: (1) data creation, (2) data gathering and management, (3) analysis and information extraction, (4) hypothesis and experiment, and (5) decision-making and action. It is better to speak of data science in which the systematic use of data via applied analytical disciplines (statistical, contextual, quantitative, predictive and cognitive models) leads to data-based decisions. For every segment in this ecosystem there is a need for appropriate professional training and job classification: the data engineer, the data scientist and the data strategist. The role of data science in healthcare is threefold (triple AIM): an increase in

experience, quality and perception of the patient; better public health; and cost savings.

2. The future hospital is a network of care function components. It is based on integrated and structured information flows that strive for the best implementation of the demand for care: the best added value, the most feasible clinical result, acceptable costs.
3. The EU GDPR (*general data protection regulation*) combines two objectives: better protection of personal data for individuals and more opportunities for business in the digital single market by simplifying the legislation. The implementation of this regulation for the individual Belgian patient – with access to his/her healthcare data via a consolidated platform – should come into effect on 25 May 2018. For more information, see the recent report by the Privacy Commission (https://www.dropbox.com/s/5e64ylub6nuds75/17179%20Big_Data_Rapport_2017%20NL.pdf?dl=0)
4. Availability, accuracy, reliability and security are essential preconditions if data science in health care is to be of added value. Anonymised data must be made available for the purposes of research in such a way that the identity of the patient cannot be revealed. This latter aspect will come under increasing pressure as a result of technological developments.
5. The current position of the Belgian legislation is as follows:
 - patients are the owners of their data and give specific people access to (parts of) their medical records, limited in time (so more of an opt-in than opt-out approach: patients must actively give consent, it is not awarded automatically, except to general practitioners who manage the Global Medical Dossier [GMD]);
 - a patient must have the right to be 'forgotten' with regard to data;
 - a patient has the right at all times to withdraw access rights that he or she has given;
 - a patient always has the right to consult all medical data related to him that have been stored somewhere and to be informed of the existence of these data;
 - health data must never be sold to third parties.

There is currently a lack of legislation to make adequately protected patient data available to parties – other than traditional care providers – that could benefit from them (for research, product development ...), without patients having to give their consent each time for a similar purpose. Account must be taken here of the European legislation that assigns an important role to the data controller. The issue here is therefore adherence to given consent, the boundaries of 'profiling' of personal data and the management of possible transfers of data outside the EU.

References

Bergström, Richard (2015), *Data protection regulation set to benefit patient health*.

Breslow, N.E. and Day, N.E. (1987), *Statistical Methods in Cancer Research*, Volume II. oxford: oxford university Press.

de Montjoye, Yves-Alexandre, Radaelli, Laura, Singh Vivek Kumar, Pentland, Alex 'Sandy' (2015), *Unique in the shopping mall: On the reidentifiability of credit card metadata*. Science 30 Jan 2015: Vol. 347, Issue 6221, pp. 536-539.

European Data Protector Supervisor, opinion 7/2015: "Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability".

European Commission DG for health and blood safety, Directorate B – health systems, medical products and innovation: "Study on big data in public health, telemedicine and healthcare", December 2016.

Mackay, Colin (2016), *Big data could benefit healthcare but raises privacy concerns* <https://www.theparliamentmagazine.eu/articles/opinion/big-data-could-benefit-healthcare-raises-privacy-concerns>

May, M. (2014), *Life Science Technologies: Big biological impacts from big data*. Science, 344(6189).

Morley-Fletcher, Edwin (2013), *Big Data Healthcare. An overview of the challenges in data intensive healthcare*. <http://www.lynkeus.eu/big-data-healthcare/>

Morley-Fletcher, Edwin (2014), *Healthy data?* horizon 2020 Projects Magazine, Issue 4. <http://www.horizon2020projects.com> <http://www.lynkeus.eu>

Roney, K. (2012), *If Interoperability is the Future of Healthcare, What's the Delay?*. Becker's hospital Review.

Zhang, B., Tsiatis, A.A., Davidian, M, Zhang, M. and Laber, E.B. (2012), *Estimating optimal treatment regimes from a classification perspective*. Stat 1, 103-114.

Websites:

Particularities of data protection in the area of public health (2016): http://ec.europa.eu/health/data_collection/data_protection/in_eu/index_en.htm#fragment3

http://ec.europa.eu/justice/smedataprotect/index_en.htm/

http://ec.europa.eu/info/law/law-topic/data-protection_en

<https://www.ehealth.fgov.be/nl/egezondheid/patienten/ehealthconsent>

http://europatientrights.eu/general_overview_patient_rights_legislation/html?LAN=E

<http://www.nature.com/articles/sdata201618>

https://www.theparliamentmagazine.eu/articles/partner_article/efpia/data-protection-regulation-set-benefit-patient-health

Data Protection: Council agrees on a general approach (2015): <http://www.consilium.europa.eu/en/press/press-releases/2015/06/15-jha-data-protection/>

<http://ned.mycarenet.be>

<http://www.vitalink.be/>

<https://www.patientslikeme.com/>

<https://www.openhumans.org/>

<https://corepointhealth.com/wp-content/uploads/continuity-of-care-document-ccd.pdf>

<http://www.obofoundry.org/>

<http://lod-cloud.net/>

<http://rml.io>

<http://www.w3.org/TR/prov-overview/>

Appendix

International examples of Big Data projects

Comet K-Project DEXhELPP – AT

The Shared Care Platform – DK

E-Estonia – National Identity Scheme – EE

AEGLE (An analytics framework for integrated and personalized healthcare services in Europe) – UK, IT, GR, SE, BE, NL, PT, FR

The Business Intelligence database system – GR

PASSI (Progressi delle Aziende Sanitarie) – IT

Arno Observatory – IT

The Swedish Big Data Analytic network – SE

Clinical Practice Research Datalink (CPRD) – UK

Sentinel Stroke national Audit Programme (SSnAP) – UK

Hospital Episode Statistics (hES) – UK (England)

The YoDA Project (Yale University open data access) – US

FDA Adverse Event Network Analyser – US

CEPhoS-LInK – FI, AT, RO, NO, SI, IT

Twitter (Adverse drug reactions and public health) – International

Flatiron – US

UK Biobank – UK

Semantic Data Platform for healthcare (SEM CARE) – DE, NL, AT, UK, ES

Integrated BioBank of Luxembourg (IBBL) – LU

Spanish Rare Diseases Registries Research network (SpainRDR) – ES

Members of the Working Group

Pascal Verdonck (KTW)

Marc Van Hulle (KAGB)

Bart De Moor (KTW)

Erik Mannens (UGent)

Rudy Mattheus (VOKA Health)

Geert Molenberghs (UHasselt)

Femke Ongenaë (UGent)

Marc Peters

Bart Preneel (KU Leuven)

Frank Robben (e-Health Platform)

KTW = Class of Technical Sciences

KAGB = Belgian Royal Academy of Medicine

RECENT POSITION PAPERS (from 2014)

25. Pascal Verdonck e.a. – *Medische Technologie, als motor voor innovatieve gezondheidszorg*, KVAB/Klasse Technische wetenschappen, 2014.
26. Charles Hirsch, Erik Tambuyzer e.a. – *Innovatief ondernemerschap via spin-offs van kenniscentra*, KVAB/Klassen Natuurwetenschappen en Technische wetenschappen, 2014.
27. Giovanni Samaey, Jacques Van Remortel e.a. – *Informaticawetenschappen in het leerplichtonderwijs*, KVAB/Klasse Technische wetenschappen en Jonge Academie, 2014.
28. Paul Van Rompuy – *Leidt fiscale autonomie van deelgebieden in een federale staat tot budgettaire discipline?* KVAB/Klasse Menswetenschappen, 2014.
29. Luc Bonte, Paul Verstraeten e.a. – *Maatschappelijk verantwoord ondernemen. Meedoen omdat het moet, of echt engagement?* KVAB/Klasse Technische wetenschappen, 2014.
30. Piet Van Avermaet, Stef Slembrouck, Anne-Marie Simon-Vandenbergen – *Talige diversiteit in het Vlaams onderwijs: problematiek en oplossingen*, KVAB/Klasse Menswetenschappen, 2015.
31. Jo Tollebeek – *Metamorfoses van het Europese historisch besef, 1800-2000*, KVAB/Klasse Menswetenschappen, 2015.
32. Charles Hirsch, Erik Tambuyzer e.a. – *Innovative Entrepreneurship via Spin-offs of Knowledge Centers*, KVAB/Klassen Natuurwetenschappen en Technische wetenschappen, 2015.
33. Georges Van der Perre en Jan Van Campenhout (eds.) – *Higher education in the digital era. A thinking exercise in Flanders*, Denkersprogramma KVAB/Klasse Technische wetenschappen, 2015.
34. Georges Van der Perre, Jan Van Campenhout e.a. – *Hoger onderwijs voor de digitale eeuw*, KVAB/Klasse Technische wetenschappen, 2015.
35. Hugo Hens e.a. – *Energiezuinig (ver)bouwen: geen rechttoe rechtaan verhaal*, KVAB/Klasse Technische wetenschappen, 2015.
36. Marnix Van Damme – *Financiële vorming*, KVAB/Klasse Menswetenschappen, 2015.
37. Els Witte – *Het debat rond de federale culturele en wetenschappelijke instellingen (2010-2015)*, KVAB/Klasse Menswetenschappen, 2015.
38. Irina Veretennicoff, Joos Vandewalle e.a. – *De STEM-leerkracht*, KVAB/Klasse Natuurwetenschappen en Klasse Technische wetenschappen, 2015.
39. Johan Martens e.a. – *De chemische weg naar een CO₂-neutrale wereld*, KVAB/Klasse Natuurwetenschappen, 2015.
40. Herman De Dijn, Irina Veretennicoff, Dominique Willems e.a. – *Het professoraat anno 2016*, KVAB/Klasse Natuurwetenschappen, Klasse Menswetenschappen, Klasse Kunsten en Klasse Technische wetenschappen, 2016.
41. Anne-Mie Van Kerckhoven, Francis Strauven – *Een bloementapijt voor Antwerpen*, KVAB/Klasse Kunsten, 2016.
42. Erik Mathijs, Willy Verstraete (e.a.), *Vlaanderen wijs met water: waterbeleid in transitie*, KVAB/Klasse Technische wetenschappen, 2016.
43. Erik Schokkaert – *De gezondheidszorg in evolutie: uitdagingen en keuzes*, KVAB/Klasse Menswetenschappen, 2016.
44. Ronnie Belmans, Pieter Vingerhoets, Ivo Van Vaerenbergh e.a. – *De eindgebruiker centraal in de energietransitie*, KVAB/Klasse Technische Wetenschappen, 2016.
45. Willem Elias, Tom De Mette – *Doctoraat in de kunsten*, KVAB/Klasse Kunsten, 2016.
46. Hendrik Van Brussel, Joris De Schutter e.a., *Naar een inclusieve robotsamenleving*, KVAB/Klasse Technische Wetenschappen, 2016.
47. Bart Verschaffel, Marc Ruyters e.a., *Elementen van een duurzaam kunstenbeleid*, KVAB/Klasse Kunsten, 2016.

The complete list of position papers and all PDFs can be viewed at
www.kvab.be/en/position-papers

DATA SCIENCE AND HEALTHCARE

The application of data science in healthcare offers possibilities in various different areas: an increase in the effectiveness and quality of patient care, better public health in general and a reduction in costs.

Big Data allows for unprecedented amounts of information to be processed in order to discover new patterns. Better use of health-related data can help us to better understand diseases, make better diagnoses, evaluate treatments and develop new ones ... For this to happen, medical data must be available in such a way that the identity of the patient cannot be revealed.

It is crucial to find a balance between protection of privacy on the one hand and the creation of opportunities for research and product development on the other. The current EU guideline (General Data Protection Regulation) attempts to reconcile these two aims. However, there is still no regulatory framework for making available medical data in a reliable and transparent way with respect for patient privacy. This Position paper puts forward recommendations for future policy concerning Big Data in healthcare.

The Academy's Standpunten series (Position papers) contributes to the scientific debate on current social and artistic topics. The Academy's authors, members and working groups write in their own names, independently and in full intellectual freedom. The quality of the published studies is guaranteed by the approval of one or several of the Academy's classes.