



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 D7.1.1: Interim strategy and charter for the big data community

Author(s): Lorenzo Bigagli, *National Research Council of Italy*
Rachel Finn, *Trilateral Research*
Martí Cuquet, *University of Innsbruck*
Laszlo Gergely, *NIIF*
Edward Curry and Umair ul Hassan, *NUIG*
Erik Stensrud and Jarl Magnusson, *DNV GL AS*
Scott Cunningham and Claudia Werker, *TU Delft*
Hans Lammerant, *Vrije Universiteit Brussel*
Guillermo Vega Gorgojo, *Universitetet i Oslo* (review)
Sonja Zillner, *Siemens* (review)

Dissemination level: Public

Deliverable type: Final

Version: 1.1

Submission date: 30 September 2016

Table of Contents

Preface.....	4
Executive summary.....	5
1. Introduction.....	6
2. Context and background.....	8
2.1 Relevant existing big data communities and projects.....	8
2.2 BYTE policy recommendations.....	11
2.3 BYTE vision and roadmap.....	13
3. Interim strategy for the big data community.....	17
3.1 Possible approaches.....	17
3.2 Value proposition.....	19
4. Interim impact strategy.....	22
4.1 Stakeholder Taxonomy.....	22
4.2 Stakeholder Engagement Strategy.....	24
4.2.1 Incentive mechanisms.....	26
4.3 Dissemination Strategy.....	27
5. Interim development plan.....	29
6. Interim charter of the big data community.....	32
6.1 Governance structure.....	32
6.2 Membership.....	32
6.3 Key goals and objectives.....	33
6.4 Types of activities foreseen.....	33
7. Future work.....	34
Annex A.....	35
A.1 BDVA.....	35
A.2 Group on Earth Observation (GEO).....	37
A.3 Research Data Alliance (RDA).....	38
A.4 Big Data Public Private Forum (BIG).....	40
A.5 United States Big Data Senior Steering Group.....	41
A.6 NIST Big Data Public Working Group.....	42
A.7 CODATA.....	43
A.8 Belmont Forum.....	45
A.9 Global Alliance for Genomics & Health (GA4GH).....	47
A.10 Big Data Europe.....	48
A.11 European Data Forum.....	48
A.12 OGC Big Data Domain Working Group.....	49
A.13 NSF Earthcube.....	49
A.14 RECODE.....	49
A.15 Scalable End-user Access to Big Data (OPTIQUE).....	49
A.16 SIM4RDM.....	49
A.17 SmartVortex.....	50
A.18 Planet Data.....	50
A.19 PRELIDA.....	50
A.20 ROBUST.....	50
A.21 Creating Knowledge out of Interlinked Data (LOD2).....	50
A.22 EarthServer.....	50
A.23 Semicolon.....	50

A.24 Partnership for Advanced Computing in Europe (PRACE)	51
A.25 CERN LHC Computing Grid.....	51
A.26 US Big Data Research and Development Initiative.....	51
A.27 Australian Public Service Big Data Strategy	51
A.28 UK Data Service	51
A.29 UN Global Pulse Initiative.....	51
A.30 Copernicus	52
A.31 The European Bioinformatics Institute (EMBL-EBI)	52
A.32 Teradata.....	52
A.33 deCODE genetics.....	53
A.34 WuXiNextCODE	53

PREFACE

The BYTE project will assist European science and industry in capturing the positive externalities and diminishing the negative externalities associated with big data in order to gain a greater share of the big data market by 2020. The project comprises three phases of work: a preliminary investigation, an exploration of present and future societal impacts, and the future agenda for big data.

This deliverable captures the results of part of the activity performed in phase three, within *Work Package 7 – The big data community* (WP7), namely in *Task 7.1 – Forming the big data community* and *Task 7.2 – Implementation plan for the big data community*.

The overall objectives of WP7 are:

1. To design and form the big data community, including drafting founding texts;
2. To prepare a BYTE project final report, including a series of guidelines supported by community members;
3. To input BYTE findings and guidelines into relevant networks.

The above tasks and this deliverable contribute to objective 1. In particular:

- Task 7.1, which involves CNR-IIA, TRI, NIIF, UIBK, INRIA, and DNV, aims at designing a cross-disciplinary community to implement the BYTE roadmap, amplify positive externalities and diminish negative externalities associated with big data, and ultimately achieve the BYTE vision. As such, it relies on the work of Work Package 4, 5 and 6, hence also VUB and TUD have contributed.
- Task 7.2, which initially involved CNR-IIA, NUIG, and UIBK, outlines a plan for the development, goals and sustainability of the above community. Following the reviewers' recommendations in the BYTE Year 1 Review report, it also addresses the community impact strategy, relying on Work Package 8 and 9 (led by NUIG and UIBK), with the involvement of TRI and all other partners, to ensure its robustness and coherence with the other BYTE work packages.

This deliverable has been further revised and updated following up to the Year 2 Review of BYTE, elaborating on the incentive mechanisms to facilitate the active participation of stakeholders in the big data community. A definitive version of this deliverable, with the final strategy and charter for the big data community, is due near the end of the BYTE project.

EXECUTIVE SUMMARY

BYTE will culminate in the launch of the BYTE Big Data Community (BBDC), a sustainable, cross-disciplinary platform that will implement the roadmap identified throughout the project, and will assist the European stakeholders in identifying and meeting the big data challenges, to finally achieve the scenario envisioned by BYTE for 2020.

Although several organisations, some supported by the EC, like the Big Data Value Association (BDVA), are gathering stakeholders together to capture big data innovations, BYTE is quite unique in examining the economic, legal, social, ethical and political externalities associated with the collection, linking and sharing of big data. In fact, BYTE is focused on alleviating societal concerns and including civil society organisations in the community, as well as policy-makers.

This document outlines an interim strategy and a value proposition for the BBDC, taking as a first-choice working hypothesis that it will channel its contribution through the BDVA. In light of this preferred initial option, BYTE has already begun a process of exploring such collaboration with the BDVA steering committee.

This document also outlines: an interim impact strategy, leveraging BYTE stakeholder engagement and dissemination activities to incentivise active participation in the BBDC; a tentative plan for the development of the BBDC, divided in three phases, with appropriate milestones until the end of the project; a tentative charter, including governance structure, membership criteria and activity types, aligned with those of the BDVA.

This interim strategy and charter will be finalised near the end of the project.

1. INTRODUCTION

As per its Description of Work (DoW), BYTE will culminate in the launch of the BYTE Big Data Community (BBDC), a sustainable, cross-disciplinary platform that will implement the roadmap identified throughout the project, and will assist the European stakeholders in identifying and meeting the big data challenges, to finally achieve the scenario envisioned by BYTE for 2020, as presented in deliverable *D5.1 – The BYTE vision*.

In the third phase of its development, BYTE has started cultivating an extensive and diverse group made up of traditional big data stakeholders, such as industrial actors, statisticians, standardisation bodies and policy-makers, computer scientists, and other science experts. Given its focus on societal externalities, also social science scholars and open data activists will be engaged, to create a shared vision and roadmap for future investments based on concrete challenges. To appropriately consider public perceptions and aspirations, the consortium will also involve representatives of civil society organisations, as well as public institutions.

Each of these stakeholders has different relationships with particular externalities. For example, industry, scientists and statisticians might be more concerned with the economic externalities associated with big data, such as innovations, efficiencies and lower entry barriers to the market. In contrast, social scientists and civil society organisations might be more interested in privacy implications, legal challenges in enabling innovation and the provision of better goods and services. Finally, policy-makers and other decision-makers are interested in boosting the economy, providing efficient services, increasing transparency in decision-making and protecting civil rights.

Accordingly, the BBDC will bring these different stakeholders together to consider these issues and attempt to find points of consensus in order to address big data societal challenges. The community will monitor progress in meeting societal challenges associated with big data, capturing opportunities, provide support where necessary and identify new and emerging externalities to be addressed.

Furthermore, the BBDC will facilitate a shared understanding of concrete problems and opportunities worth investigating, and suggest approaches for cooperative problem solving and collaborations across different disciplines and stakeholder categories both now and in the future. The community members will support and implement the BYTE guidelines and recommendations on capturing and addressing the positive societal externalities associated with use of big data, laid out in BYTE deliverable *D7.3 – Final report and guidelines*.

Besides creating a vital collaborative mechanism to build consensus on the recommendations identified by the BYTE project, the community will also help European decision-makers to assess policies and practices across Europe, and benefit from feedback from the long-term engagement of key stakeholders.

BYTE will publicise the community to a large population of stakeholders, in order to encourage them to participate, to foster further innovation and economic competitiveness in Europe's engagement with big data. Because the use, reuse and linking of large data sets depends in large part upon developments in other, related movements, the BBDC will have to link with and

input BYTE findings into related networks and initiatives in order to influence their policies, goals and processes.

BYTE will organise a workshop with the identified founding members of the BBDC and European and Member State policy-makers. The workshop participants will finalise the charter and the funding plan. Besides, they will identify and discuss strategies for inputting into related networks and initiatives, and discuss how to publicise, populate and strengthen the community. The community will be sustainable after the close of the BYTE project and at least until 2020.

The next chapter provides some background information on the context in which the BBDC is being designed and implemented, including the initiatives of most interest on big data (see Annex A for a more comprehensive listing of the projects and initiatives analysed), as well as work done in the framework of BYTE, to define recommendations, a vision, and a policy and research roadmap for big data in Europe in 2020 and beyond.

The following chapter outlines the strategy that the BYTE consortium is pursuing to create and initiate the BBDC and its value proposition. A subsequent chapter elaborates upon the strategy to maximise the impact of the community, by leveraging BYTE resources allocated for stakeholder engagement and dissemination. As requested by the Year 2 Review report, it elaborates on the incentive mechanisms to facilitate the active participation of stakeholders in the BBDC. The subsequent chapter lays out a development plan for the BBDC, from the foundation to the end of the BYTE project. The following chapter proposes a tentative organisation of the BBDC, including governance structure, criteria for membership, policies, procedures and types of activities foreseen.

Finally, the last chapter lists the work items planned to assess the community strategy proposed in this interim document, especially with respect to its impact against the values and expectations of members of the public. The results of such analysis will be used to evolve this strategy and charter into the final version, which is due near the end of the project.

2. CONTEXT AND BACKGROUND

To design the strategy and charter of the BBDC, it is necessary to take into account the context and background in which it will operate.

Externally to BYTE, there are several organisations, some already supported by the EC, designed to gather stakeholders together to assist in capturing big data innovations. These are typically focused on the technical and infrastructural elements of big data innovations, not the societal externalities that BYTE will examine. In fact, BYTE is quite unique in its interest in examining the economic, legal, social, ethical and political externalities associated with the collection, linking and sharing of big data. Nevertheless, it is important to examine also the other initiatives, to capture what works well and take it in the BBDC charter (in terms of governance structure, membership criteria, type of activities, etc.) Likewise, we are also interested in the gaps that we feel the BBDC be in a good position to fill in. Annex A provides a commented list of the main initiatives and projects related to big data in Europe and the world. The next chapter focuses on the most relevant to BYTE.

Internally, BYTE is set to define recommendations to address positive and negative externalities of big data, a vision for big data in Europe, and a roadmap to achieve and realise it, which the BBDC must take into account. The following two chapters set the BBDC in this context, to the extent possible in this interim strategy, taking into account that part of this work is still in progress (especially for the vision and the roadmap).

2.1 RELEVANT EXISTING BIG DATA COMMUNITIES AND PROJECTS

Two existing initiatives provide the context and working models for the BBDC: the American Big Data Senior Steering Group and the European Big Data Value Association. These organisations mirror the intended goals and activities of the BBDC. However, the BYTE community also differs in important ways from these organisations. Specifically, part of the remit of BYTE is to integrate citizens' perspectives, aspirations and protections from unwanted societal impacts. Neither of the model organisations appears to adequately integrate this perspective, and this is where the BBDC can provide added value to existing initiatives.

The United States Big Data Senior Steering Group (BD SSG)¹ within the Networking and Information Technology Research and Development (NITRD) Program is charged with initiating and managing the US White House's Big Data Research and Development Initiative; a \$200 million programme to improve the "ability to extract knowledge and insights from large and complex collections of digital data" and to "help solve some of the Nation's most pressing challenges".² Although this initiative was discussed at length in BYTE deliverable *DI.3 – Big data initiatives*, specific aspects of the initiative warrant repetition in this context due to their relevance for BBDC building.

¹ [https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_\(BD_SSG\)](https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_(BD_SSG))

² Office of Science and Technology Policy, "Obama administration unveils "big data" initiative: Announces \$200 million in new R&D investments", Press release, 29 March 2012, p. 1. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf

The role of the BD SSG is very similar to the BYTE project and the BBDC. The BD SSG is tasked with identifying areas of interest relevant to big data within the US, including areas of research and industrial relevance. Furthermore, like BYTE, the Steering Group will also construct a vision for big data in the US and a series of roadmap steps (described as action steps) that different government agencies can take, both individually and collaboratively, to achieve that vision. While BYTE will develop both a vision and roadmap during the course of the project, as noted above, one of the objectives of the BBDC is to implement that roadmap and keep it up to date. Furthermore, both the BD SSG and BYTE vision and roadmap(s) are using stakeholder consultation and engagement exercises to generate stakeholder commitment to the vision and resulting roadmap steps.

Furthermore, as the BBDC intends, the BD SSG is also characterised by multi-disciplinary and multi-sector stakeholder collaboration. The initiative they are managing is focused on six different government agencies and departments working together. In addition, it is intended to impact a range of different sectors, including health, defence, intelligence gathering and others. The BD SSG includes partnerships with major companies such as Novartis, Pfizer, IBM and SAP as well as major universities and programmes such as NYU and MIT. The Steering Group launches competitions and partnerships programmes with these organisations and supports programmes between these organisations and other parties.

While many of these objectives and activities overlap between the BD SSG and the BBDC, BYTE provides added value for Europe in two ways. First, the activities of the BD SSG are obviously focused on the US and will result in innovation in that space, rather than in Europe. Second, the BD SSG appears to be a partnership between government and industry, which results in top-down activities that will find the integration of citizens' perspectives and concerns challenging. Nevertheless, the focus on stakeholder collaboration and the development of multi-stakeholder partnerships to influence policy and industry activities means that the organisation is a good model from which BYTE can draw to develop the strategy for the BBDC.

Yet, Europe does have an organisation tasked with developing the European big data sector, the Big Data Value Association (BDVA)³, and BYTE will have to establish and delineate synergies between the BBDC and this existing organisation. While at first glance the objectives of the two organisations appear to be similar, the specific remit of BYTE means that the BBDC can provide added value without duplicating activities.

The BDVA is the private counterpart in the Big Data contractual Public-Private Partnership (cPPP) established in 2014 between the EC and industry and research organisations. The main role of the BDVA is providing the Big Data Value Strategic Research and Innovation Agenda (BDV SRIA; SRIA in short) and its regular updates, defining and monitoring the metrics of the cPPP and joining the EC in the cPPP partnership board. Like the BBDC will be one of the outcomes of the BYTE project, the BDVA grew out of a EC-funded research project, the Big Data Public Private Forum, which originally developed the SRIA.

The goals of the BDVA include the following:

³ BDVA, "About us", 2016. <http://www.bdva.eu/?q=about>

- Collaborating with the EC;
- Developing strategic goals of European Big Data Value research and innovation and supporting their implementation;
- Improving industrial competitiveness of Europe through innovative Big Data Value technologies, applications, services, solution;
- Strengthening networking activities of the European Big Data Value community;
- Promoting European Big Data Value offerings and organisation;
- Reaching out to existing and new users;
- Contributing to policy development, education and technology ramification in the widest possible sense and addressing ethical, legal and societal issues.

Thus, like BYTE, the BDVA seeks to develop and implement a roadmap, improve industrial competitiveness in Europe, strengthen networking between established and new members of the big data ecosystem and contribute to policy development.

Also, like both BYTE and the BD SSG, the BDVA is made up of major European and international companies relevant to the big data ecosystem. These include Siemens, Atos, Intel, Orange and SAP as well a major research centres in Europe, including Fraunhofer, Amsterdam University and the Technical University of Denmark. Associate members include SMEs and other organisations who do not have voting privileges but who take a relatively active role in the organisation. The BDVA is organised into 9 task forces, including task forces on legal, technical business and societal issues. Members are invited to join any of the task forces of interest to them and participate in discussions and the development of outcomes. Task forces are led by an individual organisation, but are open to the participation of any member.

However, the remit of BYTE means that the BBDC will differ from the BDVA in two ways. While the BYTE roadmap will include a research element and a policy element, the BDVA's SRIA is primarily a technical roadmap focused on "research, technological development and innovation".⁴ Thus, the policy element of BYTE and our focus on technological development that specifically addresses societal concerns to enable innovation sets BYTE apart from the BDVA. In addition, because BYTE is focused on alleviating societal concerns, as mentioned above, the BYTE interest in including societal organisations in the community, as well as policy-makers, will provide significant added value to already established European activities in this area.

These citizen perspectives are particularly important for BYTE given the research findings from early project activities. BYTE deliverable *D2.2 – Report on public perceptions and social impacts relevant to big data*, examined public perceptions of and aspirations for information practices relating to big data. Specifically, it identified issues that were of particular concern for members of the public and benefits they hoped to gain from big data practices. The research revealed that the main areas of concern relate to the privacy and security of data collected as well as a general distrust of those handling their data, particularly companies operating in the private sector. The main expected tangible benefits associated with big data practices were when data use produces improvements in public security or where developments in health care

⁴ Ibid.

treatment and diagnostics are achieved. According to the *Report on public perceptions*, “recognising these issues is imperative to the continuation of data processing activities and the future of big data as a value adding process”.⁵

The *Report on public perceptions* argued that transparency was a particularly important aspect of ensuring that citizens trust big data practices and agree to contribute their data to them. More transparent policies would allow companies and organisations to better inform their users about the potential societal benefits of big data processing, which would encourage citizen participation. More transparent practices would also alleviate privacy and data security concerns as people build more confidence about what is happening to their data. With respect to developing big data practice into the future, this transparency can be partly accomplished through the involvement of citizen groups, consumer groups and civil society organisations in the BBDC to ensure individual’s rights are protected and inform industry and other big data practitioners about good practice in communicating with the public. As such, it is clear that the involvement of citizen groups in the BBDC will provide added value for Europe beyond the activities already being undertaken.

2.2 BYTE POLICY RECOMMENDATIONS

WP4 (*Evaluating and addressing positive and negative externalities*) has evaluated the societal externalities of big data and came to the formulation of a broad agenda for policy-makers, consisting of updating legal frameworks, promoting big data practices through public investments, and enabling policies and an active policy to keep markets open and competitive.

Several best practices to deal with economic externalities and capture the benefits of big data were identified in deliverable *D4.2 – Report on diminishing negative externalities and amplifying positive externalities*:

- Governments and public institutions can diminish the negative impact of scarce resources by public investment in infrastructures and funding research and innovation programmes for big data;
- They can also fuel the data economy by making more data available through investments in open government data and by persuading “big actors” to release some of their data, e.g. creating incentives for corporate data sharing;
- Another important task of government is to include attention for big data and data skills in education: education policies have to address the current scarcity of data scientists and data analysts, but also promote the inclusion of data skills in a range of educational programs;
- The last practices concern changing the mind-set of corporate actors, both large and SMEs, and research institutes to better value opportunities concerning data and to adopt big data practices.

Opportunities concern both the creation of new data-driven business models and the exploitation of existing data sources. Public authorities, corporate actors and scientific research

⁵ Donovan, Anna, Rachel Finn, Kush Wadhwa, Lorenzo Bigagli and Jose Maria Garcia, *Report on public perceptions and social impacts relevant for big data*, BYTE D2.2, 30 September 2014, p. 4.

institutes can all participate and create new partnerships around data and make new big data use cases possible.

Positive social benefits can be captured with the same best practices recommended to address economic externalities. The evaluation based on the interaction perspective showed interoperability as the key enabling factor. Investment in the several dimensions of interoperability (semantic, technical, legal, organisational) helps to make big data practices possible and capture the associated benefits.

To deal with negative social and legal externalities of big data, some general conclusions were drawn. First, legal frameworks have to be reviewed in order to restore the optimal balance between the different objectives and interests concerned in the context of big data. Regulations need to be adapted to make sure they remain effective in a context with a growing amount of interactions and data flows between actors. Such evaluation can lead to different conclusions depending on the objectives of this practice. E.g. concerning copyright, we noticed that the objective of obtaining the optimal incentive for innovation is best reached in the context of a data economy by limiting the extent of the exclusive rights for the creator. Although other solutions exist which deliver more remuneration to the creators, they present too much of a restriction. The market expansion enabled by removing the legal barriers to use data provides a better result on societal scale. A similar evaluation is not true for data protection, where we deal with two incommensurable objectives. Limiting the extent of data protection in order to foster big data practices implies a strong negative impact on privacy. In this case it is more interesting to look at how the individual transaction-based elements of the data protection framework, which are difficult to scale, can be substituted by collective mechanisms. Such collective approaches can provide a more optimal balance between the protection of privacy and the objective to foster big data practices.

This points to the second element: mechanisms based on individual transactions or individual control lead to high transaction costs in a context of a larger amount of interactions. Or they become dysfunctional, or they present barriers to big data practices. Mechanisms that require an assessment or other action by an actor at each interaction (exchange of data) need to be substituted for mechanisms that enable a more aggregated control and decision making. This implies a shift from a transaction or an individual control model, where the individual decides about each interaction, to a liability-model, where on the contrary decision-making during operations happens more collectively or aggregated, and the active role of the individual gets reduced to receiving compensation or obtaining a claim towards the big data operator.

Thirdly, a specific method to reduce transaction costs is to move a large amount of the decision making to the design phase and to create standardised solutions based on 'by-design' approaches. This requires the translation of societal concerns into technical objectives and requirements that get integrated in the design methodologies. Standardisation is then a further collective process to lower transaction costs by creating well-understood common solutions. This reduces the needed information processing in individual decision making to deciding about which solution to choose from a set of options of which the impacts can be easily understood. The result should be a toolbox of technical, but also organisational and legal safeguards, which can be standardised and integrated into an overall management method during the lifecycle of the data. Privacy-by-design has been developed above mere guidelines

into a set of technical objectives with a continuously developing catalogue of threat models and solutions and into design methodologies mainstreaming these objectives alongside security objectives. Similar work needs to be done to integrate other concerns, like discrimination.

Lastly, we are concerned with socio-technical systems and problems are not purely technical or social. Best practices are often a mixture of legal, organisational and technical elements. Some of these elements can substitute for the other, e.g. encryption as a substitute for legal regulation of access, but more often an effective solution contains a mixture of these elements augmenting each other. E.g. legal protection of personal data is complemented with organisational measures like role specification linked with access rights based on a need to know and technical access control measures preventing unauthorised access.

This implies that design objectives and methodologies and standardisation efforts have to be broadened to include the whole range of legal, organisational and technical elements. This approach led to specific recommendations concerning anti-discrimination, copyright and database protection, confidentiality of commercial information (trade secrets) and data protection.

Dealing with the political externalities also needs to take the specific impact of big data on interactions into account. On the one hand this concerns an active policy to keep markets open and competitive, where the specific ways how a company can acquire a dominant position need to be taken into account and addressed with adequate tools. On the other hand, this concerns efforts by states to retain their regulative capacity. In this case adequate policies and safeguards need to be developed to preserve the balance with citizens' rights and with other states.

2.3 BYTE VISION AND ROADMAP

One of the main goals of the BBDC is to share and disseminate the BYTE vision to support European stakeholders in increasing their share of the big data market by 2020. Hence, the strategy for the community should take into account the BYTE vision, which is identified in the context of WP5 (*Foresight analysis*) along with sector-specific visions for big data. At the time of this writing, WP5 is being completed in parallel; hence this interim strategy has been developed based on a tentative definition for the vision in consultation with the WP5 team. The final version of the strategy will be updated according to the final version of the vision.

Big data impacts choices made, whether for privacy protection, business models, or appropriate legal frameworks. These choices cause a wide range of externalities and impacts - be they positive or negative. Such choices certainly weigh on the competitiveness of the big data industry and these choices influence the ability of companies to launch new services. Nonetheless, making the correct choices can be problematic.

As an example of these impacts consider the discussion around “the right to be forgotten”, also discussed in D4.2. Some may see legal decisions on the “right to be forgotten” as a way to empower citizens. Yet others may see such legal settlements as a limitation of rights - such as the freedom of information. Choices made regarding the right to be forgotten also affect the ability of citizens to hold their elected leaders accountable. Furthermore, it also forces individuals who claim their “right to be forgotten” to alert search engines, and also to justify

themselves regarding matters which they specifically do not wish to have discussed. Thus the many choices concerning the governance of big data are multivalent, create diverse impacts, and are valued differently by different stakeholders.

WP5 has circulated a design document and workshop agenda oriented towards creating the BYTE vision. The key elements of this document and agenda are five goals:

- Reviewing and systematising the impacts of big data across six sectors;
- Examining the critical assumptions underlying European big data governance;
- Identifying new and emerging threats and opportunities for Europe driven by big data;
- Creating four visions of big data, with milestones for tracking and action;
- Developing opportunities for action in hedging and shaping the big data future.

The first of these goals is to integrate the seven BYTE case studies within a common framework of problem structuring techniques and ex ante policy analysis. These techniques draw out the key actors, external forces, system relations and key performance measures from the various cases.

The second goal is to have a joint discussion concerning the major problems in big data, and the root causes behind the perceived problems. Participants in the workshop, which was held in February 2016, have discussed, and WP5 has addressed, the gaps between what is being achieved in various sectors versus what would be desirable for various actors.

The third goal is to create a cross-impact analysis, whereby various impacts in the case materials are identified, and whereby different kinds of impacts across sectors are compared and contrasted. This will enable larger forces shaping big data to be made clear, and it will enable the BBDC to identify new and emergent impacts. Furthermore, the cross-impact analysis will enable the case studies to be compared and contrasted across forces. Most importantly, the cross-impact analysis will help the BYTE project generalise the case work to new, heretofore unconsidered sectors.

WP5, and participants in the corresponding workshop, have created four rich pictures of the future, based on the case studies work and the workshop participation. All four futures were considered to be equally plausible. The objective was not to select a future, but to consider each future concurrently to select appropriate actions to shape the future, or maximise the benefit of impacts. WP5 has also considered actions to hedge against the worst possible impacts of big data. The resulting overall vision has been assessed with project and board members as part of the WP5 workshop and is being completed in parallel with this document.

WP6 (*Roadmapping*) will begin with this inventory of potential hedging and shaping actions. There will be both a policy roadmap, as well as a research roadmap. The objective of the roadmapping activity is to stage and sequence appropriate strategies to achieve the vision elements, and to present these results in a graphically appealing and intuitive manner.⁶

⁶ WP6 was scheduled to start in April 2016, after the delivery of this document, and to release the roadmap at the end of August 2016, as deliverable *D6.1 – A roadmap for big data incorporating both the research roadmap and the policy roadmap*. [The roadmap delivery has been later postponed to the end of October 2016.]

A major goal of the BBDC is to implement the roadmap defined by the project, which will assist industry in capturing current and potential efficiencies, new business models, etc. associated with the collection, analysis, linking and re-use of big data, and proactively address current and potential negative externalities before beginning a project, initiative or programme. The roadmap is being developed in the context of WP6. At the time of this writing, WP6 has not started yet; hence this interim strategy is based on a tentative definition for the roadmap. The final strategy and charter for the BBDC will take into account the final definition of the roadmap, with its possible improvements and evolutions.

WP6 will develop a policy and research roadmap for the development, use, re-use and linking of big data. It will focus on policies, investment, funding and infrastructure required to take full advantage of the opportunities surrounding big data. As part of this work it will rely on the gap analysis undertaken in *Task 5.1 – Sector-specific visions for big data*, as well as the potential policy linkages identified in *Task 5.2 – Vision for big data in society*. The roadmap will outline a step-by-step policy process for meeting the infrastructural, funding and policy needs outlined in the BYTE visions in a socially responsible way that incorporates findings related to positive and negative externalities.

WP6 will use the policy information gathered in *Task 1.2 – Cartography of data flows* to identify key policy building blocks and examples of good practice. The main outcomes of the analysis of the current big data landscape (from Task 1.2) indicate that:

- The rise of intermediation platforms is among the most important phenomena in the development of big data;
- Platforms are concentrated in terms of geography and in terms of number: there are few major platforms and most of them are American; this is true for mobile platforms too⁷;
- Platforms are major societal and economic actors (see, for instance, the rise of the term “uberisation”);
- As a result, European countries are dependent on foreign systems on big data, in particular the European environment is characterised by a dependency from US platforms.

The analysis of the current web landscape and the BYTE vision leads us to set two goals for a policy and research roadmap for big data in Europe:

1. Address the lack of major European platforms:
 - Motivation: our work studies the historical, societal and legal factors for the current lack of European platforms. A roadmap should address these factors. Examples:
 - Build incentives (in terms of tax, funding, etc.) for digital industries and especially platforms;
 - Help legacy activities (e.g. taxis) to transition from traditional businesses to platform economy;

⁷ The legal implications of this situation are also reflected in the decision of the European Court of Justice concerning the US-EU Safe Harbor framework.

- Promote the development of “government as a platform” initiatives and the cooperation with platforms;
 - Rethink data protection: find agreements at an international level (within and outside the EU): follow up the discussions on the Safe harbour for instance;
2. Fill the “institutional gap”:
- Motivation: Currently, courts and some national or thematic institutions set the agenda on big data (see, for instance, the Microsoft case in Ireland or the settlement of the right to be forgotten in Europe). Yet, we are lacking a multi-stakeholder institution that would help to govern the digital activities and build a shared vision of big data between practitioners, governments and the members of civil society.

WP6 will further take into account the policy recommendations made in WP4 on how to address positive and negative externalities of big data, discussed in the previous chapter.

3. INTERIM STRATEGY FOR THE BIG DATA COMMUNITY

This chapter outlines the overall approach pursued by the BYTE consortium to create and initiate the BBDC. This choice has a big impact on the long-term evolution foreseen for the community itself.

Many elements of the BYTE recommendations concern review and adaptation of legal frameworks. The BBDC may make such review and present its recommendations, in other words formulate policy advice. Other elements concern the development of instruments to deal with big data issues, like design methodologies, standards, etc. Although arguably putting these in practice should be the task of a technical community, the BBDC may contribute to the development process: through recommendations on a research agenda, input in standardisation, involving regulatory bodies, etc.

Which parts of the recommendations would be most relevant to the community strategy depends on who is part of the BBDC and what is the relation between the BBDC and the rest of the big data ecosystem. A significant distinction is that between a community of big data practitioners and a community of stakeholders: the former is the technical and industrial community building big data technology and applications; the latter is the wider circle of people, industry, and other parties who use the big data applications, or is subject of its use (e.g. NGOs dealing with privacy, broader industry groups, civil society organisations). As elaborated in the previous chapters, the BDVA covers mostly the practitioners' side, whereas the BBDC should complement it with other stakeholders, e.g. scientists in Social Science and Humanities, NGO, and the society at large.

The next chapter discusses three possible approaches which have been considered for forming and implementing the BBDC, with their respective pros and cons, and the selected interim strategy. Based on that choice, the following chapter outlines the value proposition of the BBDC.

3.1 POSSIBLE APPROACHES

The strategic approaches for forming and implementing the BBDC can be broadly characterized as the following options:

1. Autonomous community;
2. Umbrella organisation;
3. Contribution to an existing initiative.

Strategy 1 implies setting up a totally new and independent community to put forward the BYTE recommendations, vision and roadmap. Strategy 2 means to gather the existing efforts into some sort of unifying federation that would coordinate their efforts, as far as they are concerned with the objectives of BYTE. Strategy 3 means to establish a link and a close collaboration with one of the existing initiatives, trying to gain trust and influence on the aspects of interest to BYTE goals, so as to steer its governance accordingly. Under the latter perspective, it may be foreseeable (or even desirable, as an indication of full success), that the community eventually merge into the chosen initiative.

The following table summarises the main aspects of the three above approaches, in terms of positive/negative implications and opportunities.

Table 1 - implications of the community strategies

	1-Autonomous	2-Umbrella	3-Contributor
Strengths	Complete freedom to organise activities and goals	Scalability and flexibility	Direct and immediate influence
Weaknesses	Significant effort for creation and maintenance	Impact on members is indirect and mediated	Must adapt to existing governance, etc.
Opportunities	Build the BYTE reputation around societal issues as a core focus	Reach many stakeholders	Leverage/optimize existing resources
Threats	There are already many organisations	Mass of BYTE may be insufficient to attract enough parties	Limited willingness to cooperate and/or limited impact of the chosen initiative

Alternative 2 would allow reaching many stakeholders, but is critically dependent on the influence that BYTE may exercise on the prospective members of the community. Given the size and scope of the BYTE consortium with respect to the complex big data ecosystem introduced in chapter 2.1 and its intricate relationships with strategic industrial assets, there is a significant risk that BYTE may not be in the position to sufficiently mobilise the existing efforts, or at least a relevant fraction of them, under its leadership, although limited to a few selected areas, such as ethical or legal implications of big data. Besides, even in that case, the impact of an umbrella organisation is necessarily mediated and hence limited by the mechanisms that are actually operational for the participant organisations. In summary, the combined implications of this threat and weakness suggest that option 2 is the least preferred for the strategy of the BBDC.

We can hence focus preferably between option 1, creation of a new specific community, and option 3, joining the efforts of an existing one. The BYTE DoW seems to refer to the community as a self-standing, autonomous entity, comparable to similar ones promoted by other projects such as BIG and Planet Data. However, not only is the option to collaborate and liaise with existing initiatives not excluded, but actually the *Task 7.4 – Input into related networks* will specifically examine how the BBDC can successfully link with related networks and initiatives in order to input BYTE findings and influence their policies, goals and processes. Hence, we would consider option 3 to be consistent with the project remit, as laid out in the DoW, and in line with the project objectives.

Option 3 seems also more directly sustainable, as it allows tapping into resources that have already been allocated (of course subject to the willingness of the existing initiative to cooperate with ours). Instead, this opportunity turns into a threat for option 1, when BYTE would have to advocate for yet another initiative in the big data ecosystem, and to maintain it competing for funding reallocation. In particular, it is important to note that the EU big data landscape has drastically changed with the set-up of the Big Data cPPP in 2014, compared to when BYTE was designed and planned. Hence, option 3 may arguably be the most practical and effective, given BYTE scope, strengths, aims, and context.

Given the option of contributing to an existing initiative, the choice of the target is decisive for chartering the BBDC, as it implies its scope and type of activities, the immediate recipients of BYTE contributions, and ultimately its overall impact. Considering the scope of BYTE and the composition of its consortium, we judge that the BBDC would be able to best contribute to the following major efforts in the big data ecosystem, in order of preference:

- BDVA - the EC has officially endorsed the BDVA as its private counterpart in the aforementioned Big Data cPPP; as widely elaborated in chapter 2.1, given the remits of BYTE, the BBDC would be in the position to provide added value without duplicating activities;
- Group on Earth Observation - there are many expectations and concerns about Big Data in the geospatial sector; besides, GEO is progressively widening its scope to many societal benefit areas; BYTE consortium members are key contributors to GEO Data Sharing activities, which also address big data challenges;
- Research Data Alliance - RDA is mainly addressing researchers and innovators, from the technical but also from the social viewpoint; the Big Data Interest Group is chartered to produce a set of recommendation documents to advise diverse research communities.

Based on the above rationale, the consortium agrees there is high value for the BBDC in contributing to the BDVA, joining, and possibly ultimately merging into it. Hence, this interim strategy takes as a first-choice working hypothesis that the BBDC will mainly channel its contribution through the BDVA.

3.2 Value proposition

The EU Big Data debate (and resources) has insofar mainly targeted and mobilised the industry and the academia, such as in the case of the Big Data cPPP. As a consequence, NGOs (Non-Governmental Organisations), NPOs (Non-Profit Organisations), CSOs (Civil Society Organisations) and actors of the third sector in general, who are usually unable to divert limited resources from their core objectives to gain proactive insight on this subject (e.g. by joining/observing the BDVA), have been largely excluded from it, though definitely important stakeholders in the big data revolution.

The BBDC could support them in taking a more active part. For example, recognizing that virtually no one in our society is fully aware of the policies governing the information he/she produces, the BBDC could improve the awareness on the existence and use of personal data,

gathering and divulgating the policies of the industrial data holders (cf. Sherpa/Romeo⁸ on publisher's policies regarding the self-archiving of journal articles in Open Access repositories).

Constructive interaction between industry and society is difficult to achieve: societal organisations may be reluctant to engage with the industry, to avoid being instrumentalised; the industry may rather focus on business than on societal implications. Still, particularly in the present transition to a data-driven economy, the EU Big Data PPP, that is essentially the BDVA and the EC, should include societal actors as first-level parties. The BBDC may be the venue where this is materialized, as represented in Figure 1.



Figure 1 – the BBDC between BDVA and society

The BBDC value proposition for societal stakeholders includes the following benefits from participation:

1. Facilitate direct contact with European industry and policy makers, providing opportunities to channel instances of interest;
2. Opportunity to contribute to the BDV SRIA;
3. Opportunity to contribute to EC workplan (e.g. H2020 calls), through BBDC members who are also part of the BDVA;
4. Opportunity to steer the BBDC Research and Policy Roadmaps, equivalent documents to the BDV SRIA, and take an active role in BBDC evolution.

The BBDC also represents an added value to the BDVA. The proposed collaboration strategy could augment the reach and the positioning of the BDVA within the European big data innovation landscape. Furthermore, because integrating different types of stakeholders is a key objective of both initiatives, a combined effort will increase the impact of both the BDVA and the BYTE project. The BBDC roadmap will provide updated information in relation to the roadmap that the BDVA is continuously updating, i.e. the BDV SRIA. The BBDC can also provide practical and material support to BDVA networking events, and organise joint events to link the BBDC and the BDVA in stakeholders' minds, as well as to attract new stakeholders to BDVA events and thus augment the reach of those events. Finally, the BBDC could build

⁸ <http://www.sherpa.ac.uk/romeo/index.php?la=en&fidnum=|&mode=simple>

and administer a shared blog where stakeholders could share information across the different stakeholder categories and problem areas relevant to the BBDC and the BDVA.

The success of the proposed strategy is depending on what the BDVA is willing to accept as a contribution from BYTE and the BBDC. The BYTE consortium has already begun a process of exploring such collaboration with the BDVA steering committee. Consortium members who are existing members of the BDVA have been collaborating with BDVA counterparts to identify the extent to which the two communities' goals are aligned and identify benefits of collaboration on both sides. In the first instance, the BYTE project has identified that although the BDVA has societal and legal working groups, the organisation is currently having difficulty encouraging appropriate stakeholders to join the organisation and contribute to these discussions. Such stakeholders include social science focused academics, civil society organisations and citizen groups. The BBDC is specifically meant to engage such stakeholders, and we can augment the reach of the BDVA by involving them in synergetic activities. This may require a slight change in BDVA membership requirements, for these types of organisations to enable them to adequately participate in BDVA activities. In addition, the BBDC can benefit from the impressive network of industry stakeholders and practitioners that the BDVA has already built, as well as benefit from the organisations' existing reach and centrality within European policy-making. As such, this collaboration would produce useful synergies for European policy and innovation in this area by creating a one-stop-shop for big data expertise.

The BDVA leadership has responded positively to our proposal, and the BYTE consortium is currently in talks with them about how collaboration could be achieved. From the discussions initiated, it seems that the BDVA steering committee would encourage us to contribute on the legal/societal aspects and has already indicated that BYTE could take a leading role in the two task forces linked with societal externalities: the societal task force and the legal task force (possibly about to join into a single task force). While this solution would be highly desirable and fully in line with BYTE scope and objectives, it may require some further negotiation with the existing task force leadership.

The following chapter outlines the strategy that BYTE will use to ensure the widest possible impact of the BBDC. In light of the chosen option, chapter 5 and 6 anticipate an interim Development Plan and Charter for the BBDC. In the future months, we will continue liaising with the BDVA, to agree and detail the BBDC contributions to the BDVA operations, which will allow to finalise the strategy and charter of the BBDC, as planned.

4. INTERIM IMPACT STRATEGY

This chapter elaborates on the strategy that BYTE will assume to maximise its impact on the policies and practice related to big data in Europe. As requested during the Year 1 Review, this interim impact strategy focuses on the identification of the stakeholders to engage, how to contact them and how to incentivise their participation, both in general and at key stages, such as after the roadmap delivery (M30) and at the close of the project (M36). As requested during the Year 2 Review, further information has been added on how the BYTE consortium is going to create incentive mechanisms to facilitate the active participation of stakeholders in the BBDC.

As underlined by the project DoW, stakeholder engagement and dissemination are crucial to the success of BYTE and its community.⁹ The impact, within the scope of this document, is defined as the demonstrable contribution made by the BYTE project. Stakeholder engagement and dissemination activities lead towards such contributions to the BBDC. Integration of stakeholders across various sectors, within the BYTE project, will facilitate knowledge exchange. Sharing of information about challenges and strategies will enable learning opportunities for stakeholders; thus incentivising long-term stakeholder engagement in the future via the BBDC. Furthermore, it will help developing effective strategies for dealing with positive and negative externalities.

The primary goal of the impact strategy is to raise awareness of the BYTE project among stakeholders and other actors relevant to big data ecosystem in Europe. The impact strategy, as described here, is a comprehensive plan to achieve the planned impact of the BYTE project. The stakeholder engagement activities will lead to the creation of the BBDC. The impact strategy is aimed to maximise the number of stakeholders who become aware of BYTE activities, recommendations, and the BBDC. The tools to be employed for generating impact will not be limited to mass-mailed notification. Stakeholder engagement and dissemination will be done through tools such as workshops, public talks, social media, etc.

The specific objectives of impact strategy to encourage participation in the BBDC are:

- Identification of key stakeholders and development of a stakeholder taxonomy;
- Engagement of stakeholders throughout the key stages and events of BYTE;
- Dissemination of BYTE activities and deliverables through various channels.

The remaining of this chapter describes three parallel activities undertaken as part of WP8 (*Stakeholder engagement*) and WP9 (*Dissemination*). First and second, a set of high-level categories of stakeholders are introduced and their related engagement activities are summarised. Third, current and planned dissemination activities of BYTE project are detailed.

4.1 STAKEHOLDER TAXONOMY

The first objective of the impact strategy is met with deliverable *D8.1 – Stakeholder taxonomy*, which defines, identifies, and classifies the BYTE stakeholders. A BYTE stakeholder is any

⁹ BYTE Description of Work, Overall strategy and work plan, p.66.

group or individual who can affect or is affected by the information ecosystem in a positive or negative manner. The high-level taxonomy has 4 main stakeholder types:

- **Data Providers:** Person or organisation that provides the data to the Big Data ecosystem;
- **Data Users:** Persons or organisations that consume information from the Big Data ecosystem;
- **Enablers:** Persons or organisations that support the function of the Big Data ecosystem;
- **Secondary Stakeholders:** Persons or organisations that influence or are impacted by the Big Data ecosystem and its operations, but do not interact directly with the Big Data.

Each of these stakeholder types is applied across different sector to identify potential targets of engagement activities. The following tables provides cross tabulation of stakeholder types and sectors.

Table 2 - list of potential stakeholders for different sectors

<i>Sector</i>	Data Providers	Data Users	Data Enablers	Secondary Stakeholders
<i>Environment</i>	- Environmental Scientists	- Environmental Scientists - Citizens - Policy Makers	- IT Engineers - Data Engineers - Network Engineers	- Policy Makers - Data Protection - Officers
<i>Crisis Management</i>	- Citizens - First Responders - Scientists - Business Owners - Local Government	- Citizens - First Responders - Crisis Managers - Scientists - Policy Makers - Legal specialists - Economists - Business owners	- IT Engineers - Data Engineers - Network Engineers	- Policy Makers - Data Protection Officers - Community Groups
<i>Utilities</i>	- Scientists - Citizens - Business Owners - Local Government - Grid Operators - City Government	- Scientist - Citizens - Policy Makers - Consumers - Economists - Legal - Economists - Energy Consumers - Energy Producers - Traffic users - City Government	- IT Engineers - Data Engineers - Network Engineers	- Policy Makers - Data Protection Officers - Community Groups - Citizens
<i>Cultural</i>	- Cultural Scientist - Librarians - Historians - Archivists - Citizens	- Cultural Scientist - Librarians - Archivists - Historians - Citizens - Policy Makers	- IT Engineers - Data Engineers - Network Engineers	- Policy Makers
<i>Energy</i>	- Scientists - Grid Operators	- Scientists - Citizens - Policy Makers - Economists - Safety Officers - Human Resources	- IT Engineers - Data Engineers - Network Engineers	- Policy Makers - Data Protection - Regulators
<i>Health</i>	- Patients - Health Care	- Health Care Professionals	- IT Engineers - Data Engineers	- Policy Makers - Regulators

	Professionals - Medical Research Scientists - Citizens - Pharmaceutical Scientists	- Patients - Public Health Policy Makers - Pharmaceutical Scientists	- Network Engineers	- Economists - Data Protection Offices - Citizens
<i>Transport</i>	- Fleet owners and operators - Equipment vendors	- Citizens - Transport Policy Makers - Fleet owners and operators - Port authorities - Community Groups - Freight owners	- IT Engineers - Data Engineers - Network Engineers	- Transport Policy Makers - Regulators - Community Groups - Economists

In general, the sector-wise stakeholders are differentiated for data providers and data users, in terms of their nature of work and interest in the BBDC. Although data enablers and secondary stakeholders are somewhat similar across sectors, their involvement in a community is driven by the sector's specificities.

4.2 STAKEHOLDER ENGAGEMENT STRATEGY

The specific strategy for stakeholder engagement focuses on maximising participation of stakeholders and Advisory Boards in the BYTE project. The engagement activities are scheduled at key stages of BYTE project: foundation (M18), after the roadmap delivery (M30) and at the close of the project (M36).

The project maintains a mailing list of stakeholders within the European big data community. This mailing list has more than 900 subscribers and is constantly growing over the lifetime of BYTE project. The mailing is primarily used to communicate the status of BYTE project and future planned events, through quarterly newsletters. Additional on-demand newsletters are sent to subscribers for communicating agenda and logistic details of BYTE specific events such as workshops.

The BYTE Advisory Board serves a key role in impact strategy of the BYTE project. Periodic interaction with members of the Advisory Board enables sectors specific oversight of the BYTE vision, roadmap and implementation plan for the BBDC. For this purpose, the advisory is actively engaged for participation in BYTE workshops through invited talks and panel discussions. Additionally, a conference call was organised between Advisory Board members and project partners in last quarter of 2015. The purpose of this conference call was to engage Advisory Board members in the BYTE project by reporting on the previous activities; gathering feedback on completed deliverables and impact on community; and discussing future plans for the BYTE project. Furthermore, the board members were encouraged to participate in future workshops and facilitate further networking and stakeholder engagement within their sectors.

To discuss sector-wise externalities of big data, a virtual workshop for Advisory Board members was organised in last quarter of 2015. The workshop reported on the results of following activities of BYTE to participations:

- Case studies and focus groups in crisis informatics, culture, energy, environment, healthcare, transportation and smart cities;
- Horizontal analysis and social impacts of positive and negative externalities;
- Strategies amplifying positive and diminishing negative externalities.

The outcomes of each activity was validated with the workshop participants; therefore, facilitating stakeholder engagement. A similar approach will be continued with future planned workshops of BYTE. Besides these workshops, the project partners also engage stakeholders through presentations and talks at stakeholder specific events including national, international and European conferences. In this regard, project partners are encouraged to report and disseminate their activities on the BYTE project website through blog posts.

The following table summarises the activities that are already undertaken and will continue, as part of the impact strategy of the BYTE project. Each activity is associated with key performance indicators that are used at key stages of the project to assess progress.

Table 3 - stakeholder engagement activities

Activity	Tools	Success Criteria
<i>Identity communication</i>	BYTE Logo	
	BYTE Website	Number of unique visitors
	Mailing List	Number of subscribers
	Digital Newsletters	Number of newsletters Newsletter opening rate
<i>Events</i>	BYTE Workshops	Number of participants
	European Conferences	Attendee participation Number of BYTE presentations/talks
	International Conferences	Attendee participation Number of BYTE presentations/talks
	Local Seminars	Attendee participation Number of BYTE presentations/talks
<i>Social Media</i>	Twitter	Twitter mentions and retweets
	Facebook	Facebook shares and likes
	Blogs	Number of blog posts
<i>Print Media</i>	Newspapers	Number of news articles
	Institutional Press	Number of institutional articles
<i>Direct Meetings</i>	Onsite Interviews	Number of interviews Sector-wise participation
	Focus groups	Number of focus groups Sector-wise participation
	Conference Calls	Advisory Board participation

Invited Talks	Number of talks by AB and stakeholders
---------------	--

Another key aspect of stakeholder engagement in BYTE project is the on-going collaboration with the BDVA. To this end, BYTE project partners are directly involved in the BDVA and actively contribute towards its management and sustainability.

4.2.1 INCENTIVE MECHANISMS

To create incentive mechanisms to facilitate the active participation of stakeholders in the big data community, the project will offer a direct link between BBDC members, and especially those from the area of NGOs, NPOs and CSOs, and the European industry. With the aim to avoid duplication of tasks and efforts with the BDVA, the community plans to serve as a channel to contribute to the BDV SRIA and related initiatives, with a special focus on the societal impact of big data research and policies.

In this regard, the BYTE policy and research roadmap will make broad recommendations and define an action plan until 2020, but also undertake a deeper study of selected sectors: each year a group of three sectors will be addressed in detail to produce special recommendations and actions. This will be taken up initially by the BYTE project partners and community members, and by the BYTE community alone after project completion. As part of the dissemination strategy, these results will be fed to BDVA, as well as to other relevant networks and projects such as the ones identified in Appendix A, thus providing BYTE big data community members with the incentive and opportunity to contribute to the BDVA and EC agenda on big data by putting the accent on its societal implications. To this aim, at least an annual roadmap updating workshop will have to be organised by the BBDC, preferably in co-location with relevant civil society events or BDVA annual summits. While BYTE is still active, selected individuals will be invited to become members of the project Advisory Board. This will create immediate opportunities to directly support and incentivise their participation in the initial activities of the BBDC. The BDVA has currently 150 member organizations from industry, academia, and research; in addition, it has 10 task forces including community, legal, and societal. This significant participation, of Big Data stakeholders in BDVA, serves as another incentive to civil society organization and individuals to engage with BDVA through a common platform like BBDC.

The specific incentive mechanisms to be put in place for encouraging participation in the BBDC are:

Category	Incentive
Contribution to policy and research agendas	<ul style="list-style-type: none"> ● Contribution to the annual update of BYTE Policy and Research Roadmaps for societal aspects of Big Data ● Contribution to the BDV SRIA through membership of the societal task force
Easy access to relevant stakeholders	<ul style="list-style-type: none"> ● Direct access to Big Data industry stakeholders through BDVA

	<ul style="list-style-type: none"> ● Communication and dissemination opportunities through managed contacts of industry with minimum administrative overhead ● Networking through BBDC organized workshops and targeted events
Early access to information	<ul style="list-style-type: none"> ● First hand access to information related to EU policy and programmes for Big Data ● Industry or sector specific analysis and reports
Influence on EU workplans	<ul style="list-style-type: none"> ● Opportunity to influence EU workplans (e.g. H2020) through common members between BDVA and BBDC ● Increased visibility on national and EU levels ● Opportunity to make cavity society as first-class citizen in Big Data ecosystem

4.3 DISSEMINATION STRATEGY

The dissemination strategy aims to promote and encourage participation in the BBDC and to publicise its activities. It is divided in three phases and uses multiple channels that can be grouped in four categories. As described in the deliverable *D9.1 – The BYTE dissemination strategy*, the three strategy phases are the creation of the dissemination strategy (in months 1 to 3, i.e. March to May 2014), its implementation (for the whole duration of the project, but especially from months 3 to 36, i.e. May 2014 to February 2017) and the sustainability of the dissemination activities (in months 34 to 36, i.e. December 2016 to February 2017).

The first phase resulted in the deliverable D9.1. In the second phase, dissemination activities are performed via the four categories of channels, namely online presence, promotional materials and publicity campaigns, publications and conferences and through the BBDC. The online presence, in the form of a project website and social media channels such as Twitter, Facebook, LinkedIn, Slideshare and Mendeley, is aimed at providing a high outreach and visibility targeting all stakeholders. In particular, all results and deliverables of the project are made available on the website, and presentations are published in Slideshare. Particularly useful to promote these results and engage the target stakeholders is the Twitter channel. Especially, live updates during the project workshops and at relevant high-profile events, such as the European Data Forum or the ICT conference, are directed to establish a dialog with relevant stakeholders and engage them. Promotional materials and press releases are also addressed to all stakeholders, and especially to the general public, while publications and presentations at conferences target stakeholders more interested in research. During the formation of the BBDC, these three groups of channels will be also devoted to publicise the activities of the community, engage its members and create ties between them. To this aim, the project online presence will be used, promotional materials such as posters about the BBDC and its work will be created, and BBDC events will be supported. The fourth group of channels relies on the community that will be established by WP7.

To establish the BBDC, we will exploit the project research findings and the definition of a vision and a roadmap for Europe. As in all the project workshops, the workshop on Visions for big data and foresight analysis will invite Advisory Board members from the seven industries and disciplines and use the contact list to call for other experts and representatives in the social,

economic, legal and political fields, as well as members from industry and civil society organisations that aside from providing feedback on the BYTE vision will be engaged in the creation of the BBDC, and in particular in the development of the roadmap. To obtain an even higher impact within stakeholders, the roadmap workshop will be co-located within a major conference to increase attendance, obtain high impact, involve a broad spectrum of stakeholders in the validation of the roadmap and build a consensus as large as possible around it. Case study organisations will also be invited and encourage to participate in the validation of the roadmap with their expertise in their respective fields.

Other events will be also used to build the BBDC and attract interest. As an example, the co-organisation of the European Data Economy workshop within the SEMANTICS 2015 conference, held in Vienna, facilitated networking activities toward building the BBDC. Among other participants, several European projects aside from BYTE were present, such as BigDataEurope, RETHINKbig, and ODINE. Presentations in other events, such as the European Semantic Web Conference 2015 or the European Data Forum 2015, and networking session such as ICT here, have been also working towards the construction of the community. These activities will continue and be intensified during the last year of the project. In this last year, the final conference of the project will be organised. Aside from presenting the major findings of the BYTE project and to disseminate the roadmap, this event will be used to publicise the BBDC and the future activities that it will undertake after completion of the project.

5. INTERIM DEVELOPMENT PLAN

This chapter lays out a tentative plan for the development of the BBDC, as anticipated at the time of writing this interim strategy. The plan will be precisely defined in the future months, reflecting the refinements of the strategy (e.g. the details of the possible agreed BBDC contributions to the operations of the BDVA).

The plan is tightly connected with the findings of the other BYTE work packages, to ensure that it remains robust and comprehensive through the project and beyond. The design of the BBDC takes into account the other national, European and international current research projects, communities and initiatives dealing with big data and other issues relevant to big data, identified in BYTE WP1 (*Setting the stage on big data*). In particular, we have analysed their charters, scoping best practices and gaps to address. The BBDC is shaped around BYTE recommendations, vision and roadmap, defined within WP2-5, also capturing the contributions of the stakeholders identified and invited to participate in BYTE workshops. This will facilitate the engagement of those stakeholders in the development of the community. Besides, the BYTE Advisory Board is asked to provide input into the design of the BBDC.

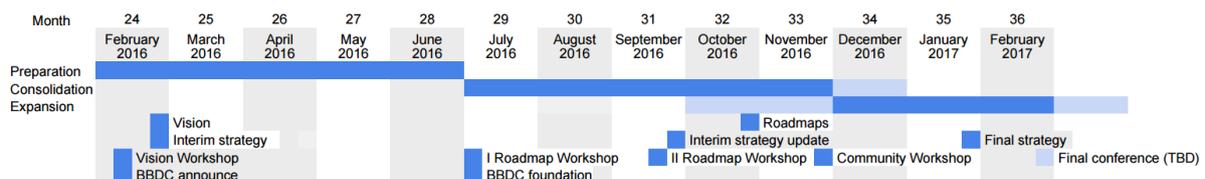


Figure 2- BBDC interim development plan

The first step of the development plan (see Figure 2) will be the preparation work for the foundation of the BBDC. This step foresees the primary involvement of the BYTE consortium, whose large size ensures that there is sufficient cross-disciplinary representation and sufficient representation of different types of organisations within the BBDC. In addition, all Advisory Board members, depending upon their availability, will be asked to act as founding members, again to ensure that the results of BYTE integrate as many disciplinary and stakeholder perspectives as possible, to assist in disseminating BYTE findings to a suitably wide range of stakeholders within the European big data network, and to encourage a diverse membership in the BBDC.

Apart from the BYTE partners, the founding members will ideally be eight, e.g. six from the Advisory Board, two from outside the project, representing their institutions (possible candidate external founding members are BDVA official Nuria De Lama, Samuel Schauss from SNCF, and Science International expert Sally Wyatt).

The preparation phase for the foundation of the BBDC will start at month 24, with the release of the BYTE vision and of this interim strategy and charter. The prospective founding members of the BBDC will have to first objective to validate, assess and improve the interim strategy, charter, and sustainability plan of the BBDC, anticipating its alignment and harmonisation with the policy and research roadmap that will be delivered at month 30¹⁰. The BBDC will be

¹⁰ The roadmap delivery has been later postponed to month 32 (end of October 2016).

officially founded during the forthcoming BYTE workshop (WP6; presumably around month 30), in a specific agenda item planned with the organisers. A possible co-location option is the European Data Forum 2016, planned at the end of June.

The second step of the development plan for the BBDC is its consolidation and enlargement to key big data stakeholders identified and engaged in the course of the project, namely among the workshops participants. Stakeholder engagement is central to the successful coordination of big data actors required by the implementation of the roadmap and the development of the BBDC will specifically enable stakeholders to implement the project's guidelines and provide a space for discussion for future collaboration and problem solving on the economic, legal, social and ethical impacts of big data innovations. It will also enable the implementation of policies and practices that will support European scientists and industry in capturing a competitive share of the big data economy whilst maintaining and protecting fundamental rights, personal data, intellectual property, etc.

This phase of development will be supported primarily by the activity conducted in *Task 8.3 – Build and strengthen the cross-disciplinary big data ecosystem*. The consolidation of the BBDC will be assessed and measured during the upcoming community workshop (WP7).

The next step of the development plan for the BBDC is its expansion to other big data actors in the BYTE contact list (social science scholars, open data activists, statisticians, computer scientists and civil society organisations). At the moment of this writing, the contact list includes about 900 entries. We will consider segmenting, ranking, selecting and identifying who to contact, depending on the overall strategy for the community and keeping track of the contacts made. This phase will be supported primarily by the activity conducted in *Task 8.1 – Mapping the relevant constituencies in the big data ecosystem* and *Task 8.2 – Create a contact list*.

WP9 will also play an important role in this phase, ensuring that we encourage the right stakeholders to participate in the BBDC, reaching them several times in different and appropriate ways, and achieving the expected impacts in terms of engagement and uptake of BYTE research findings. A particular focus of this phase of development will be the further expansion of the BBDC to representatives of civil society organisations, given BYTE commitment to consider public perceptions and aspirations. All BYTE partners will heavily recruit stakeholders from civil society organisations to participate in the BBDC. While BYTE is still active, selected individuals will be invited to become members of the project Advisory Board. This will create immediate opportunities to directly support their participation in the initial activities of the BBDC.

To this aim, we will make use of the contact list developed in Task 8.2 and the stakeholder taxonomy of Task 8.1 to personally address representative of civil society organisations, as these organisations have proved difficult to encourage to participate in FP7 activities and the analysis of WP8 will enable to specifically incentivise stakeholders in this segment. One of the main tools to do that is the organisation of workshops that also target such stakeholders. Moreover, we will explicitly ask the founding members of the BBDC to engage civil society organisations. As mentioned, around six member of the Advisory Board would ideally join as founding members of the BBDC. If available, an ideal candidate for this task would be Chris

Jones from Statewatch, who could not only bring his organisation but also engage others, e.g. EDRi (European Digital Rights) and its members, of which Statewatch is one.

Drawing from the case studies experience, and in particular with regard to the DNV GL role in the maritime case, we observed that a typical big data barrier is that data are proprietary, and are not shared to an independent third party who could add value (see for example Tables 44 and 45 and chapter 2.3 in deliverable *D3.2 – Case study reports on positive and negative externalities*). This challenge could be solved by bringing a variety of regulators into the BBDC, who can act to overcome this big data barrier of proprietary data and unwillingness to share by providing regulations that enforce disclosing and sharing data for societal safety reasons. Attracting such regulators to the BBDC would then not only provide valuable members to the community, but also contribute to engage civil society organisations. This can be done by providing a common meeting point in the future BYTE workshops, where such actors can get involved in the BBDC.

In this phase, BYTE partners will also leverage their relationship with other projects, studies, and networks, to further expand BYTE networking activities and encourage participation in the BBDC. The partners will systematically establish contacts with other relevant projects, communities and studies, to involve them in the BBDC and to integrate BYTE into the growing array of research work being conducted in Europe and internationally.

As identified in DoW chapter 2.3 and 3.1.4, BYTE partners are involved in other European projects and/or initiatives relevant to big data (RECODE, OPTIQUE, BIG, ROBUST, PRACE, etc.) and will be able to leverage these relationships to encourage participation in BYTE collaboration activities. As the BBDC is intended to be an overarching and comprehensive initiative, possible overlaps will be functional for the BBDC to act as a focal point for all different types of stakeholders and all aspects of big data.

The expansion phase of the BBDC development will culminate in the community workshop (WP7; around or before month 35). During this, the founding members of the BBDC and European and Member State policy-makers will adopt the strategy charter and the sustainability plan for the BBDC, whose final version are due at month 35. The workshop will involve all the constituents of the BBDC to express their consensus on the guidelines outlined in the final report. Finally, it will be the first opportunity to identify and discuss strategies for inputting into related networks and initiatives, and discuss how to publicise, populate and strengthen the BBDC for the future.

To maximise the impact, we will consider possible co-location options, e.g. eChallenges 2016 (normally in October-November¹¹); ICT Conference 2016 (normally in October¹²); RDA Symposium or Plenary Meeting; BDVA events like the General Assembly or a Summit. Shortly after, around February 2017, the BYTE final conference will be another opportunity for a face to face meeting where further promoting and disseminating the BBDC, perhaps to a more institutional audience.

¹¹ See the 2015 edition website: <http://www.echallenges.org/e2015/>

¹² See the 2015 edition website: <http://ec.europa.eu/digital-agenda/en/ict2015-innovate-connect-transform-lisbon-20-22-october-2015>

6. INTERIM CHARTER OF THE BIG DATA COMMUNITY

This chapter lays out a tentative charter for the BBDC, as anticipated at the time of writing this interim strategy. The charter will be precisely defined in the future months, reflecting the detailed agreements on BBDC contributions to the operations of the BDVA.

6.1 GOVERNANCE STRUCTURE

In line with the first-choice working hypothesis of this interim strategy, the BBDC will align with the appropriate elements of the BDVA governance structure. This is for two reasons: first, because the organisation of the BDVA is quite democratic and inclusive; second, because we might choose to consolidate with it in the future. Ideally, the BBDC would participate in the BDVA structure, e.g. as one of the thematic task forces, or a subgroup thereof, such as Task Force 5 (Legal) or Task Force 9 (Societal). Task Force 3 (Community) and 2 (Impact) may also be of interest for the activities of the BBDC.

We envision a quite flat structure for the BDVA members' assembly. The founding members of the BBDC and the BYTE consortium will be assigned a steering role, at least initially, until the end of the project, and possibly until 2020.

6.2 MEMBERSHIP

The BBDC is open to both individuals and organisations. The following categories of membership are defined, in alignment with the BDVA:

Category	Description	Opportunities	Engagement
Interested Party	The member is in the BBDC contact list, possibly contributes in identifying key issues in sectors of interest and discusses with peers on topics of interest	Stay informed and interact via the BBDC website, newsletter, mailing list, etc., participate in consultations	Low level – reactive Occasional telecon/meeting
Associate Member	As above, plus the member regularly contributes on focus areas, reviews outcomes from the BBDC	As above, plus option to participate in research and innovation projects (priority of choice given to Full Members)	Regular telecons Occasional meeting or annual workshop Involvement in specific areas of interest, to influence and shape the BBDC and the roadmap
Full Member	As above, plus the member commits	As above, plus priority option to participate in research and innovation projects	Signature of a letter of intent (see D7.2.1, Annex A)

	resources to the activities of the BBDC	Support BBDC management in organisational and strategic decisions	120 hours/year (0,75PM), for management / operation and/or research and consultation work Participation in annual relevant meeting, (agreed with BBDC management)
Full Member + BDVA	As above, plus the member is also a BDVA Member (Associate or more)	As above, plus option to lead research and innovation projects Take part in management, organisation and strategy of the BBDC	Complete BDVA application process with related commitments

6.3 KEY GOALS AND OBJECTIVES

As stated in the project DoW, the key goals of the BBDC are:

- Implement the BYTE policy and research recommendations (roadmap), to achieve the BYTE vision;
- Feed the roadmap back to members' network;
- Continue the work of BYTE, investigating the positive and negative societal externalities of big data and the ways to make the best of them.

6.4 TYPES OF ACTIVITIES FORESEEN

In line with the first-choice working hypothesis laid out in this interim strategy, the BBDC may organize, or participate in, BDVA activities such as:

- Yearly summits, held generally 1-2 times a year, alternated with smaller ones;
- Activity group meetings, typically held 8-10 times a year (every 6 weeks), in Brussels, and consisting of a plenary meeting and 6-8 sessions around Task Forces or emerging topics;
- Mailing lists, teleconferences, etc.

As regards the financial aspects of BBDC governance and activities, a tentative funding plan is laid out in D7.2.1, §3.

7. FUTURE WORK

This interim strategy and charter will be finalised in the third and final year of the project, incorporating the changes resulting from the preparation, consolidation and expansion phases.

In particular, in the next (ca. six) months before the foundation of the BBDC, we will continue liaising with the BDVA to agree and detail our contributions to the BDVA operations (namely to the Legal and the Societal task forces). We will also make our value proposition for the EC on the chosen strategy, as an agenda item for the Year 2 Review in April 2016. After month 30, we will work on aligning the strategy and charter of the BBDC to the BYTE roadmap for policy and research, steering the expansion phase accordingly.

The above activities will allow us to refine the strategy and to achieve the most effective definition of the development plan and charter for the BBDC.

In parallel, we will validate and refine the impact strategy, assessing our compliance with the values and expectations of BBDC members. We will coordinate with WP8 and the BYTE Impact Manager in promoting stakeholder engagement, to ensure that the Advisory Board members, case study organisations, project partners and Community members actively contribute to meet the goals of the Community. Furthermore, we will coordinate with WP9 to stimulate stakeholders' interest in the community via website, press releases, promotion campaign, journal articles and presentations at third-party conferences, to demonstrate the added value of the community, thus incentivising them to participate and increasing its impact.

The BBDC impact will be assessed at the end of the consolidation phase, after the foundation of the BBDC and the delivery of the roadmap (month 30), in synergy with *Task 8.3 – Build and strengthen the cross-disciplinary ecosystem*. The impact assessment will allow us to refine the BBDC impact strategy, in particular as regards the specific measures to undertake, should the community members not be diverse enough, active enough or numerous enough to ensure the desired impact.

ANNEX A

This Annex lists the main initiatives, projects and communities on Big Data that we have considered for drafting the strategy and charter of the BBDC (see also BYTE deliverable D1.3).

A.1 BDVA¹³

The Big Data Value Association (BDVA) is the private counterpart in the Big Data contractual Public-Private Partnership (cPPP) established in 2014 between the EC and European industry and research organisations.

Mission

The objectives of the association are to boost European Big Data Value research, development and innovation and to foster a positive perception of Big Data Value. It aims at:

- Strengthening competitiveness and ensuring industrial leadership of providers and end users of Big Data Value technology-based systems and services;
- Promoting the widest and best uptake of Big Data Value technologies and services for professional and private use;
- Establishing the excellence of the science base of creation of value from BIG DATA.

The association carries out acts, takes steps and commits to all activities that are deemed appropriate or useful in view of achieving its objectives. This is to include:

- Collaborating with the EC (including to establish a Public-Private Partnership, and to develop and implement a strategic roadmap for research, technological development and innovation in the Big Data Value and other ICT domains);
- Developing strategic goals of European Big Data Value research and innovation and supporting their implementation;
- Improving industrial competitiveness of Europe through innovative Big Data Value technologies, applications, services, solution
- Strengthening networking activities of the European Big Data Value community;
- Promoting European Big Data Value offerings and organisation;
- Reaching out to existing and new users;
- Contributing to policy development, education and technology ramification in the widest possible sense and addressing ethical, legal and societal issues contributing to policy development, education and technology ramification in the widest possible sense and addressing ethical, legal and societal issues.

The main role of the BDVA is defining the Big Data Value Strategic Research and Innovation Agenda (BDV SRIA) and its regular updates, defining and monitoring the metrics of the cPPP and joining the EC in the cPPP partnership board.

Governance structure

The Big Data Value Association AISBL is a fully self-financed non–for-profit organisation under Belgian law. There are 24 founding members from large and SME industry and research.

¹³ <http://bdva.eu/>

The BDVA General Assembly or Board of Directors runs both administrative and strategic leadership.

BDVA is structured in eleven task forces, one administrative and ten thematic; task forces form subgroups. The ten thematic task forces are:

1. Programme
2. Impact
3. Community
4. Communication
5. Legal
6. Technical
7. Application
8. Business
9. Societal
10. Skills

Membership criteria

The basic principles of membership are openness, transparency and inclusiveness. There are several categories of members:

- Interested party, simply on the BDVA mailing list;
- Associate member, participating in the Stakeholder Platform and in relevant tasks forces;
- Full Member, commitment is taken more seriously and is expected and monitored by the BDVA and the EC;
- Board of Director Applicant, helping implement general assembly decisions, lead task forces, etc.

Type of activities

- Yearly summits, held generally 1-2 times a year, alternated with smaller ones;
- Activity group meetings, typically held 8-10 times a year (every 6 weeks), in Brussels, and consisting of a plenary meeting and 6-8 sessions around Task Forces or emerging topics;
- Mailing lists, teleconferences, etc.

How sustained/funded

Membership fees per year:

- Large business Full Member 12700 EUR, Associated Member 1000 EUR;
- Small business Full Member 1900 EUR, Associated Member 250 EUR;
- Academic, research. Full Member 1900 EUR, Associated Member 250 EUR.

A.2 GROUP ON EARTH OBSERVATION (GEO)¹⁴

One of the main advocates of open data in the geospatial sector, the Group on Earth Observation (GEO) is a global effort of voluntary nature, carried out by governments and intergovernmental international organizations with a mandate in Earth observation. GEO promotes information sharing across many different scientific disciplines and applications, by implementing a coordinated and sustained observing framework, a global and flexible network of content providers (at present, it interconnects more than thirty autonomous infrastructures) allowing decision makers to access an extraordinary range of data and information at their desk: the Global Earth Observation System of Systems (GEOSS). GEO explicitly acknowledges the importance of data sharing in achieving GEOSS anticipated societal benefits: "The societal benefits of Earth observations cannot be achieved without data sharing".

BYTE relates to GEO via CNR who is a key partner of the implementation GEOSS, in particular as a member of the Infrastructure Implementation Board, which coordinates GEO Data Sharing activities.

Mission

The Group on Earth Observations (or GEO) coordinates international efforts to build a *Global Earth Observation System of Systems* (GEOSS). It links existing and planned Earth observation systems and supports the development of new ones in cases of perceived gaps in the supply of environment-related information. It aims to construct a global public infrastructure for Earth observations consisting in a flexible and distributed network of systems and content providers.

GEO focuses on facilitating access to earth observation data for nine priority areas: natural and human-induced disasters, environmental sources of health hazards, energy management, climate change and its impacts, freshwater resources, weather forecasting, ecosystem management, sustainable agriculture, and biodiversity conservation.

Governance structure

GEO is governed by a plenary consisting of all members and participating organisations. GEO meets in plenary at least once a year at the level of senior officials and periodically at the ministerial level. Members make decisions at the plenary by consensus. The Programme Board that supports the development and implementation of GEO activities. The GEO Plenary establishes Working Groups.

Communities of Practice may form within or engage with GEO in response to specified needs or shared interests. Communities of Practice may, for example, consider aspects of societal challenges, or managing and developing pooled resources. A Community of Practice (CoP) is a user-led community of stakeholders, from providers to the final beneficiaries of Earth observation data and information, with a common interest in specific aspects of societal benefits to be realised by GEOSS implementation.

¹⁴ The information in this chapter was collected from the GEO website: <https://www.earthobservations.org>; GEO on Wikipedia: https://en.wikipedia.org/wiki/Group_on_Earth_Observations; GEO on Earthzine: <http://earthzine.org/geo-and-geoss-the-group-on-earth-observations-and-the-global-earth-observations-system-of-systems/>; GEOSS Portal: http://www.geoportal.org/web/guest/geo_home_stp

Membership criteria

GEO is a voluntary partnership of governments and international organisations. It provides a framework within which these partners can develop new projects and coordinate their strategies and investments.

As of February 2016, GEO's membership includes 102 governments, including the EC. In addition, 92 intergovernmental, international and regional organisations with a mandate in Earth observation or related issues have been recognised as participating organisations¹⁵. A so-called principal and a principal alternate represent each member and participating organisation.

Type of activities

The types of meetings include workshops, symposia, Executive Committee meetings, plenary sessions, and summits. GEO writes and publishes Ministerial declarations, strategic plans and work programmes, and maintains a list of point of contacts per activity. GEO meetings are frequently co-located with other events.

How sustained/funded

Members make financial contributions to GEO on a voluntary basis.

A.3 RESEARCH DATA ALLIANCE (RDA)¹⁶

The RDA builds the social and technical bridges that enable open sharing of data in the highly fragmented global research data landscape. The RDA vision is researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.

Participation in RDA is open to anyone who agrees to its guiding principles of openness, consensus, balance, harmonisation, with a community driven and non-profit approach. It was started in March 2013 by a core group of interested agencies – the EC, the US National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation. RDA also has a broad, committed membership of nearly 3,000 individuals, from 102 countries.

RDA works through focused Working Groups and Interest Groups, formed of international experts from multiple domain, from academia, industry and government. RDA's methodology is to realise its vision through working groups that address components in 12 to 18-month projects. The RDA Big Data Interest Group (BDIG) has the following charter:

Mission

The ultimate goal of RDA BDIG is to produce a set of recommendation documents to advise diverse research communities with respect to:

- How to select an appropriate Big Data solution for a particular science application with optimal value? and
- What are the best practices in dealing with various data and computing issues associated with such a solution?

¹⁵ <http://www.earthobservations.org/pos.php>

¹⁶ RDA, "About", 2016. <https://rd-alliance.org/about.html>

Governance structure

The RDA has a Council, a Secretariat, a Technical Advisory Board, and an Organisational Advisory Board, drawn from the Organisational Assembly, the body where all members participate (only when the Assembly exceeds 36 members; otherwise the Advisory Board coincides with the Assembly).

The OAB meets monthly by teleconference and in-person at the bi-annual RDA plenaries to provide guidance on RDA business operations, oversight of Interest Groups and Working Groups, and outreach to the larger community. Specifically, the OAB is charged with:

- Providing organisational and operations advice either at the request of the Secretary General or the RDA Council or to express issues raised among the organisational membership with regard to the directions, processes, and mechanisms of RDA to the Secretary General and/or Council.
- Providing input to the Secretary General and the Secretariat on organisational needs; for example, with the goal of encouraging the formation of Interest Groups or Working Groups or other actions that could be taken by RDA to meet organisational needs.
- Interacting with the Technical Advisory Board (TAB) during the review of IG and WG proposals to provide guidance on overlap and synergies with other RDA and community efforts.
- Collaborating with the TAB to review midpoint and final Working Group products and to advise on how implementable proposed product is likely to be. OAB provides guidance on organisational needs and partnership and pilot opportunities, with the goal of encouraging broad community adoption of RDA products.
- Assisting with Plenary programming and other outreach activities.
- Developing of test-beds to support WG pilot projects and distribution of products to the community.
- Encouraging and facilitating the adoption of relevant RDA “products” among organisational members to drive broad adoption.

The BDIG has currently four chairmen.

Membership criteria

BDIG is open to all RDA members to participate. The following participants are especially relevant:

- Domain scientists wishing to utilise Big Data solutions for their research and/or applications,
- Data specialists with experience in data production, curation, analysis, and management, especially involving large volumes and varieties of data,
- Computational scientists or software engineers with special interests in data analysis techniques and algorithm analysis, especially pertaining to Big-Data-relevant technologies and tools,
- Experts, or aspiring experts, of various Big Data technologies and tools,
- Computational infrastructure and architecture experts in fields such as distributed computing, high-performance computing, and database systems,
- Data scientists with a blended interest involving some subsets the activities mentioned above, in particular with share, use, and reuse of open scientific data collections, and
- Managers involved in any combination of the activities mentioned above.

Currently, the BDIG has 138 members.

Type of activities

BDIG uses capabilities provided by the RDA platform to communicate and collaborate effectively to achieve its goals, including:

- Monthly telecons with planned agenda to discuss specific issues;
- Asynchronous collaboration using collaborative editing platforms, wiki and email list servers;
- Semi-annual RDA Plenary meetings to hold sessions for face-to-face interactions amongst members and to inform other RDA members of its on-going activities.

How sustained/funded

Major funding for RDA is provided by the following organisations:

- The Australian Commonwealth Government through the Australian National Data Service
- EC's 7th Framework Programme through the RDA Europe project
- The National Science Foundation

The majority of RDA activity is conducted on a volunteer and self-funded basis by individuals actively contributing and involved in the Council, Technical Advisory Board, Organisational Advisory Board and Working & Interest Groups. Organisation Members also provide support and are crucial to the success of the RDA.

A.4 BIG DATA PUBLIC PRIVATE FORUM (BIG)¹⁷

The Big Data Public-Private Forum project is terminated. It aimed to build the industrial community around big data in Europe and establish an infrastructure to link different parties involved in the development and management of big data. Thus, this project focussed on *technical infrastructure, technology adoption*, and the industrial community.

UIBK, NUIG and SIEMENS were project partners in the BIG project. BIG created BDVA., The industry dialogue and engagement activities initiated in BIG therefore continue in the BDVA (and via a successfully started and already twice conducted European Data Forum).

BYTE will use the link to BDVA to ensure that its efforts are not replicated and to guarantee that the two projects complement one another and address any discrepancies in an efficient manner. This link will also be used to complement stakeholder engagement and the development of the BBDC.

Mission

Big Data Public Private Forum (BIG) was working towards the definition and implementation of a clear strategy that tackles the necessary efforts in terms of research and innovation, while also providing a major boost for *technology adoption* and supporting actions for the successful implementation of the Big Data economy. Building an industrial community around *Big Data* in Europe was the priority of this project, together with setting up the necessary collaboration

¹⁷ www.big-project.eu; see also CORDIS: http://cordis.europa.eu/project/rcn/105709_en.html

and dissemination infrastructure to link technology suppliers, integrators and leading user organisations.

Foundational research technologies were analysed and assessed in BIG and technology and strategy roadmaps created, so that the business and operational communities understand the potential of these technologies and were enabled to implement appropriate strategies and technologies for commercial benefit.

As an example, the *Health Sector Forum* focussed on deployment of suitable IT infrastructure, definition of interoperable standards, and measures that encourage stakeholders in the industry to share existing databases. The latter requires a special consideration of security and privacy concerns. Another example, the *Data Curation Working Group* focussed on identifying artefacts and tools to support the data curation process, where data curation is understood as the active management and appraisal of data over its life-cycle of interest.

Governance structure

BIG was governed by Sector Forums and Technical Groups.

Sector Forums: There were five Sector Forums. They were 1) Finance & Insurance, 2) Health, 3) Manufacturing, Retail, Energy & Transport, 4) Public, and 5) Telco, Media & Entertainment. BIG thus selected a wide spectrum of sectors in their Sector Forums. The selection was initially based on several criteria, including potential in terms of business, timeframe to reach that potential (balancing short and long term approaches) and availability of data in the different application sectors.

The Working Groups were clustered around processes of the data workflow which consists of the following processes: data acquisition, data analysis, data curation, data storage, and data usage.

Membership criteria

BIG was a EU-funded project. It is not clear on which criteria the project partners were selected.

Type of activities

The main activities of BIG have been typical EU project activities, such as producing deliverables in the form of reports, white papers, presentations, scientific publications, interview series, PR material, and press releases.

How sustained/funded

BIG was a EU FP7-ICT project funded from 2012 to 2014. The EU contribution was 2.5 million Euro. The project has ended and is no longer funded.

A.5 UNITED STATES BIG DATA SENIOR STEERING GROUP¹⁸

The Big Data Senior Steering Group (BD SSG) works to facilitate and further the goals of the White House Big Data R&D Initiative. The Big Data Senior Steering Group was formed to identify current big data research and development activities across the Federal Government, offer opportunities for coordination, and identify what the goal of a national initiative in this area would look like. Subsequently, in March 2012, The White House Big Data R&D Initiative

¹⁸ [https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_\(BD_SSG\)](https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_(BD_SSG))

was launched and the BD SSG continues to work in four main areas to facilitate and further the goals of the Initiative.

Governance structure

No information could be found on this point, for the time being.

Membership criteria

No information could be found on this point, for the time being.

Type of activities

No information could be found on this point, for the time being.

How sustained/funded

In March 2012, the Administration announced the National Big Data Research and Development Initiative, which aims to solve some of the Nation's most pressing R&D challenges related to extracting knowledge and insights from large, complex collections of digital data. As part of this initiative, the Administration encouraged multiple stakeholders including federal agencies, private industry, academia, state and local governments, non-profit, and foundations, to develop and participate in Big Data research and innovation projects across the country.

To augment on-going activities and to ignite new Big Data public-private partnerships across the Nation, NSF's Directorate for Computer and Information Science and Engineering (CISE) is seeking to establish a National Network of Big Data Regional Innovation Hubs (BD Hubs). Each BD Hub would be a consortium of members from academia, industry, and/or government. This solicitation aims to establish four Hubs across distinct geographic regions of the United States, including the Northeast, Midwest, South, and West. Each BD Hub should focus on key Big Data challenges and opportunities for its region of service. The BD Hubs should aim to support the breadth of interested local stakeholders within their respective regions, while members of a BD Hub should strive to achieve common Big Data goals that would not be possible for the independent members to achieve alone.

A.6 NIST BIG DATA PUBLIC WORKING GROUP¹⁹

The US National Institute of Standards and Technology (NIST) Big Data Public Working Group (NBD-PWG) is focused on the technical challenges related to big data and is aimed at developing a Big Data Technology Roadmap.

Mission

NIST is leading the development of a Big Data Technology Roadmap. This roadmap will define and prioritise requirements for interoperability, portability, reusability, and extensibility for big data analytic techniques and technology infrastructure in order to support secure and effective adoption of Big Data. To help develop the ideas in the Big Data Technology Roadmap, NIST is creating the Public Working Group for Big Data.

¹⁹ The information in this chapter has been collected from <http://bigdatawg.nist.gov/>

The focus of the NBD-PWG is to form a community of interest from industry, academia, and government, with the goal of developing consensus definitions, taxonomies, secure reference architectures, and technology roadmap.

The aim is to create vendor-neutral, technology and infrastructure agnostic deliverables to enable Big Data stakeholders to pick-and-choose best analytics tools for their processing and visualisation requirements on the most suitable computing platforms and clusters while allowing value-added from Big Data service providers and flow of data between the stakeholders in a cohesive and secure manner.

Governance structure

Several subgroups are established. The subgroups are formed around specific technical challenges: Definitions and Taxonomies, Requirements, Security and Privacy, Reference Architecture, and Technology Roadmap. Each subgroup has a Lead Co-Chair and two Co-Chairs.

Membership criteria

The NBD-PWG is open to everyone. They hope to bring together stakeholder communities across industry, academic, and government sectors representing all of those with interests in Big Data techniques, technologies, and applications.

There are no membership fees.

Type of activities

The NBD-PWG holds weekly subgroup meetings and every three weeks for a subgroups joint meeting by teleconference.

How sustained/funded

This is unclear. There are two Points of Contact employed by NIST. It is unclear whether the Co-Chairs are funded by NIST or voluntary.

A.7 CODATA²⁰

The Committee on Data for Science and Technology (CODATA) is an interdisciplinary Scientific Committee of the International Council for Science (ICSU) that works to improve the quality, reliability, management and accessibility of data in science and technology. Founded in 1966, CODATA promotes international collaboration for open scientific data in three strategic priority areas: promoting open data policies, advancing the frontiers of data science and mobilising capacity for data science and data handling.

Mission

CODATA works to improve the quality, reliability, management and accessibility of data of importance to all fields of science and technology. CODATA is a resource that provides scientists and engineers with access to international data activities for increased awareness, direct cooperation and new knowledge. CODATA was established in 1966 by ICSU to promote and encourage, on a worldwide basis, the compilation, evaluation and dissemination of reliable numerical data of importance to science and technology.

²⁰ www.codata.org

CODATA is concerned with all types of data resulting from experimental measurements, observations and calculations in every field of science and technology, including the physical sciences, biology, geology, astronomy, engineering, environmental science, ecology and others. Particular emphasis is given to data management problems common to different disciplines and to data used outside the field in which they were generated.

In short, the reason for CODATA is to help foster and advance science and technology through developing and sharing knowledge about data and the activities that work with data, with the following objectives:

- The improvement of the quality and accessibility of data, as well as the methods by which data are acquired, managed, analysed and evaluated, with a particular emphasis on developing countries;
- The facilitation of international cooperation among those collecting, organising and using data;
- The promotion of an increased awareness in the scientific and technical community of the importance of these activities;
- The consideration of data access and intellectual property issues.

Governance

Each National, Regional, Union or Co-opted Member designates a delegate to represent it to CODATA. Every two years, a General Assembly of CODATA is held, which consists of delegates of members. Each member has one vote. The workings of CODATA membership and governance are laid out in the CODATA Constitution.

CODATA has elected Officers and an Executive Committee who govern its affairs between General Assembly meetings.

Membership criteria

- National Members: A scientific academy, research council, scientific institution or association of such institutions, which has activities in scientific data evaluation and compilation;
- Regional Members: An international body, representing a group of nations on the basis of some regional or other relationship, which has activities in scientific data evaluation and compilation;
- Union Members: Each scientific union federated in ICSU, which expresses an interest in the scientific data program of CODATA;
- Co-opted Members: An international organisation wholly or predominantly in work covered by the objectives of CODATA.

Type of activities

CODATA uses many mechanisms in its on-going effort to reach its objectives. These are:

- Task Groups;
- Working Groups;
- National Member activities;

- Conferences;
- Workshops;
- Publications;
- Co-operation with other organisations on common interests.

CODATA has four primary activities, all in support of its fundamental aim of fostering worldwide cooperation in scientific and technical data:

- Sponsorship of a Biennial CODATA International Conference on data, which attracts approximately 300 data specialists from around the world;
- Specialist meetings of scientific data experts, which address issues specific to one discipline or topic;
- Publications on data handling, data compilation, surveys of data activities, and conference proceedings;
- Sponsorship of Task Groups, Working Groups, Commissions and other groups addressing specific data issues.

CODATA runs 10 Task Groups:

- Advancing Informatics for Microbiology;
- Anthropometric Data for Special Populations;
- Data at Risk;
- Data Citation Standards and Practices;
- Earth and Space Science Data Interoperability;
- Interoperable Data Publications;
- Linked Open Data for Global Disaster Risk Research;
- Global Roads Data Development;
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries;
- Science and the Management of Physical Objects in the Digital Era.

In addition, CODATA runs 5 Working Groups:

- Nanomaterials;
- CODATA Early Career Data Professionals Working Group;
- CODATA-RDA Interest Group on Legal Interoperability of Research Data;
- CODATA-RDA Working Group on Research Data Science Summer Schools;
- CODATA-RDA Working Group on International Materials Resource Registry.

A.8 BELMONT FORUM²¹

The Belmont Forum is dedicated to research on environmental change, by means of initiatives like the e-Infrastructure and Data Management Collaborative Research Action (CRA).

²¹ The information in this chapter is extracted from the following web sites: <http://www.belmontforum.org/>; <http://www.bfe-inf.org/> (the e-infrastructures strategy); [Open Data survey](#); <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695>

Mission

The vision is to deliver knowledge needed for action to avoid and adapt to detrimental environmental change including extreme hazardous events. The priority foci are coastal vulnerability, freshwater security, ecosystem services, carbon budgets, and the most vulnerable societies.

Belmont Forum has recently adopted *Open Data Principles* for the Environmental Change Research. It recognises the crucial role of open and effective data and information exchange to its mission. The principles were developed from input by more than 120 international researchers, computer and information scientists, legal scholars, and social scientists. The principles state that data derived from global change research shall be discoverable, accessible, understandable, well managed and properly conserved. The data must be discoverable through catalogues and search engines; accessible as open data by default; understandable in a way that allows researchers to use them; manageable and protected from loss for future use in sustainable, trustworthy repositories.

The *e-infrastructures Coordination Office* will also engage the global e-infrastructure community through cooperation with a broad network of organisations such as the Group on Earth Observations (GEO) and the Research Data Alliance (RDA).

Governance structure

Belmont Forum members elect a *Steering Committee* to support and advise Belmont Forum activities between plenary meetings. Two *Co-Chairs* are elected from the Steering Committee to act as spokespersons for the Belmont Forum and to represent the Belmont Forum to external bodies. Currently, one of the Co-Chairs is from the EC.

The *Secretariat* is the administrative arm of the Belmont Forum. The Secretariat implements the decisions taken at Belmont Forum plenary meetings. The Secretariat is composed of a Director and staff. A Theme Program Office within a Member organisation runs individual Collaborative Research Actions.

Membership criteria

Belmont Forum Members are organisations that are legally allowed to mobilise resources from national or international research funds, either public or private, and are engaged in activities that address the Belmont Challenge. Members have (i) declared their commitment to actively contribute to Belmont Forum operations, (ii) participated in a Belmont Forum CRA, and (iii) contributed to the Belmont Forum Secretariat, either through an in-kind contribution or annual fee. New Members are accepted into the Belmont Forum at plenary meetings. Current Belmont Forum Members and their organisational representatives are listed below.

Currently, there are 18 member countries, represented by either their science ministries or science research foundations and research councils. The EC is a member. In addition to the Members, there are Belmont Forum Partners. Partners are organisations that subscribe to the Belmont Challenge but do not fund research and/or do not meet the criteria for membership. They may participate in Belmont Forum Collaborative Research Actions.

Type of activities

The Belmont Forum funds projects. It develops CRAs, many of which include Call for Proposals. Each Belmont Forum Member funds researchers from their own country. The Forum also holds an annual plenary meeting for members, partners, observers, and guests. Additionally, the Thematic Programme Offices organise mid-term and end-of-project meetings for each of the Collaborative Research Actions. The Forum also issues reports and white papers.

How sustained/funded

Members support the Belmont Forum Secretariat through financial and in-kind contributions.

A.9 GLOBAL ALLIANCE FOR GENOMICS & HEALTH (GA4GH)²²

The Global Alliance for Genomics and Health (Global Alliance) was formed to help accelerate the potential of genomic medicine to advance human health. Currently, it brings together 387 leading institutions from 38 countries working in healthcare, research, disease advocacy, life science, and information technology. The partners in the Global Alliance are working together to create a common framework of harmonised approaches to enable the responsible, voluntary, and secure *sharing of genomic and clinical data*.

Mission

The mission of the Global Alliance for Genomics and Health is to accelerate progress in human health by helping to establish a common framework of harmonised approaches to enable effective and responsible sharing of genomic and clinical data, and by catalysing data sharing projects that drive and *demonstrate the value of data sharing*.

Governance structure

The Global Alliance for Genomics and Health is a collaborative, global partnership led by a *Steering Committee* and an *executive team*. The role of the *Steering Committee* is to make high-level decisions about the direction, values, and working products from the Global Alliance. Steering Committee members are established through a nomination process and, as a group, reflect the diversity of perspectives, backgrounds, and geography of the Global Alliance generally. The Steering Committee appoints a *Committee Chair* and an *Executive Director* for the Global Alliance. Global Alliance partners compose a *plenary body* and collaborate to advance its mission. These partners participate in *Working Groups* and other committees. There are currently four Working Groups: Data, Regulatory & Ethics, Security, and Clinical. Professional staff, located at Host Institutions, supports the Steering Committee and Working Groups. A *Secretariat* runs the day-to-day operations.

Membership criteria

The Global Alliance for Genomics and Health is a broad and inclusive organisation that includes a growing number of the world's leading organisations and individuals in healthcare, research, disease and patient advocacy, life science, and information technology.

The criteria for becoming an organisational Member are:

²² The information in this chapter is extracted from the following web sites: <https://genomicsandhealth.org/>

- The mission, operations, and public statements are consistent with the Global Alliance for Genomics and Health Constitution and Member Agreement.
- The organisation is an established entity (e.g., corporation, not-for-profit organisation, partnership).
- The organisation has an open and professional presence (e.g., website).
- The organisation is active in the field of genomic research/medicine or related activity.
- The organisation ownership, governance, funding, and leadership is clearly identified.

Type of activities

Members identify new needs. The needs are prioritised and assigned to one of the Working Groups who form Task Teams and execute projects or Initiatives. The projects deliver tools and solutions that are Work Products put to use in Demonstration Projects. The focus is on data sharing and interoperability.

For example, the *Data Working Group* is the custodian of file formats like CRAM, SAM, BAM, VCF, BCF; they develop a Genomics API; they build a metadata schema. The *Regulatory & Ethics Working Group* has delivered Accountability and Consent Policies and Tools; a Data Sharing Lexicon that facilitates international data sharing by promoting common terms across jurisdictions. The *Clinical Working Group* aims to share clinical data and link with genomic data. One of the initiatives is related to developing *phenotype ontologies*. It aims to bring together existing international efforts to develop and promote standard language and tools for recording patient clinical phenotypes and exploiting phenotype data for diagnostics and translational research.

How sustained/funded

No information could be found on this point, for the time being.

A.10 BIG DATA EUROPE²³

Big Data Europe will undertake the foundational work for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models. The societal challenges and their Big Data focus areas are Health, Food & Agriculture, Energy, Transport, Climate, Social Sciences, and Security.

A.11 EUROPEAN DATA FORUM²⁴

The European Data Forum (EDF) is a meeting place for industry, research, policy makers and community initiatives to discuss the challenges of Big Data and the emerging Data Economy and to develop suitable action plans for addressing these challenges. Of special focus for the EDF are Small and Medium-sized Enterprises (SMEs).

The range of topics discussed at the European Data Forum ranges from novel data-driven business models (e.g. data clearing houses), and technological innovations (e.g. Linked Data Web) to societal aspects (e.g. open governmental data as well as data privacy and security).

²³ <http://www.big-data-europe.eu/about/>

²⁴ <http://www.data-forum.eu/>

A.12 OGC BIG DATA DOMAIN WORKING GROUP²⁵

The Open Geospatial Consortium (OGC) is an international industry consortium of over 515 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC® Standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services and mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.

A.13 NSF EARTHCUBE²⁶

EarthCube, which began in 2011, is a joint initiative between the National Science Foundation (NSF) Directorate for Geosciences and the Division of Advanced Cyberinfrastructure. Achieving EarthCube objectives requires a long-term effort, which NSF anticipates supporting until at least 2022. EarthCube envisions a dynamic, community-driven cyberinfrastructure that supports standards for interoperability, infuses advanced technologies to improve and facilitate interdisciplinary research, and helps educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science.

A.14 RECODE

TRI and CNR-IIA are project partners in Policy Recommendations for Open Access to Research Data in Europe (RECODE). BYTE will use this linkage to bring expertise on open access to large research data sets to the project as well as the BYTE vision, roadmap and community.

A.15 SCALABLE END-USER ACCESS TO BIG DATA (OPTIQUE)²⁷

The Optique project has a diverse approach for dissemination to Industry, academia, government, and to the public. In short we are using three ways for spreading of information; the website, the contact list and the Partner Program. The website describes the Optique project, it's mission, organisation and results, news and events. The contact list has grown to 1024 names, a diverse selection of interested companies, organisations and individuals. The Partner Program has circa 35 names, with dedicated people that are aiming to initiate new projects using Optique technologies. One such project is AHUS from the Norwegian Health sector that will use Optique solutions to do data integration as a preparation for surgery. UiO, DNV GL, and SIEMENS are project partners. The project involves leading partners in industry and academia with strong interest in big data. BYTE will leverage this link for design and implementation of case studies, accessing large data sets, establishing the BBDC and network.

A.16 SIM4RDM

NIIF is project partners in Support Infrastructure Models for Research Data Management (SIM4RDM). BYTE will harvest relations with this project through NIIF's existing relationship. They will also engage with SIM4RDM's findings to assess any externalities that may have been uncovered in their research. This link will also be used to complement stakeholder engagement and the development of the BBDC.

²⁵ <http://external.opengeospatial.org/> for further information

²⁶ <http://earthcube.org/>

²⁷ <http://optique-project.eu/>

A.17 SMARTVORTEX

TUD is an investigator in the FP7 project SmartVortex. The project involves using large-scale collection of engineering process data to support enhanced collaboration, simulation, and design. The big data experiences learned in the project will further inform BYTE.

A.18 PLANET DATA

The Planet Data project [www.planet-data.eu] will also establish a community of researchers to support organisations in exposing their data; however, this project focuses on making data open and integrates those who are producing data, rather than addressing the externalities associated with big data.

UIBK is the coordinator of Planet Data. Planet Data features a very strong research team, and BYTE will leverage on its findings in the area of large scale data management.

A.19 PRELIDA

UIBK is a partner in Preserving linked data (PRELIDA). Linking communities that often face similar challenges and tend to solve them without coordinating with or being aware of each other, it will be a joint task of PRELIDA and BYTE to raise awareness of existing solutions on both sides and to facilitate their uptake, taking into the account social factors. BYTE will also engage with this project to help support the development of the BBDC.

A.20 ROBUST

NUIG is project partners in the Risk and Opportunity management of huge-scale BUSiness communiTy cooperation (ROBUST) project. BYTE will assess the findings of this project to help aid its understanding of externalities in the sharing and analysis of data. As with other projects, BYTE will also use this project to identify relevant stakeholders and to expand the BBDC.

A.21 CREATING KNOWLEDGE OUT OF INTERLINKED DATA (LOD2)

NUIG is also project partners in the LOD2 project. BYTE will assess the projects findings to examine how it contributes to promoting the use of interlinked versions of public Semantic Web data sets, promoting their use in new cross-domain applications by developers across the globe. As with previous projects, this existing relationship will be explored to complement the development of the BBDC.

A.22 EARTHSERVER

BYTE will establish a relationship with the FP7 EarthServer project, via CNR who is part of the EarthServer consortium. EarthServer is establishing open access and ad-hoc analytics on extreme-size Earth Science data, feeding experience and lessons learnt back into standardisation, to advance OGC specifications.

A.23 SEMICOLON

Large-scale project in the Norwegian public sector, exploiting linked open data for publishing and exchanging public information, the experiences from which BYTE will accommodate.

A.24 PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE (PRACE)

NIIF takes part in this project being responsible for working out technical challenges towards storing large amounts of big data, as well as disseminating this knowledge. BYTE will draw on the projects findings as well as using its ties to complement the BBDC.

A.25 CERN LHC COMPUTING GRID

The Worldwide LHC Computing Grid (WLCG) project is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources to store, distribute and analyse the ~30 Petabytes (30 million Gigabytes) of data annually generated by the Large Hadron Collider (LHC) at CERN on the Franco-Swiss border. See <http://wlcg.web.cern.ch/> for further information.

A.26 US BIG DATA RESEARCH AND DEVELOPMENT INITIATIVE

Aiming to make the most of the fast-growing volume of digital data, in March 2012, the Obama Administration announced the “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation’s most pressing challenges. The Big Data Senior Steering Group continues to work with the White House Office of Science and Technology along with several Federal departments and agencies to help further the goals of the Initiative.

See [https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_\(BD_SSG\)#title](https://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_(BD_SSG)#title)

A.27 AUSTRALIAN PUBLIC SERVICE BIG DATA STRATEGY

This Big Data Strategy is intended for Australian Government agency senior executives with responsibility for delivering services and developing policy. It outlines future work by the Government that will assist agencies to make better use of their data assets whilst ensuring that the Government continues to protect the privacy rights of individuals. See http://itlaw.wikia.com/wiki/The_Australian_Public_Service_Big_Data_Strategy for further information.

A.28 UK DATA SERVICE

The UK Data Service is a national data service that provides research access to a range of social and economic data collections including UK census data and government funded surveys as well as qualitative and business data. It was established in October 2012 with funding from the Economic and Social Research Council (ESRC). See https://en.wikipedia.org/wiki/UK_Data_Service for further information.

A.29 UN GLOBAL PULSE INITIATIVE

Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data. Its vision is a future in which big data is harnessed safely and responsibly as a public good. Its mission is to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action.

Global Pulse is working to promote awareness of the opportunities Big Data presents for relief and development, forge public-private data sharing partnerships, generate high-impact analytical tools and approaches through its network of Pulse Labs. See <http://www.unglobalpulse.org/about-new> for further information.

A.30 COPERNICUS²⁸

Copernicus is a European system for monitoring the Earth. Copernicus consists of a complex set of systems which collect data from multiple sources: earth observation satellites and in situ sensors such as ground stations, airborne and seaborne sensors. It processes these data and provides users with reliable and up-to-date information through a set of services related to environmental and security issues.

A.31 THE EUROPEAN BIOINFORMATICS INSTITUTE (EMBL-EBI)²⁹

The European Bioinformatics Institute is part of the European Molecular Biology Laboratory (EMBL-EBI).

Mission and Services

EBI provides a range of *services* including freely available data from life science experiments; it performs basic *research* in computational biology, and it offers an extensive user *training* programme, supporting researchers in academia and industry. Below, we highlight some of their services.

EBI provides a range of services. It maintains the world's most comprehensive range of freely available and up-to-date *molecular databases*. Their services enable data sharing, queries, and analysis of the results in different ways.

The services span DNA & RNA, proteins, systems, ontologies, gene expression, structures, chemical biology, and literature.

As an example, the *DNA services* include integrated annotation on genomes (Ensembl), multiple sequence alignment (Clustal Omega), an open platform for sharing and dissemination of public-domain sequence data (European Nucleotide Archive), a catalog of human genetic variants (1000 Genomes), a catalog of curated data (NHGRI-EBI GWAS), analysis of metagenomics data (EBI Metagenomics), and a Genome-phenome Archive; close to 40 DNA-related services in total.

The ontologies services provide taxonomies and controlled vocabularies. They include a Gene Ontology, an unambiguous, universal language to describe biological phenomena; ontology lookup service; a gene ontology and annotations browser (QuickGO).

The Literature services provide interfaces to Europe PubMed Central; a patent sequence database with access to patent documents; a text mining tool.

Governance structure

EMBL-EBI is a research institute led by a Director and having around 20 research groups. They also have postdocs and a PhD Programme. EBI shares a campus with the Wellcome Trust's Sanger Institute, south of Cambridge, UK.

A.32 TERADATA³⁰

Teradata is a company providing end-to-end solutions and services in data warehousing, big data and analytics, and marketing applications.

²⁸ www.copernicus.eu

²⁹ The information in this chapter is extracted from the following web sites: <http://www.ebi.ac.uk/>

³⁰ <http://www.teradata.com/>

A.33 DECODE GENETICS³¹

DeCODE genetics, Inc. is a *biopharmaceutical* company based in Iceland. It has discovered genetic risk factors for dozens of common diseases including cardiovascular disease and cancer.

Mission and capabilities

The company was founded to identify human genes associated with common diseases using population studies, and apply the knowledge gained to guide the development of candidate drugs. deCODE has gathered genotypic and medical data from more than 160,000 volunteer participants in Iceland, comprising well over half of the adult population. Using Iceland's uniquely comprehensive genealogical records, deCODE has also put together a genealogy database (Íslendingabók) covering the entire present day population and stretching back to the founding of the country more than 1000 years ago.

deCODE attempted to set up an Icelandic Health Sector Database (HSD) containing the medical records and genealogical and genetic data of all Icelanders. This was controversial in terms of their approach to the concepts of privacy and consent. The Íslendingabók is the current output from the more ambitious HSD project.

Governance structure

deCODE genetics is a commercial company and a subsidiary of Amgen, founded in 1996. It is led by a CEO, and the top management also includes a Vice President of Research and a Vice President of Statistics. The company went bankrupt in 2009, and has been purchased several times, last time by Amgen who spun off the database to a new company called NextCODE Health which in turn was acquired by a Chinese company, WuXi PharmaTech.

A.34 WUXINEXTCODE³²

WuXiNextCode is a biopharmaceutical company based on the technology and data developed by deCODE Genetics. It provides clinical-grade next-generation sequencing services (ISO 13485-certified laboratory) and platforms aimed at reducing current bottlenecks in clinical genomics. It furthermore provides big data management and analysis solutions.

The data currently includes 40 million validated variants annotated from 350,000 whole genomes and is 30x larger than the 1000 Genomes database and larger than any public datasets. 3000 patients have been deep sequenced up to 30X. Public domain data have been curated and added to their database. The data include more than 1.5 million indels. Their approach to minimising false-positives and false-negatives is to use raw sequence data, not VCF files, in the variant call analysis process. The analysis pipeline is calibrated with all their data. The aligned raw sequence data are stored as trimmed FASTQ files.

Clinical diagnostics is supported through data mining tools like the Genome Browser and a Sequence Miner, and a proprietary Clinical sequence Analyser (CSA), designed by physicians. The CSA functionality includes variant frequency, mortality impact, and known disease-causing gene data. The information infrastructure comprises a database engine that is a so-

³¹ The information in this chapter is extracted from the following web sites: <http://www.decode.com/> ; https://en.wikipedia.org/wiki/DeCODE_genetics

³² www.nextcode.com

called Genomic Ordered Relations DataBase (GOR rather than RDBMS) promising fast response times despite the large data volumes when comparing large volumes of variants. Data security is in compliance with ISO/IEC 27022. The technologies and data are a result of 16 years of development and more than 1 billion USD invested.