



# BIG DATA Workshop

16<sup>th</sup> European Health Forum Gastein

REPORT - OCTOBER 2013



*The Big Data Workshop was organised by the European Alliance for Personalised Medicine and supported by the Lithuanian Health Forum, EFPIA, IBM and Pfizer in Gastein, Austria from 2 – 4 October 2013.*

**ORGANISED BY**



European Alliance for  
Personalised Medicine

**SUPPORTED BY**



**COLOPHON**

**Production** Vital Transformation BVBA  
[www.vitaltransformation.com](http://www.vitaltransformation.com)

**Words** Nuala Moran

**Design** Peter Koekoek

**Photos** European Health Forum Gastein  
BigStock Photo

# Harnessing big data to improve R&D and the translation of new therapies

With health care under pressure from many angles, the key issue is to ensure sustainability of health systems in the face of these challenges. Rational application of big data has an important role to play.

It can be difficult to talk about the promise of investing in big data at times of falling resources for health care, said **Leonas Kaletinas, Member of the Board of Directors of the Lithuanian Health Forum**. However, “There clearly is opportunity for countries like ours, and for patients, professionals and payers to use the basics of big data to become cost-effective, to avoid waste, to increase value and increase efficiency,” Kaletinas said opening the debate in the **Big Data Workshop**, held as part of the **16<sup>th</sup> European Health Forum in Gastein**, Austria from 2 – 4 October.

“For our system big data could be a revolution, because we could see huge changes in health care management,” Kaletinas said, in the first session, **‘Big Data and Best Practice for Public Health’**.

An important aspect for Lithuania in its current EU Presidency role is to stress the potential benefits of big data in helping to strengthen health systems. **‘Sustainable Health Systems for Inclusive Growth in Europe’** is the topic for the Lithuanian Presidency Conference in Vilnius on 19 – 20 November. Kaletinas said she hopes the conference in Vilnius will progress things so that next year, “best practice and not barriers” are the focus of attempt to move the deployment of big data forward.



Leonas Kaletinas, Member of the Board of Directors of the Lithuanian Health Forum

Two particular areas where the European Commission could help open doors to big data, are in new data protection legislation and through funding programmes to promote adoption, Kaletinas believes. For countries with limited funding and a shortage of the appropriate skills being part of a European network would be very helpful.



# Quality not quantity

It is necessary to take a measured approach to big data – bearing in mind that quality counts more than quantity, **Terje Peetso, Unit H1 – Health and Well Being, DG Connect at the European Commission** suggested. There may now be differences in scale, but medical science has always been about collecting data said Peetso. “It’s not about taking blood it’s about analysing it,” she said.

Similarly, big data should not be seen as some sort of end in itself, but should be applied in ways that contribute to public health and patient empowerment. A good place to start would be existing data repositories. “It’s like refining crude oil to add value; we should look into data we already have and get value from that,” Peetso said.

This would provide the foundation for looking into “new” data that is

accumulating. In health, it is important not to focus solely on data per se, but also on questions about ownership, trust and access, and on interoperability between different computer systems to ensure all relevant data can be assembled and analysed.

Indeed, interoperability is seen as a particular issue in applying big data to public health, with many disparate systems, both within different tiers of national health services and between member states.

The question of interoperability of medical data systems keeps cropping up, and there have been many attempts to deal with this. While it is “a monumental task” there are some compelling examples of the value of persisting, said **John Crawford, Healthcare Industry Leader Europe, IBM**. The European Union’s epSOS (European Patients Smart

Open Services) system, which not only transmits medical data from one country to country, but also translates from one language to another, “is a great example,” Crawford said.

Another is Scotland’s Emergency Care Summary database containing a summary of demographic, allergy and medication information for 5.5 million people. It enables health care professionals to access to important patient information in emergency and unscheduled care situations.



Terje Peetso, Unit H1 – Health and Well Being,  
DG Connect at the European Commission



# The Internet is a hammer

**A**melia Andersdotter, MEP (Greens/EFA, Sweden) has a deep personal understanding of the issues of ownership, trust and security in the age of big data because some of her own childhood medical information has been published in Sweden without her consent. The information is available via the Internet, which Andersdotter pointed out, “is not a place that is secure.”

For Andersdotter, there are two particular issues here: that information technologies are beyond the competences of most medical practitioners and that the Internet is outside the law. As Andersdotter pointed out, “The [US] National Security Agency can’t invade my house but they can invade my Internet.”

Having no control over the published medical information is making Andersdotter “uncomfortable” and leading her to avoid contact with health institutions, a situation which is exactly the opposite of how anyone would like to see the application of big data playing out.

As Andersdotter noted, in all other circumstances there is high confidence that medical records are secure, with no access for the courts, the police or the press. Even if an individual gives informed consent for their data to be used, the problem of security of data remains. “The Internet is for distributing information, so if there’s information you don’t want distributing, it’s not the right tool - like a hammer, the Internet solves some problems but not others,” Andersdotter said.

A further question revolves around the nature of consent. When an individual agrees their health information can be made available in some way, they do not know what they are consenting to. “The doctor inputs the data, so the patient doesn’t know if there’s a risk,” said Andersdotter, adding, “there are some serious challenges for politicians and industry to preserve the confidence of citizens.”

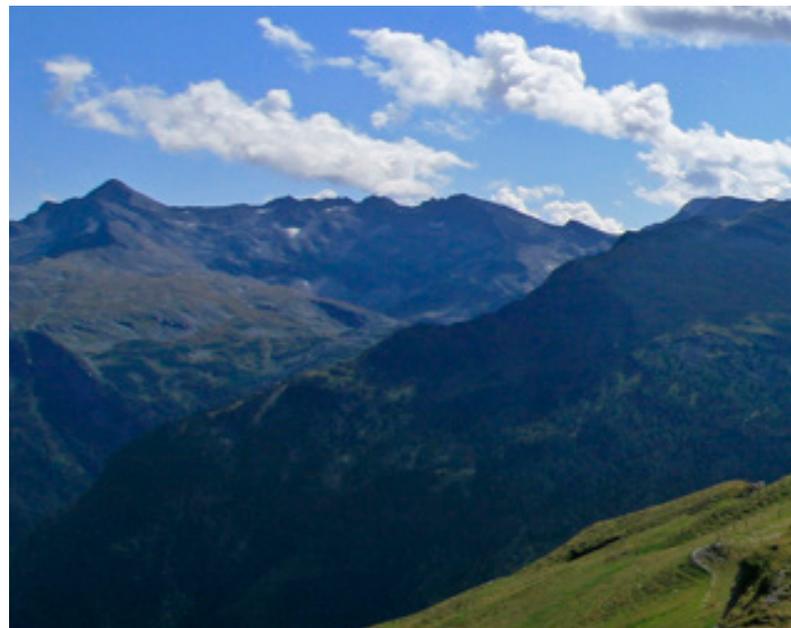
The European Commission needs to address points of failure in the system before big data moves ahead, Andersdotter said, suggesting a more granular approach to use of data may be appropriate and that the immediate focus should be on opening up data that is not personal but is nevertheless relevant to public health, such as clinical trials data and information in subscription medical journals.



Amelia Andersdotter, MEP (Greens/EFA, Sweden)

John Crawford agreed the design principles of the Internet, which are based on sharing information, make it inherently unsuitable for sensitive private data. But he noted, private data stores in the Cloud, with appropriate encryption and other security measures, could provide a solution.

The appropriate security measures depend on how the information is used, Andersdotter believes. Pseudonymisation may work with machine reading of data, but would not necessarily preserve confidentiality if the information is being read by a human. “You need to think about issues like this in implementing any technical systems,” Andersdotter said.



# The issues are on the table

**T**he challenge of big data and health informatics lies not only in capturing and storing information securely, but also in devising the tools and analyse and manage it, said **Barbara Kerstiens, Head of Sector, Public Health, DG Research at the European Commission**, opening the second session of the workshop on **'Data Sharing for Improved Research and Translation'**.

The Commission is funding research on both these aspects. A number of issues need to be addressed in harnessing the power of big data to improve research and speed up translation of research outputs to improve health both at a public health level and in the development of personalised medicines.

These include: standardisation, integration – especially to achieve economies of scale, for example in research into rare diseases; the challenge of open access, and not just making data available but ensuring it is readable and useable; the need for new statistical methods and tools; and providing the means to track clinical outcomes.

Dealing with this 'to do' list calls for international collaboration. "No



Bonnie Wolff-Boenisch, Head of Research Affairs at Science Europe and John Crawford, Healthcare Industry Leader Europe, IBM

individual country can deal with these challenges or get the benefits [alone], Kerstiens said. The European Commission is well placed to support the necessary research, building on previous investments such as the European Bioinformatics Institute and a significant number of international collaborations it has funded in this field.

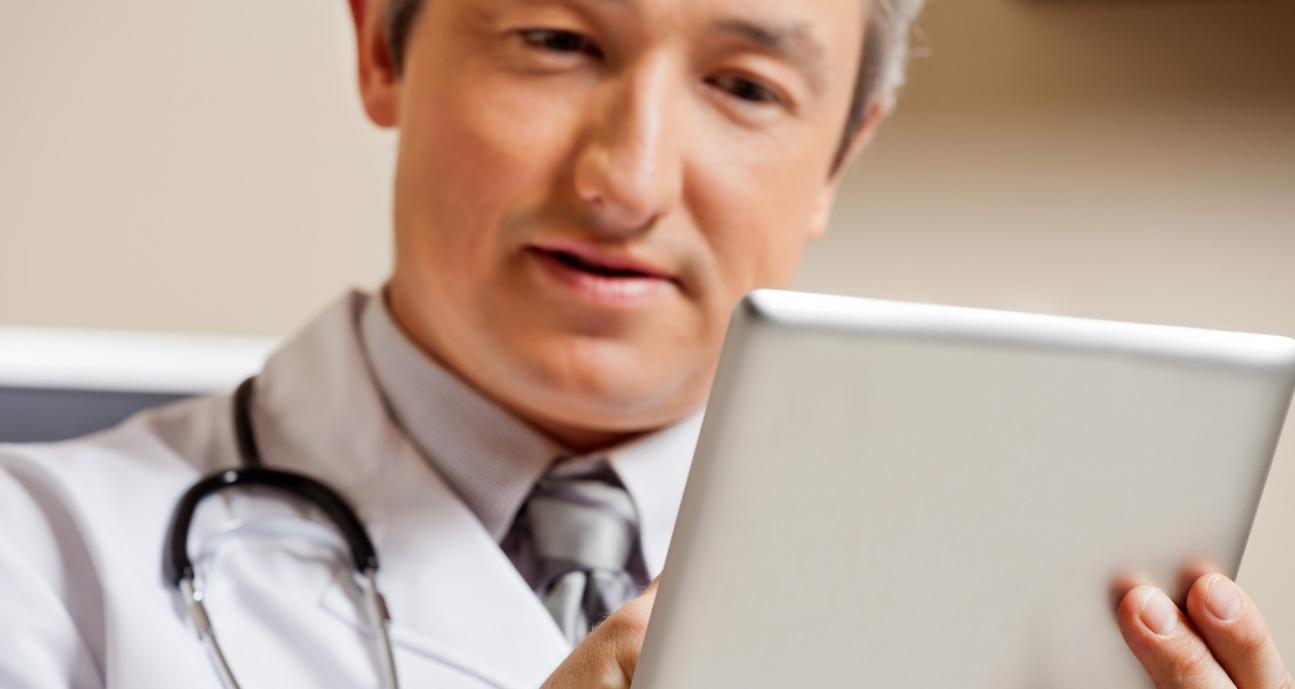
There is also a need for an EU-level public private partnership involving all stakeholders to consider all aspects of data sharing and access, to ensure there is a participant-centred approach.

From a research perspective, it is critical to avoid perpetuating data silos that are disconnected from one another, since this will limit the potential for big data analyses. "It's a work in progress

and continuing talks are needed," Kerstiens said. "Providers of data need to understand the challenges."

The Innovative Medicines Initiative is an example of an EU-funded programme that aims to improve drug development and regulation through the use of pooled data. Meanwhile, new EU Clinical Trials and Data Protection rules that are being formulated currently mean the issues relating to big data, health research and privacy, are on the table. "It is a conversation that has started and is to be continued," said Kerstiens.





# Reading the book of big data

**B**onnie Wolff-Boenisch, Head of Research Affairs at Science Europe, the body established in 2012 to represent the views of the leading research funding organisations around Europe, told the workshop that all research universities are trying to strike a balance that ensures the potential of big data is realised, but that privacy is not compromised.

Opening up data is important: “If you have bright minds accessing it, you don’t know what will come out of it,” Wolff-Boenisch said. Science Europe considers some rules are too strict and could hamper certain types of research. For example, a requirement to get informed consent for each individual piece of research would make biobanks inefficient.

Similarly, it is important not to be too prescriptive, since the key to extracting value from data is to be able to apply new tools, to “play around” and come up with new methods and approaches for converting data to useful information. “We need to be able to read the book of big data,” said Wolff-Boenisch.

The power inherent in big data is that it can provide “individualised evidence” leading to the development of truly personalised medicine, said **Angela Brand, Professor of Health, Medicines and Life Sciences at Maastricht University and Co-chair of the workshop on behalf of the European Alliance for Personalised Medicine.**

Big data will provide the means for decision support across all aspects of health care – ranging from assessing safety

and efficacy of drugs, to carrying out health technology assessments, and prevention, diagnosis and treatment - to be refocused from the population level, one size fits all paradigm, to the individual. “We need to get individual evidence,” Brand said, “and this information must be available on a just-in-basis.”

Achieving this ambition raises presents challenges around the governance and the quality of implementation of big data in health, and calls for standards for consolidating, characterising, validating and processing data. However, Brand said, its inherent diversity and complexity means health information “will always be messy”, raising the question of how to set the bar in assessing quality of implementation.

While data users should be accountable for the custodianship of personal medical information, it is impossible to guarantee complete data security, and it would be dishonest to do so. Given this, Brand suggested a more appropriate approach – to replace the requirement for individual informed consent every time someone’s data is used – would be that individual data sets are aggregated into big data algorithms.

In the discussion, delegates raised a number of other issues, including how to guarantee the quality of data going into shared databases, agreeing technical approaches to which all stakeholders can sign up, and developing sustainable business models for the deployment of big data in health.



# We are all health billionaires

It's clear that establishing trust is essential if the techniques and tools of big data are to be successfully applied to health. The key to this is transparency so that, "people who give data know what happens to it and follow it," said **Ernst Hafen of the Institute of Molecular Systems Biology, ETH Zurich**, opening the third session on **'Big Data and improved evaluation models for efficacy and efficiency'**.

Hafen suggested 'The People's Health Databank' as a model for engendering trust and transparency. This would be a safe and secure place to store data, which people trust in the same way as they trust a bank to store their money, and to transmit it to third parties on the instructions of the account holder.

Such health and genomic databanks could be run as cooperatives, with requests for data access for research purposes handled centrally, and individuals having the right to withhold data from particular pieces of research. For companies requesting information for a drug development project, there would be a charge, with the money invested back into the running of the databank.

The huge potential that big data holds to drive the development of personalised medicine makes it appear counter to the principle of solidarity that underpins Europe's health care systems. Organising the People's Health Databank as a cooperative would enshrine solidarity in the new age of personalised medicine. Individuals would share their data to get cures for themselves and for everyone else.

Hafen proposed there would be a cooperative health databank in every country, each using the same data standards so information could be shared between them.

The idea of collectively creating consent in the People's Health Databank is very compelling, believes **Adam Heathfield, Director of Science Policy Europe at Pfizer**. The cooperative model is a good one if people buy in.

As a company, Pfizer is making a concerted effort to make better use of real life data and has put a new team in place to look at this. In R&D the first step will be to build on epidemiological information to become smarter in target selection, and then overlay genomics to link genotypes and phenotypes. For existing products, big data will be used to answer questions about how well medicines actually perform in the market and to provide inputs for health technology assessments.

Post-marketing studies and health technology assessments are becoming a much bigger burden, requiring information that cannot be generated in clinical development, and big data promises to provide some relief. Issues remaining to be resolved include guaranteeing data quality and developing robust methods for framing and answering questions. "We are a long way from having the data analysis tools we need," Heathfield said.

While, as suggested in Hafen's People's Health Databank model, Pfizer is prepared to pay for access to anonymised data sets, Heathfield noted

that pharmaceutical companies cannot pay people to take part in clinical trials (though they can pay expenses). "There would be a problem of a cooperative genuinely getting consent and being paid for data, without skewing that issue," said Heathfield.

Hafen suggested this could be finessed by a gatekeeping function. For example, in a database with 10 million records, there might be 30,000 women with a BRCA gene mutation who have agreed to share information on their status. If Pfizer paid for access, the cooperative would filter the database and then approach the women and ask if they wanted to participate in a clinical trial.

The technology is at hand to apply big data to health, but there must be a public debate about the risks associated with data sharing, said John Crawford. Furthermore, there is no point in accumulating data unless it is then analysed and the results translated into action.

Health certainly fits the big data paradigm in terms of the volume of data it generates. However, it remains the



Angela Brand, Professor of Health, Medicines and Life Sciences at Maastricht University



case that the majority of this is held in text files. “You have to get it into a state where you can do something useful with it,” Crawford said.

Information held in health records may not have the other essential big data property of velocity, but other forms of health data do. In one famous example, Google claimed to have tracked the outbreak of seasonal flu before the US Centers for Disease Control and Prevention because people started using the company’s search engine to look up symptoms.

Similarly, analysis of data generated by monitoring devices in intensive care units can pick up signs of nosocomial (hospital-acquired) infections before

there are observable symptoms. Health data also fits the big data mould in terms of variability, with information often being inconsistent, incomplete and contradictory, Crawford noted. IBM’s Watson computer, with its ability to read and understand natural language and weigh evidence, is moving decision support to a new level, allowing doctors to access and interpret all the latest evidence and make better decisions as a result. This also highlights the way in which big data can shift analytics from retrospective to real time.

Summing up, **Ralf Sudbrak, Scientific Coordinator, Max Planck Institute for Molecular Genetics** noted that concerns about data protection vary from too much to too little, depending on the

individual’s perspective. “We need to make sure we are not too protective; it is necessary to find the balance between the need for data protection and the use of big data. There is a huge opportunity for benefits to patients and society.”

A pre-requisite to realising these benefits is to create the right framework for data sharing and data access to enable research. Data collection and data access can be for a number of different purposes. Given this, there is a need for harmonisation of data and harmonisation of patient records. “Data needs to be in the right format,” Ralf concluded.

