



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 D10.4: **Final, public project synthesis report**

Author(s): Rachel Finn & Kush Wadhwa
Trilateral Research

Dissemination level: Public

Deliverable type: Final

Version: 1.0

Submission date: 30 June 2017

Table of Contents

Preface..... 3
1 Project summary 4
2 BYTE expected impact..... 7
3 The community 10

PREFACE

This document is a consolidated, concise version of the BYTE final report as submitted to the European Commission as part of our final reporting process under grant no. 619551. The information presented in this document is derived from that report and is intended for public consumption.

1 PROJECT SUMMARY

The BYTE project has progressed well in meeting its objectives. This section provides an overview of BYTE work in its entirety for quick reference. The following section provide a more detailed presentation of each with emphasis on the main science and technology outcomes resulting from the work.

Year one focused on work packages 1 and 2 which comprised foundational research on big data and related positive and negative externalities. All deliverables were submitted on time or in accordance with an extended deadline, and made publicly available on the project website. The works provided inputs and variables to be considered within the case studies conducted in WP3 and a list of externalities to attend to through the course of the project.

Year two focused on work packages 3, 4 and 5, which undertook a series of case studies, performed a horizontal analysis of the case studies and developed a vision for big data. Each of these represented an incremental step. The case studies informed the horizontal analysis,

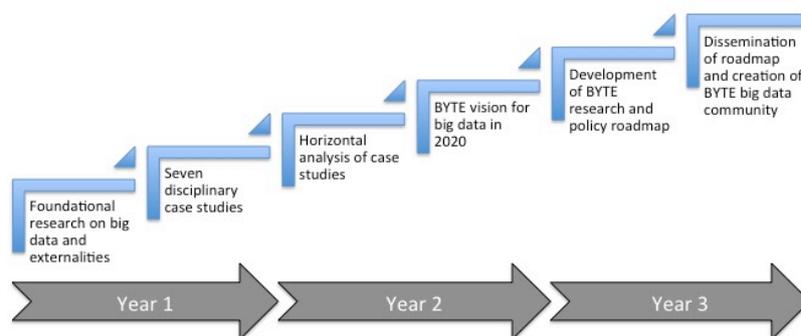


Figure 1: BYTE Workflow and methodology

where the horizontal analysis sought to identify common impacts of big data and good practice in addressing them. The output of the horizontal analysis was a series of recommendations for amplifying the opportunities associated

with positive externalities and diminishing or mitigating the roadblocks associated with negative externalities. This information was used to create a vision for big data in 2020, where the opportunities have been captured and the negative impacts have been adequately addressed.

Work package 3, *Case studies in positive and negative externalities* used this information and examined a series of variables associated with the definition of big data and the externalities associated with actual big data practice across seven different industrial and disciplinary sectors (see Figure 2). These case studies provided evidence-based information about what externalities were in evidence and how practitioners were addressing them. The output of WP3 was a methodology for conducting the case studies (D3.1, produced in Y1) and a



Figure 2: Seven sector-specific case studies

comprehensive case study report *D3.2: Case study reports on positive and negative externalities* that provides first-hand information about the positive and negative impacts identified in each of the case study organisations and the associated sector. This information fed into the *Horizontal analysis* produced in work package 4 that consolidated the case study findings and identified good practice in amplifying positive externalities and addressing negative externalities, resulting in *D4.1 Horizontal analysis and social impacts of positive and negative externalities* and *D4.2 Report on diminishing negative externalities and amplifying positive externalities* (see list of deliverables below). One of the highlights of D4.2, was the examination of legal gaps associated with the ability to realize opportunities associated with big data and potential avenues to address these gaps. Work package 5 utilized the good practice identified in the case studies and horizontal analysis to undertake a series of foresight exercises and consolidate a vision for big data in Europe for 2020. The *BYTE Vision* (D5.2) developed a series of scenarios to help decision-makers plan for the changes associated with the data economy, as well as a series of sector-specific visions that considered how prepared different sectors were for the transition to the data economy. Deliverable 5.2, *Tackling the externalities of the vision*, describes an adaptive framework to aid policy decision-making to assist in addressing the externalities associated with the BYTE vision.

The principal research activity happening in the **final year** of the project was the development of *Policy and research recommendations* (D6.1) for big data towards 2020. Partners utilized the foundational research, case studies and horizontal analysis to create a series of policy and research recommendations that are required to be accomplished in order to achieve the BYTE vision. With respect to policy recommendations, BYTE produced the following three over-arching recommendations:

- 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 recommendations are geared towards European and national policy-makers as well as institutional or industry practitioners. With respect to research recommendations, BYTE recommended that capacity should be developed in relation to the following five areas:

- Semantic interoperability
- New business models
- Data protection frameworks
- Data provenance, control & IPR tools
- Citizen research mobilisation

These five areas are spaces where skills, technology and capacity development is most important and most urgent. In addition, the project produced policy and research

recommendations in three sector-specific areas: healthcare, environmental data and smart cities.

Year three also focused on the completion of work packages 7, 8 and 9, which commenced in year two and run parallel to the work described above. Work package 7, *The big data community*, began in year two and developed a strategy (D7.1.2) and sustainability plan (D7.2.2) for building the BYTE big data community. Both documents examine three potential strategies for the BYTE big data community: acting as an autonomous organization, acting as an umbrella organization or joining with an existing, multi-stakeholder organization. The project has been in conversation with the Big Data Value Association (BDVA) regarding the third option, and plans for the big data community have been finalized in consultation with the Advisory Board and the other 35 founding members of the community to provide inputs into the BDVA on an annual basis. Work package 8, dealt with *Stakeholder engagement* since the beginning of the project and remained live throughout its duration. Project partners created and constantly updated a Stakeholder Taxonomy that mapped relevant stakeholders within the big data ecosystem with the view to invite their active participation, endorsement of research findings and support in implementing the resulting policies and roadmaps to amplify the impact of project findings. This resulted in the creation of the BYTE advisory board consultations of which intensified during the primary research in year 2. Work package 9, *Dissemination* also remained live throughout a carefully designed dissemination plan and encompassed online presence, social media, event sponsorship and participation, dissemination of findings in academic and industry events, and other publicity activities, such as newsletters, blogging, popular media articles, interviews and promotional videos. More importantly, however, were the organisation of ‘hands on’ workshops that invited the active involvement of key stakeholders in the process, development, and dissemination of the project and created a community of approximately 950 interested parties that maintained membership in the BYTE Big Data Community (BBDC).

BYTE has had the following measurable impacts on stakeholders relevant to the project and the community. The project has produced **37 peer-reviewed publications** oriented towards the scientific community. Additional publications are planned for after the close of the project in relation to Y3 findings. BYTE has sponsored 6 external events, creating visibility and brand recognition for the project. We have also given 86 presentations at 115 different events, **reaching a stakeholder mass of up to 7000 people**. We have also provided **direct input into 50 different networks**, including smart cities networks, big data networks and associations, humanitarian networks, open data networks, geospatial networks, healthcare networks as well as others. Finally, the BBDC will continue to augment the impact of the BYTE project by acting as a continuation of our project work – feeding information on societal issues directly to policy-makers and industry and ensuring that civil society organisations, consumer organisations and non-government organisations continue to have a voice in the development of big data capacity and policy in Europe.

For more information about BYTE and the big data community, please visit the project website or watch our video.

Website: <http://byte-project.eu>

Video: https://youtu.be/_8pP4UmJrL0

2 BYTE EXPECTED IMPACT

The key impact of the BYTE project is to support European stakeholders achieve the 2020 benchmark discussed in the Digital Agenda for Europe and Horizon 2020 and achieve a 30% share of the big data market. With this in mind, the project aims to provide European policy-makers, industry and scientific community with the background knowledge and tools to achieve their respective objectives.

Potential socio-economic and wider societal implications

Policy recommendations were driven by a European vision for big data and the key priorities to: (a) 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; (b) develop a pan-European data infrastructure, and (c) 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 with respect to:

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. 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. 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. To this end, urgent action 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 R&D priorities commanding immediate attention, investment and collaboration between industry and the scientific community reflected the need to address four key agendas underpinning multiple externalities to unlock the potential contribution of Big Data in the European economy. These related to the following topics:

Semantic interoperability. 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 are 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. To address such issues, urgent actions around the following priorities is required to 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. 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 and technology designs should shift focus to protecting individuals' concerns, regarding discrimination, equality and trust concerns. Public dialogue on key social concerns and how they can be satisfactorily addressed by big data practices and tools is urgently required. 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.

3 THE COMMUNITY

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 3: BYTE Big Data Community founding members

In terms of Community activities, members have agreed the following. The BBDC will organise an annual workshop/panel at the BDVA focusing on responsible innovation lessons or good practice information. These may focus on a (few) big data sector(s) or specific question(s) that are topically relevant. The focus will be chosen by the BBDC and the BDVA in conversation. Inputs will be encouraged in the form of presentations, but also white papers, opinion papers or blog posts, and participants will be asked to volunteer or be selected from the membership pool based on the topic or their level of interest. All of the inputs will be made publicly accessible on the BYTE website.

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:

- 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](http://byte-project.eu/byte-community) at the project's website: <http://byte-project.eu/byte-community>, or contact us at: community@byte-project.eu for more information