DATA PROTECTION REGULATION

PSEUDONYMISATION
TRUSTED THIRD PARTY
K-ANONYMITY

DATA DONATION
DATA INHERITANCE
ANONYMISATION

DOES NOT FIT ALL

DATA ACCESS & PROTECTION

PERSONALISED
MEDICINE

BIO-MEDICAL
LONG TERM

IN SILICO CLINICAL TRIALS

TARGETED
TREATMENTS &

DISEASE SIGNATURES

"LONG TAIL"
RESEARCH

MD-PAEDIGREE NEWSLETTER

Issue 1 - November 2013

Model-Driven Paediatric European Digital Repository

AN OWNERSHIP ISSUE

EUROPEAN RESEARCH AREA

POLICY-MAKING

BIG DATA HEALTHCARE

PRIVATE
PUBLIC

FRAMEWORK INVESTMENT
UPDATE FRAMEWORK
UPDATE ETHICAL FRAMEWORK
UPDATE LEGAL FRAMEWORK

THE 5 V'S

VALUE
VERACITY
VARIETY
VELOCITY
VOLUME

DEALING WITH THE
DATA GLUT

BIG DATA LITERACY

MORE THAN BLACK BOXES
SCIENTISTS
INTEGRATION OF DATA
DATA GLUT
EXTENT
USING CLINICAL WORKFLOWS
ADDING DATA ANALYSIS TO

TOP-DOWN
BOTTOM-UP

THE "DIMENSIONALITY CURSE"
BIO-MEDICAL REPOSITORIES

MD-Paedigree wins
Best Exhibit Award
at ICT 2013

MD-PAEDIGREE

SEVENTH FRAMEWORK PROGRAMME

European Commission

Editorial

Bruno Dallapiccola
MD-Paedigree Coordinator

Contents

Welcome to the first biannual newsletter of the MD-Paedigree (Model-Driven European Paediatric Digital Repository) project. It provides both an overview of the project and describes MD-Paedigree's role in the VPH (Virtual Physical Human) community.

Since the kick-off meeting in March 2013, MD-Paedigree has already come a long way, following on, as it does, the two highly acclaimed 6th and 7th FP projects: Health-e-Child and Sim-e-Child. It is thanks to the breakthrough outcomes of these two projects that it was possible to lay the foundations for this new, exciting, and successful cooperation within the MD-Paedigree Consortium.

One of the aims is to establish and consolidate synergies between all partners, and the cooperation with closely related projects such as the EU co-funded Personalised Medicine (P-Medicine) and Virtual Physiological Human Share (VPH-Share). Close cooperation with these two projects will ensure that MD-Paedigree developments will be highly interoperable with the VPH Infostructure on the semantic layer for effective indexing and searching of data collections.

In its first six months, the MD-Paedigree Management and Technical Coordination Board has been working hard to develop the semantic data structure through face to face and online collaboration as well as at the MD-Paedigree 1st biannual meeting in Genoa (Italy).

In addition, special efforts have been made to present MD-Paedigree at several major events: European

Prof. Bruno Dallapiccola, Project Coordina-tor and Scientific Director of Ospedale Pedi-atrico Bambino Gesù (OPBG)

Society for Movement Analysis in Adults and Children (ESMAC), held in Glasgow (Scotland) last September, this October at Paediatrics 2040 in Anaheim (USA) and at the EC's ICT 2013 Conference, which took place on 6th – 8th November in Vilnius (Lithuania). At ICT 2013, MD-Paedigree hosted a booth which won the "Best Exhibit Award", ran a packed networking session on Big Data and Data Analytics Impact in Healthcare and distributed a discussion paper on Big Data Healthcare.

Our ultimate goal is to enable a more predictive, individualised and safer healthcare in several disease areas, by establishing a worldwide advanced paediatric digital repository providing decision support to medical professionals. While we have much challenging, demanding work lying ahead, I am confident that through this groundbreaking research MD-Paedigree will open up a new landscape for paediatric research and lead to a future where children's healthcare is more effective, more personalised, and more affordable.

Project Overview

Big Data in Healthcare is not only a major challenge for IT and clinicians, but also a great opportunity. The use of massively available biomedical data will allow clinicians to simulate potential outcomes and so prevent patients from undergoing ineffective treatments. In other words, using data to develop a greater understanding of pathophysiological processes will result in significant healthcare improvements.

MD-Paedigree aims to achieve effective knowledge exploitation from knowledge discovery from clinical data in healthcare, its major challenges are: content-based image retrieval, similarity search on annotated data, correlation-based vs. hypotheses-driven knowledge discovery, model-driven analysis, the balance between data quantity and data quality and, last but not least, new innovative personalised care application scenarios.

To reach its goals, MD-Paedigree will implement innovative tools for storage, sharing, collaboration, multimodal similarity search, outcome analysis, risk stratification, and patient-specific clinical decision support. Furthermore, it will provide innovative strategies in manipulating, curating and exploiting massive amounts of sensitive medical data. Through disease simulations, requiring high performance and supercomputer resource availability, it will improve the paediatricians' diagnostic precision and offer child-specific treatment choices in Europe. In this way, physicians will be able to select highly individualised treatment options and receive on-the-spot support based on their patients' medical data. Since the models' acceptance and use is a key factor, MD-Paedigree has several healthcare centres of excellence as key partners. The whole MD-Paedigree Consortium is committed to attaining a goal-driven, federated infostructure that will enhance existing disease models that stem from longstanding academic and industrial research. The infostructure will interoperate with the VPH-Share infostructure and will be used to develop robust and reusable multi-scale models for more predictive, individualised, effective and safer healthcare in four disease areas. MD-Paedigree will integrate at the point of care through state-of-the-art user interfaces and establish a worldwide advanced paediatric digital repository.
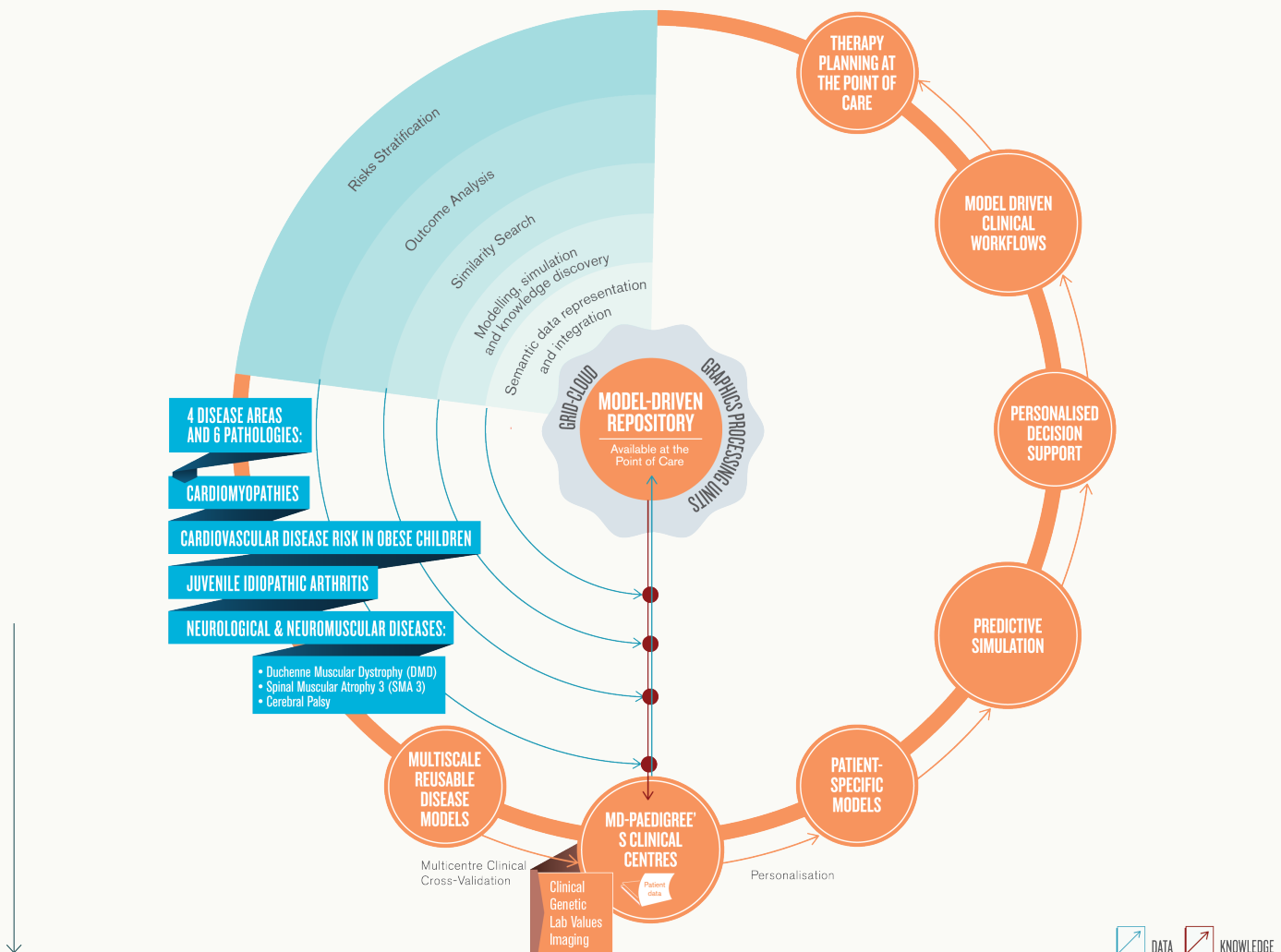
Highlights & Objectives

Facts

Giacomo Pongiglione
Head of Action 1: Clinical background activities, user requirement, clinical workflow and validation.

The aim of MD-Paedigree is to develop a model-driven data workflow digital repository and leverage information processing and a knowledge discovery evolutionary framework for personalised and predictive treatment in paediatrics. As such, it is a timely initiative meeting the big data market challenge in healthcare.

The project will build the models starting from routine clinical data in the various clinical centres. We are focussing on four paediatric disease areas:

+ **Cardiomyopathies**

+ **Obesity-related Cardiovascular Disease**

+ **Juvenile Idiopathic Arthritis**

+ **Neurological and Neuromuscular Diseases**

MD-Paedigree will integrate methodological approaches from the targeted specialties and consequently analyse biomedical data from a multiplicity of heterogeneous sources (clinical, genetic and metagenomic analysis to MRI and US image analytics, haemodynamics, real-time processing of musculoskeletal parameters and fibres biomechanical data, etc.) and specialised biomechanical and imaging VPH simulation models. It will make it possible to apply big data analytics to routine clinical data on children providing three fundamental functionalities on top of those from an advanced Electronic Health Records registry:

+ Search for clinicians to access patients like mine and for patients to find with patients exactly like me

+ Model-based patient-specific simulation and prediction

+ Patient-specific clinical work flows

The MD-Paedigree Infostructure integrates the models with knowledge discovery tools, semantic data representation and integration, outcome analysis, risk stratification, and similarity search tools.

Thanks to the development of compatibility for GPU processing and to the model-driven prediction locally executed onto live patient data, MD-Paedigree provides time critical support to physicians at the point of care. This will not only facilitate the construction and operation of new decision support tools for prediction and simulation in paediatrics and for the relevant VPH workflow models opening further collaboration with a growing number of paediatric clinical centres, but also pave the way for having functional databases assisting pharmaceutical researchers in realising model based-drug development.

Clinical application of MD-Paedigree

Cardiomyopathies

Gabriele Rinelli

Jonathan is a 12 years old boy with a Cardiomyopathy. At clinical evaluation he reported no dyspnea at rest with some fatigue at mild exercise. He was free of cardiovascular treatment, heart rate was mildly increased and blood pressure was low-normal. A dilated left ventricle with left ventricular hypertrophy was seen by the echocardiography. Systolic function was low-normal and diastolic function analysis demonstrated increased left ventricular filling pressure. The CMR demonstrated a frankly dilated ventricle with mildly reduced ejection fraction and mild diffuse fibrosis of the cardiac muscle. The mitral annular plane was dilated and mitral insufficiency was caused by leaflet tethering. He was treated according to clinical guidelines. A follow-up clinical evaluation and echocardiogram were programmed after three months to evaluate the treatment effect.

MD-Paedigree allows Jonathan's physicians to exploit the information derived from different diagnostic techniques. MD-Paedigree provides the robust, multi-scale 4D anatomical, hemodynamic and electromechanical model to integrate all available clinical and diagnostic data. Integrated information on CMR cardiac geometry and volumes is merged to the echocardiography functional one and clinical examination hemodynamic data. Beyond that, electromechanical and haemodynamic models of the heart provided by MD-Paedigree help doctors to understand Jonathan's specific mechanism of muscle dysfunction by integrating information on muscle fibrosis and systolic mechanics, and predict the impact of therapy in reducing mitral regurgitation, filling pressure and thus relieve symptoms. Jonathan's treatment is personalised and tailored using his modelled cardiac morphology and function, integrating information on heart geometry, ejection function, relaxation, ventricular inter-dependence, valve function and cardiac workload. His response to drugs is predicted by MD-Paedigree which helps physicians in prescribing the most effective treatment at the first evidence of cardiac disease, reducing the timeframe from disease evidence to medical treatment.

Obesity-related Cardiovascular Disease

Andrew Taylor

Rose is a 17-years-old adolescent suffering from an Obesity-related Cardiovascular Disease. Her waist girth is 102 cm. She underwent the fasting measurement of systemic inflammation markers, the MRI evaluation of adiposity, the evaluation of cardiac morphology and haemodynamics by echocardiography, and a cardiopulmonary exercise test. Her waist girth has significantly increased over the last 18 months demonstrating central fat distribution. Her fasting glucose is elevated and her oral glucose tolerance is impaired. Rose is dylipidemic, with increased levels of inflammation markers. Echocardiography demonstrates left ventricular hypertrophy with normal systolic function and impaired cardiac relaxation. The CMR evaluation and the applanation tonometry show a significant amount of pericardial fat paired with the mild diffuse fibrosis of the cardiac muscle and a reduced arterial compliance with increased wall stress and impaired endothelial function. The cardiopulmonary test reduced tolerance to physical activity, with increased oxygen consumption and an evident pathological blood pressure profile.

MD-Paedigree provides integrated information on cardiovascular structure and function, the quantitative assessment of fat distribution in the body, and metabolic and genetic data from laboratory tests. The electromechanical heart model allows understanding the mechanism of cardiac muscle and vascular dysfunction by integrating related information on systemic fibrosis, inflammation and cardiovascular mechanics. It also allows predicting the impact of selected therapies and weight loss for the specific cardiovascular function, fat distribution and exercise tolerance. Rose's treatment is personalised and tailored using integrated information on anthropometrics, demographic data, cardiac geometry and function, vascular compliance, genetic and metabolic profiles. MD-Paedigree's accurate estimation of cardiovascular risk, prognosis of disease development and a successful therapy prediction (on previously observed cases clinical history in the digital repository) helps selecting the most effective treatment at the first disease evidence.

Clinical application of MD-Paedigree

Juvenile Idiopathic Arthritis

Alberto Martini

Chiara and Simona are affected by Juvenile Idiopathic Arthritis (JIA). Their age at disease onset was 4 years old and they both tested positive for antinuclear antibodies. They both had an asymmetric involvement of knee and ankle at disease onset. Within 1 year both Chiara and Simona showed wrist involvement and started a treatment with second line agent. Chiara experienced a severe and irreversible structural damage progression, as revealed by a 4-year follow-up radiography. Simona developed a milder course of the disease and her 4-year follow-up radiography showed no significant signs of structural damage. Notably, at disease onset demographic, routine clinical and laboratory data did not significantly differ between these two girls. It was not possible, therefore, using traditional methods to distinguish which of them would develop a more aggressive disease.

Using MD-Paedigree to set their data in a multi-scale integrated model clinical, immunological and meta-genetic data, as well as biomechanical analysis results, outcome predictors are identified. This means that Chiara's more severe course of the disease is determined and her need for earlier treatment is established.

A comprehensive model of JIA-related changes in two joints (wrist and ankle) is available. Different imaging modalities (i.e., MRI) provide information to classify the degree of bone erosion and synovitis in both regions. For this, automated image analysis tools are developed to reduce the time necessary for performing that task and becoming independent from the individual observer. Enhanced biomechanical models are generated by adapting highly sophisticated standard models to the individual case and thus predicting locomotive changes caused by JIA. These are tested against results of a personalised gait analysis that enriches this model by providing more details about the locomotive constraints for the considered joints.

The MD-Paedigree database contains morphological changes visible in the image data and clinical, immunological, genetic and metagenetic data. Having access to such a large classified individual JIA cases repository helps to better predict the girls' disease progression and provide them with medication that is specifically adapted to them.
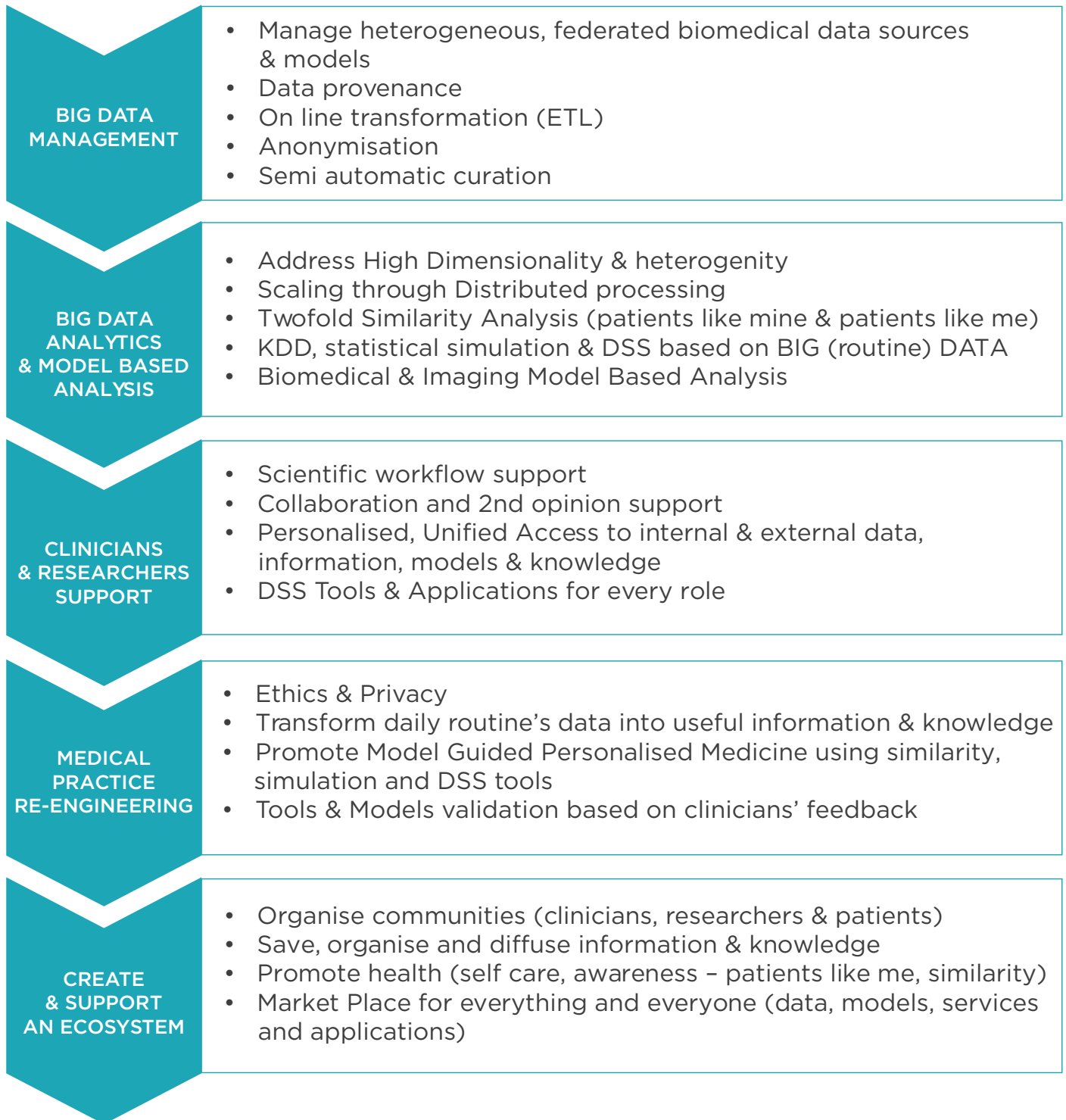
Neurological and Neuromuscular Diseases

Jaap Harlaar

François is a 10 year old boy with a Neurological and Neuromuscular Disease. His medical records show a consistently significant delay in his motor development. His current level of the Gross Motor Function Classification reveals that he can walk unsupported with difficulties while walking outside. He falls over in the playground frequently, and has a very limited walking distance, due to early fatigue. Moreover, his physiotherapist is concerned that his walking pattern may deteriorate further, eventually resulting in wheelchair dependency. François is referred to a specialist paediatric centre of rehabilitation medicine, where his walking pattern is analysed in the gait laboratory. A complete recording of François' gait pattern, using 3D kinematics, joint kinetics and muscle activation patterns, is performed along with metabolic measurements of the Energy Cost of Walking (ECW). From the gait analysis results, the physiatrist concludes that hyperactivation of the calf muscles is present, while François is walking with slightly flexed knees when standing. This positioning causes compensatory hyperactivation of muscles at other levels, resulting in an overall increased ECW. The physiatrist should decide the therapy, but she can only rely on scattered information derived from some single cases, rather than explicit design rules. Current knowledge shows that chemo denervation of the calf muscles should normalize the hyperactivation, end compensatory activation and decrease the enhanced ECW. However, it would also contribute to a higher knee flexion moment that would drive the knee in further flexion when standing, worsening his walking. This latter phenomenon would call for another therapy known to be effective to counteract knee flexion. Fortunately, the laboratory has applied the EU standards of clinical gait analysis becoming a MD-Paedigree clinical centre. An additional scan is made and all the information on François' case is anonymised and uploaded to the system. MD-Paedigree's disease modelling analysis shows that chemo denervation of the calf muscles could solve the problem, but the effect on knee flexion remains undecided, within acceptable model accuracy. However, the probabilistic model supports the treatment choice and points two matched cases where chemo denervation with additional physiotherapy proved to be successful. The physiatrist is now confident of how she will treat François.

The Infostructure Challenges

BIG DATA MANAGEMENT

- Manage heterogeneous, federated biomedical data sources & models
- Data provenance
- On line transformation (ETL)
- Anonymisation
- Semi automatic curation

BIG DATA ANALYTICS & MODEL BASED ANALYSIS

- Address High Dimensionality & heterogenity
- Scaling through Distributed processing
- Twofold Similarity Analysis (patients like mine & patients like me)
- KDD, statistical simulation & DSS based on BIG (routine) DATA
- Biomedical & Imaging Model Based Analysis

CLINICIANS & RESEARCHERS SUPPORT

- Scientific workflow support
- Collaboration and 2nd opinion support
- Personalised, Unified Access to internal & external data, information, models & knowledge
- DSS Tools & Applications for every role

MEDICAL PRACTICE RE-ENGINEERING

- Ethics & Privacy
- Transform daily routine's data into useful information & knowledge
- Promote Model Guided Personalised Medicine using similarity, simulation and DSS tools
- Tools & Models validation based on clinicians' feedback

CREATE & SUPPORT AN ECOSYSTEM

- Organise communities (clinicians, researchers & patients)
- Save, organise and diffuse information & knowledge
- Promote health (self care, awareness – patients like me, similarity)
- Market Place for everything and everyone (data, models, services and applications)

David Manset	Iannis Ioannidis and Harry Dimitropulos	Henning Muller and Patrich Ruch	Constantin Suciu
Head of the Infostructure Group-Grid-Cloud Services Provision	ATHENA - Biomedical Knowledge Discovery and Simulation for Model - guided Personalised Medicine	HES-SO - Semantic Data Representation and information access	UTBV - GPU Services Integration

News

8th November MD-Paedigree wins the Best Exhibit Award at ICT 2013 - Create, Connect, Grow in Vilnius, Lithuania. This is a considerable achievement as ICT 2013 hosted over 200 exhibits and 5500 participants from all over the world.

ICT 2013 was the 20th and largest EC event in ICT and was hosted for the first time in Eastern Europe. It is one of the most significant events in Lithuania's presidency of the European Union (EU) Council organised by the Ministry of Transport and Communications and Ministry of Foreign Affairs together with the European Commission (EC). The most prominent EU ICT researchers, representatives of new and experienced companies, strategists in digital technology, engineers, investors, high-ranking EC and EP officials, ministers and PMs from EU member countries, students and journalists were present at the event. The MD-Paedigree booth at ICT 2013 showed demonstrations of cardiac-modelling and case reasoning from Siemens, semantic case-based reasoning from HES-SO, automated data curation and model exploration from Athena all interrogated together with a common interface using the digital repository from gnúbila. MD-Paedigree's other contribution to the conference was to conduct a networking session entitled Big data and data analytics impact in healthcare. This session was standing room only, in no small part due to the discussion paper Edwin Morley-Fletcher had prepared on Big Data Healthcare. This paper was so well received at the conference that it has been quoted and re-distributed on the EC's Digital Agenda for Europe website. During the networking session, MD-Paedigree brought together a wide range of stakeholders to come together and discuss the challenges faced in improving Europe's ability to manage and exploit its healthcare data assets laying the foundation for a Big Data in Healthcare interest group. The MD-Paedigree team had members from all the technical partners as well as several clinical ones. They worked tirelessly both before and during ICT 2013 and did everything they could to give MD-Paedigree the winning exhibit.

For more information see:
http://ec.europa.eu/digital-agenda/en/news/big-data-what-it-and-why-it-important

Special Feature: Big Data Healthcare

An overview of the challenges
in data intensive healthcare [1]
E. Morley-Fletcher

Neelie Kroes Vice-President of the Europen Commission speaking at ICT 2013.

Big Data was one of the main themes at the ICT 2013 conference, as evidenced by the quote from Neelie Kroes above. For the conference, Edwin Morley-Fletcher prepared a discussion paper on Big Data Healthcare which we reproduce here.

In order to bring about the revolution in healthcare that modern ITC promises, there are legal, technical and cultural/societal barriers that must be overcome. In this draft paper we deal with the concept of Big Data as applied to healthcare, or Big Data Healthcare, and the developments it may bring. We then consider some of the current major hurdles to its acceptance in standard healthcare.

1. Big Data Healthcare

Big Data Healthcare is the drive to capitalise on growing patient and health system data availability to generate healthcare innovation. By making smart use of the ever-increasing amount of data available, we can find new insights by re-examining the data or combining it with other information. In healthcare this means not just mining patient records, medical images, biobanks, test results , etc., for insights, diagnoses and decision support advice, but also continuous analysis of the data streams produced for and by every patient in a hospital, a doctor's office, at home and even while on the move via mobile devices. Current medical hardware, monitoring everything from vital signs to blood chemistry, is beginning to be networked and connected to electronic patient records, personal health records, and other healthcare systems. The resulting data stream is monitored by healthcare professionals and healthcare software systems. This allows the former to care for more patients, or to intervene and guide patients earlier before an exacerbation of their (chronic) diseases. At the same time data are provided for bio-medical and clinical researchers to mine for patterns and correlations, triggering a process of "data-intensive scientific discovery", building on the traditional uses of empirical description, theoretical computer-based models and simulations of complex phenomena.

Big Data has been characterised as raising five essentially independent challenges:

+ volume,

+ velocity,

+ variety,

+ veracity (lack thereof),

+ value (hard to extract).

As elsewhere, in Big Data Healthcare the data volume is increasing and so is data velocity as continuous monitoring technology becomes ever cheaper. With so many types of tests, and the existing wide range of medical hardware and personalised monitoring devices healthcare data could not be more varied, yet data from this variety of sources must be combined for processing to reap the expected rewards. In healthcare, veracity of data is of paramount importance, requiring careful data curation and standardisation efforts but at the same time seeming to be in opposition to the enforcement of privacy rights[2]. Finally, extracting value out of big healthcare data for *all* its beneficiaries (clinicians, clinical researchers, pharmaceutical companies, healthcare policy-makers, etc.) demands significant innovations in data discovery, transparency and openness, explanation and provenance, summarisation and visualisation, and will constitute a major step towards the coveted democratisation of data analytics.

2. Healthcare Data Access and Protection

Three political, academic and business discussions are currently at the core of the EU debate in this area - the need to ensure that EU citizens' data are adequately protected, the need for open access to data for research purposes, and the need for Europe to develop a vibrant Big Data industry, capable of investing a growing amount of resources into break-through innovations in healthcare that the appropriate utilisation of Big Data promises to deliver. The combined impact of the effective use of Big Data in research and clinical applications has the potential to significantly enhance the health and wellbeing of the EU citizens.

These three debates are running largely in parallel with the following:

1) the loss of trust in data privacy promises, caused by recent disclosures highlighting massive intrusions by government agencies and globally acting commercial companies, as well as inappropriate data sharing by social media organisations;

2) the continuous growth of healthcare data, which awaits transformation into bio-medical knowledge by adequate data crunching and data analytics applications that will bring about innovations and improvements in healthcare;

3) the open access debate, being driven by the ever increasing size of scientific literature and the rising

cost of access to journal articles, which is further complicated by recent discussions of treating datasets as first-class objects in scholarly communication and the research life-cycle, thereby bringing them into the open access sphere as well.

2a. An ownership issue and the European gift relationship tradition

A further topic, often overlooked in reference to Big Data, is the rights of patients to their data or "the data subject's legitimate interests". There are currently instances where others use these data in a supposedly anonymised form, usually without the patients' knowledge or consent, to earn profits without letting patients participate in the wealth generated from adding their respective data to large repositories. For example, in the pharmaceuticals industry there is already a growing market for data on individual prescriptions. Regulation aimed at letting patients participate in some way in the advantages derivable from their data is a complex topic in need of extended public debate leading eventually to policy attention.

Richard Titmuss' 1970 seminal study on *The Gift Relationship: From Human Blood to Social Policy*, has long been an implicit reference standard due to its concern regarding "processes, institutions and structures which encourage or discourage the intensity and extensiveness of anonymous helpfulness in society" and its warning that "patients can be harmed, physically and psychologically, by giving themselves, willingly or unwillingly, knowingly or unknowingly, as teaching material" to scientific research and the medical profession. International agreements, such as the 1997 European Convention on Human Rights and Biomedicine (and its 2002 additional protocol), have stated that "the human body and its parts shall not, as such, give rise to financial gain or comparable advantage". Also the UK House of Lords Science and Technology Committee's 2001 report on human genetic databases concluded that: "we do not regard ownership of biological samples as a particularly useful concept [...] we prefer the notion of partnership between participants and researchers for medical advance and the benefit of others", while the 2011 report on *Human Bodies: Donation for Medicine and Research*, issued by the Nuffield Council on Bioethics, is not opposed, in principle, to money payments for people donating human bodily materials but insists on a subtle distinction between *paying for the material* purchase of a thing and *paying for the people* for their donation, rejecting the former.

Another point, raised by Anne Philips' *Whose Bodies, Whose property?* (2013), is the advocacy for introducing "a levy on the proceeds of research to be returned, either as assistance to the specific group suffering from the disease, or to the wider community".

How these complex issues are reflected in the Big Data debate is still entirely to be ascertained, even though it is clearly of paramount importance for the development of bio-medical research.

---

2   The more we remove information from patient data that can be used to identify patients, the lower the data's veracity and thus clinical value. A patient's name may not be clinically significant, but age, gender, blood type and clinical event timings can be used to help identify patients while also having obvious clinical relevance.

2b. The goal of a European Research Area and the Data Protection Regulation

Access to data is recognised by the European Commission as key to the completion of the European Research Area (ERA) and to ensuring a "single-market of data". The Commission Communication on "A reinforced European Research Area Partnership for Excellence and Growth" establishes the fundamental priorities for completing ERA. Amongst them, it includes the optimal Union-wide circulation, access and transfer of scientific knowledge, as also imbedded in data, in order to ensure optimal transnational co-operation in the frame of building ERA. In order to foster Europe's research and innovative potential, Member States shall implement the Commission's recommendation of improving research teams' access, *inter alia*, to medical record data.

The Conclusions issued by the European Council on 24/25 October 2013, however, have been rather generic, stating that "Europe must boost digital, data-driven innovation across all sectors of the economy", that "strategic technologies such as Big Data and Cloud computing are important enablers for productivity and better services", that "EU action should provide the right framework conditions for a single market for Big Data and Cloud Computing", that "the European Council calls for the establishment of a strong network of national digital coordinators which could play a strategic role in Cloud, Big Data and Open Data development", and that "the commitment to complete the Digital Single Market by 2015 has to be delivered on".

A precondition for all this to become a reality, and for Europe to assert a dynamic and independent role in this crucial area, is the speedy adoption of the Data Protection Regulation, which is still under discussion. Two basic principles in the current text are the "right to be forgotten" (when a data subject no longer wants its data to be processed and when there are no legitimate grounds for retaining it, the data must be deleted) and "informed consent" (consent should be given explicitly by any appropriate method enabling a freely given informed indication of the data subject's wishes).

2c. The right to be forgotten

The right of data subjects "to be forgotten" implies, de facto, some sort of ownership of data relating to them, and therefore the right to eventually donate or sell them. It is true, however, that when it comes to data relating to individual subjects, the preferred model for the advancement of scientific research seems to have been not to allow intellectual property (IP) rights deriving from raw datasets, while IP rights should only become attached to the analytic work performed on the data, in the same way as current IP law covers arrangement of facts, but not the facts themselves. Therefore, a distinction needs to be made between data and the products and services developed using these data. For a market in data analysis to be feasible, if ownership of such products and services should rest with the developer, the basic ownership of the data on which it was based should rest with the data subject,

but no direct IP rights should be derived from the raw data as such.

2d. Data donation and data inheritance

What remains unclear is how the "right to be forgotten" affects these products and services. Suppose there is some patient data based product (e.g. a disease model that has been developed by analysing and in some novel way aggregating a dataset). If some data subjects whose data makes up part of the underlying dataset request the deletion of their data, must the effects of their data be removed from the developed product? If the product has been sold or licensed, or if the data was in fact totally anonymised, which in itself is difficult to prove, this may no longer even be possible.

The demands for informed consent and de-identification imply the implementation of appropriate counteracting measures to prevent deductive disclosure, i.e. the ability to re-identify data based on some inferences either by aggregating more data or by querying the available dataset. A simple and effective approach consists of setting a minimal threshold on query engines so that queries returning less than a minimum number of cases do not inform the user on the real count (K-anonymity). Another approach is the injection of noise into datasets for prevention of re-identification. Nevertheless, with both of these approaches, there is clearly a balance to be struck between protecting privacy and maintaining the utility of the datasets. As more and more diverse person-related datasets become available, data that were once regarded as 99.99% anonymised may no longer qualify for this score.

From an ethical perspective, data subjects should be adequately informed of the current and future health care advantages that they may derive by making their data available for biomedical exploitation. Yet, how can a data subject grant consent for as yet unknown purposes? There is a need, in this sense, for what could be defined as "enhanced privacy" or "enhanced consent", based on spreading the awareness of the personal and social significance of anonymised individual patient and personal data for preventative and predictive purposes in healthcare, and for promoting "data donation" mechanisms. This could be combined with "data inheritance" mechanisms, which are automatically applied after a certain period has elapsed from the data subject's time of death, unless they have explicitly opted-out prior to death. Certainly, a different treatment should be considered for the parts of these data whose the public availability could have detrimental consequences for the relatives of the deceased, such as genetic information.

2e. Anonymisation and enhanced privacy

"Enhanced privacy/enhanced consent" could also focus on the subject being able to define restrictions under which the consent becomes void (e.g., when the study involves the development of armaments) and should be coupled with the concept of "personal data portability": an individual should be able to export or delete his or

her data from the system at the end of a relationship with a particular service provider or researcher. In an age of intense and pervasive innovation, the speed of technological development is likely to outpace the legislative process, so there is a need to constantly update legal frameworks to implement forward-looking policies and laws, allowing for flexible regulations to keep step with technology. It may be that the more pragmatic approach is to legislate against misuse of data as opposed to prescribing allowable uses. Article 81 of the Data Protection Regulation states that Union law or Member State law dealing with processing of personal data concerning health shall provide for suitable and specific measures to safeguard the data subject's interests and fundamental rights versus the necessity for the purposes of preventive or occupational medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services, and where those data are processed by a health professional subject to the obligation of professional secrecy, or another person also subject to an equivalent obligation of confidentiality.

Article 83 states that within the limits of the Regulation, personal data may be processed for historical, statistical or scientific research purposes only if the following hold:

(a) these purposes cannot be otherwise fulfilled by processing data that do not permit or no longer permit the identification of the data subjects;

(b) data enabling attribution of information to an identified or identifiable data subject are kept separately from all other information under the highest technical standards, and all necessary measures are taken to prevent unwarranted re-identification of the data subjects.

This highlights the crucial importance and desirability of de-identification of data, but recognizes that it is not always possible to carry out the necessary research if the data are fully de-identified.[3]

---

3   Note that a 100% anonymisation of patient data is impossible, or they may be rendered useless. Whether, for all practical purposes, data are anonymised with a probability of 99.9%, 99%, or 95% will depend on the availability of other data on individuals (which is hard to predict for the future), computer power and related factors. This seems to be a largely unexplored field. Of great interest, however, is the legal framework adopted by the p-medicine project, based on the following pillars:
    1. Pseudonymization
In clinical care pseudonymization should be the norm, but, when in need of using these data for research or to upload them with semantic links to a data warehouse, a second pseudonymization should take place, performed by a trusted third party (TTP) which is in case of p-medicine is the CDP (Center for Data Protection; http://www.privacypeople.org). This pseudonymization is done using a tool called CAT (Custodix Anonymization Tool).
    2. Contracts between data providers and data users
These contracts bind users legally to use the data only according to the research for which the data are requested. Fines are defined if somebody tries to re-identify a subject.
    3. National differences (which might disappear after the approval of the new Data Protection Regulation)
Today there are different laws regarding data protection in different member states throughout Europe, where the usage of data is different.
    4. Informed consent is seen as a fallback in this legal framework
Data are regarded as de-facto-anonymous if they fulfil the above mentioned points.

It must also be noted that full anonymisation is in principle impossible, depending on the data recorded; e.g. both DNA (fingerprinting) and images (e.g. 3D) do always theoretically allow the identification of a specific patient. To address this, wherever possible, data "itemising" must be considered so that there is not a complete image/genotype record, and therefore identification of the individual is more difficult, although in principle not impossible.

The debate on data protection and open access should come to an ethically-based consensus agreement, allowing for the views of minorities to be respected, if the right of citizens to appropriate data protection is to be adequately balanced against their right to further improved healthcare based on patient data-facilitated clinical research. This balance is crucial if legislators wish to avoid overprotection of the rights of a minority becoming detrimental to the delivery of effective healthcare for the majority.

To this end, it must be ensured that it is not just the nature of the data that influences the level of protection afforded, but also the intended use of those data, and the potential risks implied by their usage. EU citizens have shown time and time again in surveys that their concerns over data security relate not to the use of their data per se, but to who will use it, and to who might have access to it in the future.

In a UK survey conducted in 2009, it was revealed that an overwhelming number of UK citizens were willing to donate their health data where such data would be used solely for the advancement of medical knowledge. Concerns over the use of data were confined to fears that their data would be leaked to their employer or insurance company. The nuances of EU citizens' concerns shows that there is a need to ensure that in crafting the solution to the data protection vs. open access and Big Data analytics dichotomies, the specificities pertaining to each of these issues are appropriately identified.

Denying health researchers access to data necessary for potentially life-saving research, the results of which EU citizens may someday need, is no guarantee that their data is any more secure. On the contrary, it will serve only to hinder the progress of that research, delaying, or even preventing, the development of new treatments in demand by EU citizens.

2f. Striking a balance between between privacy, security, and innovation

Today in Europe there is an increasing call by some for unambiguous data protection regulation that could ease new scientific research instead of hampering it. Regulators should understand that where

and when strong, consistently audited data security

measures are applied, the benefits of medical research in all probability outweigh the putative risks associated with the use of patients' data by health professionals. A harms-based approach, focusing more on the regulation of possible misuse of data, rather than on limitations of usage, is likely to be the most appropriate approach for striking a reasonable

balance between privacy, security, and innovation.

Article 70 of the new draft Data Protection Regulation suggests that the "indiscriminate general notification obligation should be abolished, and replaced by effective procedures and mechanism which focus instead on those processing operations which are likely to present specific risks to the rights and freedoms of data subjects by virtue of their nature, their scope or their purposes. In such cases, a data protection impact assessment should be carried out by the controller or processor prior to the processing, which should include in particular the envisaged measures, safeguards and mechanisms for ensuring the protection of personal data *particularly as specified in the enhanced consent* and for demonstrating the compliance with the Regulation".

Article 71 significantly specifies that "this should in particular apply to **newly-established large scale filing systems**, which aim at processing a considerable amount of personal data at regional, national or supranational level and which could affect a large number of data subjects".

This paper proposes the inclusion of a concept of enhanced consent where the data subject is able specifically to exclude certain data usage whilst allowing data utilisation for the benefit of, for example, healthcare research, alongside maintaining and ensuring that consent can be withdrawn and data completely deleted. As described in the Regulation, all data management will be under the auspices of a controller[4] to ensure compliance, and thus the rights of the citizen are protected and the benefits of the utilization of Big Data in research, innovation and business development in healthcare are retained and developed.

3. Dealing with the Data Glut

With the ever-increasing volumes of data being produced outstripping[5], and arguably being driven by, the computing power available to analyse them (quantum computing aside possibly), we already are faced with the reality that we have too much data. As of 2011, this "data glut" was estimated to be 150 exabytes (150 billion gigabytes) for healthcare globally. To make sense of so much data, where sense is to be found, will require innovative analytical techniques that can make it possible to efficiently search, process and analyse these massive datasets. Some handle may be gained over the torrent of data by reducing the dimensionality of a dataset. Feature selection methods, whether selecting features on the basis of existing medical knowledge or on statistical techniques, can be used to map a dataset with many feature dimensions to one with significantly fewer, thus creating a manageable search space. This simpler search space is then used for querying and analysis, with the full dataset only referred to when necessary, allowing for the application of goal-

oriented search techniques such as Model-Driven Analysis. Fractional Factorial Design uses this sort of approach to concentrate search efforts in areas of a multi-dimensional dataset that have been selected by searching a lower-dimensional one that it has been mapped to.

Following a top-down system level approach, Feedback System Control (FSC) has also been proposed recently to reduce the number of experiments in *in silico* clinical trials. FSC was shown to efficiently hone-in on an optimised drug combination with 10^2–10^6 times fewer experiments than a typical high throughput ("brute force") approach. As opposed to collecting all measurable data and trying to find a needle in a haystack, the FSC scheme is a goal-oriented method, which uses the phenotypic outcome to tune the intervention of engineering systems, achieving system-in-system integration.

These various techniques allow us to find correlations, patterns and structures in overwhelming volumes of data, giving them value. They reinforce the fact that **data do not possess inherent value in the absence of a means to make sense of them.** They are meaningless until analysed for significance, visualised within a context or compared to other data. This means that the value of a dataset will vary according to its context. The corollary of this is that the value of a dataset is the sum of its value in each analytical context. This fits neatly with the concept that it is the research results, services and products generated from data that will provide the value in a Big Data economy.

3a. Statistical and mechanistic methods within the moving boundaries of the "dimensionality curse"

Ultimately, data will be as useful as the knowledge that can be derived from them, implicitly or explicitly. Two diverging modelling approaches need, therefore, to be taken into account here. On the one hand, the bottom-up approach, based on statistical models that can identify patterns and correlations between observed variables in large datasets, and that are necessarily dependent on the data to which they are trained. These models will not reveal fundamental causality between variables or dynamic aspects, as would be needed for understanding complex biological processes. In the era of Big Data, however, *a priori* knowledge is supposedly not enough, but given the fact that valuable information exists in various data sources, these can be additionally used for enriching the global amount of available information. Thus, Big Data (bottom-up) knowledge discovery and data mining (KDD) can remain the primary goal, even if, to address such a challenge, it becomes necessary to move beyond classical – one source, one modality – statistical simulation models, and there is the need of analysing and combining information from different data sources (e.g. biomedical data & literature) and modalities (e.g. clinical variables & images or streams).

On the other hand, following a top-down approach, mechanistic models, based on an understanding of the behaviour of biological systems' components, are

---

4 Who in turn needs to be controlled or audited.

5 For instance, in genetics DNA gene sequencing machines based on Big Data analytics can now read 26 billion characters of the human genome in seconds.

obviously more difficult to build and generally larger in scale, but are more robust to noise and to local inaccuracies, and eventually provide, when executed properly, unprecedented power for extrapolation and prediction in domains in which the other techniques fail. This is why, approaches such as those advocated by molecular systems biology, or more recently by the Virtual Physiological Human across all scales, need to be fully exploited in healthcare. The development of such models, requiring identification of the system nonlinearities and model structure and estimates of the system model parameters, implies, however, a computation time directly related to the number of terms and increasing dramatically with the model order. A trade-off between accuracy and complexity becomes therefore the way to avoid the "curse of dimensionality" leading eventually to a computationally intractable combinatorial optimisation problem.

Furthermore, when studying highly complex and inherently non-linear physiological systems, it would not make sense to assume that their model order and structure can be sufficiently well known *a priori*. In fact, in biological systems analysis the main objective is precisely to gain insight into the underlying structure and the system identification is based on a learning process, associating parameter estimation with structure selection algorithms, aimed at finding the simplest possible model capable of revealing the unifying rule relating the components and behaviours of the system under study. However, since patterns discovered through KDD enrich already existing *a priori* knowledge, and the latest statistical simulation models can incorporate *a priori* logic (e.g. formulating constraints, imposing dependencies, etc.), a sensible goal is to combine both bottom-up and top-down approaches.

The search for parsimonious and computationally efficient methods enabling the determination of the simplest model structure, has additionally led to the development of methods that can cope with partial and/or incomplete mechanistic knowledge, such as the nonlinear, autoregressive, moving average exogenous (NARMAX) structure detection approach, pioneered for many years by the signal processing and complex systems research of Stephen Billings and others at the University of Sheffield.

This method is based on black-box parametric system identification and leads to a mathematical model on the basis of input/output causal relationship calculated with non-linear difference equations. It has demonstrated a remarkable capability of accommodating the dynamic, complex and nonlinear nature of real-world time series prediction problems, successfully achieving the goal of determining the model form and estimating the numerical values of the unknown parameters, and eventually validating those results which have shown sufficient accuracy. In the analysis of physiological systems this mathematical modelling can allow to integrate very large data sets into a consistent information framework, which can further be used to arrive at novel mechanistic insight and predictions.

3b. The value of bio-medical Big Data repositories

Given the fact that data are "experience goods", according to Arrow's paradox it should be true that once access to them has been awarded, their value should be significantly reduced because of the inherent non-rivalry and non-excludability characteristics of open access information. However, Big Data repositories coupled with analytics can be utilised in a variety of excludable ways and their value is not critically influenced by Arrow's paradox, while the "experience goods" challenge mainly concerns the availability of "enough" descriptive information about the data, their structure, process of collection, and possibly "teasers" about the analyses or outputs from the database.

However, it remains true – as stated in the OECD Report on "Supporting Investment in Knowledge Capital, Growth and Innovation" – that in order to reap this growing value there will be the need not only for clinicians and researchers to acquire Big Data analytics skills and services, but also to develop a framework for data repositories which adheres to international standards for the preservation of data, sets common storage protocols and metadata, protects the integrity of data, establishes rules for different levels of access and defines common rules that facilitate the combining of datasets and improve interoperability. These frameworks could, someday, render some of today's data protection rules and procedures invalid.

Such repositories can allow "testing" of the prior knowledge of clinicians, who identify the data features deemed to be key for specifying a patient's treatment, versus the correlations that big data crunching may highlight, possibly leading to further knowledge discovery. Indeed, by statistically and semantically reasoning on the data, existing pathophysiological patterns may be revealed and inputted as a first step in a fractional factorial and model driven research process supporting physicians in their iterative and interactive quest to discovering new knowledge.

The goals here are: to be able to provide model-driven patient-specific predictions and simulations and consequent optimised personalised clinical workflows, to allow for advanced similarity search among patients, such that clinicians can find "the patient like mine", and to get support through risk stratification and outcome analysis. Eventually it is hoped that specific pathophysiological patterns ("disease signatures") can be detected, refined and made available to other clinicians and researchers in the form of pattern libraries. These pattern libraries, identifying homogenous groupings among patients and model similarities, could be shared between researchers and clinicians to allow for data intensive pathophysiological diagnoses. Allied to the above is the potential to revolutionise health communications by making it possible, on the basis of semantically advanced repositories, to use social media among patients aware of sharing highly similar conditions ("patients exactly like us"), empowering them to bridge the gap with the clinicians, especially in the case of paediatric patients and their parents.

4. The "Long Tail" potential in bio-medical research

The Pareto Distribution, introduced in 1906 by the Italian economist Vilfredo Pareto, described a feature of social distributions which has been a recurrent mantra in organizational studies. Presented in simplified form as Pareto's Principle, or the "80/20 Rule", it states that 20 percent of something would normally be responsible for 80 percent of the results.

In the economic arena, this traditional 80/20 concept has begun to be reversed, as highlighted by the innovative insights in Chris Anderson's book, *The Long Tail* (2006). This developed the concept that, when transaction costs are greatly lowered, "the biggest money is in the smallest sales", whereby a series of small niches cumulatively achieve a much larger amount than the traditional focus on selling the preferred 20% of the items.

Internet businesses such as Amazon have demonstrated this, because they have infinite shelf space and a radically different cost structure to "bricks and mortar" retailers. The long tail has been extremely lengthened as a result. Consumers really can find and choose whatever they want, no matter how popular or sought after the item is. While such retail niches were not economically viable in the past, they can now better fulfil the market.

An analogy to epidemiological studies can apply here. What works for business transactions can also work for clinical interactions. Big Data applications in medicine radically change the capacity for going beyond the average patient population, searching for specific cohorts of patients fitting into very peculiar niches of their own.

Such an approach can show the way to truly personalised medicine. Applying the Long Tail insight to bio-medical research and to drug discovery, will lead to people receiving treatments and drugs specifically targeted to their own genomic, proteomic, and metagenomic characterisation. Tailoring treatments, drugs and research to everyone's individual needs is precisely what the long tail's approach is about. Thus focus on the long tail in healthcare should allow medicine to better address all those diseases and ailments suffered by a relatively small number of people or by a large number of people whose common conditions have numerous underlying causes.

Big data allows for "long-tail medicine" drugs with enhanced personalised information content, based on customized algorithms tackling the individual disease conditions that can be best addressed only by personalised treatment. Notably, such conditions may often be the most serious. For example, cardiovascular disease is responsible globally for more morbidity, mortality and economic burden than any other disease. Despite growing knowledge about its various causes and risk factors, it remains difficult to tackle in a preventative sense. A key element of this difficulty results from the complex nature of the disease, arrived at by a multitude of pathways, influenced by genome, metagenome and environmental exposures. Thus, no single intervention or set of interventions

applied in a blanket fashion to the population has been able to tackle this disease adequately. Personalised interventions, resulting from a big data focus on the disparate underlying genotypic and phenotypic drivers of the disease offers a possible solution to this.

There are also "orphan" diseases, which affect relatively small numbers of people. The low prevalence of these diseases results in little or no direct research investment being made by industry to understand them or to develop new treatments for them. The biopharmaceutical industry in particular is still largely focusing on a "one size fits all" approach, but "one size" medicines do not fit all patients, and the same is true of the R&D process. The limitations of this approach have become increasingly clear, starting from the need of increased paediatric knowledge discovery, as exemplified at a basic level by the Paediatric-use marketing authorisations which are promoted by the European Medicines Agency.

Data sets from a sub-population or from longitudinal clinical data have the potential to expedite the development of targeted therapies in terms of both patient population and disease. This model of research is in its infancy but has tremendous implications for medical knowledge discovery and for future drug development.

5. Big Data Literacy in Healthcare

In parallel to the legal and technological challenges we have outlined above, and to the fact that the healthcare domain is known for its ontological complexity, variety of medical data standards and variable data quality, there are also healthcare cultural changes that will be required to fully capitalise on Big Data Healthcare as a movement. This especially revolves around the need for medical staff to be educated using examples of its success. Clinicians will need to understand and be persuaded of the potential of Big Data Healthcare, as well as its limits. To maximise the acceptability and utility of Big Data Healthcare solutions in the clinical arena, clinicians should therefore be involved in the development of these solutions from their inception. Thus, a clinically led development ideology will ensure that technical know-how and innovation translates into clinically useful tools that fit more naturally into clinical workflows. Clinicians and engineers must work together to translate and extend their existing and advanced data analysis technology, that is the clinically trained human mind, into targeted big data analytical approaches that will achieve clinically useful outputs. Although engineers and clinicians have long collaborated successfully, development work in Big Data Healthcare will require particularly intimate reciprocal understanding by each disciplinary culture of the other. This will require further cultural development in both areas. This should be achieved at a grass-roots level by a greater emphasis of clinical informatics in medical curriculums and similar exposure of developing bioinformatics engineers to the unique challenges that medical bioinformatics faces. At the same time, decision support systems will need to be more than black boxes but be capable of showing clinicians why

they advocate particular courses of action, providing the necessary assurance that advice is based on sound principles.

Diffusion of medical technologies is necessarily a lengthy process. This means that these processes of education and culture shift should begin now, preparing new generations of clinicians and bioinformatics engineers for forthcoming data intensive methodologies and collaboration.

The following points will need therefore to be fleshed out:

+ Management of big data

+ Seamless end-to-end big data curation

+ Data discovery, profiling, extraction, cleaning, integration, analysis, visualisation, summarisation, explanation

+ Use of big data

+ Appropriate use of big data – avoiding over-reliance

+ Responsible use of automated techniques

+ Communicating big data findings to patients

+ Integrating data analytics into clinical work flows

+ Data (clinical) scientist

6. The consequences for policy-making

When Jean Monnet launched the European construction in 1950, he based it on coal and steel, as these sectors were of strategic importance at the time. In an updated spirit, Nellie Kroes has suggested that the unifying vision for Europe should presently be based on two concepts: "being wired" and "being digital". Sometime earlier, she had also stated that "knowledge is the engine of our economy, and data is its fuel".

It is hard to disagree with her. However, this vision brings with it the need to go beyond the usual proliferation of EC-jargon acronyms that hint at barely specified future goals.

In line with the targets set out in the *Europe 2020 Strategy*, the issue of Big Data – as already mentioned - has now been singled out as a political priority at the 24-25 October European Council. There is consequently much discussion about creating a *European Big Data Analytics Service* (EBDAS) network, aimed at fostering a sustainable, smart and inclusive *European Big Data economy*, while the *Digital Agenda for Europe (DAE)* has been assigned the overall task of creating a *European digital single market* through a number of measures directed at the use of data sources in Europe. All this will need to pass through the *EU Open Data strategy* (conceived of as an amendment of the *Public Sector Information Directive*), the *Data Value Strategy* (addressing the entire data life-cycle), and the crucial *Data Protection Regulation*.

At the same time, due to the staggering, and ever growing, size and complexity of bio-medical big data, computational modelling in this area is likely to provide the most intriguing insights into the emerging complementary and synergistic relationship between computational and living systems. Bio-medical research institutions and related industries, as well as the whole healthcare and pharmaceutical sectors, must therefore be considered as key stakeholders in the European process leading to the next generation of data-centric systems. Whether these are systems capable of learning from data, or data-analysis products and applications capable of translating medical knowledge discovery into widespread medical practice, they will put predictive power in the hands of clinicians and healthcare policy makers.

First of all, this requires the swift definition of an updated legal and ethical framework that adequately protects citizens on a European-wide basis. This is a necessary premise for allowing private and public European investment to be directed towards this crucial area of innovation. Doing so will support the entrepreneurial initiatives aimed at commoditising once complex big data analytics, foster massive clinical uptake, and reinforce European competitiveness in medical developments. FP8 European research funding needs to be directed accordingly.

MD-Paedigree Newsletter - Issue 1

Editorial Board

+ Beatrice Bressan - Lynkeus
+ Mirko De Maldè - Lynkeus
+ Callum MacGregor - Lynkeus
+ Almerico Bartoli - Lynkeus

Guest Authors

+ Bruno Dallapiccola - OPBG
+ Gabriele Rinelli - OPBG
+ Andrew Taylor - UCL
+ Alberto Martini - IGG
+ Jaap Harlaar - VUmc
+ Giacomo Pongiglione - OPBG
+ Edwin Morley-Fletcher

Subscription

Disclaimer

Bambino Gesù OSPEDALE PEDIATRICO

ISTITUTO GIANNINA GASLINI
Istituto Pediatrico di Ricovero e Cura a carattere Scientifico

Universitatea Transilvania din Brașov

UMC Utrecht

VU VRIJE UNIVERSITEIT AMSTERDAM

Fraunhofer

UCL

empirica

Hes·so
Haute Ecole Spécialisée de Suisse occidentale

KU LEUVEN

TUDelft Delft University of Technology

gnúbila

motek medical improving human performance

Inria informatiques mathématiques

SIEMENS

BMR Genomics

SIEMENS corporate research

The University Of Sheffield.

SAPIENZA UNIVERSITÀ DI ROMA

Athena Research Center
Research and Innovation Center in Information, Communication and Knowledge Technologies

LYNKEUS.
STRATEGY CONSULTING
POLICY ADVICE
DATA ANALYTICS