Project acronym:	BYTE
Project title:	Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities
Grant number:	619551
Programme:	Seventh Framework Programme for ICT
Objective:	ICT-2013.4.2 Scalable data analytics
Contract type:	Co-ordination and Support Action
Start date of project:	01 March 2014
Duration:	36 months
Website:	www.byte-project.eu

Deliverable D7.3:

BYTE Final Report and Guidelines

Author(s):	Rachel Finn and Kush Wadhwa, *Trilateral Research* Stephane Grumbach, *Inria* Anna Fensel, *Universität Innsbruck*
Dissemination level:	Public
Deliverable type:	Final
Version:	1.0
Submission date:	28 February 2017

TRILATERAL RESEARCH LTD
72 HAMMERSMITH ROAD
LONDON
UNITED KINGDOM
W14 8TH
FEBRUARY, 2017

THE BYTE HANDBOOK

POLICY RECOMMENDATIONS AND RESEARCH ROADMAPS

THE HANDBOOK FOR EUROPE'S BIG DATA FUTURE

Summary of the BYTE project

The BYTE project is a multi-disciplinary, multi-sectoral study investigating the social, economic, political and legislative impacts of big data in Europe. It sets forth a comprehensive policy and capability planning framework for Europe in order to respond to the evolution of Big Data futures. Our goal is to provide guidance to government officials and managers to seize the opportunities arising from Big Data in ways that maximise social benefits and minimise potential negative externalities. Our recommendations are driven by extensive engagement with data and domain experts across European countries, along the lines of a set of research methodologies designed jointly by 11 expert organisations in 10 European countries.

Figure 1: Research Design

In particular, we conducted case work in seven domains – humanitarian crisis management, culture, energy, environment, healthcare, maritime transportation and smart cities. Evidence was gathered through mixed methodology case studies comprising documentary analysis and primary research in the form of semi-structured interviews with data experts and focus groups with domain experts in each sector. These tasks extracted and analysed positive and negative impacts relating to economic, social and ethical, legal, and political spheres for each sector (see D3.2) and conducted a cross-sector comparison exercise (see D4.1) to reveal commonalities and understand common root causes (see D 4.2).

In addition, the project used foresight research to predict the future impacts from the introduction of Big Data. This work also enabled the credible generalisation of study results to a number of sectors outside the current scope of the project. This was important for developing a comprehensive policy and research roadmap to enable European institutions to coordinate their response to its challenges. Full results are reported in D3.2, D4.1, D4.2, D5.1 and D5.2, whereas a full list of recommendations can be found in D6.1 under 'Research' at the project's website: http://byte-project.eu/

Socio-economic and political externalities

BYTE engaged more than 100 experts across seven domains to understand the potential impact and implication of Big Data in their sector.

The crisis informatics case study focused on the use of social media (mostly Twitter) to support crisis management and the delivery of humanitarian relief. The culture case study focused on the creation of an open access 'digital library' of copies of European cultural heritage works. The energy case study involved big data analytics in oil & gas exploration and production in the Norwegian Continental Shelf. The environment case study examined the sectorial opportunities arising from amassing authoritative sources of environmental data (such as prominent earth and space observation portals) with crowd-sourced data. The healthcare case study considered the implications of big data for disease identification and health care innovation utilising genetic data. The smart cities case study examined value creation from urban, infrastructure and citizen data. The maritime transportation case study sought to analyse the use of big data in the shipping, yet due to embryonic state of its development and the lack of perceived value from key stakeholders in the sector and strong proprietary sentiments regarding information sharing, the case study was excluded from further consideration.

Lessons learned from cross-impact analysis in 7 different sectors, from more than 100 expert participants

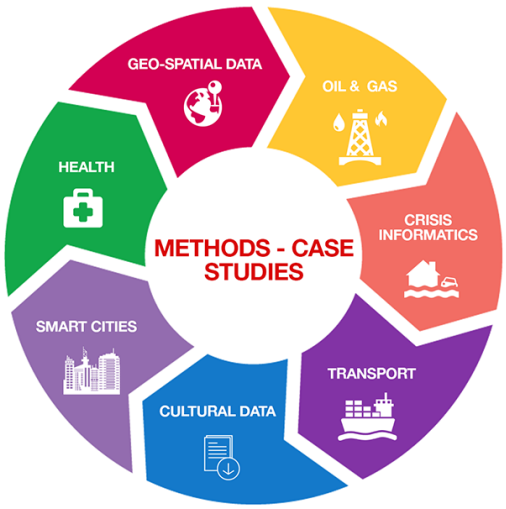

Figure 2: BYTE case study domains

Case study analysis found that data practices require higher levels and regularity of information exchange. Often the boundaries between involved parties become blurred as the locus of value creation and information ownership is dynamic between stakeholders. While greater openness and interoperability underpinning big data practices give rise to a number of positive externalities, it also compromises the effectiveness of existing regulatory and protective practices and industry norms raising a number of potential negative externalities.

Cross-impact analysis revealed the following economic, societal & ethical and political themes requiring attention and further consideration (see figure 3 below for a summary). We present the legal externalities separately in the next section along with our set of legal recommendations.

Economic Externalities

Economic benefits have been a key driver for the introduction of big data made possible by gains in operational efficiency, data-driven innovation and the generation of new knowledge. Indeed, operational efficiency benefited from better informed decision-making, resource optimisations, service quality and operational performance. Open data and standardisation fostered data-driven innovation, despite a level of resistance by private organisations to open proprietary data sources. In addition, new scientific knowledge was generated by aggregating public insight and analysing large datasets. Inevitably, the shift to the big data era will render some of the existing business models and jobs obsolete raising concerns regarding unemployment, though new opportunities will arise from the creation of big data sector for people and organisations. Finally, large investments in big data infrastructure rely primarily on public funding, at least in the beginning, competing inevitably with other pressing social agendas. However, governments maintaining a core role the creation of a big data digital platform might be necessary to balance out the competitive advantage of large US digital platforms and give small European SMEs a chance to innovate and compete.

Social/Ethical externalities

Improved quality and speed of decision-making was a common positive contribution of big data across cases. This, in turn, can be translated into better governance of environmental resources, improved public services, such as crime prevention, or personalised services by private organisations. While, people benefit from better services, big data raises concerns regarding the risk of discrimination and possible inequalities. Biases inherent in problem definition and software algorithms may lead to unintended discrimination, while new bases for prejudice may arise from big data derived health, behavioural or other personal information. Methods to redress such biases are currently in development with the view to be integrated in auditing tools and design approaches. Anti-discrimination legal frameworks are also in development as described in the following section. In addition, while big data promises higher citizen participation through digital means, pre-existing inequalities due to lack of skills, access, infrastructure or language may hinder its realisation.

Political externalities

Our analysis has revealed a digital divide of geostrategic implications. Currently, private US companies are better positioned to exploit public data, particularly US digital platforms with global reach, such as Facebook or Google. While the benefits of their services can be

enjoyed both by companies and individuals globally, it nevertheless raises concerns regarding the monopolisation of big data ecosystem and the privatisations of public utilities. In addition, their dominance in the digital sphere stifles innovation by European SMEs and public organisations and/or creates totally dependent (i.e., locked-in) value chains. Political tensions also arise from European citizens' being subject to potential surveillance outside the boundaries of EU, as well as private data misuse, especially sharing with third parties without consent. This raises potential concerns regarding power imbalances between the two continents. Such platforms can command monopoly powers due to their proprietary technical standards. Hence, the use of open source, open protocols and open data and data portability could counteract their impact as gatekeeping and control mechanisms. This, however, may have ambivalent effects; strengthening their position of dominance instead of counteracting it. In this case capturing the monopoly rent through providing the data through licensing may be a better solution.

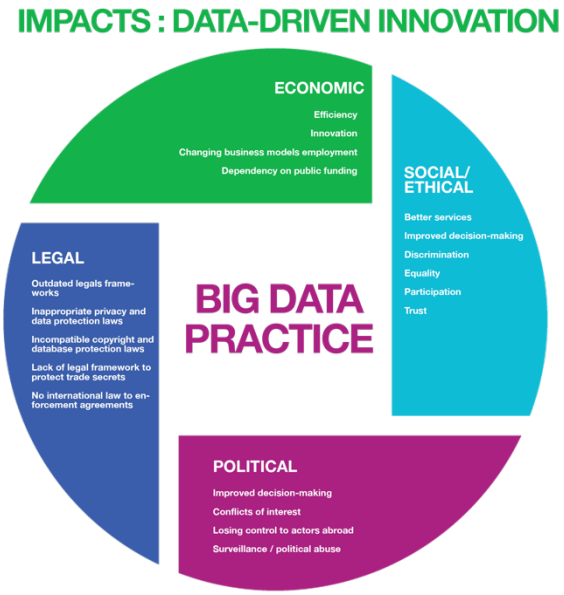

Figure 3: BYTE case study externalities / impacts

Legal issues & recommendations

Legal frameworks lag technological big data developments, raising information transaction costs and impeding the development of appropriate best practices. From an individual's perspective, existing privacy and data protection laws, for example, currently require consent or authorisation for every single instance of data exchange causing delays in amassing large amounts of data and/or analysing their regular data flows. Privacy concerns arose as a critical factor to be appropriately addressed to safeguard public trust in the dig data era. While several issues regarding potential privacy violations were in evidence, case study research showed that people and organisation have tools at their disposal to enable them to adapt their responses to manage their privacy risks. From a business perspective, protection of proprietary company information and trade secrets were raised as common company concerns. Existing copyright and database protection laws are incompatible with big data sharing across organisational boundaries. They are too restrictive in that they do not recognise the fundamental need for simultaneous large scale data exchanges on a regular basis. From a governance perspective, the translational nature big data highlights the lack of international agreements on data management and hence the inadequacy of extraterritorial law enforcement practices.

Recommendations suggest legal developments in three key areas to improve interoperability and restore public trust and legal certainty.

Privacy & Data protection	Promote development of standard evaluation
	Broaden privacy-by-design for legal and organisational safeguards
	Include collective mechanisms of data protections frameworks
Copyright & Database protection	Drop the sui generis right on database protection
	Limit copyright by excluding data mining
	Enable collective licencing agreements
Trade Secrets	Develop new standards around data sharing
	Develop best practice toolbox

Figure 4: Legal recommendations

First, changes in privacy and data protection mechanisms to guard against known privacy risks. To this end, privacy-by-design principles to encompass the design and development of socio-technical solutions, and the integration of such privacy mechanisms into them need to be amplified. Also, the development of collective mechanisms of consenting and licencing need to be enabled and hardwired into big data solutions. Finally, enabling the auditability of such practices, along with a strengthening of data protection authorities, will improve transparency and foster public trust. **Second**, legal developments to protect proprietary company information without impeding big data practices should be developed. For example, dropping existing laws on database ownership and amending copyright laws to exclude text and data mining, yet maintain liability and accountability frameworks would be a useful step forward. Legal provisions should also cater for collective licencing to decrease transaction costs in information exchanges. Third, new standards for the protection of trade secrets should be considered along with the development of a practical toolbox of legal, organisational and technical means to regulate data flows in accordance. Finally, the harmonisation of legal frameworks at an international level will be required to facilitate adequate law enforcement.

Key policy recommendations

Policy recommendation were driven by a European vision for big data and the key priorities:

- **Develop an EU wide data policy, with clear rules about protecting citizens' autonomy and empowering them to claim for the respect of their privacy rights;**
- **Develop a pan-European data infrastructure, and**
- **Utilise big data for the greater good, with a fair distribution of benefits and resources management responsibilities.**

These raise expectations from policy makers and regulators for immediate attention and action translated into key recommendations:

EU wide data policy:

A prerequisite to achieving the big data vision is an EU-wide data policy to facilitate intra-European and extra-European data transfers and foster transparency and innovation. Aspired long-term outcomes include: the signing of a pan-European agreement on a EU single digital market, with a balanced relationship to the US; the utilisation of big data in the provision of everyday citizen services, with balanced public/private partnership participation; EU leadership in open data underpinned by innovation leadership in privacy-enhancing technologies and supported by a network of civil actors engaged in continuous improvement of data governance. To this end, immediate action on the following objectives is required to:

- **Build the legal and technical foundations of a EU single data economy,** to ensure a uniform level of data protection across European countries and hence enable the movement of people and the development EU-wide services.
- **Design international agreements on data management** to allow big data as a global business.
- **Invest in government-as-a-platform,** where governments provide the architecture on which citizens and organisations can build services.
- **Engage in public dialogue on data management** the spectrum of public, private, non-profit and citizen organisations to educate users and enable citizens to understand privacy protection options and commit institutions to transparent best practices.
- **Foster open data and abolish data silos** in public organisations to enable the innovation of new services and ways of interacting with citizens.

Pan-European data infrastructure:

A prerequisite to the big data vision is a pan-European data infrastructure to ease the dependence of Europe in US digital platforms by developing European data standards and capabilities rivalling those of China and US, and to foster innovation by European companies and institutions by gaining leadership in Artificial Intelligence (AI), development of big data services and interactive e-government platforms. With respect to the development of such a pan-European data infrastructure the following key challenges were identified: a) lack of common data standards and formats; b) centralisation of public investment on data infrastructures and imbalance on informational power across actors and regions; and c) resistance in transitioning to new digital business models by traditional sectors, including government agencies.

To this end, urgent action is required to:

- **Promote agreement on data standards,** to enable seamless communication and collaboration between different actors.
- **Define a capability development strategy and action plan and an investment strategy** to transition traditional private sectors and public agencies to new digital business models to ease resistance and accelerate big data adoption.
- **Invest in R&D in emerging technologies** such as AI to gain geostrategic leadership and unlock the innovative potential of Europe.
- **Invest in transitioning traditional sectors and public agencies** to the big data infrastructure.
- **Cover existing data-related skill gaps** by hiring, developing public/private partnership and/or investing in education to ensure an adequate pool of European expertise in big data.

Big data for the greater good:

Core to the development of a big data future for Europe is stakeholders' commitment to serve causes for the greater good, dominated by two key agendas climate change and citizen participation. With these in mind, BYTE has envisioned six long-term goals. First, to address issues of citizen participation there are aspirations for a) the integration of data technologies into mainstream curriculums; b) the implementation of reliable and transparent security procedures; c) transitioning of traditional sectors to the digital era, d) eliminate digital divides in smart cities and e) adapt social and taxation systems to reflect the new realities. Second, to address climate change there is a single but compelling aspiration for Europe to lead on climate change management. The key challenges identified centred around citizen participation, such as: privacy and security concerns from the existing trade-off between values and procedures such as freedom and surveillance; decreasing levels of citizen participation in democratic processes; and finally anticipated unequal allocation of big data resources. These exacerbate social inequalities and unequal gains distribution from big data transition.

To address such issues, urgent actions around the following priorities is required to:

- **Promote digital literacy,** focusing especially on non-expert members of the public who are accessing data platforms.
- **Define a strategy, a capability development action plan and an investment strategy** to transition traditional private sectors and public agencies to new digital business models to ease resistance and accelerate big data adoption.
- **Invest in data infrastructures at all levels of government,** to enable citizens to benefit from data-driven innovations and push responsible environmental practices.
- **Develop a capability plan to support local authorities** in using data-driven innovations for resource management and decision-making.
- **Develop a plan to manage the big data transition effectively,** including support plans for early adopters and capability plans for those that may lag behind.

The Big Data Enablers

Semantic interoperability

New Business Models

Data protection frameworks

Data provenance, control & IPR tools

Citizen research mobilisation

Research Roadmaps

While BYTE has documented a comprehensive list of R&D topics for prioritisation, the following four key agendas underpin multiple externalities and should command immediate attention and investment.

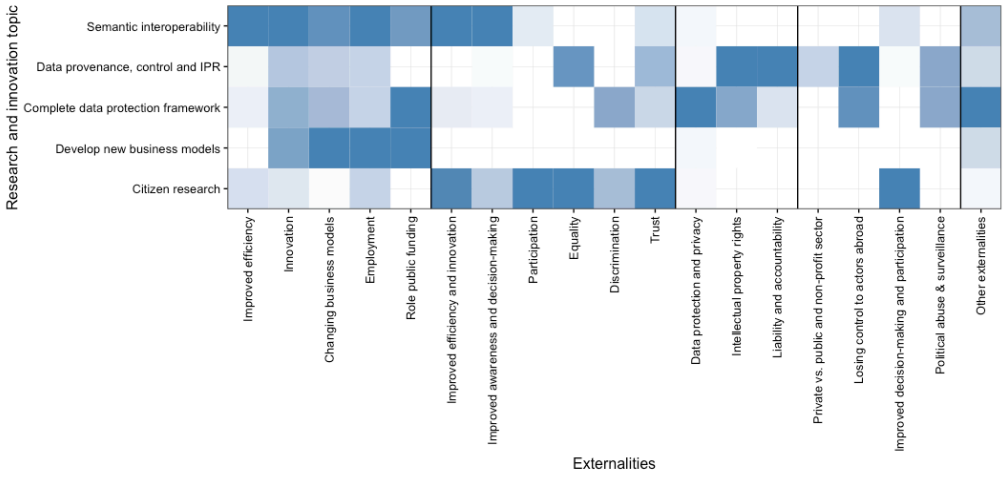

Figure 5: Immediate externalities requiring attention

Semantic interoperability

Due to the high heterogeneity in data sources and channels, such as sensor data, social media data and user-generated data, innovations in integration and fusion are needed to enable advanced data sharing and analysis.

Issues with format conversion lead to intelligence loses, which is currently recovered by reengineering practice. New policies and simple technologies need to be developed to ensure interoperability among different formats and make them easily adoptable. Also, semantic search, schema matching and mapping, as well as ontology alignment have to be addressed. Commercially viable and easy-to-use reporting tools need to be developed that automate semantic annotations. These will have a direct positive effect on the development of new services. Its potential was particularly prominent in health care, and smart cities infrastructure management.

Developing New Business Models

New business models with closer linkages between research and innovation will be required to capture opportunities for economic growth. Within the project several directions have been identified, though this is just an indicative, rather than exhaustive, list. **First**, the exploitation of open source big data to foster and develop new business models. Novel models, such as R&D partnerships where competing organisations share data and

cooperate around tasks, such as data curation, that do not affect their competitive advantage. Exchange of data and expertise in new types of public-private partnerships needs to be further studied and cultivated. **Second**, new data-as-a-service B2B business models can be created around a big data sector. **Finally**, the implication of technological developments such as blockchain may facilitate the development of new business models, for example, around multimedia data mining due to its implications for increasing accountability and trust. Traditional business models may see a resurgence of their lifecycle due to big data. For example, sea data can be mined to improve the efficiency of fishing.

Complete Data protection framework

A key obstacle hindering the uptake of big data is the inadequacy of a legal framework to mitigate against perceived risks. Appropriate legal tools must be developed to ensure that data protection is not an obstacle for big data practices. Developments should take into consideration the need for scalable transaction models; the security challenges which are now in non-relational data stores; and the granular access controls required to allow sharing data on a fine-grained level. The new sources of data create new ways of possible data misuse. Our legal frameworks should shift focus from protecting data to protecting individuals' concerns. To address discrimination, equality and trust concerns for example, privacy-by-design methods should be extended to include anti-discrimination, ensure transparency and sufficiently attribute accountability.

Public dialogue on social concerns can be addressed by big data practices and developing and incorporating tools into the big data design and practice is required between policy makers, regulators and other stakeholders to devise appropriate legal frameworks. For example, data security and privacy issues hinder data exchange in healthcare and need to be addressed by advances in the complete data protection framework. Data storage, processing, access and protection concerns should be openly discussed to update legal framework and guidelines and design acceptable technical solutions to unlock the potential of big data in healthcare management and ensure that the protection of citizens come first.

Data provenance, control and IPR

Digital data attract new rights and require new rights statement initiatives. Core to ensuring such rights is data provenance that documents the provide a historical record of data and its origins. This is particularly relevant for the digitalisation of objects, e.g. an art artefact and has implications for open data and data generated by the internet of things, where data is distributed among different physical locations and where often the appliance and software manufacturers are the organisations that grab the data. It is also important for assigning permissions and digital rights at the data level and for providing context to data curation and data management.

New theoretical models and methodologies for data management are required to ensure that permissions around data transportability are auditable and to ensure accuracy and control over permission management, and guarantee the integrity of data and the confidentiality of the data originator. This will address citizen concerns regarding trust on the one hand, and business concerns regarding intellectual property rights, including those of scholars and contributions.

Citizen Research

Our crisis management case study demonstrated the huge potential benefits such an approach may have and the willingness of people to participate and be actively involved in certain processes. Further research is required to understand how to amplify social engagement mechanisms in socio-technical change across social agendas. For example, citizen science, i.e. crowdsourcing of scientific tasks may be used to increase data accuracy and large scale data curation. New methods of routing tasks to participants based on their expertise, demographic profiles, and long-term teams, and develop open platforms for voluntary work. Smart cities, for example, represent the obvious space for community engagement on local political issues, in which crowd-sourced applications for data collection and analysis can be pursued. Citizen research based on self-monitoring and self-sensing can contribute to several social objectives, most notably preventive medicine and well-being. The environment sector also shows great potential for citizen engagement taking the form of crowd-computing, pervasive-computing, or crowd-sourcing. Tools need be

developed to increase participation citizen science and enable the development of scientific models and simulations. Citizen science leadership by the public sector can ensure data and services remain public goods available to all.

Domain Roadmaps

Due to their overarching societal impacts for Europe, this section outlines capability development roadmaps for three key domains – healthcare, environment and smart cities. For each section we present domain specific policy recommendations and a more detailed research roadmap. This sequences research priorities and outlines a logical investment plan and a timeframe for committing funds and human talent to achieve their intented goals.

Healthcare

The vision for the healthcare sector involves the opening and freely sharing of healthcare data to improve the efficiency and accuracy in the provision of preventive, curative and rehabilitative medical services.

Healthcare priorities:

Data integration

Data management compliance

Real-time analytics

Big data developments will benefit the healthcare sector through improvements in decision making and event detection, new data-driven innovation business models, privacy-aware data practices, and legal and technical developments on anti-discrimination. For example, development of new standards and the adoption of new tools, such as digital clinical records, will ensure data quality and allow the advancement of clinical research. Investments in technologies and infrastructures and public/private partnerships will determine the timing of their achievements. For example, investments in data storage and processing facilities along with encouraging the development of techniques such as data mining will advance clinical and medical knowledge. Partnerships between national health systems and data companies will also accelerate big data benefits. Finally, educating data scientists and integrating big data analysis into mainstream medical curricula will ensure a stream of data literate medical experts to cope with increasing big data demands.

Health data are very sensitive, hence auditing security mechanisms to protect individuals from possible misuse of their information is paramount. Raising citizens' awareness and opening a public debate while investing on technologies such as privacy enhancing technologies is necessary to allow people to trust the data transition in the health sector. The necessity for maintaining the accuracy in health data throughout data management is also undebatable. These two key principles are reflected in the prioritisation of research agendas below (see figure 5 for a summary).

Research priorities for 2017 l Prepare data for data integration

Data quality is paramount for ensuring reliability and gaining physicians trust in the development of data-driven health services. To this end, R&D innovations should focus on a) the identification and redressing of low data and b) the development of easy-to-use reporting tools that include semantic annotations so that unstructured and semi-structured data can be handled efficiently without misrepresentations.

Research priorities for 2018 l Develop frameworks and tools for legally protected data management

Data security and privacy issues hinder data exchange in healthcare, and need to be addressed in advance. Hence, a complete data protection legal framework as well as technical solutions that can embody and implement it are required. While these are already in development, for example advances in cryptographic mechanisms applicable to cloud and big data, such as attribute-based encryption, and developments in privacy-by-design, security-by-design, anti-discrimination-by-design. Nevertheless, intensification of research is required in each field as well as in their intersection to ensure its reliable implementation.

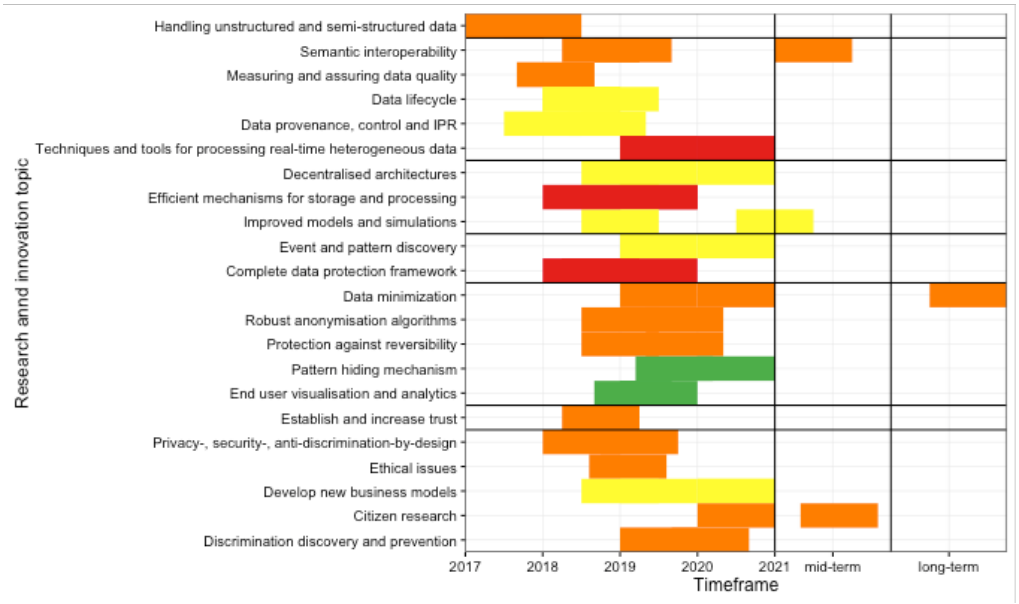

Figure 6: Healthcare research roadmap

Research priorities for 2019 I Prepare for auditable, real-time analytics

Developments in real-time analytics and auditability of big data systems will dominate the horizon post-2018. Innovations in tools and techniques for real-time analytics will enable the use of user-generated content (e.g. from blogs, forums and social media) and personal sensors to track disease outbreaks that will prove invaluable in responding to epidemics. Data minimization tools will ensure that sharing with third parties can take place in a need-to-know basis, eliminating risks around private data misuse. As disease detection and management become increasingly algorithmic, developments in legal informatics and algorithm accountability is required to increase discrimination detection as well as our ability to swiftly redress it.

Environment priorities:

Big Data market creation

Privacy for participation

Real-time analytics

Environment

The vision for the environment sector is an overall change in environmental management so that it becomes a driving force and key social objective. Our policy recommendations involve actions to be taken to utilise big data tools towards this goal.

Big data developments will benefit the environmental sector through improved decision making & event detection, such as tracking environmental challenges; data-driven innovation & business models, to enable safe and environmental friendly operations, and citizen participation in environmental research and management. Hence, the environment section will benefit from investments in: a) the development of data standards to allow data exchanges between a network of actors; b) data storage/processing facilities, data mining techniques and global partnerships to model the current situation, and identify tools to improve it and monitor progress, d) sustainable computing paradigms and green computing to reduce its very own environmental footprint. Our research roadmaps reflect the need for mentality change around environmental resources management (see figure 6 for a summary).

Research priorities for 2017 I Create the Big Data environmental market

Key to the establishment of environmental management as a driving force and key social objective, is linking environmental management to economic development. To this end, development in data-as-a-service economic paradigm can create new opportunities for economic growth by encouraging the development of new products and services based on open environmental data. New business models within this value ecosystem will constitute an environmental sector adding and spring boarding European economic recovery. To this end, data licensing and ownership is key to the creation of the new marketplace, as is the valuation of data to drive data creation and data mining. Inherent, in this effort is the data provenance, control and IPR, as discussed above.

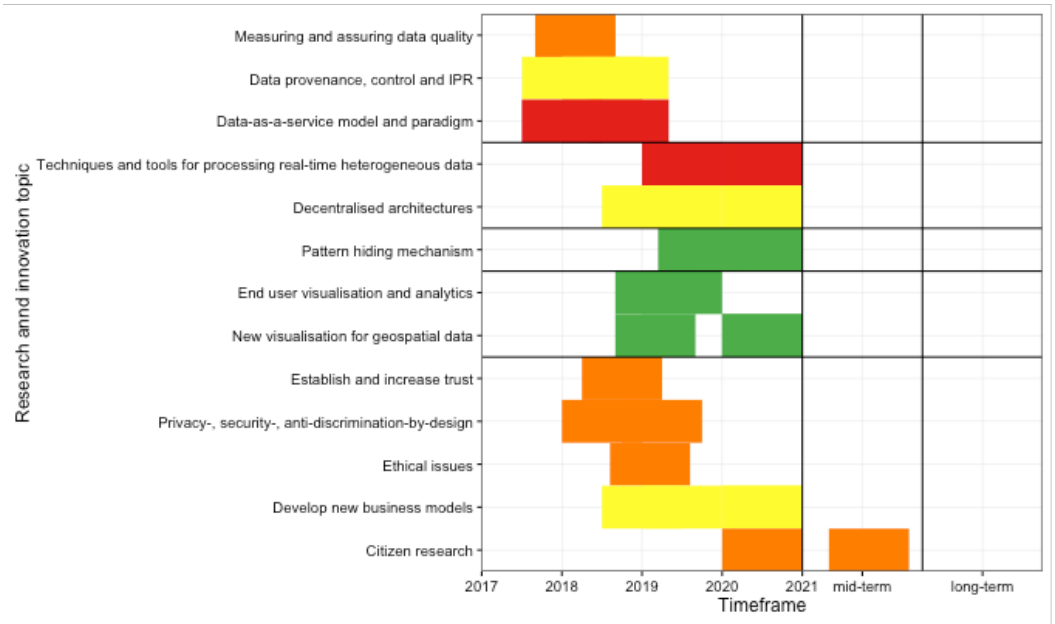

Figure 7: Environment Research Roadmap

Research priorities for 2018 I Ensure privacy to increase citizen participation

A key new resource for environmental management is personal information and overall citizen engagement in data collection, contextualisation and analysis, and monitoring as well as in environmental resource manage. Hence gaining citizens trust is paramount. Hence addressing data security and privacy issues that may hinder participation and environmental data exchange need to be addressed in advance. The means of achieving these have been iterated above (see **Data provenance, control and IPR).**

Research priorities for 2019 I Prepare for auditable, real-time analytics

Developments in real-time analytics and auditability of big data systems will dominate the environmental horizon post-2018 too. Innovations in tools and techniques for real-time analytics will enable the use of user- and sensor-generated content to track environmental trends that will prove invaluable in pre-empting and responding to environmental disasters. Data minimization tools will ensure that sharing with third parties can take place in a need-to-know basis, eliminating risks around private data misuse. As disease detection and management become increasingly algorithmic, developments in legal informatics and algorithm accountability is required to increase discrimination detection as well as our ability to swiftly redress it.

Smart cities

The vision for smart cities is the creation of value in public service provision in its broad sense from urban data generated by citizens and businesses in their interaction with the urban infrastructure.

Smart cities present both political and technological challenges, as they signify a deeper, systemic digitalisation of everyday life, with overarching implications other sectors, such as the environment and health. Smart cities will benefit from all aspects of big data developments, but in particular from investments in: a) data storage/processing facilities, and data mining that help local governments and institution understand the local operations and identify the right course of actions; b) private/public partnerships to help local policymakers retrieve useful datasets, model and influence social processes and behaviours, and monitor resources allocation and consumption; c) local governments as platforms to simplify two-way interaction with citizens; d) improving digital literacy among citizens to foster citizen participation in open data initiatives and citizen science, e) safe and trustworthy data infrastructure and in a democratic debate on data to gain citizens' acceptance of smart cities and relevant technologies and finally, f) new economic models

for the fair distribution of the cost and benefits of the local sharing economy to ensure social cohesion. In addition, the sector has great demands for data-intensive engineers and the development of standards for city data that can provide an equal footing for European SMEs.

The following research priorities were identified and prioritised (see figure 7 for a summary):

Smart city priorities:

City-as-a-platform

Real-time analytics

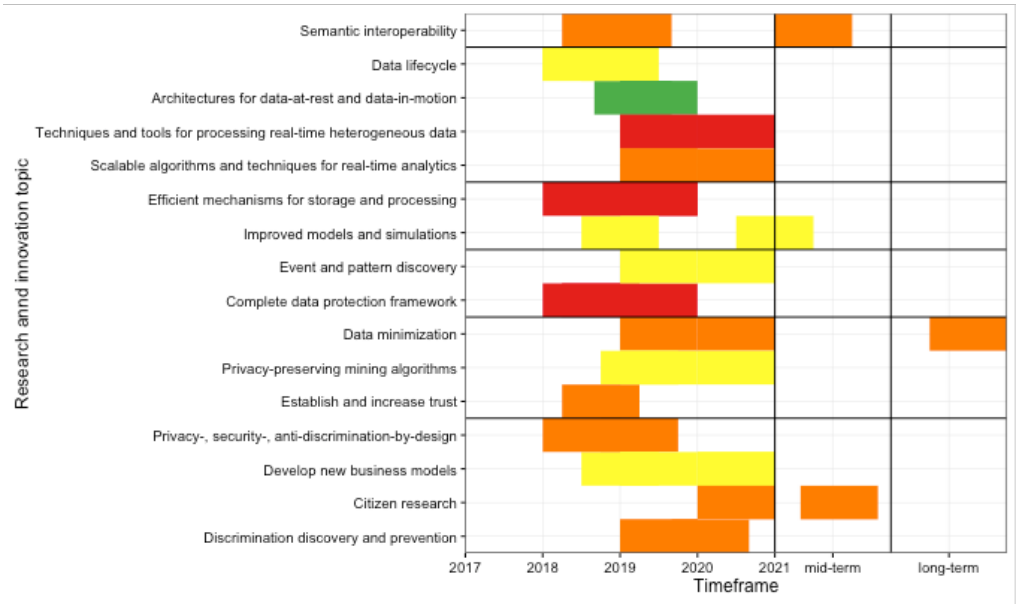

Figure 8: Smart Cities Research Roadmap

Research priorities for 2018 I Turn the city into a trusted platform

Developments in efficient mechanisms for storage and processing are required to foster innovation from open and government data by the engaging citizens and start-ups, which will naturally be drawn to such cities. Data security and privacy issues hinder citizen participation in smart cities, and need to be addressed in advance. Hence, a complete data protection legal framework as well as technical solutions that can embody and implement it are required. While these are already in development, for example advances in cryptographic mechanisms applicable to cloud and big data, such as attribute-based encryption, and developments in privacy-by-design, security-by-design, anti-discrimination-by-design. Nevertheless, intensification of research is required in each field as well as in their intersection to ensure its reliable implementation.

Research priorities for 2019 I Prepare for auditable, real-time analytics

Developments in real-time analytics and auditability of big data systems will dominate the environmental horizon post-2018 too. Innovations in tools and techniques for real-time analytics will enable the use of user- and sensor-generated content to track environmental trends that will prove invaluable in pre-empting and responding to environmental disasters. Data minimization tools will ensure that sharing with third parties can take place in a need-to-know basis, eliminating risks around private data misuse. As disease detection and management become increasingly algorithmic, developments in legal informatics and algorithm accountability is required to increase discrimination detection as well as our ability to swiftly redress it.

BYTE Big Data Community I Get involved

For those with an interest to contribute to the future of Big Data in Europe, a new community has been established to ensure that industry and other big data practitioners are aware of and able to integrate societal concerns into existing and emerging big data practices. The community is open to civil society organisations, academics and legal experts and provides them with a platform to feed their concerns into industry and other practitioners, in order to transition into a responsible data-driven economy in Europe.

Figure 9: BYTE Big Data Community founding members

So far, the community has undertaken a series of case studies to understand the positive and negative impacts of the "big data revolution" on European society and has defined a vision for big data in Europe in 2020, a policy and research roadmap to achieve it. Going forward we focus on four key objectives:

"Civil society as important as industry, academia and policy makers in the EU strategy for big data"

- Build the community by reaching out to civic societies, patient groups and consumer groups as well as individuals.
- Engage in deeper discussions on what and where are the gaps and challenges of big data to influence the European Big Data Value Strategic Research and Innovation Agenda
- Influence the European strategy for the EU "Big Data revolution", by recommending funding choices, good practices, specific research and policy needs to be filled to the EU commission.
- Become instrumental in implementing best practices, policy and research roadmaps through pilot projects, map issues and initiatives and define strategies for each one, exchange success stories, develop frameworks for citizen engagement, conduct international surveys, and the like.

If you would like to be part of the BYTE Big Data Community follow the link to Join our Community at the project's website: http://byte-project.eu/byte-community, or contact us at: community@byte-project.eu for more information

Appendix A

This appendix presents a brief overview of the key legal, social and ethical and technical recommendations for each domain.

Healthcare sector Recommendations

With respect to the Healthcare sector the following key recommendations in each category are proposed:

Legal recommendations

Key for the promotion of big data within the Health sector is privacy and data protection. To this end, changes in privacy and data protection mechanisms should be develop to guard against known privacy risks. We suggest to:

- Broaden privacy-by-design to encompass the design and development of socio-technical solutions for legal and organisational safeguards. Better integrate them into existing risk assessment and design strategies.
- Include collective mechanisms of consenting and licencing into data protection frameworks and ensure they are hardwired into big data solutions.
- Promote the development of standard evaluation enabling the auditability of such practices, along with a strengthening of data protection authorities to improve transparency and foster public trust.

Social and ethical recommendations

The health sector will benefit from socio-economic developments that enable better collaboration amongst network actors and between network actors and citizens. To this end, we suggest to:

- Develop partnerships between global platforms and agreements between national health systems and data companies.
- Initiate large-scale citizens' consultations on privacy enhancing technologies to raise awareness and allow people to trust the data transition in the health sector.
- Educate data scientists and integrate big data analytics into mainstream medical curricula to facilitate the adoption of big data mining techniques by the medical sector

Technical recommendations

The health sector will benefit from technical developments that increase the confidence of citizens and medical professionals in big data technologies. To this end, we suggest to:

- Invest in data storage and processing facilities and in developing data mining techniques to advance clinical and medical knowledge.
- Develop techniques and tools for processing real-time heterogeneous data from blogs, forums and social media to track disease outbreaks, side effects of drugs, etc.
- Devise efficient mechanisms for storage and processing that can competently cope with data volume and heterogeneity without compromising data quality and reliability to raise users' confidence.

Smart Cities Recommendations

With respect to the smart cities sector the following key recommendations in each category are proposed.

Legal recommendations

Key for the adoption of smart cities are also changes in privacy and data protection, including mechanisms to guard against new privacy risks and potentials for discrimination due to new sources of information on citizens' lifestyle. To this end, we suggest to:

- Develop a new data protection and data minimisation framework to protect individuals against new ways of possible data misuse made possible by new sources of data in the smart city sector.

- Develop privacy-by-design, security-by-design, and anti-discrimination-by-design frameworks focusing primarily on protecting citizens' interests and rights.

Social and ethical recommendations

Ways to motivate more collaborative ways of working with the private sector and of interacting with citizens is fundamental for the uptake of smart cities. To this end, we suggest to:

- Encourage partnerships between data companies and local authorities will help the later derive value from data mining of useful datasets in terms of modelling and influencing constituents' activities and better monitoring and management of city resources.

- Invest in turning local governments as platforms, improve digital literacy and an open data strategy, and open a democratic debate on open data and privacy enhancing technologies to improve the interaction between citizens and their local governments.

- Develop new economic models that ensure the fair distribution of the cost and benefits of the local sharing economy to ensure social cohesion.

Technical recommendations

Smart city initiatives will benefit from technical developments that increase the confidence of citizens and government staff, but also to facilitate agile and responsive city management systems. To this end, we suggest to:

- Invest in data storage and processing facilities and encourage the development of techniques such as data mining will help local governments and institution understand the territory they govern and identify the right course of action.

- Build a safe and trustworthy data infrastructure to guarantee the acceptance of smart cities.

- Invest in the development of techniques and tools for processing real-time heterogeneous data and scalable algorithms and techniques for real-time analytics to develop agile and rapidly responsive city management systems.

Environment sector Recommendations

With respect to the environment sector the following key recommendations in each category are proposed.

Legal recommendations

The protection of trade secrets and intellectual property rights is fundamental for motivating data exchanges amongst network actors. To this end, we suggest to:

- Develop new legal frameworks to protect proprietary company information suited to big data practices and amend copyright laws to reduce threats to intellectual property rights and private data accumulation and ownership.

- Develop international legal frameworks and standards to protect trade secrets to regulate data flows amongst international actors and safeguard law enforcement to improve liability and accountability.

Social and ethical recommendations

New global partnerships utilising data-as-a-service paradigm is fundamental for making inroads and developing global solutions in the field. To this end, we suggest to:

- Partner with global platforms to help understand environmental issues at a global level. This could also assist in monitoring the footprint of users' activities and influence these activities.

- Eventually, investments on sustainable computing paradigms, such as green computing could help reduce the impact of the data transition on the environment.

- Develop new business models based on data-as-a-service to exploit new opportunities for economic growth and diminish inequalities to data access between big data players and the rest.

Technical recommendations

Technical developments that increase the manageability of data volumes and accuracy of developed models and assessment of interventions are required. To this end, we suggest to:

- Develop data standards that facilitate the exchange of data between a network of actors to promote the development of partnerships.

- Invest in data storage and processing facilities and data mining techniques to model the current situation and identify the adequate tools to improve this situation.

- Invest in the development of techniques and tools for processing real-time heterogeneous data that help gather public insight by identifying environmental trends and statistics.

Trilateral Research, Ltd.

72 Hammersmith Road I W14 8TH
London IUnited Kingdom
Kush Wadhwa I Kush.wadhwa@trilateralresearch.com
Rachel Finn I Rachel.finn@trilateralresearch.com