EIT Foundation Young Leaders Group:

nurturing and accompanying tomorrow's innovation leaders

Six ideas utilizing big data to tackle the challenges of our future

SOLVAY
asking more from chemistry®

Vodafone Institute
for Society
and Communications

Table of contents

Foreword

Nowhere else I feel so much optimism to believe in the idea of an innovative Europe than when looking at this upcoming generation of mobile, open-minded, and talented young people. As I truly believe that the source for a European future for sustainable growth can be found in the minds of our young people I am extremely delighted that our Young Leaders' initiative was one of the first to be implemented by the EIT Foundation.

This initiative responds to the Foundation's commitment to nurturing and accompanying tomorrow's innovation leaders. It brought together talented young professionals, entrepreneurs and students, including from the EIT and its Knowledge and Innovation Communities, who during a short period of six months have pondered over and come out with concrete solutions, business and innovation opportunities for the new era of big data.

With the Young Leaders Group I believe we found some of our future leaders, showing that the potential of creativity, cooperation and innovation is out there. Isn't it denotative, in how little time those Young Leaders came up with innovative ideas? The programme endorsed my belief that we need to enable those young talents to come together in groups far away from their business routine and give them time and space to think freely and share their thoughts.

Looking back on a long history of traditional corporate culture, this programme has once more proofed the importance of opening up the sectors to foster exchange. But not only was it the variety of professional background which was of extreme value for the success of this programme. It was also the participants' cultural background. Bringing together young talents from different places around Europe is not only crucial to bring Europe closer together but as it has been shown, brings up a new energy of ideas and

innovation. Definitely this is what we need for a European sustainable future.

My gratitude to this first Young Leaders' group for proving that this could be possible; to the EIT with whom we have been working in full alignment and complementarity; and the Foundation companies who have trusted in the project and made all this happen.

To the people reading this brochure, I hope you enjoy the insights and feel inspired to take them forward.

Christian Jourquin
Chairman of the EIT Foundation

Nowadays it becomes clear that facing today's biggest challenges needs innovative approaches. Innovative in this respect does not necessarily mean "new" but it rather means "different": Sometimes the answer is already out there but not yet shared due to "siloes" of language, background and thinking. Bridging this gap between academia, corporate business and entrepreneurship and sharing knowledge and views could unleash an untapped potential. However, in theory this idea is simple, in praxis multidisciplinarity and diversity can be rather challenging, and often enough do not live up to their rhetorical promise.

When we came up with the idea of a Young Leaders' Programme we stuck to the belief that when applying a different approach by bringing together talented people in interdisciplinary groups with innovative training methods they will come up with fresh ideas: even on questions far apart from their daily business.

That was our rather optimistic – some might call it naïve – starting point. Brought to life within a very short period of time the idea became true – in the EIT Foundation's first Young Leaders Group. When I met all these highly talented people at the kick-off event I felt that our optimism might stand the test of reality. Now, five months later, I am deeply convinced that thanks to the group's initiative and determination the initial idea succeeded. Listening to our future leaders, I am pleased that the feedback reflects that everyone found the programme to be of genuine value. Most of the Young Leaders believe they have gained new technical skills as well as entrepreneurial and social skills during the programme. It was the diversity of the participants that led to a new dynamic and an atmosphere of innovation. Ideally, the Young Leaders will benefit from these inspirations and take them back to their networks and future careers.

Obviously not only the Young Leaders got new inspirations but also did the speakers, both young entrepreneurs as well as corporate seniors. And also we, as members of the EIT Foundation community, were inspired by the young leaders' intellectual vigor and ability to open up new avenues of thought. And this is a true added value for us as member companies being involved in this programme. I trust that during the Annual Innovation Forum the Young Leaders may also inspire others when sharing their thoughts; not only in terms of their innovative ideas but also with regard to the value of the diversity of this group.

Dr. Mark Speich
Director of the Vodafone Institute for
Society and Communications

The EIT Foundation Young Leaders Group initiative:

nurturing and accompanying tomorrow's innovation leaders

The Young Leaders Group is one of the three initiatives established by the **Foundation of the European Institute of Innovation and Technology (EIT).**

The EIT Foundation was established in 2010 as an independent, philanthropic organization under Dutch law. It is a group of like-minded organizations dedicated to promoting a culture of innovation and entrepreneurship in Europe. Its goal is to complement and enlarge the impact of the EIT, inventing the future by bringing together some of Europe's brightest minds at events, in youth talent programmes and similar initiatives.

The EIT Foundation focuses its activities on the:

› **Promotion of entrepreneurial education** and bridging the gap between academia and business.

› Creation of a **new generation with an entrepreneurial mindset** ready to drive the development of sustainable innovation;

› Development of an **international network** of talented professionals engaged in education, research, innovation and business development.

All EIT Foundation activities are implemented in cross-cutting areas of interest to the whole innovation community, and are intended to benefit a wide group of actors, independently of their sector of activity.

The EIT Foundation activities are directly proposed, managed, financed and implemented by its members.

Currently there are ten world-leading companies involved in the EIT Foundation. They are: Agfa Gevaert, Akzo Nobel, Alcatel-Lucent, BNP Paribas, Ernst & Young, Google, Intel Labs Europe, MOL, Solvay and Vodafone Institute for Society and Communications.

The EIT Foundation **Young Leaders Group initiative** seeks to bring a more entrepreneurial and innovation-focused approach to traditional education and professional training, combining them with a specific mentoring scheme. By benefiting from the participants' various backgrounds both in regards of expertise, discipline and origin, the Young Leaders Group will address the big innovation challenges and lead the way to a high calibre European network of future leaders and thinkers.

The main objectives and benefits of the programme are:

› Establish an **international and interdisciplinary network** of young high caliber academics from European universities, and young entrepreneurs;

› **Investigate and analyze** relevant innovation challenges and exchange views across sectors and disciplines;

› Provide a **written contribution** for presenting at the Annual Innovation Forum;

› **New teaching approach** by learning from entrepreneurship and leadership strategies, as well as management experiences of renowned experts.

This publication shows the results of the first edition of the Young Leaders Group initiative. The initial six-month pilot programme, carried out with the active support of **Vodafone Institute for Society and Communications** and academically steered by **Imperial College London** was around the theme of **data-driven innovation.**

30 participants were selected within the internal networks of the EIT Foundation community. They are young entrepreneurs coming from the EIT and its Knowledge and Innovation Communities, business players and high potentials from the academic world.

The programme was launched in November 2012 with a two days kick-off event at the premises of Alcatel-Lucent in Antwerp. Six groups were constituted to work on specific challenges around the topic of data-driven innovation. Following an intensive peer-to-peer work period, Young Leaders met again in the beginning of January 2013 to further discuss their ideas in a four days Winter Academy at Imperial College in London. During the Winter Academy, experts from diverse areas and backgrounds shared their views and experiences on data-driven innovation and supported the Young Leaders in further developing their ideas.

After the Winter Academy, and with the support of Imperial College London, each group prepared the papers of this publication via virtual teamwork. The papers highlight the relevance about the innovation challenge at stake, provide an overview of what is already been done internationally to address this question, and come up with concrete ideas and recommendations on how to better address and tackle it. The aim of the papers is to give a new and inspiring input from a "Young Leaders'" perspective, and thus, to give a fresh and forward-thinking vision for the future.

For further information on the EIT Foundation and its activities please consult http://eitfoundation.org/

Big data in healthcare: What options are there to put the patients in control of their data?

Amitte M. Gulamhussen

Robert Hirt

Marc Ruckebier

Jonathan Orban de Xivry

Guillaume Marcerou

Jeroen Melis

Abstract: *Despite of the significant improvement in life expectancy over the past 50 years there are several growing challenges that are putting pressure on the sustainability of EU healthcare systems. This unparalleled situation is offering the right momentum to rethink business models on healthcare systems which require a new strategic approach shifting from a volume to a performance based model- value for money.*

Currently big data does not efficiently contribute for improving healthcare. In an increasingly empowered digital society, big data has now reached every sector in the global economy. Personal data has become a new form of "currency", nevertheless, this "currency" is not flowing in the healthcare sector because **patients are not in control of their medical health data making impossible to build the trust that is needed.**

To reach the full potential of **big data in healthcare** *at the EU level, with a significant economic output in the ecosystem, it is essential to establish a working model of balanced interests, identifying each participant's priorities, challenges and value opportunities focusing around the core of all innovation – the patients' data – overcoming some barriers that are creating inertia, such as privacy protection, legal conformity, standardisation, interconnectivity, benefits and governance.*

The group's recommendation is to create a bottom-up model, focusing on a patient centric approach - the patients' data - from an exclusive to an inclusive ecosystem. Promoting transparency between doctor-patients relations and implementing a new market model – the "banking model" system which will promote patient trust and value perception, key features to successfully implement a new competitive pan-European market. It is necessary to enable individuals to understand and manage, in a secured environment access, the use and value of such individual information in an ecosystem, where the interests of the different stakeholders, the fundamental rights of the patients and the legal constraints are fully taken into account and respected. As patients are the primary source of Patient Health Record data (PHR), it is essential to design a user experience that is well accepted. A chip card model similar to a bank card is proposed to materialise the digital data with the physical object.

The previous recommendations are fully aligned with the core principles and strategic actions of the EC Health Strategy: prosperity, security and solidarity.

An EU Governing Body should drive gradual implementation of these solutions.

1. Introduction- currently big data does not efficiently contribute to improving healthcare

Despite of the significant improvement in life expectancy over the past 50 years[1] there are several growing challenges that are putting pressure on the sustainability of EU healthcare systems which require a new strategic approach. The worst global financial and economic crisis in decades is challenging countries to invent new ways to increase their efficiency, shifting from a volume to a performance based model- value for money. Over the 1995-2010 period, health spending per capita in OECD countries grew three times faster than the income per capita (9.5% GDP). Nearly 7% of GDP comes from government sources. If left unrestrained, health spending could soon exceed realistic limits beyond what governments, social security or family budgets can afford. Long-term trends[2] in society and demography[3,4] are equally challenging, with changes in disease patterns[5], namely, the rapidly changing structure of sicknesses and ageing related diseases. These facts will require investing in the prevention of chronic and re-emerging diseases, illnesses related with lifestyles, as well as promoting primary care and further integrated and innovative approaches.[6,7] Alongside, some market forces are driving healthcare into transformations, namely, increase of complexity for new revolutionary technologies or treatments and for new market entrants with new approaches to healthcare delivery. On the top of that, there is an absence to address the complex needs of frail elderly of a system that is shifting from a local or national to a

---

1 OECD work on Health 2013-2014; 2012
2 The new agenda for business, WBCSD World Business Council for Sustainable Development; 2010
3 The European Union and the BRIC countries; 2012
4 Demography report 2010 Older, more numerous and diverse Europeans, Eurostat; 2011
5 Health statistics – Atlas on mortality in the European Union; 2009
6 OECD 2012 Preface: High-Performing Health Systems; 2012
7 Europe 2020 Strategy – towards a smarter, greener and more inclusive EU economy?; 2012

trans-national level[8], still, with an inexistent patients' empowerment. In the meantime some opportunities are starting to flourish like personalised medicine, where genomic and proteomics will have a key role on early detection of specific diseases and more effective therapies are expected.

Health is central in people's lives. To support this core value, a new EC Health Strategy[9] was designed (2008-2013 framework) since a cooperative action at Community level is indispensable to support effective policies in healthcare to improve citizens' wellbeing. The political importance of health in all policies reinforces theEU 2020 strategy for growth and jobs and the Citizens' Agenda, emphasising the links between health and economic prosperity, recognising people's right to be empowered in relation to their health and healthcare.

The strategy proposes four core principles: shared health values; "Health is the greatest wealth"; health in all Policies (HIAP) and strengthening the EU's voice in global health,

Underpinning three strategic objectives:

1. **Solidarity:** fostering good health in an ageing Europe;

2. **Security:** protecting citizens from health threats;

3. **Prosperity:** supporting dynamic health systems and new technologies.

Healthcare sector produces unlinked thousands of data. The information is produced as "single-dots" since it is fragmented in four major pools that provide such data: clinical data; payer activity (claims) and cost data; pharmaceutical, medical products and R&D data; patient behavior and sentiment data.[10] Despite the numerous publications that are currently dealing with the Personal Health Record (PHR) they are analyzing its potentials and limitations from various and often opposing perspectives. To reach the full

potential of big data in healthcare at the EU level, with a significant economic output in the ecosystem, it is essential to establish a working model of balanced interests, identifying each participant's priorities, challenges and value opportunities focusing around the core of all innovation – the patients' data – overcoming some barriers that are creating inertia, such as privacy protection, legal conformity, standardisation, interconnectivity, benefits and governance.

In an empowered digital society[11] where information and data are increasingly valued, big data has naturally reached every sector in the global economy as a key tool for success. Personal data has become a new form of "currency"[12], nevertheless, this "currency" is not flowing in the healthcare sector because patients are not in control of their medical health data, making thus impossible to build the trust that is needed.

This unparalleled situation is offering the right momentum to rethink business models on healthcare systems. Under the Young Leaders Programme(YLP) promoted by the Foundation of the European Institute of Innovation and Technology (EIT), the multidisciplinary group of IT, legal, telecommunications and laboratory professionals dedicated on the subject big data in Healthcare, considers that **currently big data does not efficiently contribute for improving healthcare**. Those findings were corroborated after a **wide spectrum dialogue of the team's network across EU countries**, namely, patients with and without chronicle diseases, doctors, pharmaceutical marketing managers, IT professionals, consultants, lawyers, researchers and telecommunication professionals.

8 Redefining Value and Success in Healthcare Charting the path to the future; 2012

9 White Paper- Together for Health: A Strategic Approach for the EU 2008-2013; 2007

10 Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute; 2011

11 The Facebook, YouTube, Twitter, Blogs are many examples where the end-users cease personal data despite the fact that they are aware of the risks generated because users fell that they are in control of their data or because the perceived benefit seems much bigger than the incurred risks

12 The value of our digital identity, Boston Consulting Group; 2012

2. Are you talking to me?
Patient data is an asset owned

Similar to any customer information, patient health data has become an asset – used for analytical purposes to gain insight into various aspects of health for better healthcare, better medical treatments and also for commercial exploitation.

The patient however is not in control of their own data and often struggles to even get read access to such information or related use of the individual health record. This generates a currently patient "exclusive" ecosystem around the PHR.

To overcome the barriers many patients feel at present, and to encourage open information sharing in a secured environment, opportunities must be found to put the patients back in control over their own data and enable individuals to understand, manage access and use of such individual information in the ecosystem of interested parties. It is suggested to put the patient in the core of such ecosystem and enable full visibility of the access and use of such data, leveraging legal or commercial models to even monetise health data with the patient as the main beneficiary of such structures. Any serious attempt to provide a European wide capability for better use of health data needs to provide a clear incentive for the sole data asset holder – the patient.
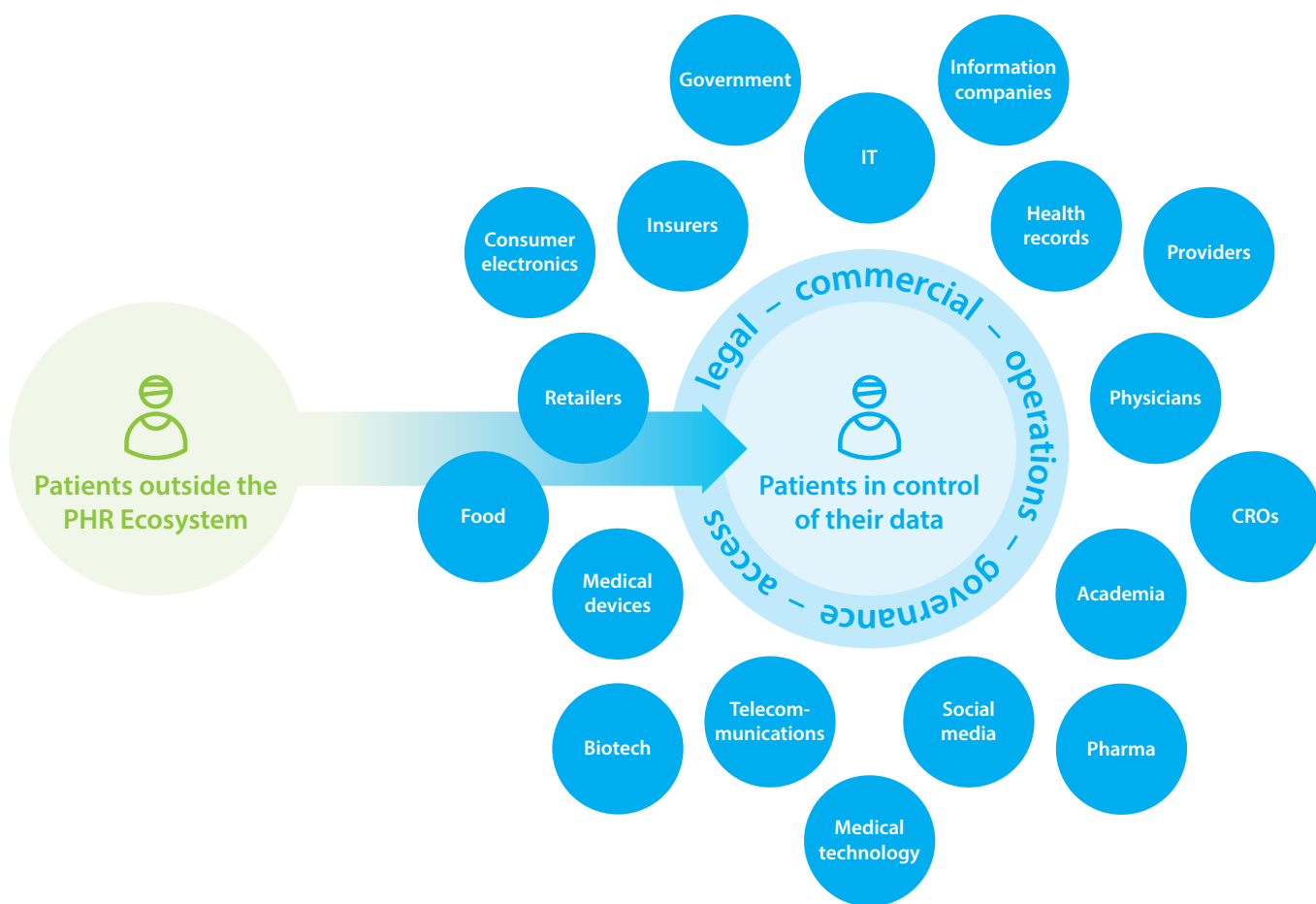

Figure 1: From patient exclusive to Patient Centric PHR Ecosystem

Alongside the overall goal of a patient centric ecosystem, it is important to understand the holistic environment and identify each participant's priorities, challenges, risks and value opportunities. With the patient in control, new barriers will be created, which will challenge the commercial, ethical or legal interests, as well as the ambitions of many other parties.

In order to nurture this idea in such a rough environment, a fair balance has to be reached for everyone participating in this approach, to be clearly incentivised by balancing gain with fair sacrifice, risk with profitable future outlook on the opportunities a European wide PHRs can offer. Figure 2 illustrates a balanced decision framework: the green segments represent the region of benefiting organisations, whereas the red area would contain organisations making a loss. Commercially, a too decentralised compromise environment will most likely not lead to a wide and positive acceptance of the proposed compromise.

Moreover, it is necessary to remember that despite its value, the processing of health data creates ethic and privacy concerns that must be addressed. Health data is considered sensitive data. Therefore, processing of such data is prohibited unless some specific clauses are respected under the influence of antidiscrimination laws in the current European legislative context.[13]

In a cross border perspective, several additional complications may arise: the insufficiently harmonised consent requirements, or ambiguous legal basis, lead to some obstacles in health data exchange within the EU.[14]

In the current legislative context, a solution could be elaborating a governance model for all ethical and privacy related aspects based on a prior study of the highest requirements at the EU level, regarding prior consent of the data subject, purposes of the data processing, recipients of data, data retention period, security and confidentiality requirements. Even though, the efficient implementation of the resulting governance model could always put local and national specificities and patients in a delicate position of the considered ecosystem.

Figure 2: Illustrative balance framework

Putting the patient in full control of their data, as the central piece of the ecosystem, and allow him/her to choose to disclose the data s/he wants to the participants s/he wants is a way to efficiently circumvent most of the legal issues related to privacy and data protection.

The following section will detail the various aspects that need to be taken into consideration for such a European capability with the foremost intention to provide the patients full control over the access and use of their own health data.

13 Directive 95/46 article 8
14 Opinion of the European Data Protection Supervisor on the data protection reform package (www.edps.europa.eu/)

3. Barriers to the system – the market model

Existing publications are usually focused around two different types of barriers to the introduction of an electronic healthcare system:

› **Data management:**
  focusing on storage and standards

› **Access management:**
  focusing on use cases

There is however a third fundamental dimension that should be taken into account:

› **Market model:** what market design is needed in order to lift existing barriers and to create a functional ecosphere for all stakeholders?

One could basically consider both, monopoly and competition based market approaches, at either a national or pan-European level. Looking at three major success drivers, a competition based pan-European market design seems the most adequate. A competition based on model would gain best acceptance by the majority of stakeholders.

Since contribution and collaboration of stakeholders is key, their individual influential potential should be equally spread. That is best achieved in an open market that develops optimal outcomes through competing offers. A monopoly on the other hand could lead to the impression that the

potential is not equally spread and thereby lead to a blockading attitude in some of the stakeholders.

A competition environment is also expected to develop quicker than a pre-defined monopoly market design, and it leaves room for opportunities to be explored. Speed of development is key for market players in order to gain first mover advantages.

Finally, and since the competition market design appears to incentivise the market participants best, market scale could be a decisive criterion in order to open up investment opportunities and attract new entrants. Therefore a pan-European market that allows players to act on an international level is expected to raise a higher amount of sponsoring investment funds than localised solutions within the EU countries.

	Monopoly		Competition	
	Local	**Europe**	**Local**	**Europe**
Stakeholder management	⊖	⊖	⊕	⊕
Speed of development	⊖	⊖⊖	⊕	⊕
Sponsor	⊖	⊖⊖	⊕	⊕⊕

Figure 3: Assessment of market models

After defining our innovation's core - patients' data - we consider now what commercial / market model could bring such ideal to life. When looking at the core requirements previously discussed, a striking similarity to the banking sector becomes obvious.[15]

In order to give data control to the patient s/he needs to obtain fundamental data management rights and opportunities. If compared to a bank account, the patient should be able to put his/her data into a dedicated account from where it is managed. While many existing PHR approaches focus on one centric national or European PHR database, this may not be the only working system.

Very much like in private banking, such accounts could be offered by a variety of institutions, as well as completely independent new bodies which could be found in the market. This may even include government bodies that could be perceived as a commercially neutral entity within the competitive environment (preferred by those who consider their data as highly sensitive or even a basic startup solution for the newborn). The striking aspect of such system would after all be the characteristic of competition.

If data is an asset and has a value that belongs to the patient, he/she should receive compensation for granting access to certain areas of that asset – just like earning interest on savings. Compensation however is not necessarily monetary – it might just be the contribution to something that is perceived as ethically or morally beneficiary. Consequently competition around personal data storage accounts could arise in a number of ways, the most striking being a monetary compensation, i.e. interest. Account providers could furthermore differentiate themselves through beneficiary supplementary services, like better or additional healthcare plans and options.

Competition – under the regulatory constraints of data protection and access regulation as described – would deliver a self-regulating stimulus to the market. Patients would deposit data where they feel safe, fairly compensated, where they can manage their data transparently and with an ease of use.

If not satisfied there would always be the option of data withdrawal and transfer into other accounts or potentially even back to offline files. In the context of the processing of health data with potential major risks for the privacy of the patients, the compliance to privacy regulations and the implementation of maybe even highest standards should be considered as a competitive advantage. According to the UK's Data Protection Authority (ICO) and its Commissioner Cavoukian *The 'payoff' to privacy- respecting organisations is (…the…) ultimately, enduring competitive advantage. In a world of increasingly savvy and inter- connected customers, an organisation's approach to privacy may offer precisely the competitive advantage needed to succeed.*

Regarding other forms of compensations and compared to other data driven business models like Facebook, where customers give away personal data and receive a limited personally perceived value out of social networking, the benefit in the PHR market model is transparency over the actual value of different data sets. For example, a patient may receive an offer to disclose certain blood values in participation of an academic study on HIV in return of a monthly monetary refund. While accepting such offer the patient may still choose to refusing participation in another study on brain tumors that offers a monthly refund of less value. Personal data would now become individually compensated by its personal value – whatever the compensation scheme may be (not necessarily monetary as in this example).

It is now essential to consider how such market could be structured. Basically there are three levels in the value chain (Figure 4).

a. Infrastructure

Any PHR solution that offers interconnectability between institutions requires a fundamental infrastructure to build upon. At this level, competition is not an indispensable requirement. As in many other infrastructures like railway, highways or even telecommunication copper networks, one shared network is theoretically enough to serve the population. What is crucial though, is a strong regulation and standardisation in order to achieve

a common level of data protection as well as the technical capability of interconnection between all participating players.

b. Platform providers

On top of the infrastructure, various platform providers could provide data storage accounts, as well as the user interfaces that allow data management. This is basically where the idea of a banking-like model would develop its full potential.

c. Analytic insight generators

In order to create benefit on the basis of PHR data it must be shared (to the extent allowed by its owners) with third parties. These are various like insurance companies, universities, research organisations, pharmaceutical companies (…) (Figure 1- Ecosystem). They would interact with the patients through the platform providers who act like a broker. The patient still decides what data to distribute to which institution and whether it is supposed to be anonymous or personalised, always considering the individual added value that is offered in return.
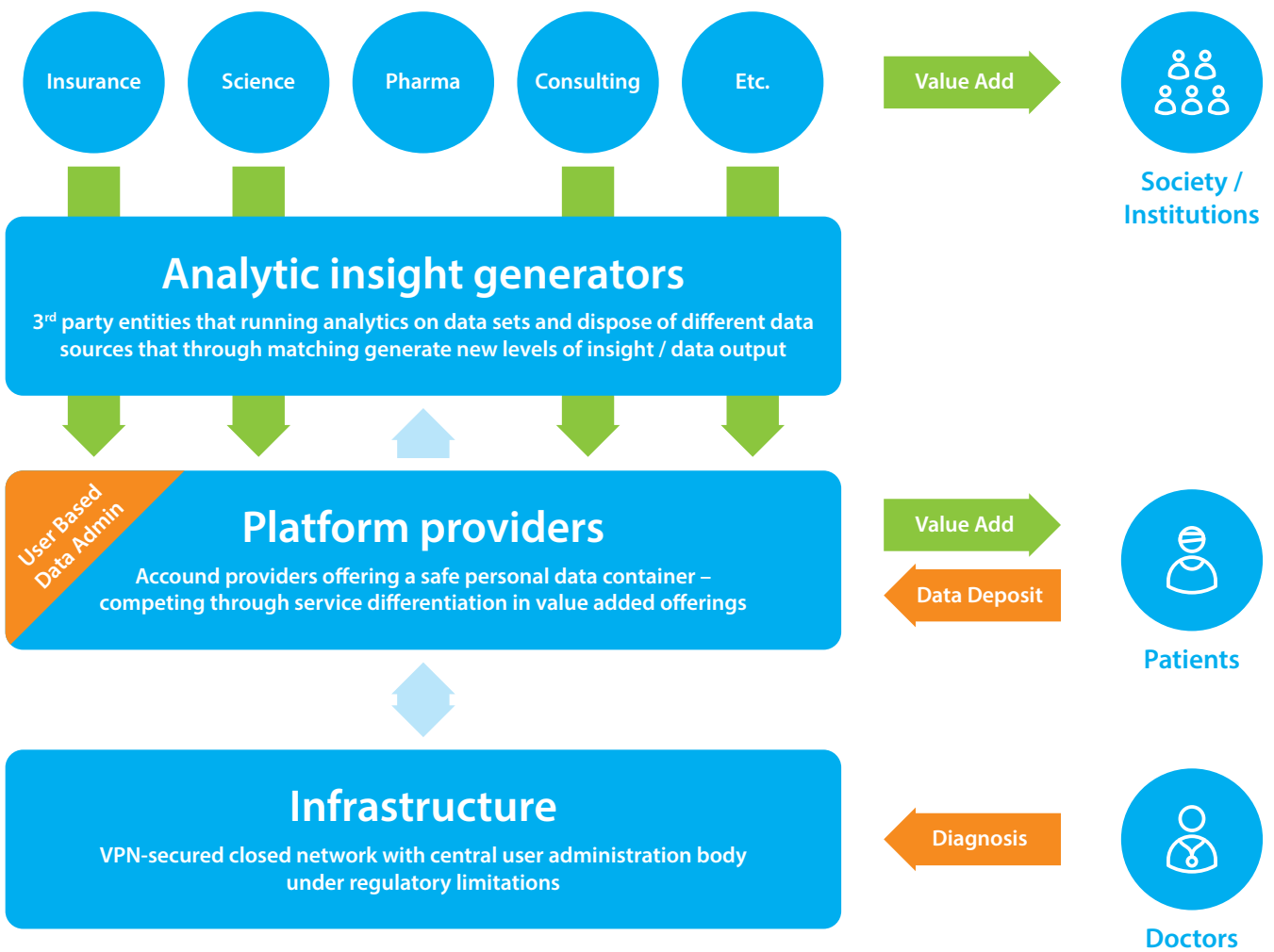

Figure 4: Market scenario based on a "Banking Model"

4. Building trust: how can we put patients at the center

To be able to put the patient at the core of the system it is important to create a platform giving access to patients and doctors with a specific set of key features (Figure5). These features are mandatory and should be considered as an intrinsic part of any platform, in order to guarantee patients have full control over their own data. They include: emergency access rules (a); access rules (b); doctors own their entries (c); hide data (d); data portability (f); withdrawal (g) easy and secure access (f).
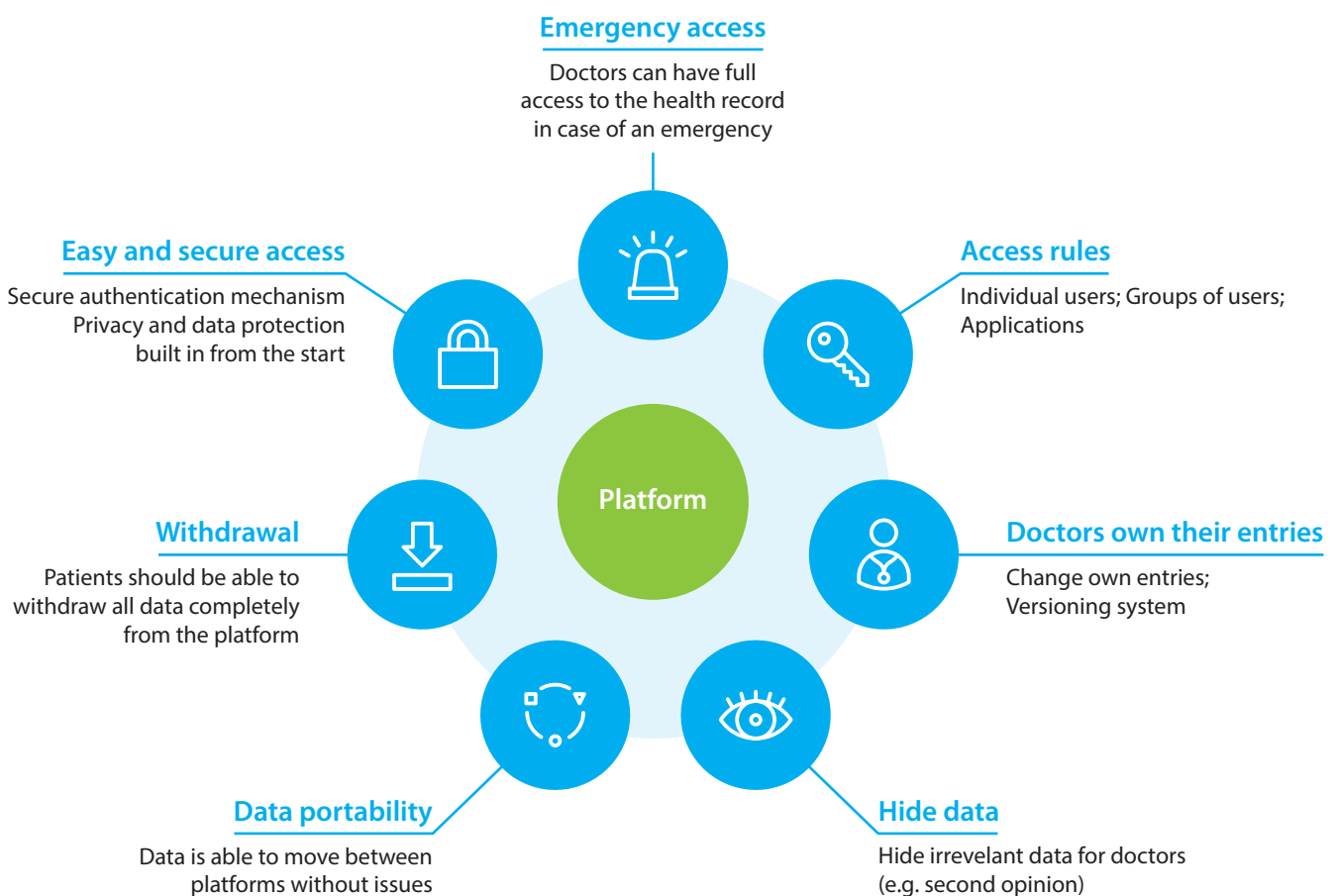

Emergency access
Doctors can have full access to the health record in case of an emergency

Easy and secure access
Secure authentication mechanism Privacy and data protection built in from the start

Access rules
Individual users; Groups of users; Applications

Platform

Withdrawal
Patients should be able to withdraw all data completely from the platform

Doctors own their entries
Change own entries; Versioning system

Data portability
Data is able to move between platforms without issues

Hide data
Hide irrevelant data for doctors (e.g. second opinion)

Figure 5: Intrinsic platform's features

In this case, there should be a possibility for the patient to verify which kind of information was checked.

However, in the current legal framework such access could only be possible under very strict conditions. As mentioned, Article 29 of Data Protection Working Party (a group gathering all European data protection authorities) does not consider that a general consent can be sufficient to allow the processing of health data or any data considered sensitive data under the Directive 95/46.

The proposal for the adoption of a new regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) drafted by the European Commission could offer an opportunity to find a solution.

Concerning health data, this draft regulation aims to provide further possibilities on the processing of personal data if justified by a number of legitimate reasons for the benefit of individuals and society as a whole (…)[16], in particular in the context of ensuring continuity of cross-border healthcare. Therefore, even if some commentators regret that this Regulation does not go far enough on the way to harmonisation[17], it seems like the right tool for the regulators to allow such emergency access (and more globally to think about the potential and barriers of big data in health care in general). Therefore the ongoing debate on this new regulation should be attentively followed in order to anticipate any change of the regulatory framework in the near future.

b. Access rules

In order for any third party to use or read the patients data (or part of), it will be necessary that the user gives his explicit consent to such access. This access will always have to be restricted in time depending on the purpose for which it is granted by the patient to the concerned third party. By doing this, the patient's data is never longer exposed more than necessary.

Granting a general practitioner reading rights would be an example of granting an individual user reading rights. When going into a hospital and seeing different departments and specialists it should be considered to grant reading rights information to the whole hospital during the time of a visit.

Applications that rely on data they collected from users to provide a useful user experience should also be allowed to read data from the patient's medical record.

In all hypotheses, clear information shall be given to the patient in order to put him in a capacity to assess the situations in which somebody will be able to access to his data. As recommended by the Article 29 Data Protection Working Party[18], no general agreement of the patient could be considered as valid and the individual concerned shall be given, in a clear and understandable manner,

accurate and full information of all the relevant issues, in particular, the nature of the disclosed data, the purpose of the access to the data, all categories of persons able to access the data and possible transfers.

c. Doctors own their entries

When a doctor adds information to patient's file s/he will control the entries. The Doctor can change them whenever s/he wants. According to the health standards, there would also be a versioning system that holds all versions of data entered. If necessary, the different versions of a diagnosis can be tracked.

d. Hide data

Patients may choose if they want to hide sections of their PHR. For instance, when going for a second opinion to a new doctor it might be advisable to hide some parts of the patient record.

e. Data portability

When users have the ability to withdraw their data it is logical that this data should be easily portable. This will add the benefit that a user can move their data without much effort from one service provider to another.

f. Withdrawal

In order for the patient to have a genuine free choice and be able to withdraw his/her consent without detriment, the patient shall be clearly informed and regularly reminded of his/her rights and notably of the 'right to be forgotten' and 'right to delete'.

If a patient decides to withdraw his/her data from any platform it should be possible for each category of access rights s/he has been granted.

g. Easy and secure access

From a legal/data protection perspective, when dealing with data processing with such risk potential, the compliance of the system is extremely dependent on a high level of

data security. The European data protection authorities require demonstrating that any access by unauthorised persons is virtually impossible and prevented.

The way of accessing this data should happen in the most secure way, while at the same time the system should be unlimitedly available for authorised professionals within the limits of the rights they are granted. A possibility to reach high security standards could be to give chip cards to patients with a type of security level currently used by online banking. It should be at least a two or three factor authentication.

Every action that is performed on the patient's record would have to be recorded in an access log giving the possibility to check if needed who had access to the patient's data at a certain time. Erasing this log should be possible.

Privacy and data protection concerns should be embedded from the early stage of conception to their implementation, use and disposal. Such concept usually referred to as *privacy by design* should be applied throughout the entire life cycle of the system. The aim is that security and privacy would become the default setting of the system, preventing privacy risks instead of remedying to it once they already happened.

5. Sharing the benefits-perceiving the value

One of the root causes of the non-adoption of global PHRs by patients is the lack of perceived benefit. Privacy issues in other domains have shown how the cost of losing privacy can be balanced by the value of perceived benefit. For instance, while strongly rejected at first, video cameras in the street have been widely accepted by the general opinion because of the increase of perceived safety. Therefore when creating a system to enable PHRs at a global level, the focus should not only be put on reducing perceived cost but also on increasing perceived benefit.

16 Whereas 122 and 123 of the Proposal for a Regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [2012/0011 (COD)]
17 Opinion of the European Data Protection Supervisor on the data protection reform package (www.edps.europa.eu/)
18 00323/07/EN WP 131 Working document on the processing of personal data relating to health in electronic health records

While evident when stated as such, this statement does not seem to have been guiding previous attempt in building efficient PHRs. In these systems, a lot of effort has been introduced in reducing the financial costs, the time required by the patient to import data and to use the system, etc. The usual assumption is that the perceived benefit does not need to be increased because having data centralised is a benefit as such and indeed, most stakeholders know how a centralised database would improve their practice if they had access to it: physicians would be able to make better diagnostics and prescribe more efficient medications; pharmaceutical companies would be able to improve their drugs; health insurance companies would be able to improve cost-effectiveness of treatments, etc. But while these benefits will ultimately have positive consequences on the patient, they are not sufficiently perceptible. One might argue that improving diagnostic and medication is a very direct benefit to the patient, but patients (at least in European countries) currently perceive treatment as being the responsibility of the healthcare community. Once a disease is diagnosed, the patient relies on his/her physicians and the clinical community to treat him. Therefore an improvement in diagnostic and medication while directly benefiting to the patient, actually internalises a problem which otherwise is external (responsibility of the physician).

In order for the patients to completely endorse a global PHR, they must receive a share of the benefits. All stakeholders who see a benefit in centralised data should find a way to share benefits with the patients in a direct, individualised, measurable and possibly short term way in order for them to be well perceived. In this case, the "banking model" described previously could meet the needs of this benefit sharing. Indeed, as it opens the way for third parties to directly interact with the patients, deals concerning data could be proposed to the patient. As this would be done in individualised way, the value could be tailored to the needs of the individuals and not enforced to the whole population.

Several categories of benefit for the patients could be envisaged.

Short term, very individual and direct benefits such as:

› **Increase in the ease of access to the medical community:** this includes the possibility to do online or over the phone advices and prescriptions, the possibility to get advices anonymously, etc.

› **Increase in provider choice:** this includes the possibility to get advice from multiple physicians and to change more easily from one to another or from one treatment option to another.

› **Money incentives:** this is the most obvious benefit when thinking about the banking model. Data could be sold for research studies, they could be sold to insurance companies for premium rebates,…

Mid-term, potential benefits for individuals such as:

› **Health monitoring:** while treatment is seen as the responsibility of the health care community, good health is usually seen as the responsibility of individuals. Centralising health related data could allow personalised analytics, prevention and behavioral change propositions.

› **Increase in safety:** this includes the possibility to give easy access to health providers in emergency situations, to have all data available when abroad, etc.

And long term global benefits such as:

› **Greater good:** the possibility of choosng which to which research studies the data will be given to, much like fundraisers exist today.

Some of the above benefits could be possible even without adopting the "banking model" system; nevertheless, it will enable new means of proposing value to the patient. The working group's recommendation is that by giving full access control to the patients in a competitive market, they will perceive these benefits more than in other systems. The proposed benefit categories are not meant to be exhaustive. The banking model will allow creativity and innovations to emerge from all stakeholders enabling new ways of interaction with the patients.

6. Implementation steps

The implementation of the suggested solution largely depends on the establishment of an appropriate regulatory framework. As suggested, the adequate market design with its stimulating incentive mechanisms is crucial in order to unlock the potential of a variety of use cases within the ecosystem (each requiring an individual implementation).

Since the suggested market design requires a regulatory framework it is now recommended for the EU regulatory bodies to evaluate the institutionalisation of such scheme. Political parties, member state representatives, market / business and stakeholders need to be heard in order to initiate the decision making process that could lead to such regulation. The political dimension of this implementation process is not within the scope of this document and should be tackled in-depth in a separate evaluation.

7. Conclusions

Currently big data does not efficiently contribute for improving healthcare because **patients are not in control of their medical health data**. The "personal data currency" is not flowing **making impossible to build the trust that is needed**.

Health is central in people's lives. This increasingly perception may be seen in the number of app downloads directly or indirectly related with healthcare.[19]

The group's recommendation is to create a **bottom-up model**, focusing on the core of all the innovation, the **patient's data - from an exclusive to an inclusive ecosystem**.

The "zamboni effect"[20] is a paradigmatic example of the actual exclusive healthcare

---

19 In Google play store there are more than 100 apps related with Healthcare. Just five apps have more than 20 Million downloads (twice the population of Portugal).
20 New York Times: A Controversial 'Cure' for M.S; Paul Tullis; October 26, 2012

system patients' perception. Promoting **transparency between doctor-patients** relations and implementing the **presented "banking model system"** will promote **patient trust and value perception**, which are key features to successfully implement a new **competitive model**. It is necessary to enable individuals to understand and manage, in a secured environment access to allow the use and valuation of such individual information in an inclusive ecosystem, where the **interests of the different stakeholders, the fundamental rights of the patients and the legal constraints are fully taken into account and respected**. As patients are the primary source of PHR data, it is essential to design a user experience that is well accepted. A **chip card model** similar to a bank card is proposed **to materialise the digital data with the physical object**.

 The PHRs may include a range of data, including demographics, personal and family medical history, medication and allergies, chronicle diseases, immunization status, laboratory test results, radiology images, vital signs, personal stats like age and weight, smoker status (…). In order to have a fully EU wide system, the **standardisation**[21] codification and the **centralisation**[22] level to be achieved are vital to have a reliable and on-time data.

The recent evolutions of the debates on the new EU regulation on the protection of personal data and notably the draft report issued on January 16, 2013 by the Committee on Civil Liberties, Justice and Home Affairs[23], should be reinforced and seems to follow the same idea when insisting on the fact that *"health data, which is extremely sensitive, may only be used without the consent of the data subject if it serves an exceptionally high public interest and in this case must be anonymised or at least pseudynomised using the highest technical standards"*.

The previous recommendations are fully aligned with the core principles and strategic actions of the EC Health Strategy:

Solidarity – Understanding patients in 360º will improve transparency between the doctor-patient relations, promoting health throughout the lifespan. A segmented monitoring and advanced statistical approach to patient profiles would help to identify those that are in risk of developing a specific disease linked to social, economic or environmental factors. It will open new windows to customise actions- "tailored solutions"-identifying the best treatments for specific patients and what the most effective preventive actions are, tackling inequities in health.

Security – Pandemics or major physical and biological incidents and bioterrorism pose major threats to health. Interconnected and standardise codification of symptoms, prescriptions, exams (…) will promote an on-time and reliable access to data, enhancing EU Community's coordination to respond rapidly to health threats globally.

Prosperity – Capturing the PHR data will accelerate the development of new drugs. Imagine that a statistical population having a chronicle disease (e.x diabetes, depression, hypertension), take the same active pharmaceutical ingredients (API) for years but none of them have a specific type of diseases. This type of information will be very important to understand expanded or adverse effects that some molecules or analogues may have helping the R&D development. The proposed healthcare competitive model will promote health with the development of new technologies, namely, information and communication technologies (ICT), genomics, biotechnology and nanotechnology fostering a competitive and sustainable future for Europe.

A EU Governing Body should drive the gradual implementation of these solutions.

---

21 This topic is discussed throughout several projects Eurorec, Ramit and OpenEHR foundation SemanticHealthNet and EHR4CR
21 Towards a sustainable data warehouse approach for evidence based healthcare, N. Stolba, PhD thesis, 2007. An example of such system is the American healthcare insurance companies and Health Care Cost Institute that were merged and provide data to research institutions to reduce treatment costs.
23 Draft report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))

Getting prepared for a globalised digital world: The value of big data in Higher Education

Nadja Dokter

Kate Hofman

Julia Manske

Richard Ramsden

Halit Ünver

Abstract: *Today's Higher Education institutions are challenged by rapid advancement of new technologies, labour market demands for more entrepreneurial graduates, increasing global competition for attracting the best talents and individuals who expect to be able to work, learn, and study whenever and wherever they want to. Leading Higher Education institutions are only slowly responding to these challenges, largely failing to leverage the enormous potential of promising emerging tools and methods which have the potential to make education more relevant and fit for educating students in a globalised digital world.*

This paper describes how big data can be utilised to promote individualised learning, improve teaching, as well as facilitate cross-organisational mobility in times of multiple options. The education team of the Young Leaders Group proposes an integrated solution - the World-wide Higher Education Network Portal – to deliver benefits to students and universities alike in a globalised digital world in particular focusing on the two aspects of learning improvement and mobility.

1. Key trends and challenges - getting prepared for the globalised world of complexity

Today`s world is characterised by global interconnected economies, mobility, volatility, and rapid advances in technology. Innovation, which fuels new job creation and economic growth, is driving global competitiveness and the effective creation, use and dissemination of knowledge is the key to success. The use of the internet and its applications continue to change the way we live, learn and interact with each other. In fact, the requirements on and demands of individuals set by these dynamic conditions have dramatically changed throughout the last decade. This affects all facets of academic education.

Employers - public and private - expect young talents to have a combination of subject-specific knowledge, as well as transversal skills and competences. Individuals who can apply their skills directly to projects, who can create added value by developing new processes, products or organisational forms are more and more valued. Intra- and entrepreneurship becomes a key competence. Employers are putting more and more emphasis on what potential employees know, understand and are able to do in practice, rather than concentrating on formal qualifications, moving from an occupation-focused to a skills and competence-oriented approach promoting employability.

Higher Education (HE) institutions are increasingly challenged to cope with and apply the enormous existing and emerging technological options in order to provide and transfer such skills. Nowadays, people expect to be able to work, learn, and study whenever and wherever they want to in order to balance demands from home, work, school, and family which pose logistical challenges. This requires access to information, tools and resources online and the proliferation of virtual platforms to learn and share information. These needs, as well as the increasingly essential access to social media and collaboration networks, have risen to the level of expectations.[1] Individual lectures will slowly disappear; academics become facilitators of learning rather than direct sources. In such a world of abundant information and relationships made easily accessible via the internet, sense-making and the ability to assess the credibility of information are paramount. Mentoring and preparing students for the world in which they will live and work is becoming a crucial element of education. In addition, learners expect up-to-the moment analysis and commentary, as well as the possibility to acquire knowledge and skills the way which best fits to their personality and learning styles.

While not so long ago, a student`s learning journey was quite predictable, with relatively few options or alternatives to choose from while completing a study programme. Nowadays students are required to undertake internships and study periods abroad, engage in extra-curricular activities and focus on those subjects relevant for their future job. In addition, group work and communication is becoming more and more important and students are increasingly evaluated not just on the overall outcome, but also on the success of the group dynamic.

Leading HE institutions, which are preparing individuals for today`s knowledge society and are providing personalised development and highly qualified people for economies around the globe, are gradually moving away from traditional input approaches, such as front-teaching and localised learning, to a

---

1 NMC (New Media Consortium) (2012): Horizon Report 2012.

more learning outcomes approach driven by such demands. As technologies such as tablets and smartphones have proven useful in the higher education context, educators are leveraging these tools, which students already use, to connect the curriculum with real life issues. Students already spend much of their free time on the internet; learning and exchanging new information — often via their social networks and intuitions can capitalise on the online skills learners have already developed independent of academia. The developments in online learning offer different opportunities than physical campuses, such as increased collaboration while equipping students with stronger digital skills. A combination of face-to-face and online learning, when designed and implemented carefully, will enable students to travel to campus for some activities, while using the network for others, profiting from the best of both environments.[2]

However, all this is challenging most European academic institutions as mastering these developments goes hand in hand with a huge complexity and high administrational efforts. Most institutions are far away from making use of the promising opportunities and there is still a long way to go to explore, catch up and invest in order to prepare today's students for a globalised digital world. Closing this gap is one of Europe's vital tasks.

2. Addressing the challenges - what's data got to do with it?

To better tackle the challenge of meeting the new society's needs, Higher Education institutions need to gain a better understanding of the demand of their "customers". HE institutions are already collecting an array of data about students' individual choices on subjects, learning paths, evaluations, grades, mobility and many more. Especially the introduction and proliferation

of virtual/e- learning environments, Content Management Systems (CMSs) and Management Information Systems (MIS) has led to a substantial increase in available data.[3] An increasing number of student activities are digitalised and are stored in students' identity cards and data processing centres of Higher Education institutions, which enables the tracking of interactions and transactions. As more learning moves online, the amount of data that is produced will exponentially increase - each click leaves a trail of data, which may be analysed.[4]

At higher levels, data about curricula, accreditations, new learning methods and material is also available. In Europe, the Bologna process has made educational systems become more comparable, compatible and coherent systems, in which information about the content of courses, lectures, exams, degrees, credit points, additional qualifications are made available and are stored in a repository at university level.

While in other areas such as marketing, medicine, physics and biology the value of these data-sets has already been identified and – even more importantly – is being utilised, educational institutions have been inefficient in their data-use, often operating with substantial delays in analysing readily available data and feedback. Organisational processes, such as planning and resource allocation, often fail to use large amounts of data on effective learning practices, student profiles, and needed interventions.[5]

This data can be used to gather valuable insights to make Higher Education Institutions more efficient and effective in their delivery, using a set of analytical methods and tools. For instance, the data mentioned above can be used to the benefit of learners to improve their learning experiences and support personalised learning, improve teaching, as well as facilitate cross-organisational mobility.

To improve the quality, effectiveness and

efficiency of the learning process a new promising field called "Learning Analytics" is emerging.[6] Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts for purposes of understanding and optimising learning and the environments in which it occurs.[7] It combines advances in data mining, interpretation and modelling to better understand teaching and learning and tailor education to individual students more effectively. A combination of the availability of big datasets, the emergence of online learning on a large scale and political concerns about educational standards has stimulated the development of this field.[8] The main opportunities for Learning Analytics are to unveil and to contextualise hidden information out of the educational data and prepare it for the actors in the educational process. In practice, learning analytics can be applied in different ways, i.e. through the use of e-textbooks an analytical tool can track a students use of the textbook, including page views, time spent on a page, highlights made and notes taken and may notify the professor of each student`s engagement. Using data from course management systems to determine whether the use of an online grade book can serve to help lower achieving students estimate their final grades more accurately. Learning Analytic tools can be applied to support peer-to-peer instruction and provide real-time feedback during class.[9]

Similarly, data can also be used to facilitate cross-organisational mobility by applying the method of matching. Up to today every student who wants to gain international experience has to search for partner universities across Europe for his special study and has to apply to programmes like Erasmus for international exchange. A matching system could provide a service for students by delivering e.g. every semester exchange opportunities with other students.

Matching in general is a very prominent method when comparing and fitting data or allocating resources. Research that

---

2 NMC (New Media Consortium) (2012): Horizon Report 2012.
3 Ferguson, R. (2012): The State Of Learning Analytics in 2012.
4 Watters, A. (2012): Top Ed Tech Trends of 2012.
5 Siemens, G.; Long, P. (2011): Penetrating the Fog: Analytics in Learning and Education.
6 Greller, W.; Drachsler, H. (2012): Translating Learning into Numbers.
7 As defined by the Committee of the 1st International Conference on Learning Analytics and Knowledge, see i.e. http://www.learninganalytics.net/?p=126
8 Ferguson, R. (2012): The State Of Learning Analytics in 2012.
9 NMC (New Media Consortium) (2012): Horizon Report 2012.

demonstrated how to find optimal matches between people or institutions 'trading' in commodities that money can't buy, such as for example student placements, has been awarded the Nobel prize in Economics (Lloyd Shapley of the University of California and Alvin Roth of Harvard University in Cambridge).[10] Allocation mechanisms should be efficient if rational people are acting, who know their best interest and behave.[11] In case of students' mobility and experience, students sometimes don't really know what could be the best or rational decision with regard to their individual learning path. Matching students with students according to their learning ability to build diverse learning groups, exchange university locations for a semester or to match a student with internship places are some practical examples of how to apply matching in the Higher Education context.

3. Benefitting from the data - a big data solution

Data helps to cope with the upcoming complexity. Existing technology makes it possible to analyse huge amounts of data synergising and translating it into actionable knowledge. As mentioned, the underlying data sources and applicable methods to add value to Higher Education already exist, but they need to be leveraged and the above-mentioned examples give an indication of the potential benefits.

In order to enhance these benefits, especially for the students who are main customers of Higher education, the following section presents a possible futuristic scenario of how big data could be used to improve the teaching and learning process and enhance cross-organisational and international mobility. This scenario relies on the assumption that data sources from different Higher Education institutions in Europe and beyond are combined in one platform/data base. A single platform which could provide information which is currently not available due to missing capacity and missing capability (especially when it comes to information shared globally). We,

the authors of this proposal, believe that to make the most of big data it is crucial to combine all the generated data to be used by one service - a so-called "Worldwide Higher Education Network Portal" - the WHENPortal.

The WHEN-Portal

We believe that in order to cope with the challenges of our future, universities have to adapt to the upcoming trends and must understand the need of revising their self-conception. This includes providing students with skills they actually need for future jobs and which are crucial for a mobile and complex future. The following technical solution follows this mission giving the university the opportunity of becoming a better service provider meeting their "customers'" demands and needs. However, this solution does neither imply to replacing universities nor universities' employees, but rather presents opportunities for simplification of complex administrational procedures, as well as benefits for optimising the learning and teaching process. This would allow universities to focus on students instead of focussing on administration.

The portal describes two specific user case scenarios rather than the technology itself.

The WHEN-Portal for teaching and learning

One aspect of the portal is the application of Learning Analytics to improve teaching and learning:

With real-time analysis of student behaviour with regards to interacting with the institution's VLE (Virtual Learning Environment), library and online learning resources and attendance records, the WHEN-Portal can provide both teachers and learners with advice on learning patterns. An administrator would be able to highlight where a certain student looks set to follow to the most common path towards failing or dropping out, and the WHEN-Portal provides the platform for informed and effective intervention for the student:

"In their second year of study, students who attended a similar number of lectures to you, and also only logged on to the VLE once a week were attaining an average of 55% in their final results. By attending an additional seminar each week, you could change your learning profile to match that of a student who is likely to achieve 65% in their final results."

…or for the teaching staff:

"Student X did not attend your lecture on The Tempest and hasn't accessed any critical theory on the play through the library or online resources. Would you like to send them a personalised recommendation they attend Thursday's seminar on Colonialism and the role of the other in Shakespeare's plays?"

The benefits for students are to support them in the specific learning process through "reflective" visualisations that compare their performance to the overall performance of a course. In addition, they can be provided with personalised recommendations for suitable learning resources, learning paths or peer students. This could help them identify where they might want to add additional learning into their curriculum to develop specific skills that they aren't acquiring.

Teachers can use course monitoring systems to inform them about knowledge gaps of particular pupils, enabling them to focus their attention on those pupils and thereby reducing the risk of dropouts. They can also yield information from groups of students to better adapt the curriculum to their needs, such as focusing on topics which students do not yet master, and more.[12]

The portal may offer some opportunities for students to tailor their learning to personal preferences, but its value for teaching and learning is much more about enabling the relationship between teachers and students to ensure that everyone gets the most out of the learning experience.

The WHEN-Portal for cross-border and cross-sectorial mobility

10 http://www.nature.com/news/a-nobel-for-the-art-of-matchmaking-1.11607
11 http://www.nobelprize.org/nobel_prizes/economics/laureates/2012/popular-economicsciences2012.pdf
12 Greller, W.; Drachsler, H. (2012): Translating Learning into Numbers.

The WHEN-Portal not only provides students and employers with the above mentioned skills, it can also help students identify exchange or internship opportunities that might be of interest to them. Acting as a marketplace, the portal would be the first place the students would look to understand what skills and experience employers were looking for, and by understanding the career paths of previous students, could make informed decisions about their education (for example in the UK, *"It is worth me accepting an offer from Newcastle University and paying an extra £3000 per year for tuition fees because with my degree I'll be able to command a 10% higher starting salary."*).

Improved mobility would also empower students to take advantage of international partnerships and collaboration. By providing up-to-date visualisation of research and administrative activities, students could identify options for placements at other universities with shared research aims. Going beyond current information sharing through publication of results, the WHEN-Portal would let students find out existing research projects that might correlate with their own work, and choose to take some credits for their study at that institution. In an extension of the Erasmus programme, European collaboration between institutions would become flexible and intuitive – so that both students and institutions could get the best results.

Similarly, employers could create "employment hotspots", geographic locations where specific skills or knowledge were required, and students would be able to view those hotspots and see if they meet the requirements.

The international exchange of students is facilitated by the European Credit Transfer System (ECTS), in which a student gets credit points for his/her courses, exams and additional lectures s/he is passing during his study.

From a big data perspective every lecture, exam, degree, credit point and other data is stored in a repository at university level. A to be defined matching algorithm could be used to match students across Europe, in an ideal case a perfect match, by merging all data from every student in Europe in one big database. The algorithm would search for best matching mechanism for students across Europe to enable them to "swap" their roles. In case of 'perfect matching' two students are academically equal. The advantage of this "real" swapping is that the HE institutions do not need additional capacity for international exchange students. The second advantage of this approach is that the EHEA (European Higher Education Area) process will be standardised much more efficiently, if a system or database models all relevant information of students, syllabuses and module catalogues. The compatibility of ECTS points of modular lectures and courses will be reproduced by the system. Matching is defined first by academic similarity and second by their personal interests to enable also social integration in the partner country. A Facebook example is *"May be you know…"*. The same process could be used to improve internship programmes. Matching students with job or internship positions of companies by offering a big data perspective of their students in common and for every student at individual is also crucial in future for HE institutions. Companies are able to pick out students for a special purpose or position, and students' applications are additional ways to get in touch with companies.

As we see a continued diversification in the provision of educational content, teaching and assessment, the role of the institution becomes even more central in pulling together all the relevant stakeholders to ensure a holistic presentation of higher education for students. The WHEN-Portal creates a community of learners who can easily and comfortably access the information they need to succeed at their studies. The portal captures and processes online and offline student learning behaviour, giving institutions valuable insight into what and how their students are learning and where there might be opportunities to improve the student experience.

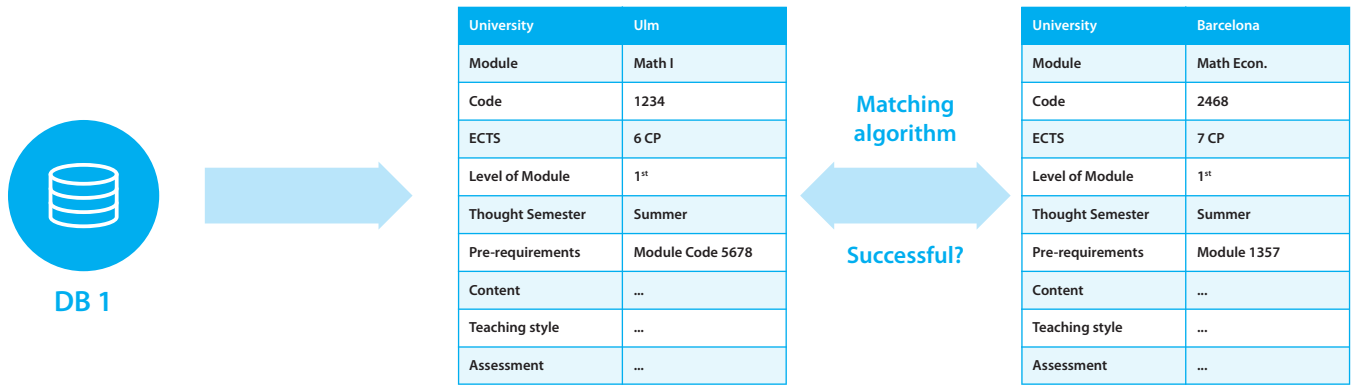## Implementation of the WHEN-Portal

Technical implementation

Through the WHEN-Portal, multiple data sources will appear to be seamlessly linked to the end user communities of students and members of HE institutions. In reality, there would be a multifaceted set of systems integration that would need to occur for the WHEN-Portal to provide meaningful information for students.

Assume Chemistry Student A at University of Ulm from Germany is in the 4th semester and has 75 ECTS-Credit points in total. All relevant academic individual data (level, module list, degree list, ECTS-List…) of Student A is stored in DB 3. A comparison of relevant data of Student A with his general syllabus stored in DB 2 provides all rest modules of Student A for successful completion. DB 2 stores also all syllabuses at different universities in Europe. So it is possible to compare which syllabus at which university in Europe fits most with the syllabus of student A. After having found the most similar few syllabuses at some universities, DB 1 compares the compatibility of the modules at the local and remote university, while language constraints has to be overcome simultaneously.

Figure 1 outlines an example of how different data sources for students at HE could be combined in the WHEN-Portal to match students for improved mobility and collaboration.

University	Ulm
Module	Math I
Code	1234
ECTS	6 CP
Level of Module	1st
Thought Semester	Summer
Pre-requirements	Module Code 5678
Content	...
Teaching style	...
Assessment	...

Matching algorithm

Successful?

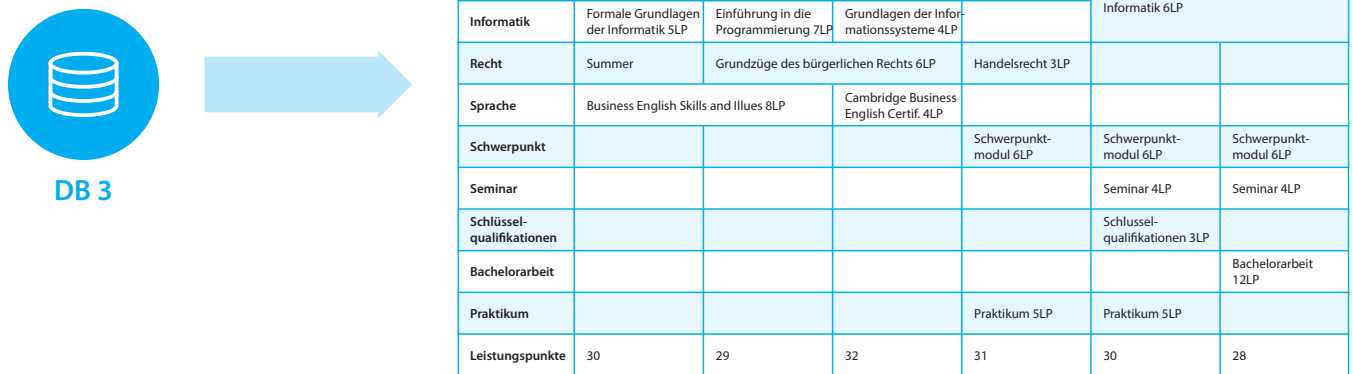| University	Barcelona
Module	Math Econ.
Code	2468
ECTS	7 CP
Level of Module	1st
Thought Semester	Summer
Pre-requirements	Module 1357
Content	...
Teaching style	...
Assessment	...

DB 1

Database 1

All lectures, courses at different HE institutions in Europe are stored with their module descriptions in DB 1. A matching algorithm is used to compare "same" modules at different places.

DB 2
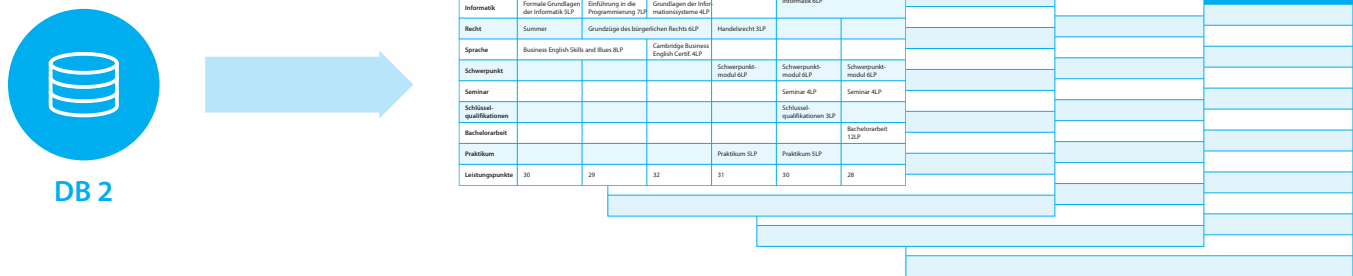
Database 2

All syllabuses for different studies at different places are stored in DB 2. A matching algorithm searches for similar syllabuses for same study at different places.

DB 3

Prüfungsbereich	1. Semester	2. Semester	3. Semester	4. Semester	5. Semester	6. Semester
BWL	Grundlagen der BWL 6LP	Externes un Internes Rechnungswesen 9LP			Wahlpflicht BWL/VWL 12LP	
			Investition und Finanzierung 9LP			
VWL	Grundlagen der VWL 6LP		Makroökonomie 6LP	Mikroökonomie 6LP		
Mathematik	Mathematische Grundlagen der Ökonomie 18LP		Stochastik und Wirtschaftsstatistik 14LP		Wahlpflicht Mathematik/ Informatik 6LP	
Informatik	Formale Grundlagen der Informatik 5LP	Einführung in die Programmierung 7LP	Grundlagen der Informationssysteme 4LP			
Recht	Summer	Grundzüge des bürgerlichen Rechts 6LP		Handelsrecht 3LP		
Sprache	Business English Skills and Illues 8LP		Cambridge Business English Certif. 4LP			
Schwerpunkt				Schwerpunkt-modul 6LP	Schwerpunkt-modul 6LP	Schwerpunkt-modul 6LP
Seminar					Seminar 4LP	Seminar 4LP
Schlüssel-qualifikationen					Schlussel-qualifikationen 3LP	
Bachelorarbeit						Bachelorarbeit 12LP
Praktikum				Praktikum 5LP	Praktikum 5LP	
Leistungspunkte	30	29	32	31	30	28

Database 3

Individual syllabuses are stored in DB 3, where all academic degrees and modules of a student are stored.
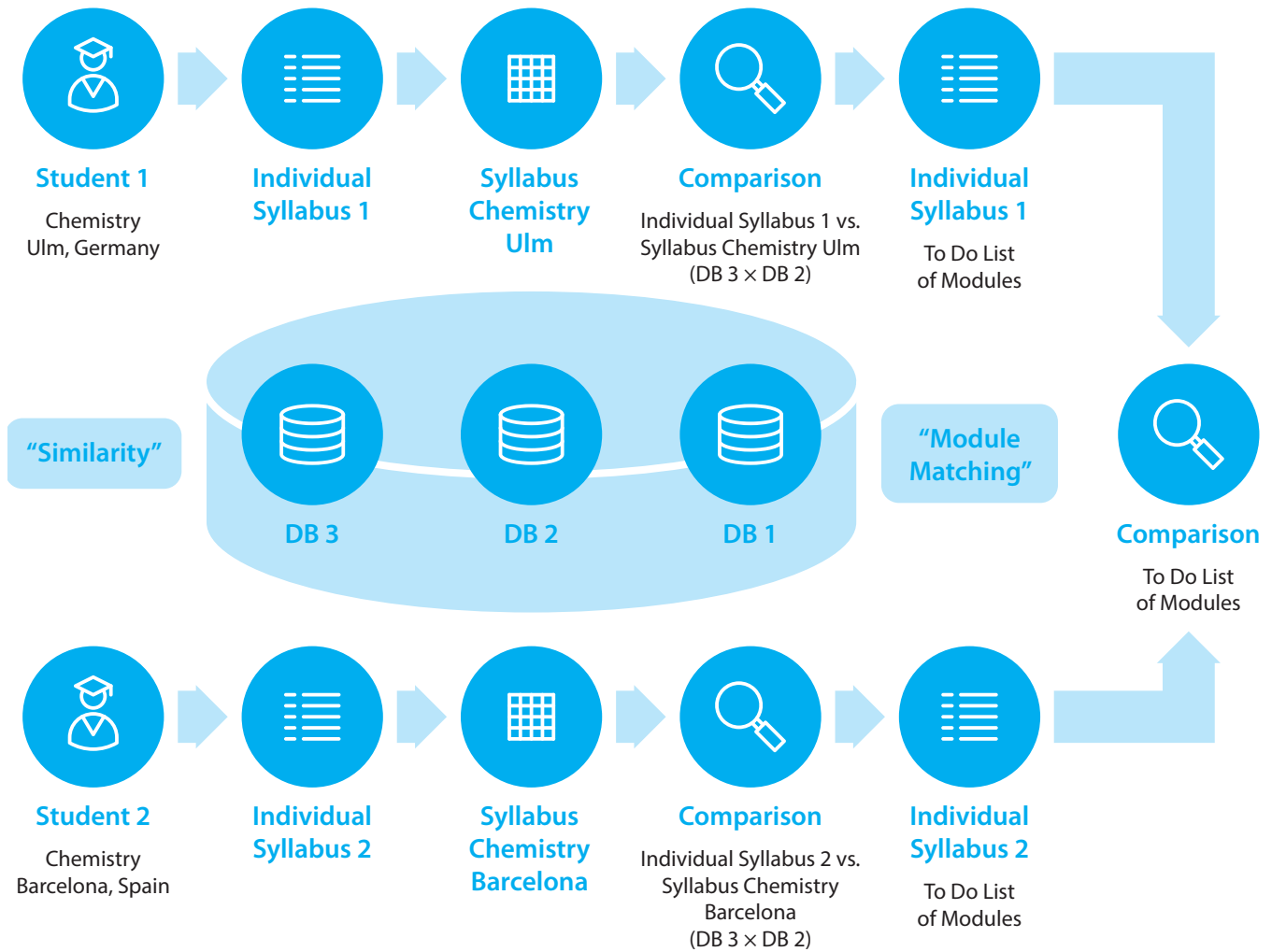
Figure 2: Example of student matching process.

Figure 2 gives an example of how to match two chemistry students.

Implementation challenges

Despite the huge promises of learning analytics, matching and learning mobility, it also raises substantial technical, privacy and ethical considerations. If the WHEN-Portal is to be effective, the following challenges need to be addressed:

The related **technical challenges** are the ones of compatibility of educational datasets, or the comparability and adequacy of algorithmic and technological approaches. It has to be identified what kind of educational data is available and what data can mean in different specific fields. Heterogeneity, timeliness, complexity and scale are some important issues for data-driven analysis to be considered. These data-driven issues have to be addressed in detail and metadata has to be explored. The whole data analysis pipeline has also to come into question. One other challenging issue is to think about how historical data about the organisational process and the individual progress of a student could be linked with degree or syllabus monitoring systems. From a big data point of view all historical data of students and their performance can be helpful to predict with data mining methods (e.g. regression analysis, clustering, classification) how future students will progress in their field of study, to give them both machine an human feedback. The algorithms which will predict results are very important and finding 'best' and stable algorithms are main technical

challenges. Standardisation and integration of different data sources and educational systems are other challenges to be addressed. However, most European universities do not yet have the required IT systems or the staff to implement this.

Privacy and ethical questions influence the acceptance and the impact of Learning Analytics. Among these are questions of data ownership and openness, ethical use and dangers of abuse, and the demand for new key competencies to interpret and act on Learning Analytics results.[1]

It is essential to ensure that Learning Analytics is not solely about extracting value from educational data and to fast-track potential insights, but to add value to improve teaching and learning. Students should not merely be the object of analysis, but have control over their data and learning.[2] This means that the implementation of Learning Analytics require careful design and implementation

to be successful and beneficial. For this a unified European regulation is required. At the moment such systems cannot be implemented and will be hampered by individual legal issues.

Awareness: In order for it to be effective, the WHEN-Portal must engage a critical mass of students and institutions so that the data it gathers and the analysis it generates are meaningful and valuable. This will require an intensive programme of awareness building across all stakeholder groups. For many first-time students, engaging in a social platform may be a natural activity, but for lifelong learners or staff, an inclusive approach to training and education will be vital to ensuring take-up and proper usage of the platform. To make this become reality, local alumni organizations should be energised to engage in future educational efforts. As examples communities of private universities could serve. In addition, national stakeholders, as local governments, ministries

or universities boards should be engaged and provided with financial means to enable e-learning and e-analytics.

4. Conclusion

The WHEN-Portal is one solution going beyond current European reality. However, if European Higher Education institutions want to remain competitive and attract and retain the smartest minds, the use of big data and analytics are crucial. There are several actors who will have to engage in this endeavour – HE institutions, national and regional governments, entrepreneurs who develop solutions, as well as institutions at European level. The first step will be to foster a wide-spread uptake of information and communication technologies by all European academic institutions to support and improve teaching and learning, as well as administrational processes. This will be a pre-requisite to develop and apply platforms as "WHEN".

5. References

Ferguson, R. (2012): The State Of Learning Analytics in 2012: A Review and Future Challenges. Technical Report KMI-12-01, Knowledge Media Institute, The Open University, UK; online available: http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Greller, W.; Drachsler, H. (2012): Translating Learning into Numbers: A Generic Framework for Learning Analytics, Educational Technology & Society, 15 (3), 42–57; online available: http://www.ifets.info/journals/15_3/4.pdf.

NMC (New Media Consortium) (2012): Horizon Report 2012. Higher education edition; online available: http://www.nmc.org/pdf/2012-horizon-report-HE.pdf.

Siemens, G.; Long, P. (2011): Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review, vol. 46, no. 5.; online available: http://creativecommons.org/licenses/by-nc/3.0/.

Watters, A. (2012): Top Ed Tech Trends of 2012: Data and Learning Analytics, Inside Higher Education, December 2012; http://hackeducation.com/2012/12/09/top-ed-tech-trends-of-2012-education-data-and-learning-analytics/

http://www.learninganalytics.net/?p=126

http://www.nature.com/news/a-nobel-for-the-art-of-matchmaking-1.11607

http://www.nobelprize.org/nobel_prizes/economics/laureates/2012/popular-economicsciences2012.pdf

---

1 Greller, W.; Drachsler, H. (2012): Translating Learning into Numbers.
2 Watters, A. (2012): Top Ed Tech Trends of 2012.

Data as a service
for social change

Chris Woods

Tommaso Alderigi

Niclas Ehn

Nina Gumzej

Pep Salas

Hjalmar Nilsonne

Abstract: *The concept of big data has failed to live up to its original promises of increasing our quality of life and contributing significantly to our economy. The value of this big data is in its statistical analysis and not in obtaining specific access to an individual's personal information. This paper proposes a Data as a Service company which has a technical approach, enabling third parties to obtain provides paid for statistical analysis of data. It is recommended that:*

› *Any additional revenue generated by such a service could be used to offset the current investment occurring in smart meters;*

› *Future EU policy considers the creation of such privately funded ecosystems;*

› *Efforts to allow the public and their legislators to recognise the anonymity and benefits of statistical analysis are encouraged.*

1. The problem – social change and data access

Big data; the storage, cross referencing, correlating and searching of vast amounts of data has been hailed as the next frontier for innovation. Claims have been made about the large leaps in productivity and social benefit which will come from the insights such data allows us to have. However, in order to get there, many challenges towards this vision need to be overcome. For example, concerns on commercial sensitivity of data, privacy and many other legal issues involved in sharing access to data may influence data owners' reluctance to provide access. In real life these data sets are not shared without a clear and mature business strategy and legal controls implemented, and thus the cross referencing needed to provide the insights is without this not possible. This in turn may cause big data analysis being reserved only for those large companies and government organizations who own particular datasets.

The big data ecosystem is not delivering the insights it once promised. If we don't change this, the vast opportunities for increasing our understanding of the world and increasing our quality of life will be lost.

Examples of insights

Health and shopping: Imagine combining shopping data with health care records, correlated past shopping habits with serious disease trends. It would be possible to gather significant data on food types or lifestyle indications which are the causes of many serious diseases. For example the probability of developing diabetes could be detected well before it has any physical impact. Likewise, foods or shopping habits which are statistically linked with occurrences of cancer could also be detected.

Commuter transit data and accurate support: Imagine combining the details of thousands of commuters, by gathering the mobile phone data with public transport data. Allowing public transit to run on an anticipated demand system with flexible and scalable timetables.

The impediments

In the examples above the data currently exists. However access to the data sets in question is limited either due to commercial or various legal concerns. There is no underlying economic model or viable value chain which connects the data producers with application developers / researchers across marketplaces and domains. Companies are not economically encouraged to share data. In fact companies typically regard the data they hold as a competitive advantage.

Consider personal data. As the examples above show its analysis can have multiple benefits. However, in general private data collected from individuals who do not wish to share their data with third parties cannot be used in this manner. The data should only be used for the stated purpose of its original collection and cannot be shared with any third party. At the EU level this also relates to a fundamental personal data protection right. Otherwise, it might be possible to use social data if properly anonymised.

Typically when companies do share their privileged data with a third party organization, they conclude legal agreements to that effect in order to ensure that a third party uses such data only for the purpose the company owning the data allowed it and also to ensure security measures are applied by such third party. For security purposes sensitive data typically does not leave the originating organisation. These factors combined result in significant impediments to open sharing of data and therefore prevent dual benefits or direct economic boost for ICT and the improvements to society and our productivity.

The solution

If we could tackle the two challenges observed, of commercial incentive and legal compliance, we could create an economic and social model that securely liberates big data and expands its value (use). Our model in many ways complements the goals set by the EU in the Digital Agenda for Europe as it provides for a sustainable economic and social model that encourages the opening up of data. It enables the true societal values of data be recognised; encouraging the creation of exciting new products and services to benefit us all.

To do this we explore a novel technical approach which allows the processing required to power the social benefits described above, while addressing the legal requirements. We also explore how such a technology would be deployed successfully and what business entities are needed to impact an existing industries' value chain.

2. Energy example: providing access to data

Imagine combining geographic and temporal electricity consumption data with data about traffic flows. It should be possible to use electricity as an indicator of activity; when people rise in the morning they take showers, make coffee, turn lights on. They do all of this before leaving their residence. But monitoring for these power draw events, and matching the delay between when they start and when the traffic builds it, should allow building a model which can predict traffic flow based on the observed electricity consumption.

The energy market is an example of a data driven industry where cross references and collating of big data can lead to new innovation and social improvement. To explore this further we focus on the challenge of releasing data on real time geographic electricity.

Existing industry structure and data flow

Within the energy market's value chain the Distribution System Operators, (DSOs) are responsible for delivering electricity and ensuring both the quality and security of supply. In order to perform these actions the DSOs' infrastructure is heavily instrumented. The DSO must capture and manage the fluxes of data generated by this instrumentation. Within the energy marketplace a DSO has a huge impact on the energy ecosystem, to ensure a competitive energy generation DSOs are heavily regulated by the EC and each Member States' regulations.
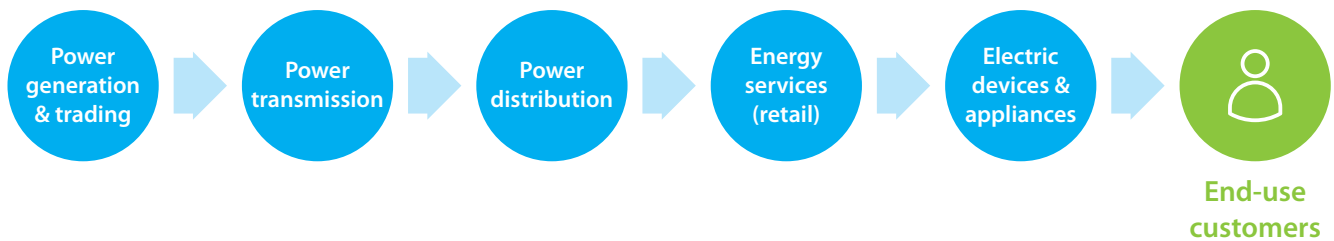

Power generation & trading → Power transmission → Power distribution → Energy services (retail) → Electric devices & appliances → **End-use customers**

Figure 1: The current electricity value chain

DSO is the entity in the value chain monitoring power consumption and connecting consumers to energy suppliers. They allow consumers to switch between suppliers quickly and act as neutral market facilitators. DSOs' unique status as heavily regulated entities is reflected in issues affecting data mobility.

3. Identified issues affecting DSO in data-sharing and developer in data-consuming

Legal conformance and fear

› Cost of legal compliance affecting DSOs

Non-anonymised private data is a huge source of potential value. However there are costs associated in creating IT systems which implement the requirements imposed by the EU when dealing with private data. These costs increase when the data sets being managed contain data which is subject to intellectual property conditions imposed by third parties e.g. data sets composed of purchased data. Further costs are incurred when additional legislation imposes controls on the data. Such additional legislation is typically industry specific, for the DSO this is legislation pertaining to the regulation of its business.

Costs

› Infrastructure costs affecting DSOs

Increasingly DSOs have access to smart meters which are deployed in consumer homes and provide regular timely updates on power consumption. However putting in place the necessary infrastructure to support regular timely (near real time) updates of power consumption is expensive to do and most DSOs are not yet to make this investment.

› Affecting developers

The energy industry generates a vast amount of data. The infrastructure needed to deal with the raw data is immense and the costs does deter data consumers.

Business challenges

› Data providing not an important element in DSO business model

DSOs currently manage and share data, but only for the fulfilment of their core business - the provision of electricity and the creation of an electricity market. Creating additional revenue by sharing the data to others both inside and outside of the energy industry has not been identified as a business opportunity. There is no clear value chain in this marketplace which the conservative DSOs can reference to.

› Logistics and charging models for data access affecting developers

There are 31 major DSOs located across Europe which account for 70% of all of Europe's energy consumption. Obtaining access to this data involves individual contact and negotiating with each of the 31 DSOs. Each will have a different attitude to providing access and charging structures. The 31 negotiations and management of multiple charging structures is costly.

Before proposing any new solutions it is vitally important that we consider existing solutions, and how they attempt to increase access to data for the overall social benefits in the energy market.

Data sharing within the energy market

The sharing of data between DSOs and energy supply companies via a central data broker organization, or hub, currently exists in a number of European states. These provide personal information describing a customer's electricity consumption data. Some examples include:

Norway: The NUBIX Model	**Netherland:** Energy Data Services (http://www.edsn.nl)	**Denmark:** DataHub (http://energinet.dk/EN/ El/Datahub/Sider/DataHub.aspx)
Slovenia: the «Perun» Portal	**Czech Republic:** Energy Market Operator (OTE)	**France:** The SGE Model

The hubs above have the sole purpose of allowing consumers to switch energy provider more easily, which is of course, great value for the consumer and instigates his interest in access to such data. They do not consider releasing data for any other purpose. This is in line with developments at EU level with respect to the Energy Efficiency Directive that is aiming, inter alia, towards enabling consumers' free access to their energy data, as we will show below.

Green Button- USA

Green button is a US energy industry response to the White House call to action. The initiative allows energy consumers to download a copy of their energy consumption data from their energy supplier. Green Button programmedefines the format of data supplied in and this common format allows developers to build against it. In order to use the system a consumer must log into their energy suppliers web site, download their own consumption data and then find a web application which can work with Green Button format and upload to their chosen application. To date 32 energy suppliers have implemented the Green Button features allowing consumers to download their data, and there are currently 240 applications which can consume this data.

Tendril - USA and Europe

Tendril is a Colorado based company with operations in Europe with Essent, which has created an energy management ecosystem based on individual's power consumption data. To achieve this Tendril supplies a set of electronic devices which monitor energy consumption within the home. These devices upload the data they capture to a cloud service. Tendril's cloud service then provides this data to just over 400 application developers via a Tendril specific API. Developers must also here receive explicit authorisation to access Tendril consumer's data.

Midata - UK

Midata is a UK government-led initiative which is similar to Green Button, towards opening-up of access to customers' transactional data. Unlike Green button Midata's remit extends beyond energy data and also includes banks, mobile telephone companies and payment companies. Data is also here only available on consumer's request, and it would need to be released in an electronic machine readable format. However unlike Green button the exact format the data should be supplied in is not stipulated. While a number of companies have joined the Midata initiative to date the only energy company involved is EDF. Currently Midata initiative is voluntary; however, the UK Government has recently proposed legislation (10.1.2013).

Existing open data sharing:
Public Sector Information (PSI)

EU has been encouraging Member States to make PSI available for use by third parties as open data. Some progress has been made in this and necessary legislation and other related initiatives are pursued to ensure that access to and reuse of PSI is made even more efficient and implemented as such EU-wide. Revised EU PSI Directive to this effect is expected to be adopted during 2013. There are other initiatives and EU developments towards opening up also of data currently not included in the PSI initiative, such as public transport data (European Multimodal Travel Planning and Information Services and related Intelligent Transport Systems - ITS action plan and ITS Directive).

4. EU Directives on disclosing user data within the energy market

By mid-2014 EU Member States will formally be bound to comply with Energy Efficiency Directive (2012/27/EU) requirements that include enabled free access of consumers to their own energy consumption data (input and off-take). The motive behind this is consumers' benefit as a provided opportunity for them to compare deals on a like-for like basis. Consumers would also have the option to choose a third party to access data on their behalf (for this purpose).

Potential solution: An EU-wide DaaS (Data as a Service) provider

A solution is to create a company (DaaS) which acts an EU-wide DSO specific data service provider. The new company develops innovative technology and installs it inside the DSOs. The DaaS receives queries from developers and then distributes queries to the participating DSOs. The DaaS ensures that correct legal controls are applied.

The technical solution: inverting data sharing

Typically when developers think about big data they assume they need direct access to the dataset in order to analyse it. However this is not always true. Typically the developer is interested in answering a question - normally investigating the relationship between data points within a data set. The answer to this question is statistical analytics produced by a developer's query or algorithm.

5. Innovative technical approach

The DaaS innovative approach is to allow the developer to craft a query, allow that query to execute on one or more DSO's data sets, while preventing the developer from gaining access to any of the original datasets. Only statistical analytics produced by the query is returned to the developer. In generating the statistical result, the technology audits the request and the statistical result and verifies that DSO's previously set rules have been followed. Any specially protected data processed in accordance with applicable legislation would never be accessed (statistical analytics). Both the query and the audit happen inside the DSO. Technology solution ensures that original data is not transferred outside the original institution, allows the DSO to maintain control over data and what is exposed, and ensures that data processing and exposure meet applicable legal requirements.

The DaaS has a central data centre connected to each of the hardware/software deployments installed inside each institutions. It receives queries from developers and may distribute the query to one or more institutions. Once the hardware/software deployment inside each institutions responds, results are collated and returned to developer. Transactions are monitored for billing purposes.

6. How the DaaS addresses identified issues

Legal conformance and fear

› Solution to the cost of complying with legal obligations
Affecting DSOs & developers

The DaaS can amortise the costs of complying with various legal obligations across many data consumers. The DaaS can do this by ensuring that this business logic complies the local legislation in the states in which it operates. This business logic works by building upon the DaaS' understanding of the data points exposed by the DSO and the queries submitted by the user. Such a solution would allow an automated validation of the query, the generated data and a comparison against local legislation. Once this is in place the DaaS provider can offer a certification of compliance with local applicable laws (of course, applicable legislation will also depend on types of data sought to be processed) and actively prevent local legislation from being breached. This solution would also offer the developer and the DSO the confidence required to participate with the DaaS provider.

› Controlling third party data use and other DSO concerns of data release affecting DSOs

Both DSOs and developers provide profiles of the types of data they provide and require respectively. The DaaS is allowed to limit the DSOs data made available to the developer. Developers provide a description of the use cases for which they require access to data. This enables the DSOs to understand the benefits of data use (use case) and either price, or decline access to their data accordingly. The developers can select one or more of the DSOs that accepted their request. The DSOs effectively compete against each other to service the request, which helps to control the pricing.

Costs

› Solving Infrastructure Costs
Affecting DSOs

The DaaS provides a direct economic incentive which encourages DSOs to capture real time data. It does so by allowing DSOs to resell secure and controlled access to that data. Thereby an offset is provided to the costs of administering the legally enforced smart meter roll out.

› Affecting developers

The innovative DaaS approach ensures that the query the developer submits is executed not by the developer, but by the data provider - the DSO. This approach moves the computation cost away from the developer and places it with the DSO. The DSO in turn can amortise the cost of such infrastructure across each of the data supply agreements it enters into.

Business Challenges

› Addressing: Data providing not an important element in DSO business model

As DaaS starts to operate it will generate a library of requests for access, use cases and price points which can be used as a description of market size and encourage DSOs to share data.

› Addressing: Logistics and charging models for data access affecting developers

The DaaS provides the developer with a single point of contact for all DSOs across the EU and allows the developer to quickly determine what data is available at what cost.

Funding the DaaS provider

The DaaS Provider generates revenue as a small percentage on top of the fees charged by the DSOs. Additional revenue is generated by offering enhanced or premium advanced analytical services – predefined queries which would be expensive for a single developer to pay for, but which work with economy of scale.

Augmenting the energy value chain

The introduction of the DaaS Provider alters the traditional energy value chain by adding additional sources of revenue generation and specifically now links DSOs with application developers:
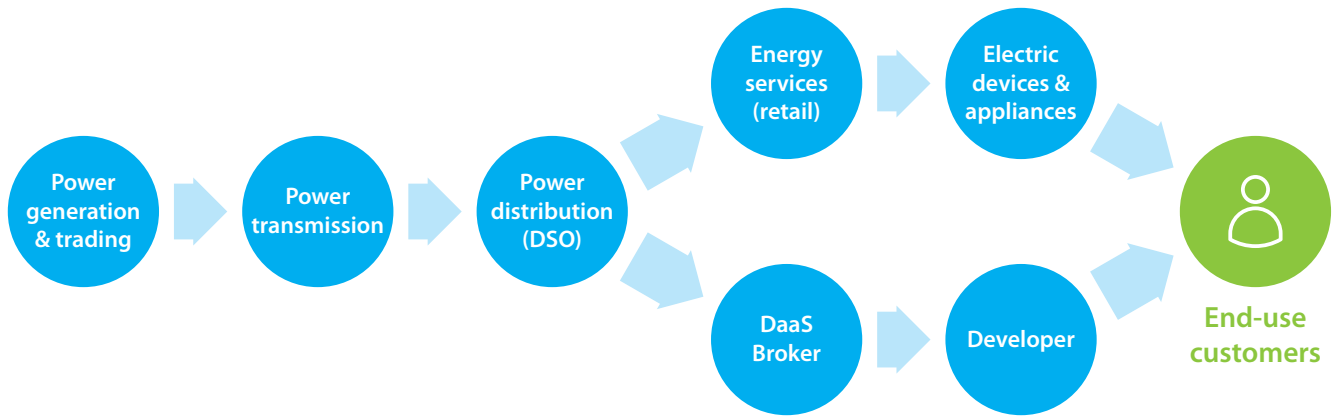
Figure 2: The current electricity value chain

Implementing the DSO DaaS provider

Identifying entry points into the market
There are roughly 100 small to medium sized DSOs in the EU, however top 31 account for 70% of all EU powered consumers. Some DSOs in the EU are already familiar with collaborating with each other to share data for their day-to-day business operations. Initial engagement would begin with speaking to DSO data hubs and the DSOs co-operating with them.

Commercial enterprise
The energy market is traditionally very heavily regulated and while there is tendency suggesting additional regulation can create the DSO DaaS provider there is also evidence, in DSO Data hubs created in Norway and Sweden, that a commercial enterprise is practical. In our opinion legislation alone would result in a slow roll out of an EU DSO DaaS, as each state would need to pass necessary legislation. Rather, commercial enterprises would have freedom to operate across national borders.

8. Conclusion

Big data is still a nascent industry, which needs to be nurtured. There is still huge potential for growth, innovation and insights into how we live today. This provides us with a chance to study our own communities from new angles, and to develop new technologies that can have have enormous impact on our quality of life and improve our society as a whole.

However, currently there still exist technical, legal and cultural challenges towards realisation of its full potential. Empowering organisations to liberate the statistical value of the datasets they own by sharing access to it, rather than to any raw data itself, is vital. Consumers, whom data analysis is ultimately intended to benefit, also need to become more empowered and a proper ecosystem established, which preserves their individual rights. This would altogether provide a concrete foundation for further innovation, economic growth and significant social impact.

Further recommendations

The DaaS provider concept shows that innovative new technology can help address some of these concerns and can lead to innovation. However, the EU's approach to policing the industry would have a huge impact on innovation and our society in the coming years.

Ensuring competition	Any emerging industry can be dominated by early entrants reducing competition and effectiveness of market forces. Legislation will be necessary to address this. The recent EU Commission's decisions on antitrust issues help to provide some guidance on possible solutions: 1. Dividing this new market in macro-geographical area. 2. Dividing this new market into specialised areas (traffic, smart grids, urban re-engineering). 3. Dividing this new market into macro-geographical and specialised areas. Unfortunately (1) & (3) reduce the size of the marketplace and endanger the ability of DaaS to amortise costs. The best solution is (2), which splits the market up while still providing for the scale necessary for successful DaaS companies to exist (and compete).
Innovation in charging models	As acceptance of the new industry occurs and competition increases DaaS companies will compete against each other new pricing models. We can foresee the following: 1. Negotiated Fee: Based on predicted value of the data supplied (as described above). 2. Success Fee / Revenue Share: Share of the revenues of the developer. 3. Standardised Fee based on consumption: Fixed Euro per query pricing. 4. Subscription Fee: A flat monthly fee for DaaS services.
Changing business practices	As the revenue from DaaS access grows relevance of this market to the data supply companies will increase. Suppliers of data to DaaS will increasingly optimise their business models to allow them to collect additional data, and to better leverage data already captured as sources of additional income.

Insight-on-time: Scaling up entrepreneurship in Europe

Jonas Vermeulen

Sebastien Petillon

Wouter Haerick

Abstract: *Too many start-ups fail. Studies reveal that 60% of the start-ups quit within 3 years after their birth. Therefore, if we want to impact economic growth and job creations, we should dare to rethink the way we can support our startups that have growth ambitions. A main reason for failure is the lack of understanding of the real market needs, while overspending on product development. In this vision paper, we address the opportunity of using big data, to remove this lack of understanding about the customer needs.*

We envision a revolutionary solution where each start-up can have its fully customised "Insight-on-time" dashboard web page, visualising the latest trends and figures in the market, the ecosystem of competitors, suppliers and customers, and providing suggestions of potential new customers, or to be aware of competitors that were unknown to the start-up, etc. While many tools exist to track the competition and the market, we discuss the immaturity of these tools, and the opportunities for further data-driven innovations. We explain that European startups and national support organizations will have the highest interest but have limited power to change the game. We see however multiple roles for the European Commission to facilitate unleashing of the full power of the presented concepts.

1. Few start-ups make an impact on jobs and growth

Too many start-ups fail. The Global Entrepreneurship Monitor (GEM) estimates that worldwide around 100 million businesses are started each year, and that this involves at least US$360 billion of informal investments and US$32 billion in venture capital [GEM, MOYA]. Also, the annual firm death rates and birth rates are almost equal. So, there is almost one business failure per start-up. The figure below indicates that 60% of the start-ups quit within 3 years after their birth.
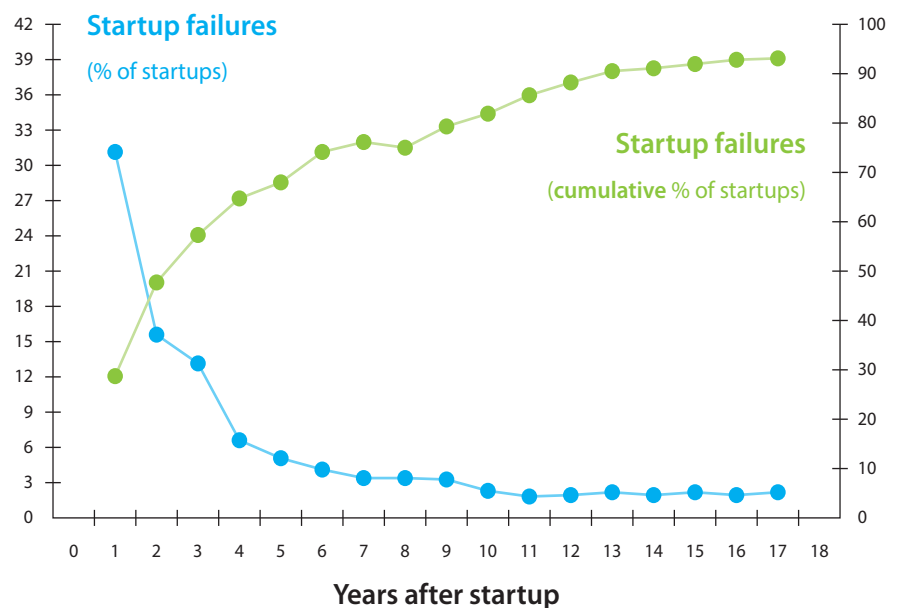

Figure 1: Percentage of failing start-ups [FAI]

If we want to impact economic growth and job creation, we should focus on supporting those start-ups that have growth ambitions. A clear distinction needs therefore be made between people that are voluntarily pursuing a great business opportunity and those who are doing it because they need a job. The group of "opportunity entrepreneurs" is rather small, and comprises about 5% of the people aged 18 to 64 in countries with well-developed well-fare systems [GEM]. Also in this group, failure rates are extremely high.

2. "Opportunity entrepreneurs" can't pinpoint reason for failure

It is known for years now, that it is not easy for starting entrepreneurs to identify what exactly is slowing down their growth. A former US Minister of Science, Barry Jones, once compared this with *"patients who walked into a doctor's surgery and could not say where their pain was"*. In the meantime, many national entrepreneurial support programs have successfully implemented agencies that act as "advisors" for small businesses, guiding them to the right resources (finance, technology, partners, …). However, we think they are not yet acting as "doctors", pinpointing the weak areas and prescribing potential solution. We also think it is essential that support agencies should be given the tools that help them know the patient (=the start-up company) with minimal efforts.

Looking at the opportunity entrepreneurs, it is remarkable that from the 93% that scale early on, and thus seam very successful, none break the 100k revenue per month threshold [Start-up Genome Project – Stanford – study done on internet start-ups only]. Additionally, the Start-up Genome project reveals that successful companies tend to grow 10 times slower at the start. Indeed, many examples exist of high growth companies that started rather slowly, but

developed a highly unique, valuable offering before a management team scaled the business (e.g. the Belgian "Le pain quotidien" bakery concept, that struggled hard before it was bought and then scaled up).

Typical reasons for failure are [MOYA2]:

› Failure to clearly define and understand your market, your customers, and your customers' buying habits;

› Inability to anticipate or react to competition, technology, or other changes in the market space;

› Failure to price your product or service correctly;

› Overdependency on a single customer, or on your early customers;

› Uncontrolled growth: slow and steady grow typically wins from going after all the business you can take. When you want to take it all, you start being less selective about customers and products, which in the end impacts your profit.

4. A three-fold challenge

Can we drastically increase the success ratio of start-up companies, and revolutionise the way they execute and continuously refine their understanding of the market?

This main challenge in fact touches an important societal aspect, as well as entrepreneurial and technical aspects.

Societal challenge

For society it is key to foster the "opportunity entrepreneurs", which pursue economic growth, job creation and welfare. The main challenge here is: Can we transform our society support for entrepreneurs towards support organizations that focus their efforts on guiding "controlled growth" (cfr uncontrolled growth as explained in the previous section)?

With respect to Europe, there is the additional challenge linked to the geographic fragmentation of the market. We should, at least in the digital version of society, aggregate all information of European business, making it a single, virtual market.

In the "recommendation" section of this paper, we will also address challenges with respect to the privacy of business information, information that by accident sometimes gets publicly available.

How society currently supports successful growth in start-ups

Today's national and European entrepreneurial support programs recognise the entrepreneur's lack of expertise on the business development side, and have successful programs to facilitate this process. For example, the European Enterprise Network (EEN) employs more than 4000 experts that successfully support innovative companies in finding funds, writing business plans, finding business partners, etc. While these support organizations mainly target the key moments in the start-up life-cycle, no or little support is oriented towards generating early warnings and early insights to circumvent imbalance in the company's growth plan. We believe the support is not yet structured around providing frequent feedback on how well you grow your market, and on how well you spend your time and money at the right customers. Rather, these programmes facilitate an entrepreneur during pre-defined key-milestone events, such as for the event of "finding a business partner abroad", "finding capital", "finding technology", and "snap-shot alike competitiveness auditing".

Entrepreneurial challenge

Entrepreneurs face two sub-challenges: (1) The high cost to continuously gather information on their market, and (2) the lack of information that is immediately relevant/applicable to their market segment. Can we therefore support the individual start-up with automated, realtime and relevant feedback on the evolving market at a fraction of today's cost to gather updated market information? In particular, can we provide realtime information on competitor's latest deals, success stories from similar companies in other regions, new features from similar products, etc. Here, it is crucial to give entrepreneurs an easy-to-use dashboard. In a single webpage, the position of the start-up in its target segment should become visible. This view should be customisable according to

the needs of a company. The cost for accessing these data should be affordable for any start-up.

How entrepreneurs currently incorporate (self-criticism about) their business plan

If we look at how entrepreneurs draft their business plan, it is still a manual, time-consuming exercise. When drafting the plan, starters use many different data sources to collect the relevant market data:

› Free text search (eg. using Google);

› Buy market reports, or hire a consultant (e.g. Gartner, Forrester, ...);

› Business-to-business website portals also can ease the search for potential customers or partners. Examples of such websites are: Kompass. com, B2BYellowPages.com, Business. com, Jayde and Zibb. Kompass. com looks like a very useful product, but it only tracks 4 million companies, whereas an estimated 1 billion companies exist worldwide;

› Use their personal network to collect information about their target market segments;

› Attract a market specialist in your board of advisors or board of directors.

Given this is a very costly process (time and/or money), this is not an exercise that is done regularly by start-up companies, but is usually done only once. As such, starters lack information about how the market is changing during the execution of their plans. The assumptions from the original plan are thus not revisited frequently.

Technical challenges

From a technical point of view, today's technology is just not ready yet to process the big data and transform it in the final data that is valuable. We identified different technical challenges.

Today's big data analytics just don't work (and are bad in simple maths)

Current flat data analytics are still very

immature. Enhanced classification technologies, and topic modelling algorithms are being studied. However, we should challenge the R&D big data community to solve tough business-to-business questions, such as:

› Calculate the total revenue of a group of similar companies (or a market segment).

Although the current "information retrieval" algorithms are applying very complex and probabilistic models, they are really bad in answering the deterministic, simple math questions.

We should expect magic from artificial intelligence

The "magic" one-button solution to show all relevant data for your company does not exist. An important challenge is defining the minimal interactions needed with the entrepreneur to make an accurate model of its start-up company, product and market. By using natural (non-intrusive) user feedback, we can train the artificial intelligence algorithms, and make its output far more accurate and relevant.

The deep web? We only search the surface…

The main search engines, such as google. com, have been mostly focussing on the data relevant for the general public. As such, these engines favour indexing the most popular and most recent data. However, to meet the business-to-business needs, very particular data sources can be of great value to a company. A challenge here is: Could entrepreneurs therefore influence the indexing of their B2B search engines, to include data which is highly relevant for them?

The main issue we want to address, is very well summarised by Rohit Shukla, advisor of OECD (global Organization for Economic Co-operation and Development)

"Most start-ups make the mistake of falling in love with their product or service. Ultimately, it is this lack of self-criticism that causes many companies, start-ups and their more mature counterparts, to fail." Rohit Shukla

It is this lack of "understanding the market", and "understanding the customer needs", we want to address in order to drastically decrease failures of promising startups.

Could we therefore better support the founding entrepreneurs in a sense that they are not caught by surprise (e.g. being forced to sell their company) but are armed to remain healthy and to engage the right people at the right time to scale the business.
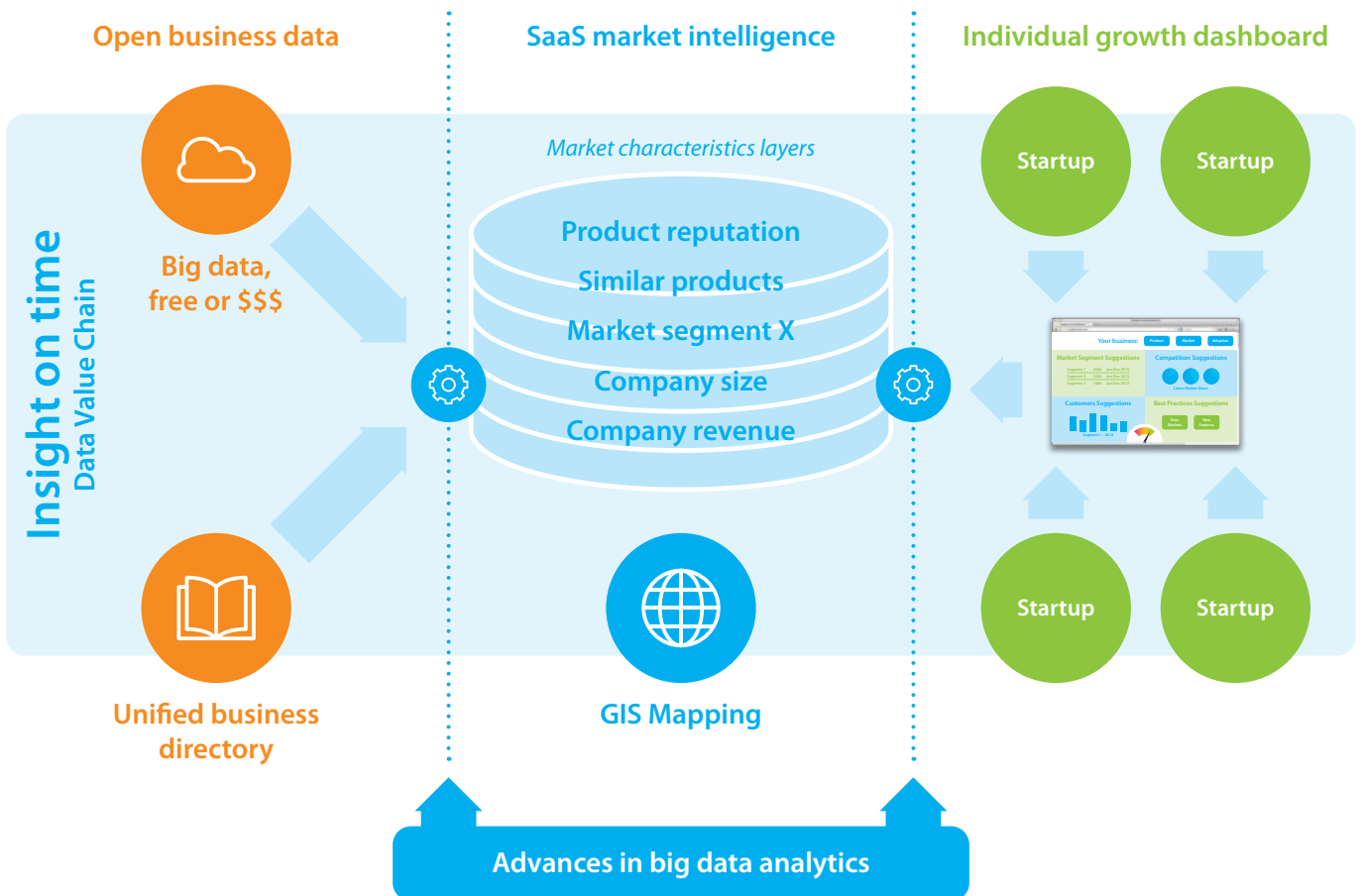

Figure 2: The Insight-on-Time value data chain

5. A new era of entrepreneurship

We envision a revolutionary solution where each start-up can have its fully customised "Insight-on-time" dashboard web page, visualising the latest trends and figures in the market, the ecosystem of competitors, suppliers and customers, and providing suggestions of potential new customers, or to be aware of competitors that were unknown to the start-up, etc.

From a technical point of view, this solution can be broken down in multiple modules or stages that take-in, process and analyse raw data, before coming up with the business-relevant information for a particular start-up.

Towards universal open business data

The raw data that can be used as inputs can take different forms:

› The various commercial yellow pages that exist in the different countries, e.g. www.kompass.com, www. www.gelbeseiten.de, http://www. thephonebook.bt.com, http://www. goldenpages.be/ to name a few.

› Through databases that the different chambers of commerce publish, where sector-specific information is listed in. Here policy makers could facilitate to evolve towards a standardised, global format for these databases. These efforts should then result in a common "Open Universal Business Directory", that should be accessible through an open API.

› Closed data sources, such as analyst reports, or manually edited data. While this could be a very valuable addition to the other data sources, it typically doesn't come for free. Hence, as a start-up uses the Insight-On-Time dashboard, it will have to choose whether to use commercially obtained data, and to use it as a payable service.

› Using, web site crawling, the data of every company could be augmented with the information that it has listed

on its public web site. Typical information such as company location, company size, product portfolio could be gathered from this.

Market intelligence data layers – as a service

The second stage in the "Insight-on-time" solution, encompasses the market intelligence data layers. The data layers are created by aggregating and analysing the various data sources. A first step consists of normalising the data. This task consists of removing duplicates, enriching the data of every company by associating data from various data sources to each other, and to index it so it can be used later as part of any Artificial Intelligence algorithm.

As an example, big data analytics could be used to enrich the normalised company data with supplier-customer relations, with market-segment information, with product catalogues, etc. Each of these aspects can be represented as an additional layer on top of the normalised company data. By exposing these data layers through open APIs, one can monetise these value-adding data layers, and contribute to a value-driven data economy. In the picture, we refer to this as "Market Intelligence as-a-Service".

Individualised dashboard

The last stage of the proposed concept is the dashboard itself, which is the graphical front-end and is the only piece that enterprises see. Rather than a cluttered and complex site that is visually unattractive, it has to be a simple yet effective interface.

The associator engine

In the background of the "Insight-on-time" dashboard, an advanced big data analytics engine is needed. We refer to this engine as the "Associator Engine". This engine implements the real business logic, and is making use of self-learning techniques to extend on the basic information that every start-up provides about itself. It therefore uses the information exposed by the Universal Business Directory, and learns about similar companies, similar profiles, and uses this knowledge to provide meaningful advice to the start-up.

A simple example

As a simple example of the workflow a start-up would go through:

1. The start-up enters some of its basic company information, or it enters its corporate website url, after which a profile is generated;

2. By entering the top 5 customers, the Associator also learns about the customer base;

3. Based on the data from the Universal Business Directory, it starts generating additional lead customers, and can prioritise those based on similarity, location, market segmentation, etc.;

4. By requesting feedback on the value of the suggested leads, better guesses could be made for the next ones;

So, as a snowball it grows the potential customer base, but only with relevant companies, using supervised learning techniques.

4-quadrant dashboard

At the moment of writing, we see four very valuable pieces of information that can be extracted:

› **Segmentation:** By entering your corporate website url, it makes a profile of your company and your product/service offering. By entering your top 5 customers, it also suggests a market subsegment. Through interaction, the entrepreneur can adapt the characteristics of the subsegments. For each subsegment, potential profits are extracted and presented for the next year. With this feedback, the company can adapt its marketing plan, or start thinking about product pricing and yearly profits;

› **Customer attack plan:** Based on your company profile, and the profile of your market segments, the tool starts generating additional lead customers, prioritised! As a snowball it grows your potential customer – using the universal business directory - but only with relevant companies, using supervised learning techniques. The dashboard helps the entrepreneur to target the most suited customers to develop a sustainable growth plan;

› **Competition:** Through similarity search techniques, similar companies that play in the same market segment, can be generated, based on a first list of 5 competitors entered by the entrepreneur. In this section, the announced sales deals of the competitors are shown in real-time. With this information, the dashboard re-calculates projected market share. By clicking on a competitor, the news alerts of recent deals are shown;

› **Opportunity Suggestions:** Based on the products or services competitors offer, a key missing feature can be suggested. Also, the dashboard can suggest best practices from a very similar company acting in a non-competing geographic region.

The dashboard is offered in a Software-as-a-Service model, which means it is accessible over the internet, and is offered as a service for which a start-up has to pay on a monthly basis (if not offered for free), and for additional data sources to be used in its analysis.
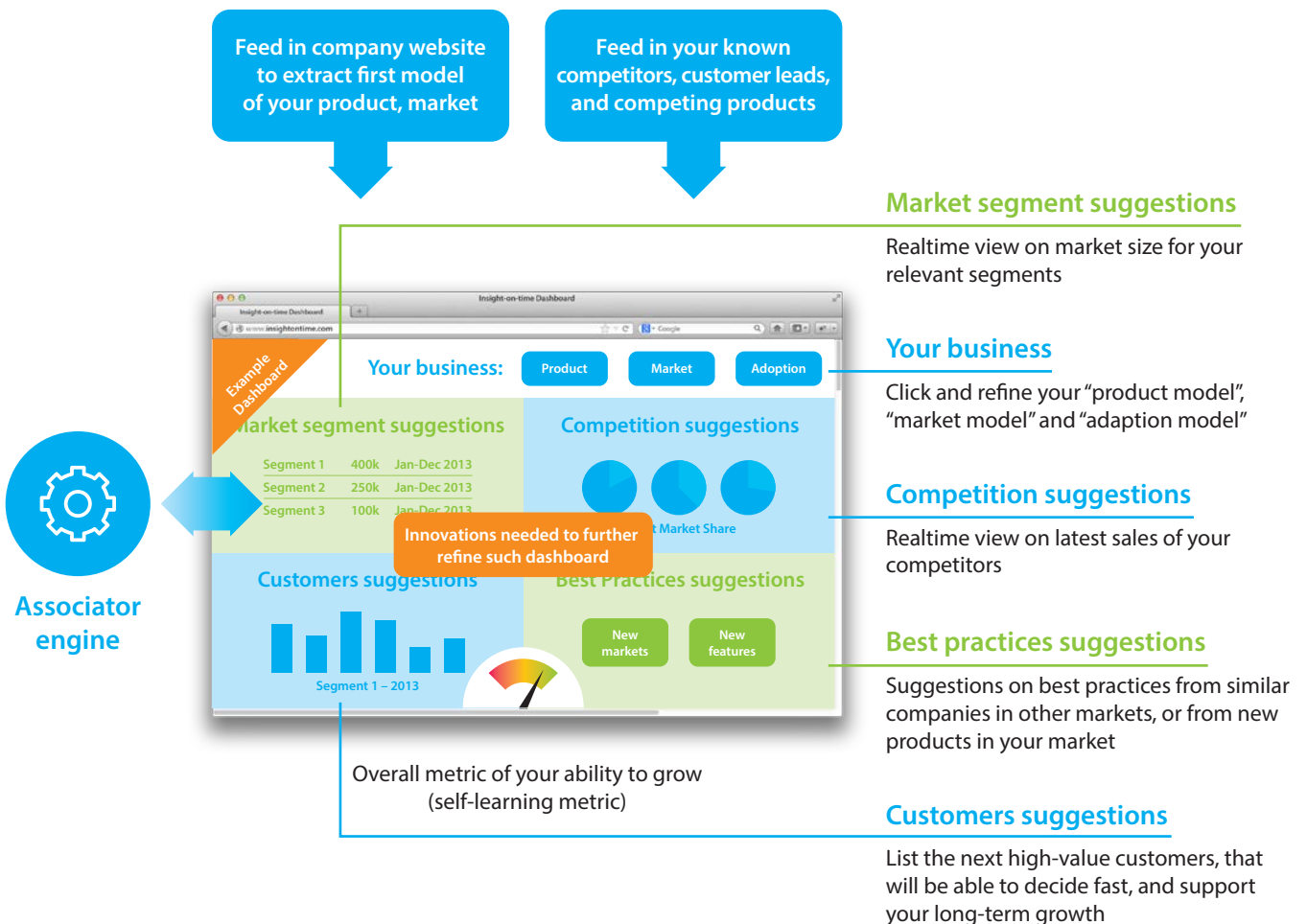

Figure 3: An example "Insight-on-time" dashboard web page

6. Assessment: Current tools are still immature, and create overload of data.

To track the competition and the market, different tools yet exist varying from "competitor or market portals", "competitor email alert services", "software spying the social networks" to get an idea of your (competitor's) brand reputation. Examples are: SocialMention.com, Radian6.com, Google Alerts, and Open Site Explorer. The social media tools are the most advanced, but do not focus on B2B interactions. While these tools provide useful information, and are a first step towards on-time insights about your market, they do not implement the tough, final, analytical step that will help start-ups to criticise their market approach.
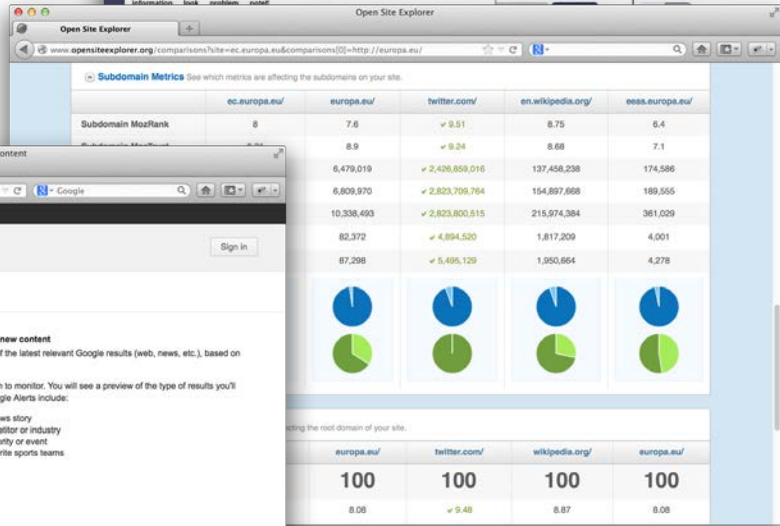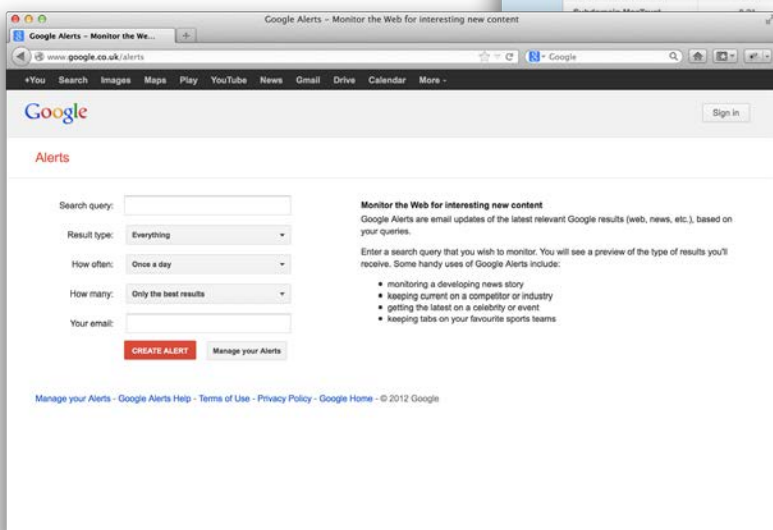
SOCIAL MENTION
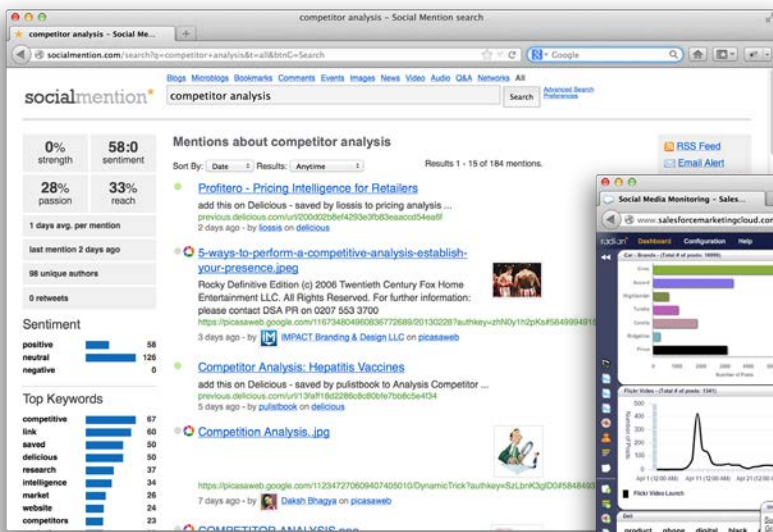
RADIAN5

GOOGLE ALERTS

OPEN SITE EXPLORER

Figure 4: Examples of current tools to monitor competition

These current tools do work for brand reputation and sentiment estimation of the big brands, but they are rather useless for start-ups entering (non-English speaking) B2B market segments. With respect to the email alert services, this is just another source of data overload which is hard to overlook, and to see the overall message after for example one month. To impact the success ratio of start-ups, we are looking for solutions that work towards the overall objective: early indication of "uncontrolled" growth, and suggestions to adapt to changing market conditions. Temporal analysis of past news alerts is here a key task, instead of the manual analysis you still need to do when all the news is in your email box.

7. Recommendations for the European, national and regional levels

Considering all stakeholders, and after performing a brief stakeholder analysis, we learn that European startups and national support organizations will have the highest interest but have limited power to change the game. In particular, for European startups that want to address the European market as a whole, we think the interest can be high. Currently, many European startups focus on the US as first market, because of the ease of identifying distributors and partners for a large market. As we further explain below, we think the European Commission may have both a high interest to foster their startups, and also have the power to ignite change. Apart from the power of European institutions, we think that big data Startups may also have a high interest
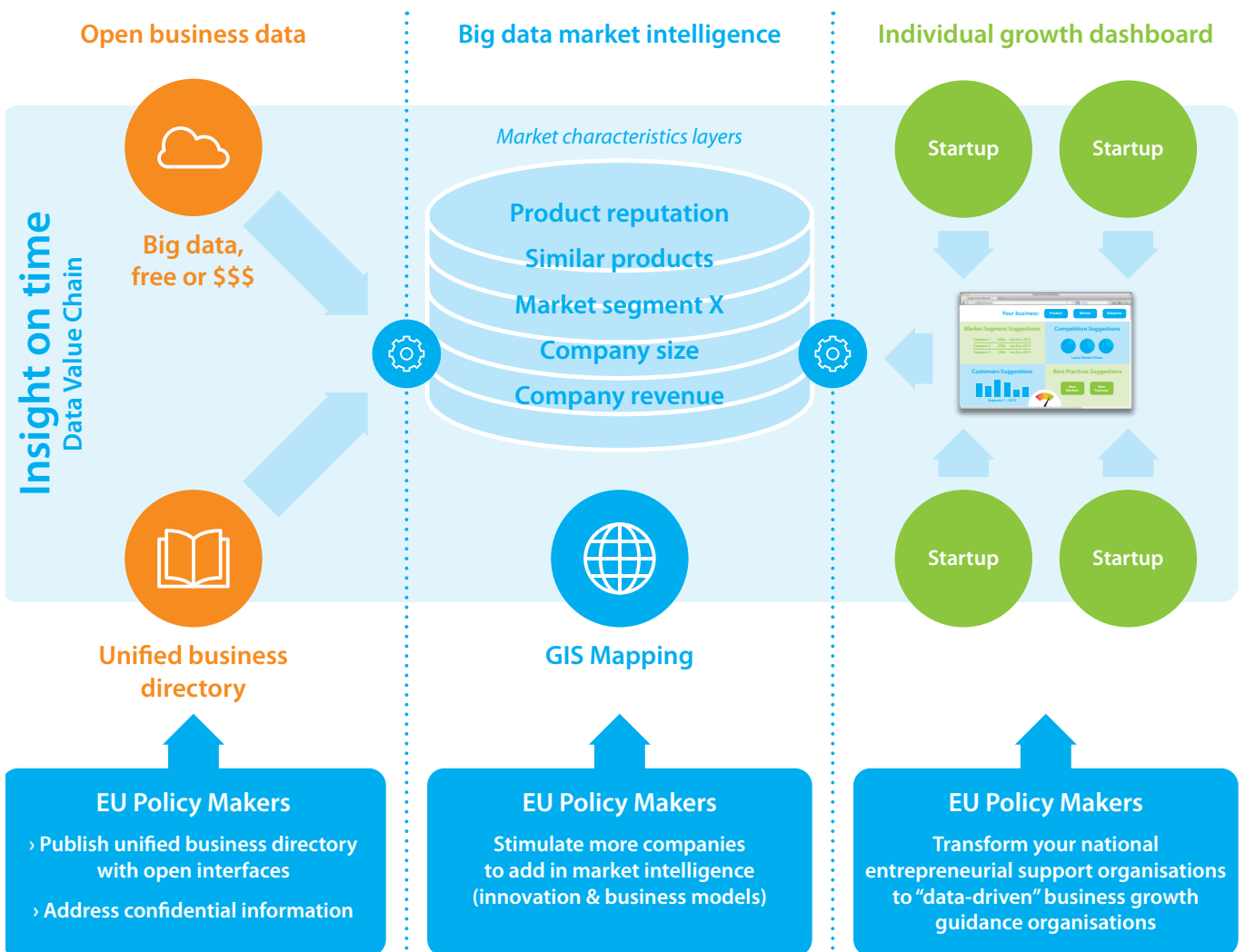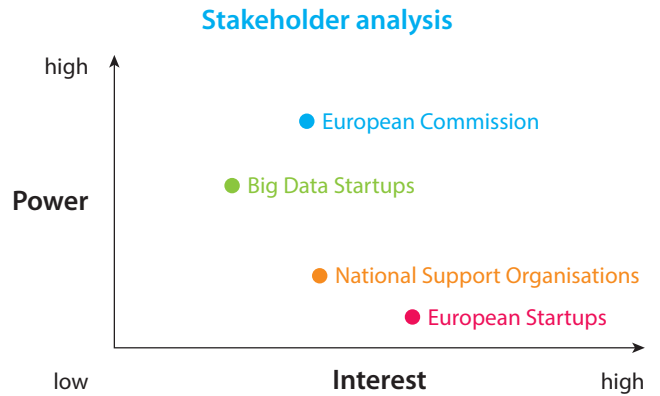
and high power to come with valuable business models and "Insight-on-time"-like solutions to support the growth of other startups. We believe new business models can arise, where at a fraction of the current cost, startups can benefit from a more realtime and transparent view on their market.

To unleash the full power of the presented concept, it should be embraced by the nation's policy makers. For each of the three stages depicted in the previous section, we have identified key recommendations towards the policy makers. Below, we summarise the different recommendations in one picture.

Figure 5: Summarising recommendations to policy makers

41

Recommendations related to "Open Business data"

Lobby for universal business directory

As mentioned previously, some information needed for an "Insight-on-Time" system is already available in open big data. Listing of European companies by business category, city, country, and more, are available at national or European level on business directory websites. Some of these directories are fairly detailed and unified. But in order for an "Insight-on-Time" system to be valuable to European entrepreneurs, more fields of information would be needed than what is currently available. The most elaborate unified business directories are not free (such as www.kompass.com). Regional chambers of commerce usually offer richer business directories, but they are not unified and not easily accessible to a tool like "Insight-on-Time".

European institutions could help develop further the unified European market, at least virtually, by making a single unified market space. This could be done by making sure databases from regional chambers of commerce are made available with their full content in a unified format. Since it involves so many chambers of commerce, it seems that the initiative and implementation of such a data unification need to come from the European level. This unification would include the need to use the same language, the same search fields and the same listing format.

In addition, the EU could work with European chambers of commerce to add more fields in information to existing directories, making the data even more valuable to a market analysis.

Address the issue of "confidential" information accidentally added to big data

But besides helping with the creation of an Insight-on-Time tool at different levels, the European institutions could have another role to play in such concept. An aspect we have not mentioned yet in this paper is the ownership of the business data present in the Open big data. Some information that are available in the big data might have been originally created as copyrighted, proprietary or even confidential information. In this case, would a software like Insight-on-Time be allowed to use this data, e.g. after crawling SlideShare.com? Would it depend on the country where the information was extracted from or published? Can a confidential information – that is "leaked" in the big data - be used for a Insight on Time tool? … It seems necessary for the European Institutions to be involved and regulate what is not yet clearly legislated at the European level.

One way to help respect confidentially and data ownership could be for the Insight-on-Time tool to include some analytics that block data if it includes key words such as "confidential" or "copyright". But here again, the analytics behind this classification would have to be elaborated. And if the protected data actually provide added-value, it should ideally offer the option of purchasing this data to whoever owns it.

Recommendations related to "market intelligence data layers"

Stimulate research & innovation on advanced big data analytics

As we have highlighted in the previous sections, one of the biggest barriers to making a tool like Insight-on-Time is the lack of efficient big data analytics. The analytics technologies available today don't seem to be "smart" enough to make the Insight-on-Time tool efficient. It is key that such a tool can have good self-learning algorithms, as well as avoid redundancies in the search results. This seems to be one of the reasons why online alerts tool today – such as Google alert - are just too inefficient today.

The reason why this is a major barrier to creating an Insight-in-Time, is because it is a complex problem. In fact, if this can be overcome, the applications could go well beyond the use of Insight-on-Time, and it could have an impact on many other industries. For this reason, we suggest that the European Union stimulates more research and innovation on big data analytics. Topics to stimulate are accurate detection of companies, product and market segmentations. Techniques that need further development are: Named Entity Recognition techniques, Named Entity Disambiguation techniques, exploration of gazetteers of known companies, products, graph-based representations of companies that occur together, time labelling of relevant press articles, linking with business social networks (LinkedIn, Yammer, etc). Advances are also needed in the domain of similarity search, where more accurate similarity metrics should be developed, for example to automatically detect competitors. Also the domain of temporal behaviour analysis could be further advanced to detect 'bursts' of interest, i.e. sudden increases of interest in a company or a competing product.

Leading universities, already focussing on data analytics, could be identified to fund additional research. We could also imagine that the European Union organises forums and conferences on this topic in Europe. Finally, this could also be achieved by directly funding European start-ups specialising in this field, which would help creating this Insight-on-Time venture, as well as stimulate other European start-ups.

Recommendations related to "individual dashboard for entrepreneurs"

Stimulate other companies to develop similar market insight tools

Another way Europe can boost the implementation and efficiency of such tool is to encourage start-ups to create Insight-on-Time tools. One could imagine that the EU could organise a programme where start-ups would compete in the creation of such a tool, and the winning company (or companies) would get some funding to start the venture.

But the EU could also facilitate a business model where many start-ups would create and sell data aggregation. Thus, these start-ups would not offer a dashboard, but aggregated data to feed to the dashboard of Insight-on-Time. This business model would quickly increase the relevancy and efficiency Insight-on-Time by improving the aggregated big data market Intelligence.

Figure 6: One common goal: having fun while stimulating growth

8. Towards fast implementation

The "Insight-on-Time" tools will benefit to individual entrepreneurs, to the supporting nations (by increasing their start-up success ratio), and therefore boost the economy. If Europe takes a leading edge on this technology, this will benefit in particular the European economy. The implication of the European Union therefore doesn't have to stop at facilitating a European business directory, or at stimulating innovation on big data analytics. We could also imagine that this tool is developed by European institutions themselves, or by companies funded by the European Union. This could fasten the development if it is included in upcoming initiatives from for example CIP or EIT. This would also allow the EU to make this tool only available to European companies, giving Europe a competitive edge in start-ups efficiency and success rate. We suggest some more, open discussions on these open questions, and as a team, are willing to further contribute.

Simultaneously, the entrepreneurial support organizations can opt-in and connect to the Insight-on-time community. The automated models for the company, its product and its market, are a welcomed source of information to start discussing how to sustainably differentiate. By doing this, and by obtaining, The support organizations could also start redefining their own roles, given a new context where they have access to relevant and realtime insights on the past and future growth of the start-ups they support.

As such, all stakeholders involved can converge to the common goal: Increase the success ratio of start-ups through controlled growth.

9. References and sources

Global Entrepreneurship Monitor [GEM], http://www.gemconsortium.org/

Start-up Genome Project – Berckely & Stanford, http://start-upgenome.cc/
http://visual.ly/why-start-ups-fail

OECD, Entrepreneurship and Business Statistics, www.oecd.org/std/entrepreneurshipandbusinessstatistics/

OECD Timely Indicators of Entrepreneurship, stats.oecd.org/Index.aspx?DataSetCode=TIMELY_BDS_ISIC4

The World Bank Group, (Founders - Simeon Djankov, Prof. Michael Klein, dr. Caralee McLiesh)
http://www.doingbusiness.org/about-us

Eurochambres - 1.800 local and regional Chambers in Europe, http://www.eurochambres.be/

Enterprise Europe Network, http://portal.enterprise-europe-network.ec.europa.eu

Worldwide business start-ups – Moya K Mason [MOYA], http://www.moyak.com/papers/business-start-ups-entrepreneurs.html

What causes Small Business to fail – Moya K Mason [MOYA2], www.moyak.com/papers/small-business-failure.html

Failure rates of start-ups – R. Peacock, W. Reynolds [FAI], http://www.sbeducation.info/downloads/sbfail.pdf

Some tools: www.liveplan.com, www.SocialMonitor.com, www.ifttt.com, www.google.com/
alerts, www.opensiteexplorer.org/, www.kompass.com, www.business.com

Supporting and enhancing a sense of community in smart cities through big data

Jessica McCarthy

Guillaume Delalandre

Peter Mechant

Romain Muller

Geert Ysebaert

Abstract: *Cities play an ever increasingly important role in the lives of vast diversities of people. Hence, (big) data in and about cities, becomes more important because it can enable the commitment of cities to competitiveness and sustainability and to become 'greener' and more 'liveable'. Thus, city big data can function as a growth engine, it can help to bridge the digital divide and to increase quality of living for city inhabitants.*

This document proposes a vision or blueprint on big data from a 'smart cities' perspective. It points to several issues that will have to be addressed to capture the full potential of big data in a smart city context and suggest an interdisciplinary action research agenda starting from the guiding principle that big data should act as a catalyst to stimulate citizen engagement in becoming active generators of data and information.

We state that in order to create such a bidirectional exchange of (big) data, a certain level of trust and cohesion between citizens of smart cities and city governments is needed. Based on the phrase 'Sense of Community' we create a conceptual framework that addresses challenges and solutions related to big data and influence, membership, need fulfillment and emotional connectivity. This framework should materialise in concrete actions and policies for European data rich or 'smart' cities, encompassing amongst others; the creation of a Data Friendly City Label, the appointment of city data Ombudsman, and the development of a corpus of initiatives merging communities and big data to address city challenges.

The set of ideas described in this paper is certainly not exhaustive but aims at inspiring and informing public authorities about the possibilities offered by the new technologies related to big data within the boundaries of a city. The presentation of such a blueprint by the European Commission – added to the promotion of cities already engaged on the path to become smart cities – can guide European cities to foster bidirectional exchanges with their citizens and face future challenges while making the most of today's innovations related to big data.

1. Introduction

Worldwide, population has been steadily concentrating in cities. A 'tipping point' was reached in 2007 when the majority of people worldwide lived in towns or cities for the first time in history[1]. In addition, we also witnessed a substantial increase in the average size of urban areas[2]. In short, cities play an ever increasingly important role in the lives of vast diversities of people. Hence, (big) data in and about cities, becomes more important because it can enable the commitment of cities to competitiveness and sustainability and to become 'greener' (with smart energy, smart environments and smart mobility), and more 'liveable' (with smart health, smart education and smart living/working). Thus, city big data can function as a growth engine, it can help to bridge the digital divide and to increase quality of living for city inhabitants[3].

Great amounts of data created by city inhabitants, businesses and governments or generated by sensors are flowing onto the w eb and are stored in ever growing data farms. At the same time, mediated communication is becoming increasingly participative and personalised by means of social computing.

This puts the user in the driving seat (see e.g. phrases such as 'user as innovator'[4] or 'open innovation'[5]) but also emphasises the importance of data. This increasing attention on data is visible in discourse (see e.g. the hyped phrase 'Data is the Next Intel Inside'[6]) and the importance of concepts such as Open, linked and big data. While these three concepts are interconnected they each carry a specific meaning. Open data is an idea that certain data should be freely available. Linked data describes a method of publishing structured data in such a way that it can be read automatically by computers and that it connects to related data that wasn't previously linked. Big data refers to a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools because of the volume of data being collected, the speed at which data are being generated/collected (velocity), and their different types (variety)[7].

This document proposes a *vision or blueprint* on big data from a 'smart cities' perspective. It points to several issues that will have to be addressed to capture the full potential of big data in a smart city context and suggest an interdisciplinary action research agenda starting from the guiding principle that big data should act as a catalyst to stimulate citizen engagement in becoming active generators of data and information. Smart cities need big data to become available from and for their citizens, they need the buy-in of their citizens as there is a win-win possible between citizens and city governments by unleashing the potential of big data. In short, 'smart cities require smart citizens' if they are to be truly inclusive, innovative and sustainable[8].

Considering the above, this paper focuses on

---

1 http://esa.un.org/unup.

2 http://web.unfpa.org/swp/2007/english/chapter_1/urbanization.html.

3 Caragliu, A., Del Bo, C., & Nijkamp, P. (2009). Smart Cities in Europe. Paper presented at the 3rd Central European Conference on Regional Science (CERS).

Dolente, C., Galea, J. J., &Leporelli, C. (2010). Next generation access and digital divide: opposite sides of the same coin? Paper presented at the European Regional ITS Conference.

4 Pierson, J., Mante-Meijer, E., Loos, E., &Sapio, B. (Eds.). (2008). Innovating for and by users. Brussels: COST 298 – OPOCE.

5 Chesbrough, H. (2005). Open Innovation: The New Imperative for Creating And Profiting from Technology. Boston: Harvard Business School Press.

6 O'Reilly, T. (2007). What is Web2.0: design patterns and business models for the next generation of software. Communications & Strategies, 65(1), 17-37.

7 Smith, J. A. (2012). Big Idea – How Big is Big Data? Retrieved from http://blog.drjerryasmith.com/2012/09/14/big-idea-how-big-is-big-data.

the problem statement of *"how to foster bidirectional exchange of data between citizens and city governments?"*. In our opinion, such a bidirectional exchange of (big) data can only be achieved if there is a certain level of trust and cohesion between inhabitants of smart cities and city governments. Thus, engagement and trust become key factors in supporting and ensuring bidirectional exchange of data between citizens and city governments pointing to the importance of citizens' 'sense of community'. Based on Lochner et al.[9], who reviewed the concept of social capital and related constructs, we choose to focus on the psychological 'Sense of Community' as an important driver for bidirectional exchange of data and for making data the 'new oil for the digital era'[10].

A framework created by McMillan and Chavis[11] that describes 'Sense of Community' using four constructs – membership, influence, needs fulfillment (integration) and shared emotional connection – will be used to structure and guide our proposal for a blueprint for big data processes in smart cities, see Figure 1.

Firstly, the document will address challenges and solutions related to big data and influence. One concrete and tangible example of empowering citizens in their relation with Big (Open) Government Data is setting-up so-called 'hackatons' where students, programming hobbyists and developers are challenged to create applications on top of the data. They are brought together and allotted a fixed timeframe to develop a prototype or mock-up of an innovative application within a city context. For example in the city of Ghent, hackatons called AppsForGhent[12] put Ghent citizens at the steering wheel of the innovation process as they themselves actively create and develop innovative applications.

Secondly, the construct 'membership' and its relation to big data is elaborated on. We will argue that membership, or having the feeling that one belongs to a community, is

Influence **Membership** **Needs fulfillment** **Emotional connectivity**

Figure 1: The four pillars of a Community

beneficial for bidirectional exchange of data and for making the life of cities' inhabitants more comfortable and safe. One example of an initiative supporting this 'feeling of membership' in a smart cities context is the city game ZWERM developed in the context of the EU ISP-CIP project SMARTiP[13]. The city game, amongst others, supports neighborhood cohesion by distributing ZWERM RFID cards to the participants in the game and by providing city furniture which can act as a physical rallying and meeting point.

The next section of the document will describe how 'need fulfillment' reflects positively on thebidirectional exchange of data in a city context and on setting up a sustained relationship. For example, the aforementioned SMARTiP project, provides in its UK pilot a renewable heat crowdsourcing toolkit that allows people and organizations to share a wide variety of experiences in relation to renewable heat technologies and thermal insulation in the home. In a similar fashion the city of Ghent created a 'heat map'[14] which shows its citizens the efficiency of their thermal roof isolation and addresses their needs for more (information on) energy efficient housing.

Finally, the document also covers the construct 'emotional connectivity', the final dimension of 'Sense of Community' that can foster bidirectional exchange of data between citizens and city governments. We will stress city governments' role in creating a virtual information platform and physical information centres where people find information on the city's data ecology and policy. In short, cities should put a face on their data (processes) in order to ensure emotional connectivity with and amongst the citizens in, and visitors, of the city.

8 Carter, D. (2009). Smart Metropolitan Areas Realised Through Innovation & People – SMARTiP, CIP-ICT-PSP-2009-4 project proposal.
9 Lochner, K., Kawachi, I., & Kennedy, B. P. (1999). Social capital: a guide to its measurement. Health & Place, 5(2), 259-270.
10 Kroes, N. (2012). Speech on ePSI conference in Rotterdam March 2012, http://www.youtube.com/watch?v=9Jq4Qy1UeAE
11 McMillan, D. W., &Chavis, D. M. (1986). Sense of community: A definition and theory. Journal of Community Psychology(14), 6-23.
12 See http://appsforghent.be/
13 See http://www.smart-ip.eu/, http://www.zwermgent.be/
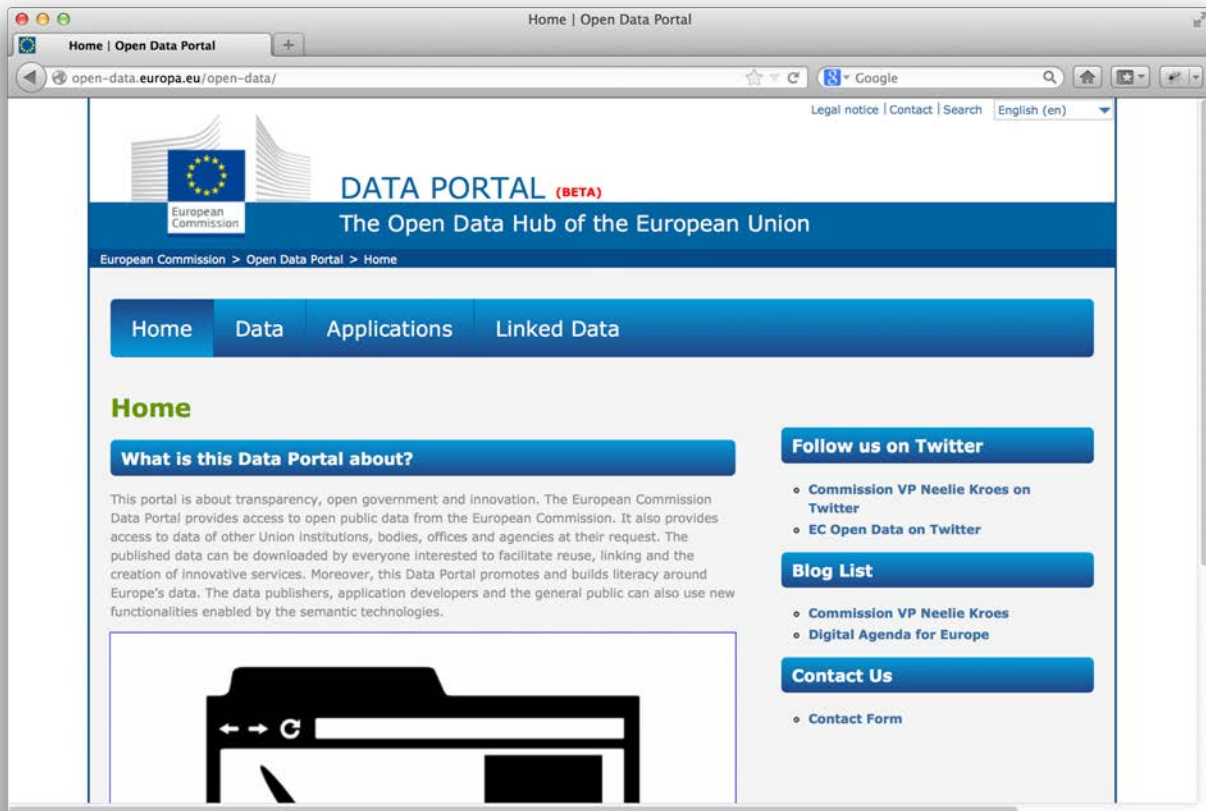14 See http://warmtefoto.gent.be/

2. Influence

As described by McMillan and Chavis, Influence is a key factor which must be considered when building a 'Sense of Community'. This section of the paper will explore how we can consider the traditional elements of influence within a community to create "digital community influence in a bi-directional manner between the citizen and the government".

By design, cities are traditionally not structured in a physical or digital integrated manner; cities are more often not structured and managed in siloed domains such as e.g. energy, transport, environmental, health, waste, water each with its own bespoke data systems where data sharing is not considered a priority. Currently this situation is changing and we are experiencing a rapid and seismic shift towards 'open data' initiatives which is being driven both by EU legislation and local government initiatives that encourage government agencies to publish their data in an open and structured manner and in turn engage with academia, industry and citizens to create applications and services based on these open data sets. In her recent blog, Vice-President of the European Commission Neelie Kroes states: "Imagine we had a resource available that could stimulate new innovations, a market worth tens of billions of euros, and increase the transparency and governance of public life"[15] this conveys the huge potential of open data initiatives we must work to unlock this digital goldmine. The EU open data journey began in 2003 when the EU adopted the Directive on the reuse of public sector information (PSI Directive), which allows for a common framework on how a government body should make their information available to the public. In addition, the European Commission has recently presented an updated Open data digital strategy for Europe setting out guidelines for open data initiatives[16]. Recently the open data hub of the European Union has been launched[17].

15 See : http://blogs.ec.europa.eu/neelie-kroes/opendata/
16 See : http://ec.europa.eu/information_society/policy/psi/index_en.htm
17 See : http://open-data.europa.eu/open-data/

The aim of the initiative is to all the data contained in the portal to be "downloaded by everyone interested to facilitate reuse, linking and the creation of innovative services". The European data portal also looks to build literacy around European data and allows data publishers, application developers and citizens to develop new functionalities and services based upon these data sets. Releasing Open data requires effort from government departments and therefore must start with a compelling business case. The main business case drivers include economic development opportunities and improved cost saving and efficiencies for government departments by harnessing the skillset of 'technical-savvy citizens' by use of app development competitions and hackathons. As described in the introduction section of this document hackathons are a mean to provide both professional and amateur developers with citizen contributed and open government data, with the aim that these events will fuel the creation of innovative applications. The event itself is a so-called 'hackaton' where developers are challenged to innovate and create the best application. All participating teams are brought together and allotted a fixed timeframe to develop a prototype or mock-up of an innovative application within a city context. Clearly, hackathons can empower people, as they put the citizen or 'user' in the driving seat of the innovation process; they themselves actively create mash-ups and new applications, both for commercial and non-commercial purposes, which lead to and foster bi-direction exchange of data between the citizen and the government.

There are many other initiatives which foster the bi-directional exchange of data between citizens and government, by taking advantage of sensors which are embedded in citizens smart devices such as cameras, PS, accelerometers, microphones, etc. One such example is Noise tube[18] an application that allows citizens to turn their mobile phone into an environmental sensor and participate in monitoring noise pollution within urban environments. Such apps provide clear motivation for citizens, local governments, city planners & researchers to participate. They are often open source code projects which in turn allow for growth, influence and participation amongst the developer community. This way of working drives innovation as anyone who wants to contribute to the project and progression of the application development can.

In order to create and drive the bi-directional flow of information between the citizen and government, influence must also be bi-directional. Governments must buy into the potential value of making this data public. However this alone is not enough. Governments must influence and encourage the citizens and communities to develop these applications and services via incentivisation strategies such as free transport or a museum pass and also adhere to EU Legislation and commit to release the data in a structured and organised manner. The citizens and developers can also increase their sphere of influence by providing feedback on what data sets should be released as a priority and highlight which data sets are the most useful to the government. As these data sets create most interest, bi-directional communication is a key factor in order to create and maintain bi-directional influence within this community.

3. Membership

Membership is one of the four pillars necessary to build a sense of community among the population of a city. Using new technologies available, a city government can meet the five different attributes that constitute membership: create a common symbol system, a sense of belonging and identification, emotional safety, personal investment and boundaries[19].

A concrete idea tocreate a sense of membership among the inhabitants of a city is the setup of a unique SmartCard for all the public services offered by a city and the creation of incentives directly linked on the possibilities offered by this card.

What is a SmartCard? One of the technologies available today for the city government is the RFID Card, also called "SmartCard". RFID stands for *"Radio-Frequency IDentification" and it refers to "any electronic device that uses radio waves to facilitate the communication of data for the purpose of identification, and sometimes to locate and/or sense the condition(s), of animate and inanimate objects"*[20]. RFID is already used in commerce (product tracking, payment by mobile phone…), transportation, passports… When a card is equipped with a RFID chip, it is called a"SmartCard".

Several cities have already implemented SmartCards to access their public transport system[21]. This system presents different advantages compare to conventional ticket readers: it allows ticket validation without direct contact to a reader, fast access to the services and lower maintenance cost of the card reader. The London Oyster card is a good example of the additional services that make an RFID card "Smarter": this card is able to store Travelcards; Bus, Tram season tickets and "pay as you go" credit, itautomatically works out the cheapest fare for all the journeys in one day[22]. Visitors and inhabitants of London both have an interest to use the services offered by this card.

Another good example of use of an Smart-Card at city-level can be found in Amsterdam. The city of Amsterdam developed a specific RFID card for tourists: the card gives visitors unlimited access to the public transport system, free entrance to museums and attractions plus different discounts[23].

The SmartCard technology can be used to create an identification system and a symbol at city level if a city government decided to implement a unique card for all its inhabitants based on the RFID technology. Concrete examples of successful implementations already exist for specific services or group of persons. The scope of application of a

18 See : http://www.noisetube.net/#&panel1-1
19 Ibid.
20 Definition of the RFID Network (http://rfid.net/basics).
21 Examples: Brussels Capital Region (http://www.stib.be , Paris (http://www.navigo.fr/), Berlin (http://www.bvg.de/index.php/de/index.html), London (https://oyster.tfl.gov.uk/oyster/entry.do)...
22 "What is Oyster?" http://www.tfl.gov.uk/tickets/14836.aspx
23 See: http://www.iamsterdam.com/en-GB/experience/deals/i-amsterdam-city-card

RFID card can be extended by city governments to create a unique "SmartCard" that gives access to all the public services offered to the citizens by a city (public transports, swimming-pools, libraries, social services…).

A SmartCard at city level can create a feeling of membership through different advantages:

› It is the key and a symbol – in a material way – to access public services;

› It makes the access to the public services easier (validation without direct contact, a single subscription to the services…);

› It gives valuable information to the city governments about the habits of their citizens through aggregated data. Indeed, today the different services available within a city are built on separate infrastructure and there is a need for interconnected services. A unique SmartCard for citizens is the key to ensure data concerning public services is integrated.

A city SmartCardis also more than a simple way to access public services: it creates the potential to be used by the city governments as a tool to create incentives toward citizens for motivating them to share their data.

Indeed, a city often seeks cooperation from its citizens to resolve issues or offer better services.

It is possible to give a concrete example to illustrate the "traditional approach": if a city wants to improve the way waste is collected, the employees of the city can measure in which street the amount of waste is the higher and set up a more efficient itinerary and calendar for the waste truck. It is a slow and costly approach.

If, during several days, the citizens indicated to the authorities how much waste they threw out, the same data would be collected by the city but at lower cost.

Facebook page of Zwerm project

24 See: http://hplusmagazine.com/2010/03/25/gamification-turning-work-play/
25 See: http://www.enterprise-gamification.com/index.php/start/3-examples/53-making-surveys-more-fun
26 See: http://www.forbes.com/2010/10/28/education-internet-scratch-technology-gamification.html?boxes=Homepagechannels
27 See: http://enterprise-gamification.com/index.php/en/facts
28 Gartner Says By 2015, More Than 50 Percent of Organizations That Manage Innovation Processes Will Gamify Those Processes (See: http://www.gartner.com/newsroom/id/1629214)
29 See : https://www.zwermgent.be/

In order to convince citizens to give this type of information, a city could grant discounts and/or rewards on the SmartCard. Those discounts/rewards could take the form of free accesses, cheaper subscriptions, possibilities to invite a friend…

The sense of belonging plays an important part in membership and the members of a community have to feel personally invested in the activities and challenges their community faces. The existence of a SmartCard makes it easy to grant discounts and rewards to the citizens, to motivate them to help the city governments.

Another idea to create a feeling of membership among the inhabitant of a city and to encourage them to share their data is the "gamification". Gamification is already used today in different non-game contexts (training programs[24], surveys[25], education[26],…) and it is known as a tool which improves users' engagement[27]. A research from Gartner identified four principal means of driving engagement using gamification techniques[28]: velocity of feedback loops increased; clear goals and rules of play (empowers players to achieve goals); a compelling narrative engages players to participate and achieve the goals of an activity; challenging but short-term and achievable tasks.

A "gamified" SmartCard at city level can be offered by city governments and it is possible to think of different types of game-based incentives:

› Different colors of SmartCards based on points or cooperation levels of a citizen with his city;

› A number of points allowed (depending of the data a citizen shares);

› Setup a ranking of the most cooperative citizens or neighborhood;

› Access to new avatars or gifts based on the ranking;

› …

In the city of Ghent, the ZWERM project illustrates how it is possible to support the sense of membership within a community thanks to RFID Cards[29]. In this game, citizens represent their neighborhood and have to collect points with their SmartCard (SWERM-kaart) and increase the points of their neighborhood by walking along trees equipped with RFID tags or whistle at sparrows located in the city.

A unique SmartCard to access public services, the set-up of incentives to stimulate citizens to share their data and a gamification of the SmarCard are three ideas that can improve the feeling of membership between inhabitants of a city.

4. Need fulfillment

Whether at personal level, job level or city level, need fulfillment plays an important role in setting up a sustained relationship. At a personal level, an intimate relationship is established if people feel that they are fully and correctly understood by the other person and if they experience this as a positive effect. At job level, job satisfaction is determined by personal job involvement and establishing an intrinsic motivational level. Likewise, at a corporate company level, a sustained relationship is maintained by fulfilling the customer's needs.

Also at city level, one wants to create an dynamic relationship between the city and its citizens, where the aforementioned characteristics play a crucial role. This includes a sense of affiliation with others, recognition, trust and a sense that the use of big data is fulfilling some of their needs. People typically will have less objections in opening up their personal data if they perceive and acknowledge that they will get something positive in return. Similar to other relations, repeated interactions will lead to a more trusted dynamic relationship which lead to needs fulfillment. This means that setting up a smart city involves a continuous effort to interact positively with their citizens, such that they perceive that their individual contribution has a positive net effect on the overall community.

One of the key elements in stimulating open data is that citizens will have access to the data that has been made public and the new services that have been developed. As not all citizens are software experts, attention needs to be paid to create understandable visualisations of the data.

Open access to data will encourage app developers to be innovative and develop new software for the community. The proposed city blueprint should include safety measures on how to secure the shared data. Typically, end-users want to safeguard that their data is being used in applications for which it was originally not intended. To this extent, it is key that there is open communication from the city to the citizens, such that they are regularly informed on how the data is used to build a smart city. Also, regulation bodies and legislation might need to define specific laws that prohibit the misuse of personal data (typically if it can be proven that the use of personal data is at a disadvantage of the person that made it available for other purposes).

The city blueprint should contain guidelines on how city governments could stimulate investments in sustainable and smart projects. The various projects could be in the domain of energy consumption, e-government, education, health-care, traffic and mobility, telecommunication infrastructure, and so forth. Each project which is meeting the criterion that it fulfills the need of a majority of the citizens should be considered and will likely face a growing support of the citizens.

As an example, let's have a look to Belgium, where the technology federation Agoria has investigated the political agreements for the next legislation of 15 major cities. Out of the different projects, the city Ghent seemed to be leading by example (source: www.standaard.be). The city has made a detailed plan to reduce the CO_2-emission by 20 percent by 2020, which is clearly fulfilling the expectation of citizens to strive towards a sustained environment. As first steps into this direction, Ghent has made investments to visualise the loss of heat and energy in the city by means of establishing an energy heat-map.

---

30 See : http://warmtefoto.gent.be
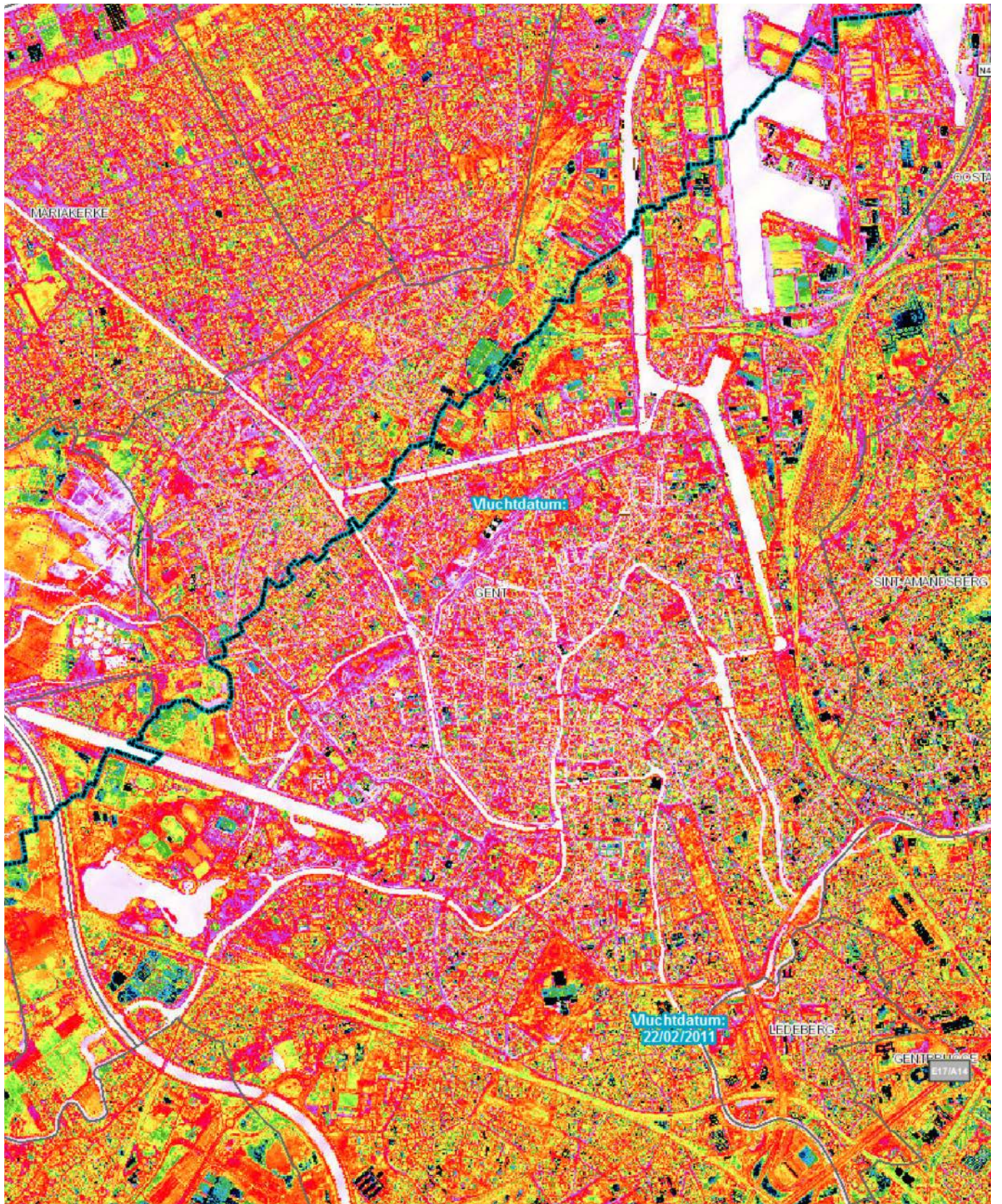
Heat-map Ghent

5. Emotional connectivity

At a personal level, building an emotional connection requires time and trust for the two parties to feel in confidence to share information about themselves. It has to be a two way street so that there is a sense of reciprocity in sharing information which relate to personal and even intimate spheres. Honesty is also a key enabler since this guarantees the fact that the connection is reliable and trustworthy. Building an emotional connection fundamentally requires the means to communicate to each other, i.e. to have the communication means to share information.

Transposing the emotional connection at the level of interaction between a city and an individual, it is worth noting that this is a factor which is a measured known concept in the United States via the Gallup Soul of the Community survey which measured the attachments of citizens to their communities. Peter Kageyama an international expert on community development states that "what makes people emotionally attached to their cities are things that make a city fun, like the arts, culture, design, landscaping and events". However, politicians do not wish to appear frivolous and insensitive to fiscal challenges, so they often say "no" to these initatives, whilst this emotional attachment has economic and social benefits. So how could emotional connectivity be created between a citizen and his/her city government and how could big data help in this respect together whilst also creating a sense of fun?

The proposed answer is: *by putting a face on the data.*

On the image of Space Invaders tiles which flourish in a great number of cities around the world, communities within cities should be recognisable through a graphical design which create a sense of belonging and recognition whilst also creating a sense of fun amonst the community. The use of these designs should be associated to the pride of the communities' members in creating value by collecting and/or analysing data for the benefits of their community or more largely to their cities. In the example of the city of Ghent, those participating to the Heat Map should be tagged as part of the community Heat Map for less CO2.

The storing of data shall be performed by

the city government and availability of this data needs to controlled to end users. These communities should be clearly identified as addressing an issue faced by the city in an innovative manner through the use of big data. To create visibility for these communities, next to the sticker NO AD on the mailboxes a Data Friendly Logo(s) of the communities to whom the mailbox belongs could be added. This should facilitate and lead to interactive discussions in the hallways or the lifts of any apartment buildings.

City challenges

City government

Data created

City infrastructure

Ombudsman

Community

Platform of
exchange and
analysis

› Support the branding

› Identify and reward
the ambassadors

› Better identify the challenges

Data available

Advantages of the bi-directional exchanges in the context of the emotional connectivity

The role of the city government would be to create a virtual information platform (internet site) and possibly physical information centres where people could get information on these communities, which data the community has agreed to disclose, which actions the community supports, how the citizen could contribute and what are the results achieved to date by the community. The city should also nominate a Community data liaison officer who's remit is to promote the initiatives and results of the communities achieved through the sharing of data and as well act as an Ombudsman in case of reported misuse of data which were shared by the citizens with the city governments.

The virtual platform could also be used as a tool for the communities to exchange information between their members and amongst themselves. And the analysis of this information by the city government could lead to the identification of new needs and opportunities. So instead of having interest groups trying to influence the city governments in their decision, the analysis of the organic development of this platform of exchanges would allow a more fine grain analysis of the needs (not biased by partisan interest) and eventually propose concrete means to those communities members to solve the issues. As such the communities would now be able to voice in their innovations to to create better conditions of living. This platform would also allow the identification of 'champions', i.e. those citizens who are dedicated in their actions to change things in terms of tapping in the potential of using big data. Those 'champions' could then become the ambassadors of the Data Friendly initiative within their network.

In order to build trust in those monitored exchanges between communities' members and for the storing of data by the city, the city government should issue a code of conduct to preserve the confidentiality of the data and their controlled use, in particular preventing the crossing of data which could lead to the identification of individuals without their consent.

6. Conclusion

Within this document we wanted to emphasise the impact and repercussions large-scale data harvesting and processing has – and will have – on the life of ordinary citizens (in their private spheres as well as in their professional contexts). As huge amounts of big data represent a strategic and economically relevant asset which might result in a centralized power and in information asymmetries in which public and private entities (who collect, manage and create data), intermediaries (e.g. ISPs) and entities who have the skills to manage an overdose of information are at the helm, we argued for the needs to foster a bidirectional exchange of data between citizens and city governments. After all, the transition from relational, SQL-driven databases ('small data') to custom-built, NoSQL systems ('big data') has occurred behind the scenes out of view of the city inhabitants.

In this document we stated that in order to create such a bidirectional exchange of (big) data, a certain level of trust and cohesion between citizens of smart cities and city governments is needed. Based on the phrase 'Sense of Community' we created a conceptual framework that addresses challenges and solutions related to big data and influence, membership, need fulfillment and emotional connectivity. This framework should materialise in concrete actions and policies for European data rich or 'smart' cities, encompassing amongst others; the creation of a Data Friendly City Label, the appointment of city Data Ombudsman, and the development of a corpus of initiatives merging communities and big data to address city challenges. The European Commission could play a major role in this process through the implementation of a blueprint which focuses on smart cities and smart citizens. The set of ideas described in this paper is certainly not exhaustive but aims at inspiring and informing public authorities about the possibilities offered by the new technologies related to big data within the boundaries of a city.

The presentation of such a blueprint by the European Commission –added together with the promotion of cities already engaged on the path to become Smart cities– can guide European cities to foster bidirectional exchanges with their citizens and face future challenges while making the most of today's innovations related to big data.

KNOIT
A Knowledge Capture and Dissemination System

Katalin Babos

Paul Docherty

Márta Domok

Sergi Figueres

Tabajdai Réka

Abstract: *This text describes a system for enhanced knowledge sharing within or between organizations, called KNOIT. The idea is elaborated based on an analysis of already existing solutions. Furthermore, end users' as well as organisational benefits are discussed and potential routes of commercialisation defined.*

1. Introduction

The initial idea of this project is to help knowledge workers (particularly scientists) working either in industrial, commercial or academy spheres to share their knowledge and experiences in an efficient and effective way. A number of web sites owned by publishing companies or created by scientists can be found where scientific papers or patents are published. In some cases the search on these sites is free but usually one might face some kind of difficulties when trying to access these data due to their origin, the way of publishing, copyright, etc. Further, it is often the case that it is hard to access the data even within a company. Extant websites and the conferences organised on a given topic provide some opportunity for the scientists and experts to share their activities, results and successes with others, get updated about the worldwide trends and implement new ideas and concepts into their work. However, and in spite of the latent potential in this kind of sharing information, current methods are not effective as they aim to be. This is due to the conditions of accessibility or, for instance in a publication, because the set of data is not complete or only derived data can be found.

The aim of our project is to help the knowledge workers and experts both from the academy and industry to share effectively their knowledge and learning in a well structured and comprehensive way. We would like to make it traceable how the raw data was gathered in the first instance, and then transformed into information and latterly knowledge (**value creation**), critically making transparent its reverse process

(**accessibility**). Namely, if one faces the outcome or the result of a development (knowledge) one could search for the different stages of the development process (information) or even the raw data that initialised that specific development.

Sharing knowledge is a problematic issue because there is no exact, implicit and universally consistent methodology accepted for this purpose. Whilst one part of the knowledge may be searchable and measurable, lies in written form or consists of data which were obtained by measurements, other domains frequently reside only as latent (or tacit) knowledge as a result of a long intellectual process. Knowledge can be summarised by statements and facts but this substantiation of facts is not implicit. It does not show how these facts were derived, what the starting points were, what was added to them during each development phases and how the outcome emerged.

This lack developmental documentation is at its most critical failing when transfer of knowledge between parties is the desired outcome. Should a third party want to adopt knowledge and utilise it, they would go through the same process - face the same phases, raise the same questions and give the same answers - which the originator has already gone through. It is very easy to realise that all this might have serious consequences on a company's ability to leverage success, act and react, profit, and maintain a competitive proprosition.
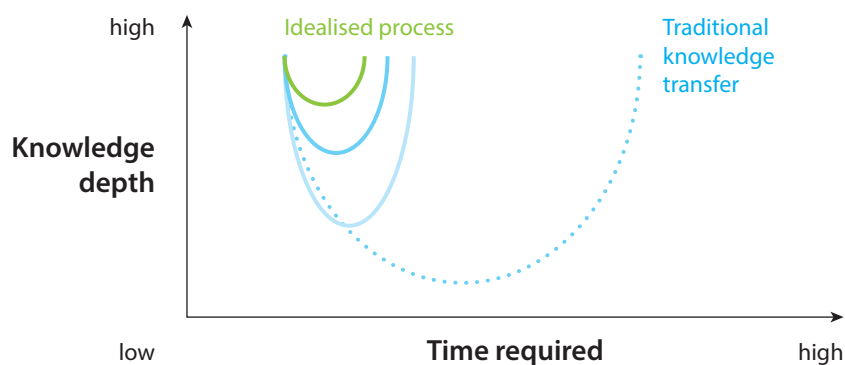

Figure 1: Knowledge dissemination between two entities. In an idealised process, dissemination of knowledge between two parties operates through simple articulation of the conclusions of a data-process. However, this is rarely possible – traditional methods of knowledge transfer generally require an unbundling of the knowledge – information – data journey and later re-establishment of the conclusions. However, in Knolt it would be possible for the user to choose the depth of detail which they require to understand the articulated message.

2. Definition of the problem

Based on the introduction phase it can be seen that we could realise these inefficient way of working and we decided to try to find a proper solution for this. At first, a brief question was framed, which can summarise the identified problems.

'How can we develop a robust system for portability of logic to improve organisational efficiency?'

During our work the aim was to analyze this problem from as many aspects as possible. To further define and scope the topic, a commonly used framework (*'the Five W's'*) was adopted, featuring five subquestions:

Who can develop this robust system?

During our conversation it was concluded that almost everybody can contribute to this development. Within a company not only the individuals but the working teams and the company itself can take part in the development. Moreover. universities, governments, international institutes can also be involved in this solution.

What this 'robust system' really is?

This system is a new *'standard'* method, with the help of which logic sharing will become much more efficient. In our approach *'standard'* means an externally validated process throughout the whole system. This does not necessarily imply that this system is based on ISO structure (or on other international standard structure), but we realised that later on an ISO system can be diffused easier. It is also very important to highlight that this system is more than just a data repository. By building semantic links of data during the knowledge journey it captures the decision making process and thereby builds transparency of process.

Where can this robust system be used?

As mentioned above, this system can be used by knowledge workers in any kind of organization. However, this tool is likely to provide greatest organisational efficiencies between data-driven transnational teams. In our context transnational means globalised, that the teams (stakeholders) work in different places in different time.

When is it practical to use this robust system?

This system should be designed to facilitate collaborate between remote partners, to exchange complex ideas, concepts or technologies. Moreover this system allows stakeholders to build diversified relations and trust. So if an organisation has to face similar problems, it will be important to solve them as soon as possible, for example with the system we offer.

Why should this robust system be used?

This system is suitable to improve organisational efficiency by speeding adaption and build trust. With the help of this tool it is possible to improve the organisation's competitive benefits. It makes it possible to operate in a well definied methodology which also reduces risk.

2.1. Existing solutions

As described above, knowledge sharing, company efficiency improvement etc. are key drivers across all organisational sectors. Many different solutions to these issues have been created in the last decade, driven by the increasing availability of electronic resources. Therefore, a limited study of existing solution types (representative of their solution category) was conducted. The aim of this research was to understand a baseline of capability, and establish the main advantage of a future system in comparison. This paper introduces the most relevant applications we investigated.

Mendeley: *'This is a free reference manager and academic social network that can help you organise your research, collaborate with others online, and discover the latest research'.* Mendeley is a good solution for making research work much more efficient within the organization, but is focused primarily on academic literature, and lacks the social networking feature required.[1]

LinkedIn: This application is used by professionals to exchange information, ideas and opportunities. One can stay informed about people and industry, find required information and knowledge and in control of his/her professional identity online.[2]

Google Apps: *'This environment is a cloud-based productivity site that helps individuals and teams connect and get work done from anywhere on any device. It's simple to setup, use and manage, allowing you to work smarter and focus on what really matters.'* Google Apps consists of many applications like: Gmail, Calendar, Drive, Docs, Sheets, Slides.[3]

Our research revealed that all these applications have some disadvantages, which may be eliminated with our solution. By design, Mendeley appropriates only for research coordination. It facilitates collaboration with colleagues, but only throughout the pdf files. LinkedIn fits perfectly for professionals to expand their online network, but it is less suitable for data & document sharing. Google Apps includes many of the advantages that Mendeley and LinkedIn have, although it uses different applications. Contrary our solution KNOIT is an individual application which integrates all the beneficial features of the above mentioned applications.

---

1 http://www.mendeley.com/
2 http://www.linkedin.com/home?trk=hb_home
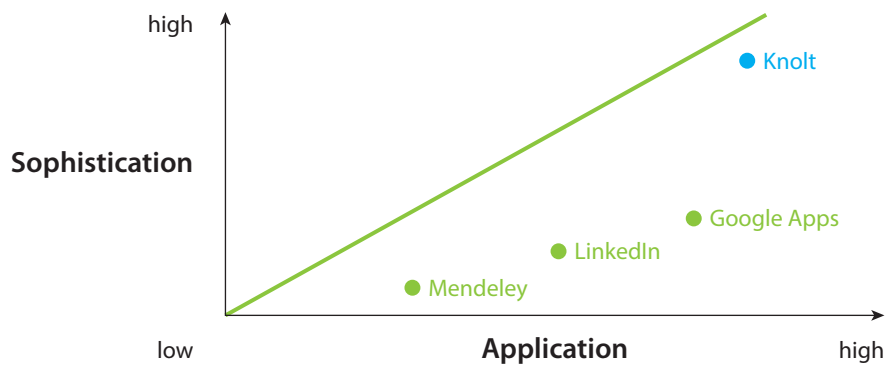3 http://www.google.com/enterprise/apps/business/

Figure 2: Existing solutions compared.

3. The solution: KNOIT

Of primary concern to the user is finding the easiest way for the user to enter data into the system. If the user is required to manually upload and annotate material, such as user profiles, project details (active data capture), there is a negligible advance over current solutions. However, should this data be captured and used in a user-passive manner[4], barriers to systematic adoption and utility are considerably lowered.

An example of user-passive data collection is Google and their Now product which abstracts user information, interfaces and activies in a user-passive manner. An example from the Now product is the filtering of user data (particulary email) for high-relevency information, such as travel plans, flight bookings and hotel reservations. The Now application uses this key information to present the user with a contextual information interface, providing travel updates (delays, cancelations), airport information (gate numbers) and city mapping. By collecting the user data in a passive manner, the user has an extremely small (or even zero) barrier to system adoption.

Similarly, the KNOIT system would use automated passive information capture, collecting social data, such as a people-network of those a user communicates with about work, their discussion topics, recent research activities, academic papers, and everything the

user decided to share. The programme would collect and analyse them, and transform into a structure to be traceable and understandable. Importantly, the application would allow users to set a privacy ranking to data, allowing the choice of accessibility and levels of access (e.g all data, or just data structures and conclusions). Of course, such privacy / security settings could be preconfigured in an enterprise manner.

So, in a nutshell, it is only necessary to choose some criteria and then the computer does the work and makes the data package sharable. It makes up the references and creates a work history, so anybody who reads it can easily understand the choices and way of thinking. This journey can build up trust between clients and it makes the spreading of products or techniques more efficient. The standardised way of introducing is another way of building trust.

The second step is the question on how the package can be found on the site. The solution was inspired by the information presentation startup Prezi. The focus of Prezi is on end-user presentation and mapping of information in an appealing and attractive manner. Similarly, KNOIT would present information packages in a variety of manners, such as a matrix with a fractal structure. When accessed by the user he/she would initially see highlighted main topics in a matrix format. After choosing a favored topic the user would 'zoom in' another matrix which

---

4 http://www.google.com/now/
5 http://www.prezi.com

contains linked sub-topics in the main topic – the user can 'zoom' in and find the most searched point. In every step the user would see the privacy rankings of linked points and efficiency rankings given by the users. They would see how many users have read that topic, and how many other topics are related. By presenting this information in a contextual and visual manner users will quickly discover semantic links between data sources.

Further, there user's interaction with the knowledge base would be monitored by automated passive data collection routines where allowing users to map their searching path and location in the system. By doing so, common paths and routes would become more established, building both trust and transparency in knowledge journeys.

The whole system has three supporting pillars: easy way to create a database and see through it, build up trust between clients, and save up time to take over this process from product, ideas or techniques to the target points and to realisation. However, the priniciple requirement of organization adopting such as system is likely to end-user time-savings, which are best understood from the perspective of such a user:

Firstly, the user does not have to waste your time to uploading semantic data, your working histories, preferences. You do not have to explain everybody your way of thinking, because within the system everybody can see it (**trust building**). You do not have to elaborate every detail of process with your co-workers, because if they are connected you can see their works and they can see yours, and you can be all integrated in an easier way. So within the company the communication can be faster (**employee monitoring and self monitoring**). The target points do not have to waste time to try out that your product or technique is trustworthy or not, because they can see your efficiency rankings, your way of thinking, your working history and your preferences instantly (**co-operation**). You do not have to waste your time to search or ask focus groups about people's pretences because if they are in the system you can see it. You would know the best way to advertise your products in different companies, organizations or in different countries.

Last but not least, you do not have to roll up and down in the site, because it is a visualised system which make the searching process more convenient and it can be easily transferable to the touch-pad system.

4. Business plan & business model

As any new project that wants to be converted into reality, and wants to become a smart tool for organizations and companies, KNOIT has behind it a business plan and a business model. Starting as an elevator pitch, we would say that KNOIT allows teams and organizations to share and transfer knowledge, converting data to knowledge. KNOIT, thus, created the right context for sharing key insights, relevant data and useful hints.

The market size is huge. Any organisation, company, NGO, government can use KNOIT and can help improving their organisational efficiency. Obviously, KNOIT will use an organic growth and will start for small medium organizations, just to get some success case. With that objective reached, it would be easy to start collaborating with a blockbuster.

The principal research contributor at Wikibon.org, Jeff Kelly, focused on trends in big data and business, says that the big data market 'is on the verge of a rapid growth spurt that will see it top the $50 billion mark worldwide within the next five years'.[6] And it is a very fragmented market: for example, the market leaders by revenue, IBM, Intel and HP, just have less than a 3% of the total amount. So, a tool like KNOIT, that can help them being more efficient while sharing data, and big data, and converting them to knowledge, will be a fundamental partner to these organizations. Summarising, we can say that the potential market for KNOIT is incredibly huge.

Relating to the business model, answering the question '*how do we make money?*', we can say that we have different value propositions and different strategies to make money from KNOIT. It will be a commercial task, to negotiate with the different organization which option do they prefer, to decide

whether one or the other will apply. Following, a list of the different options that will be used at the different phases of the project:

1. **Commissions.** Charging each organization a determined amount of money per operation realised through KNOIT.

2. **Members.** Establish a monthly quote per user or per organization (that allows each organization to use KNOIT for a determined number of users).

3. **Freemium – Premium.** Some actions can be done for free, using a freemium model, and some others, that give more value to the user, have to be realised by a premium model. That will also be a strategy to get new members and start making KNOIT well-known all over the world. At the first steps, users will starts becoming members of KNOIT for the added value and for the freemium model. At the second step, and appreciating the value proposition, a % of the users will start paying for the services and using a premium model.

4. **Sponsorships.** A big data company can sponsor the KNOIT project, becoming partners.

5. **Alliance** with a global organisation, like the United Nations or the World Bank. That can also be developed to build trust and brand awareness all over the world.

---

6 http://www.wikibon.org

Our young leaders

30 talents joined the first Young Leaders Initiative to share thoughts, get new input on big data and work on their individual data-related project ideas which are reflected in the texts beforehand. They are coming from all over Europe with various professional backgrounds.

Tommaso Alderigi

Tommaso is an entrepreneur in the renewable energy sector. He has a business background focused on energy and finance obtained at top European business schools.

Katalin Babos

After graduating as a Chemical Engineer in 2012 at the University of Pannonia, Hungary, Katalin started to work with MOL as a technology developer for the Environmental and Corrosion Protection Department. This year she changed to the Research and Innovation Department and is responsible for an ongoing R&D project.

Guillaume Delalandre

Guillaume (L.L.M. International Business Law) is a Project Officer at BNP Paribas Fortis in the Operations/Engineering department. He studied law, international relations and political-sciences in Europe and Canada, and worked freelance as an IT specialist. His work is today focused on the architecture of securities transfer system.

Paul Docherty

In 2010 Paul joined AkzoNobel as a Project Leader, where his interests in data-driven science, statistics and informatics lead him to taking a role as a knowledge Management Leader. Prior he worked for Astra-Zeneca in the Infection category, targeting diseases such as Hepatitis C and Dengue fever, as a Medicinal chemist. Paul holds a PhD in organic chemistry.

Nadja Dokter

Nadja works at the European Institute of Innovation and Technology (EIT) as a policy officer dealing with the EIT's knowledge triangle integration agenda. Prior to this, she worked at the European Commission, Directorate General for Enterprise and Industry, on EU innovation policy and support to innovation. She has a background in business administration and innovation management.

Márta Dömök

Márta Dömök is a chemist working at MOL Plc. as an R&D expert being responsible for managing R&D projects in MOL Upstream. Prior to joining MOL she worked at the University of Szeged (Hungary) where she got her PhD degree in chemistry and spent a year at the Technical University of Munich (Germany).

Niclas Ehn

Niclas is an entrepreneur at Expektra, Sweden, in the smart grids arena. He has a background as energy systems engineer with a past experience as business analyst with companies as Adeptra and Fico.

Sergi Figueres Moret

Sergi is a social entrepreneur; he is co-founder and CEO of Worldcoo, a platform for co financing and micro financing of engineering projects. Sergi holds a Master in International Relations and a degree in Law and Business Administration and has worked for different banks in marketing and international areas.

Amitte Gulamhussen

Amitte M. Gulamhussen has a Technological Chemistry Degree at the Sciences Faculty (University of Lisbon). After an Erasmus period at the ICT-Prague in Organic Chemistry and Catalysis he went to the Nuclear and Technological Institute (ITN-Portugal) to work on Electric and Magnetic Materials in 2005. He joined Solvay Portugal in 2007, where he has been working as a Laboratory Manager.

Nina Gumzej

Nina works at the University of Zagreb Law Faculty with her current position being Senior Assistant- Lecturer at the Chair of Legal Informatics. Prior to this she worked in the largest Croatian telco where she mostly worked on novel privacy and data protection challenges and was the first company-appointed Data Protection Officer. Specially inspired by the Young Leaders project exercise Nina will continue with further academic research in data-driven innovation with a focus on consumer protection (including privacy) open data strategies.

Wouter Haerick

Wouter received his Master degree in Electro-technical Engineering from the Ghent University, Belgium. In 2009, he obtained a doctoral degree in the field of computer science. At this university, Wouter Haerick currently establishes smart energy innovations programs with leading industrial companies. Wouter Haerick obtained an MBA at Vlerick Management School and is co-founder of a brokerage company.

Robert Hirt

Robert Hirt has a Masters Degree in Mathematics and Physics from the University of Munich (LMU). He has helped clients drive business performance by efficiently processing and using information in his job at various Systems Integrators. In 2012, Robert joined the IT Advisory practice at Ernst & Young in the Centre of Excellence for Enterprise Intelligence across EMEIA, focusing primarily on providing insight on information and supporting trusted business decisions through exploitation of relevant information.

Kate Hofman

Kate recently left her career in management consulting to set up her own business selling commercial aquaponic farms for rooftops on London. She is passionate about sustainable food production for cities and interested in the role that big data will play in developing the city of the future and the people who will live their. She has an MSc in Environmental Technology and Business from Imperial College, London and is an alumni of Climate-KIC.

Julia Manske

Julia works as a Vodafone Institute Fellow at Goodroot, a research and consulting agency for the social sector. In her position she mainly supports the Vodafone Institute of Society and Communications, a Think and Do Tank, with academic expertise in the field of ICT. Prior she worked in the area of Corporate Responsibility and Sustainability. Julia studied Social Anthropology in Munich and Mexico-City.

Guillaume Marcerou

Guillaume is a French attorney at law. He finished his Master in Intellectual Property Law and Information Technologies Law in 2006. After his studies, he worked for two years as an in-house lawyer in the media & entertainment sector before he passed the bar exam in 2008. Guillaume is now working as an attorney in the Business Law - IP/IT team of Ernst & Young Law Paris where he assists clients implementing their IP strategy and securing their IP assets.

Jessica McCarthy

Jessica is a senior research scientist with Intel Labs Europe. She is currently with the Energy and Sustainability Lab where she is working on ESLs sustainable connected cities research agenda. Jessica is actively involved in the European Framework Programme(FP) working on various projects such as SLA@SOI and is currently Intel's lead researcher on PLANTCockpit where Intel's focus in on improving energy monitoring and analysis in our manufacturing environment.

Peter Mechant

Peter works at the research group for Media and ICT (MICT) at Ghent University and iMinds. He has been involved in research projects focusing on eculture, Web 2.0 and online communities and has published in journals such as Observatorio, International Journal of Web-based Communities and Contemporary Social Science and coauthored papers in journals such as New Media and Society, Behaviour and Social Networking.

Jeroen Melis

In June 2009 Jeroen Melis obtained his master in Industrial Engineering, focusing on electronics and ICT. A couple of months and a bicycle trip to Spain later, he started to work at Agfa Healthcare in Belgium where he began at the SIV (Systems Integration & Validation) department. One year later he joined a small team that would focus on streamlining complex upgrade processes of Agfa's PACS solution.

Romain Muller

Romain works at the EIT as KICs Project Officer. He has been involved in various projects (including FP7 projects) and change management assignments in the public and private sectors. Romain holds a Master in Electrical and Electronic Engineering.

Hjalmar Nilsonne

Hjalmar is an entrepreneur and Co-Founder of Black Silicon Solar that focuses on a proprietary technology that will make solar cells cheaper to manufacture. He has a Master in Industrial Engineering and management from KTH, Sweden and studied in Japan.

Jonathan Orban de Xivry

Jonathan Orban de Xivry finished his M.S. in Electrical Engineering in 2006 and his Ph.D. in Medical Image Processing in 2010 at the UCLouvain, Belgium. After working two years bringing parts of his research to the clinics, he now joined iMagX, a public-private partnership focusing on bringing innovative imaging solutions in proton therapy where he is manager of the research projects

Sébastien Petillon

Sébastien obtained his Master degree in mechanical engineering in Poitiers, France. After graduating, he joined the plastic sector of Solvay as a Research Scientist in Brussels. Two years later, he transitioned to a role of Technical Support Engineer in the Specialty Polymers business unit of Solvay, and moved to Atlanta, USA, where he has been located since. In the last few years, he progressed into sales & marketing functions, becoming a Technical Marketing and Business Development Manager.

Richard Ramsden

Richard works as an Associate Technologist at the Technology Centre at International Paint Ltd for AkzoNobel. He is mainly involved in chemical and polymer research, as well as rapid concept screening and optimization by using Experimental Design. Richard was also the architect of the first big data product on the coatings industry, utilizing vessel movements and environmental data analysis to predict and specify coatings for optimum. He holds a Masters' degree in Chemistry.

Marc Ruckebier

Marc Ruckebier obtained the equivalent of a master of business degree at the Universities of Düsseldorf and Alicante. He has worked in various positions, such as financial consulting and sales coaching for Deutsche Bank, financial controlling for HSBC Trinkaus & Burkhardt, Interconnection & Roaming (wholesale) and finance decision support for Vodafone. Since 2010 Marc Ruckebier has been working in his current role as Head of Issue- and Reputation-Management for Vodafone Germany.

Pep Salas

Pep is co-founder and CSO of ENERBYTE, an IT Company that integrates consumers in the information flow of the energy supply chain and manages information for the benefit of suppliers, Distributors, Consumers and society. His background is engineering and he is a PhD candidate in technology and humanism at the Universitat Politècnica de Catalunya.

Reka Tabajdai

Réka is currently a PhD. student at the University of Szeged, Hungary, at the Department of Physical Chemical and Material Science investigating on chatalysis, surface, colloid and material science. She is also working and teaching at the University of Szeged.

Halit Ünver

Halit is a Research Associate at the University of Ulm and at the Research Institute for Applied Knowledge Processing (FAW/n), which is a Think Tank for Globalisation and sustainability. As an IT Engineer and an Economist, he is engaged in the dissertation topic 'Global Networking, Communication and Culture – Conflict or Convergence'.

Jonas Vermeulen

Jonas Vermeulen received his Master degree in Telecommunications Engineering from the University of Leuven, Belgium, in June 2006. After graduating, he started working at Alcatel-Lucent as solution architect for digital television projects. He has then transitioned to New Zealand as a project manager, to govern the deployment of an IP-based communications system. After his return to Belgium, he has been part of the technical sales team.

Chris Woods

Chris is an Enterprise Researcher at Intel Labs, Ireland. His background is in cloud and mobile software and has worked for and with companies including Microsoft and Nokia.

Geert Ysebaert

Geert (PhD Applied Sciences) works at Alcatel-Lucent as team lead of the DSL transport team, working on next generation access products. He started with Alcatel-Lucent as research fellow, where he worked on various DSL products. Later on, he took positions in respectively the DSL experts team, the product management team and acted as network systems team lead. He was involved in various government projects and holds several international publications and patents.

The EIT Foundation was established in 2010 as an independent, philanthropic organisation under Dutch law. It is a group of like-minded organisations dedicated to promoting a culture of innovation and entrepreneurship in Europe. Its goal is to complement and enlarge the impact of the EIT, inventing the future by bringing together some of Europe's brightest minds at events, in youth talent programmes and similar initiatives.

The EIT Foundation focuses its activities on the:

› **Promotion of entrepreneurial education** and bridging the gap between academia and business;

› Creation of a **new generation with an entrepreneurial mindset** ready to drive the development of sustainable innovation;

› Development of an **international networ**k of talented professionals engaged in education, research, innovation and business development.

Contact details

Jean-François Dehem
Secretary General of the EIT Foundation

info@eit-foundation.eu
http://eitfoundation.org/

Our partners