DO→IT

c

D 1.9 interim/short-term strategic guidance document

IMI2 Project ID – DO->IT

Big Data for Better Outcomes, Policy Innovation and Healthcare System Transformation

WP1 – Programme Strategy and Coordination

Lead contributor	9. Swedish Institute for Health Economics (IHE)
	fredrik.moen@ihe.se
Other contributors	20. GlaxoSmithKline (GSK) Myriam Drysdale, Marianne Cunningham and Bart Barefoot
	15. Bayer (BAY) Tino Melzer, Michael Arend
	35. Technology, Methods and Infrastructure for Networked Medical Research (TMF) Irene Schluender
	23. Janssen Pharmaceutica NV (JANSSEN) Michel Van Speybroeck, Anthony Rowe
	28. Pfizer (PFIZER) Jeremy Hayter, Jerry Lanfear
	8. Imperial College London (ICL) Joachim Marti
	1. London School of Economics and Political Science (LSE) Carmen Fenollosa, Sahan Jayawardana
	7. Health Services Management Training Centre of Semmelweis University (SU HSMTC) László Bencze
	3. National Institute for Health and Care Excellence (NICE) Heather Stegenga
	2. Novartis (NVS) Stephan Korte

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116055. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA

Due date	30/06/2017
Delivery date	19/10/2017
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V 1.1	19/10/2017

Document History

Version	Date	Description
V1.1	21/07/2017	First Draft
V1.2	13/09/2017	Second draft
V1.3	19/10/2017	Final version

Contents

Acronyms and abbreviations

AD	Alzheimer's Disease
BD4BO	Big Data for Better Outcomes
CSA	Coordination and Support Action
CTTI	Clinical Trials Transformation Initiative
Do→IT	Big Data for Better Outcomes, Policy Innovation, and Healthcare Systems Transformation
EAB	Ethics Advisory Board
EFPIA	European Federation of Pharmaceutical Industries and Associations
EHR	Electronic Health Record
ETL	Extract-Transform-Load
EXAG	Expert Advisory Group
GDPR	EU's General Data Protection Regulation
HTA	Health Technology Assessment
ICF	Informed Consent Form
ICHOM	International Consortium for Health Outcomes Measurement
IHE	Swedish Institute for Health Economics
IMI	Innovative Medicines Initiative
IP	Intellectual Property
IPD	Individual Patient Data
KPI	Key Performance Indicators
LSE	London School of Economics and Political Sciences
NICE	National Institute for Health and Care Excellence
OECD	Organisation for Economic Co-operation and Development
PHSFF	Policy Health Stakeholders Feedback Forum
ROADS	Real World Outcomes Across the Alzheimer's disease Spectrum
RWD	Real World Data

Acknowledgement

Myriam Drysdale, Marianne Cunningham and Bart Barefoot (GlaxoSmithKline) provided the summary of the results from the online questionnaire that is presented in appendix 3 and referred to in chapter 3.
Fredrik Moen from the Swedish Institute of Health Economics (IHE) contributed to subchapter 4.1 - BD4BO programme objectives. The contributors to chapter 5 were Tino Melzer (Bayer) and Irene Schluender (TMF) from DO→IT work package 4.
The input for subchapter 6.1 - Data management and collection, was extracted from Deliverable 2.1 - Interim guidance document for existing BD4BO projects on approach proposed by the Distributed Data Network working group. This deliverable was conducted under DO→IT work package 2, and the lead contributor was Michel Van Speybroeck (Janssen). The other contributors were Michael Arend (Bayer), Jeremy Hayter (Pfizer), Joachim Marti (Imperial College London), Anthony Rowe (Janssen) and Jerry Lanfear (Pfizer).
The contributors to subchapter 6.3 – Branding and communication, were Carmen Fenollosa (LSE) and László Bencze (Semmelweis University) based on the work done under DO→IT work package 3. Heather Stegenga, Senior Analyst – Science Policy and Research programme, National Institute for Health and Care Excellence (NICE) contributed with The Toolkit for BD4BO projects in subchapter 6.2. Carmen Fenollosa (LSE) provided the input for the subchapters 6.4 - Stakeholder engagement and 6.5 - Project management. Stephan Korte (Novartis) contributed subchapter 6.5 - IP management for BD4BO.

Executive summary

Big Data for Better Outcomes (BD4BO) is a part of the public-private Innovative Medicines Initiative-2 (IMI-2). BD4BO is a research programme aiming to promote the development of value-based and outcomes-focused healthcare systems in Europe through the use of big data. Big Data for better Outcomes, policy Innovation and healthcare system Transformation (DO→IT) is the coordination and support action for the current BD4BO programme. The support action has several components, including the definition of overall programme strategy for BD4BO, to manage and share knowledge from and between the different disease specific projects, act as a central point of collaboration and communication across BD4BO projects and provide minimum data privacy standards for informed consent forms and supporting materials. The objective of this interim strategic guidance document is to give an overview of the current "big data" landscape and discuss the current environment in which BD4BO projects are operating, including identifying gaps, challenges, and success factors. It further includes a discussion on how the current BD4BO-projects are addressing the identified challenges and success factors and provides guidance notes based on the work done by DO→IT work packages.

Issues and success factors

The overview of issues and success factors was based on findings from a scoping review of grey literature published by relevant international organizations and networks; an online questionnaire sent to key stakeholders, e.g. pharmaceutical industry, international organisations, patient groups, academia, HTA/payers, regulatory agencies; and seven structured interviews with participants in key projects on big data in healthcare. Eight overarching themes were identified during this process:

The importance of having a *clear regulatory framework* was highlighted. A lack of clarity regarding legal aspect is a key challenge in many big data projects. Trust is important with respect to privacy concerns and there is a need for finding a balance between facilitating the use of big data and the protection of individuals.

Structures for data control vary based on source of data. There is also variation between countries in how this is dealt with at the national level, with national data sets often lacking. These structures need to be clear since it affects how and to what extent stakeholders interact and collaborate regarding data sharing.

There is an overall lack of governance models for *managing and sharing data* both within and across countries. Poor interoperability between datasets is identified as an important challenge. The key to improving data quality and interoperability is to support the development of standards. Adopting standards will lead to the creation of structured data to facilitate more effective health care outcomes research. A key component in data standardisation is the definition of and use of minimum data sets.

It is important to *align the healthcare process and the research process* to really involve all partners. This will also help in identifying how value from the process can be generated to both the scientists and health care providers.

Involve all stakeholders in an early phase of the project and identify and understand their different interests and perspectives and to establish a common vocabulary to facilitate the communication between different stakeholders. It is important to have a transparent agenda early in the project as to ensure alignment between stakeholders. It is also important to manage expectations both from inside and outside the project.

The key role of patients. Patients and patient groups are central stakeholders in outcomes-based health care projects in general as the end-goal is improved patient outcomes. Patients have two roles as they are both the source of generating data and at the same time the end receivers of the potential benefits of big data solutions and it is necessary to determine the specific role of a patient group in the project. In more general terms, it is important to increase understanding and awareness among

patients and public about their rights to their own data and the social benefits of using health data.

Sustainability. There is a need for sustainable data ecosystems at country and EU-level. Current project based approach at EU-level need to gradually be replaced with self-sustaining structures. To facilitate sustainable systems there is a need for strategic partnerships to align interests of different stakeholders. To ensure sustainability it is of importance to disseminate results and learnings from the projects, e.g. through international meetings and workshops. Public-private partnerships are highlighted as solutions for sustainable funding of big data initiatives.

Big data has a fundamental role in developing *value-based and outcome-focused healthcare systems* but there needs to be a development of methods in using them.

Activities of the BD4BO projects

There are currently three projects operating within the BD4BO programme – HARMONY, ROADMAP and BigData@Heart. The BD4BO projects plan to address the legal and ethical challenges related to managing patient data by analysing the legal landscape on issues pertaining to collecting, using and sharing patient data with the objective of developing guidelines and protocols that incorporate the best practices for sharing, linking and aggregating sensitive medical data. These activities will be supported by external ethics/governance advisory boards. The BD4BO projects will also take several steps to deal with the technical and governance challenges related to data harmonisation. They will build upon previously developed open-source informatics infrastructures like the OMOP common data model to harmonise data from multiple sources. Novel computational and methodological tools will be developed to analyse large-scale observational data and some of these tools will be validated through pilot studies.

The BD4BO projects will also engage with decision making bodies to identify guiding principles for the use of real-world evidence through feedback forums and advisory groups. These platforms will be used to discuss and refine these guiding principles to ensure a broad scope and applicability of the project outputs to support decision-making across regulatory approval, HTA recommendations, reimbursement approval and patient access. The importance of patient engagement is also recognised and the BD4BO projects will rely on focus groups to collect patient attitudes towards data integration and sharing. Finally, the BD4BO projects will take active measures to ensure the sustainability of key project outputs by incorporating sustainability factors into their governance frameworks and adopting best practices from previous projects.

Recommendations from the DO→IT work packages

The four work packages[1] within the DO→IT consortium operate with the objective of supporting and broadening the impact of the individual projects within the BD4BO programme. The objectives of these work packages include providing operational and strategic support, providing methodological guidance for the collection and management of data in real-world settings, giving the BD4BO projects an effective means of communication and collaboration with health care system stakeholders and developing minimum data privacy standards for informed consent forms and providing explanatory information on the use of such documents.

Informed consent forms

The introduction of two new regulations, the Clinical Trials Regulation EU No 536/2014/536 (CTR) and the General Data Protection Regulation 2016/679 (GDPR) is changing the legal framework for data protection in medical research. Many core questions have been left open for interpretation or are still subject to national legislation; thus, pan-European harmonisation may continue to present a

[1] WP1: Programme Strategy and Coordination, WP2: Knowledge Integration and Management, WP3: Communication and Collaboration, WP4: Minimum Data Privacy Standards for IFC:s and Supporting Materials

challenge for innovative medicines research activities.
Within DO->IT Work Package 4 the focus therefore is first the development and continuous alignment of minimum data personal data protection standards for informed consent forms (ICFs) and the generation of supporting materials to provide and to enable the use of patient health data and human biological samples for both R&D and policy development purposes in the future. Activities include developing content structure recommendations for ICFs, the latter including considering new approaches such as charts and images. The work package also deals with interpretative issues of GDPR. In order to monitor ethics issues, two external advisory committees where members will be independent of the DO→IT Work Package 4 consortium has been created.
The GDPR comes in force in force on 25 May 2018, and it is therefore premature to issue recommendations at this point. Aligned positions, recommendations and explanatory documents are being developed during the DO→IT project duration. The final data protection section will be made available to the BD4BO projects as soon as possible.
Data management
A key consideration for a consortium in making secondary use of the data is how they plan to access pseudonymised/anonymised individual patient data (IPD) for the purposes of enabling the most flexible data analysis approaches. An approach that has been implemented with success is to keep access to IPD only at the data controller level and "bring the analysis to the data" instead of pooling data together in a data warehouse to subsequently perform the analysis on the pooled data. This federated model with only data controllers processing IPD and other consortium members processing summarised data has been demonstrated as the generally best approach to facilitate analysis across distributed network of data providers, by balancing the risk of data misuse against the need of technical innovation. This model addresses a lot of the key concerns around sharing of IPD but it might be incapable of addressing certain research questions (e.g. when requiring pooled data sets for statistical analysis). In these situation, a physical pooling of pseudonymised/anonymised IPD is the only viable approach. A central third party (the data coordinator) can assume here responsibility for data harmonization and provide pooled data sets to the community of analysts.
The recommended working practices have implications for the type of Data Management Technology that is selected and deployed. DO→IT provides a primer for the selection of technology to ensure compatibility with existing technology investments of the IMI. Briefly, this implies using OMOP/OHDSI for clinical data using a distributed model, TransSMART for omics-data and RADAR Tools for continuous remote biosensors.
It is recommended that the discussion of data management is driven in partnership with data controllers of background sources. These data controllers will need allocation of sufficient resources.
Branding
DO→IT has developed a graphic identity for the BD4BO programme that is intended to be used along with the distinct graphic identity of the individual BD4BO projects. The graphic identity is accompanied by guidelines on how to properly use the logos, fonts, colours and other visual elements.
DO→IT has also developed a communication plan and an outreach plan which aims to raise the profile of the BD4BO programme and expand the dissemination activities to audiences beyond the BD4BO projects' target groups. The communication plan covers the communication activities to be implemented by DO→IT on behalf of the BD4BO programme and these activities will complement the individual and independent communication activities of the BD4BO projects. The outreach activities as defined by the outreach plan will identify and prioritise health care system stakeholders and will set objectives and milestones for engaging with external stakeholders. The effectiveness of the programme's communication and outreach activities will be assessed through pre-defined key performance indicators.

Stakeholder engagement

Effectively and efficiently engaging with all relevant stakeholders is a critical success factor for a BD4BO project. Three key recommendations are given: 1) undertake a stakeholder mapping exercise, which should identify institutions and existing initiatives that are relevant to the activities of the BD4BO project; (2) assemble an external advisory board comprising of representatives from regulatory bodies, patient groups, health insurance funds, health ministries, academic institutions and health care providers to provide strategic input to the BD4BO project; (3) ensure that resources are used efficiently by reaching out to other BD4BO projects to identify overlaps and opportunities for synergies.

IP management

The IMI Intellectual Property (IP) provisions/rules govern the IP regime of all projects supported by IMI and apply equally to all partners in the projects. The IP provisions are designed to promote the creation and exploitation of knowledge and to reward innovation, while ensuring that the assets and interests of the project partners are respected.

The IMI Programme Office offers impartial advice to all partners during negotiations on IP and ensures that the resulting agreement is in line with the IMI IP provisions. The neutral role adopted by the IMI along with the flexibility of the IP provisions have allowed IMI project partners to share resources and knowledge in unprecedented ways. IP issues are agreed before the launch of the project ensuring that the knowledge developed and shared within the project will be used appropriately.

The base of all IMI2 IP regulations is the legal framework of the HORIZON2020 (H2020) programme which covers all H2020 research and innovation actions. Due to the special nature of IMI2, the H2020 legal framework was adapted in specific areas, one particular area being IP. The IP provisions for all IMI 2 projects are set out in Articles 23 to 31 of the IMI 2 Model Grant Agreement.

Project management

An IMI2 funded project will have a work package dedicated to the overall management of the project. Due to the complexity of the projects and the large number of partners involved, project management resources will be needed at the task, deliverable and work package level in addition to the resources allocated for overall project management. Three key documents will be used throughout the project: 1) the Grant Agreement; (2) the Consortium Agreement; and (3) the Description of Action. It is also recommended that every project develop a project management handbook containing information on project management procedures relating to governance, internal communications, administration and finance. This handbook should be distributed among the project partners. It is recommended that key performance indicators are developed to measure the projects' performance over time. Finally, it is suggested that all project coordinators and project managers devote time to understand the financial implications of running an IMI2 project. IMI has produced webinars and presentations to aid in this task.

1. Introduction

During the last decade, the interest in utilising big data to improve patient outcomes, quality of care and the efficiency of the health care system has increased. To fully leverage the potential of big data in health, it is necessary to consider its use across the full chain of healthcare, from the clinical development process to the end use by providers and patients in collaboration. This requires collaboration across different stakeholder groups: patients, health care providers, payers, authorities (regulators and HTA bodies), academia, and industry.
In a report recently issued by the European Commission, big data in health care is described as:

"Big Data in Health refers to large routinely or automatically collected datasets, which are electronically captured and stored. It is reusable in the sense of multipurpose data and comprises the fusion and connection of existing databases for the purpose of improving health and health system performance. It does not refer to data collected for a specific study." [1]

Big data is often said to be characterised by the 3Vs: Volume (quantity of data), Velocity (the frequency of incoming data) and Variety (different forms of data and from different sources). Sometimes another two Vs are added; Veracity (trust of data) and Value (return on investment).[2]

Big Data for Better Outcomes (BD4BO) is a part of the public-private Innovative Medicines Initiative-2 (IMI-2). BD4BO is a research programme aiming to promote the development of value-based and outcomes-focused healthcare systems in Europe through the use of big data. The programme consists of several topics addressing enablers for the transition towards an assessment based on real world evidence, utilising registries, electronic health records (EHR) but potentially also more novel sources of data (including patient reported data utilising wearables etc.). Topics include disease-specific areas, of which Alzheimer's disease, haematological malignancies and heart diseases are already ongoing.
The Big Data for Better Outcomes, Policy Innovation, and Healthcare Systems Transformation (DO→IT) consortium coordinates the IMI2 Big Data for Better Outcomes (BD4BO) programme. The DO→IT consortium is co-lead by the London School of Economics and Political Science (LSE) and Novartis, and consists of 35 public and private partners. The support action has several components, including the definition of overall programme strategy for BD4BO, to manage and share knowledge from and between the different disease specific projects, act as a central point of collaboration and communication across BD4BO projects and communicate minimum personal data protection standards as applicable for informed consent forms and supporting materials.
The overall objective of this interim strategic guidance document is to give an overview of issues and success factors in using big data and discuss the current environment in which BD4BO projects are operating, including identifying gaps, challenges, and success factors. Chapter 2 describes the data sources used for this document. Chapter 3 presents an overview of issues and success factors in using big data based on the findings from a scoping review of grey literature, an online questionnaire, and structured interviews directed to internal and external key stakeholders related to the BD4BO programme. Chapter 4 includes a discussion on how the current BD4BO projects (HARMONY, ROADMAP, BigData@Heart) are addressing the success factors and challenges for big data projects identified in this document. This discussion will help to put the activities on these BD4BO projects within the context of the current big data and real-world evidence landscape. Chapter 5 presents an overview of the work done by DO→IT work package 4 on the development of minimum personal data protection standards for Informed Consent Forms (ICFs). Finally, chapter 6 includes guidance notes

for the current and future BD4BO projects based on the work done by DO→IT work packages. These guidance notes will address critical areas such as data management and collection, stakeholder management, intellectual property (IP) management and project management to help BD4BO projects run in a well-functioning manner, enabling them to achieve their objectives while ensuring synergies are achieved across BD4BO projects.

2. Data sources for this report

The information in Chapter 3 on issues and success factors in using big data is based on 1) a simplified scoping review of grey literature, 2) an online questionnaire, and 3) semi-structured interviews.

The simplified scoping review of grey literature focused on websites of relevant international organisations and networks, including consulting firms. Only grey literature from the last ten years was included. From these organisations and networks, a selection of documents related to big data in healthcare was identified. The purpose was to get an overview of big data in healthcare, including opportunities and challenges, and strategies. The documents had a general approach and gave a broad outline of the issues. None of the documents went into detail on specific projects, though in some documents projects were briefly presented. Some of the documents were results from meetings or workshops. The overview of the documents resulted in identification of two main themes: (1) legal aspects, and (2) governance and management. Legal aspects could be e.g. personal data protection and intellectual property rights. Governance and management could include e.g. structures for data control, standards of data sharing, collaboration and interoperability (standards to ensure compatible formats). Another theme found in the documents was stakeholder engagements. (see Appendix 1 for a list of the identified documents).

The web-based online questionnaire was developed by IHE together with LSE, NICE and other partners in the project. The purpose was to collect input on success factors and challenges of big data initiatives on healthcare. In April 2017, an e-mail with a link to the online questionnaire was sent to key stakeholders (e.g. pharmaceutical industry, international organisations, patient groups, academia, HTA /payers, regulatory agencies) identified by EFPIA. The questionnaire was sent to 128 respondents with a response rate of 37 percent. Of the 47 responders 37 responded that they or their organisation had participated in projects involving external partners on collecting, analysing and/or utilising the analysis of big data related to healthcare. Ten responders had not participated in this type of projects or did not know/preferred not to answer.

The questionnaire included questions on the stakeholders' experiences from big data projects in healthcare, e.g. objectives of the projects, key success factors, challenges, and patient perspective. Recipients that responded that they had not directly participated in big data projects also received questions on big data, but then on a more general level, see Appendix 2 and 3 for the questions from the online questionnaires and a summary of the results from the questionnaires.

In addition to the scoping review and the online questionnaire 7 telephone interviews were conducted by IHE, LSE and NICE. The interviewees had participated in projects on big data in healthcare that had been identified as relevant based on the criteria: big data, healthcare focus, applied, and outcome focused. The interviews were semi-structured and included questions on success factors, challenges, project sustainability and patient involvement, see Appendix 4 and 5 for the questions from the interviews and a list of the interviewees.

In Chapter 4, information about the planned activities for the current BD4BO projects (HARMONY, ROADMAP, BigData@Heart) was obtained from the Description of Action (DoA) for each project. The text in Chapter 5 on informed consent forms were provided by DO→IT work package 4.

The guidance notes in Chapter 6 were submitted by the DO→IT work packages. The BD4BO programme priorities were based on the IMI call text for the Coordination and Support Action. It was supplemented with the results from a survey conducted among key DO→IT partners on the mission and vision for BD4BO (Task 3.3.1). These partners represent HTA and government bodies, academia, patient organisations and industry. The guidance on data management and collection was developed under a work package 2 deliverable that summarised a set of operational and technological

recommendations based on previous IMI data and knowledge activities (Distributed Data Network working group). The branding and communication and stakeholder management guidelines were based on the work done by work packages 3 and 1. Guidance on intellectual property management was developed by work package 1 under deliverable 1.6. The guidance note on project management was also developed by work package 1 under deliverable 1.7 on BD4BO programme coordination.

3. Overview of issues and success factors in using big data and context of BD4BO projects

Summary
This chapter was based on findings from a scoping review of grey literature published by relevant international organizations and networks; an online questionnaire sent to key stakeholders, e.g. pharmaceutical industry, international organisations, patient groups, academia, HTA/payers, regulatory agencies; and seven structured interviews with participants in key projects on big data in healthcare. Eight overarching themes were identified during this process:
The importance of having a *clear regulatory framework* was highlighted. A lack of clarity regarding legal aspect is a key challenge in many big data projects. Trust is important with respect to privacy concerns and there is a need for finding a balance between facilitating the use of big data and the protection of individuals. *Structures for data control* vary based on source of data. There is also variation between countries in how this dealt with at the national level, with national data sets often lacking. These structures need to be clear since it affects how and to what extent stakeholders interact and collaborate regarding data sharing. There is an overall lack of governance models for *managing and sharing data* both within and across countries. Poor interoperability between datasets is identified as an important challenge. The key to improving data quality and linking is to support the development of standards. Adopting standards will lead to the creation of structured data to facilitate more effective health care outcomes research. A key component in data standardisation is the definition of and use of minimum data sets. It is important to *align the healthcare process and the research process* to really involve all partners. This will also help in identifying how value from the process can be generated to both the scientists and health care providers. *Involve all stakeholders* in an early phase of the project and identify and understand their different interests and perspectives and to establish a common vocabulary to facilitate the communication between different stakeholders. It is important to have a transparent agenda early in the project as to ensure alignment between stakeholders. It is also important to manage expectations both from inside and outside the project. *The key role of patients.* Patients and patient groups are central stakeholders in outcomes-based health care projects in general as the end-goal is improved patient outcomes. Patients have two roles as they are both the source of generating data and at the same time the end receivers of the potential benefits of big data solutions and it is necessary to determine the specific role of a patient group in the project. In more general terms, it is important to increase understanding and awareness among patients and public about their rights to their own data and the social benefits of using health data. *Sustainability*. There is a need for sustainable data ecosystems but this is also defined as an important challenge. To facilitate sustainable systems there is a need for strategic partnerships to align interests of different stakeholders. To ensure sustainability it is of importance to disseminate results and learnings from the projects, e.g. through international meetings and workshops. Public-private partnerships are highlighted as solutions for sustainable funding of big data initiatives
Big data has a fundamental role in developing *value-based and outcome-focused healthcare systems* but there needs to be a development of methods in using them.

Rising healthcare expenditure, increased focus on value-based and outcomes-focused healthcare and evidence-based medicine, together with HTA agencies' request for follow-up of clinical data, all have contributed to increased demand for big data in health. Big data is used in many sectors, but using big data in healthcare adds extra complexity involving legal issues on personal data protection. The number of stakeholders impacted by such issues is also significant. The interest for big data in

health care is ever increasing. A search on Google (2017-06-12) combining the search terms “big data” AND “healthcare” gives 26 900 000 results.

There are many sources for big data in healthcare, for example electronic health records (EHR), disease-specific patient registries, medical monitoring devices, social media as well as the traditional clinical study outcomes. The supply of data and information is increasing and at the same time the technological progress has made it easier to collect and analyse data from multiple sources.

As mentioned in the introduction the definition of big data is broad and can be used in many types of projects with different objectives and including many groups of stakeholders. There is a wide range of projects on big data in healthcare which shows the versatile use and large possibilities with big data. This is reflected in the answers to one of the questions in our online questionnaire asking about the objective(s) of the last project responders had been involved in (multiple responses possible). More than half of the responders in this study (56%) stated that one objective of the last project was to improve academic research. This was followed by improving patient safety and pharmacovigilance (47%), improve quality of care e.g. improve outcomes (44%), improve access to outcomes data (44%) reduce uncertainty of the value or effectiveness of a treatment related to support HTA etc. (34%), improve effectiveness of drug development (43%), facilitate outcome-based payment models (22%) and early detection or prevention of diseases (19%). (Figure 1)

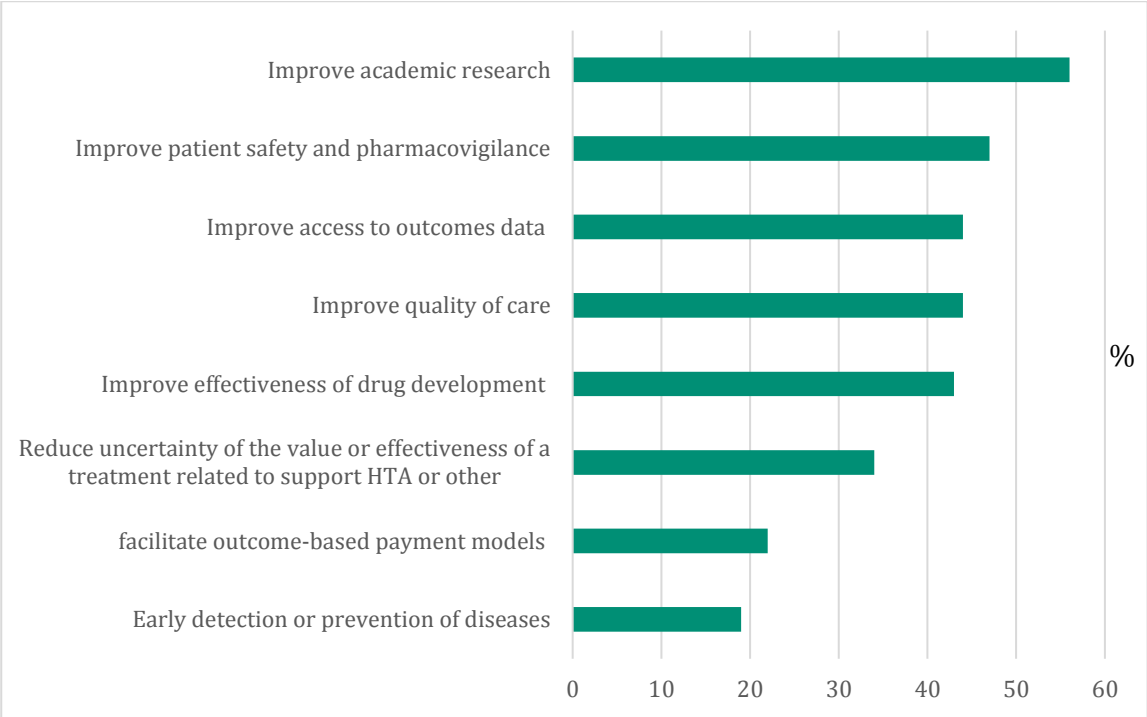


Figure 1. The objectives of the last project on big data in health care as reponded in the on-line questionnaire to key stakeholders (multiple responses were possible)

The broad and varied use of big data in healthcare also means that these projects have to deal with varying needs, demands and problems, thus requiring different approaches and solutions. There is no *one solution* that would suit all big data projects in healthcare, but each project is specific. This is important to keep in mind when reading this document. However, in 2013 the Organisation for Economic Co-operation and Development (OECD) identified 8 key factors that would maximise societal benefits and minimise risks when using health data [3]:

- Health information systems support health care quality and system performance monitoring and improvement for better health care and outcomes
- Public consultation
- Clear means for data subjects to express choices regarding data use
- Permit data sharing subject to data protection safeguards
- Concentrate data within accredited data controllers
- Fair and transparent project approval process
- Follow best practices in data de-identification
- Periodically review governance mechanisms

In the following chapter, issues and success factors in using big data and context of BD4BO is presented, including the experiences related to big data projects as identified in the on-line questionnaire and interviews conducted with people involved in previous big data initiatives (presented in Chapter 2).

3.1 A clear legal and regulatory framework

Big data projects in healthcare involve using data that could be considered sensitive if considered personal data or have an impact on data protection rights otherwise. Thus, having a clear legal and regulatory framework favourable to scientific research and not limited to purpose-specific consent narrowly construed is a prerequisite for projects of this kind. Thus, within each project there need to be partners who have knowledge of these issues and are able to seek relevant legal expertise to help frame the projects in a compliant manner. One person involved in a big data initiative that we interviewed explained the overarching role of a legal framework as:

> *"If there is no legal framework to exchange data between organisations, hospitals for instance and otherwise, a research platform then there is no interoperability, and you cannot exchange the data. So, it doesn't make sense to invest a lot in technology to make it work on a technical level if there is no legal or regulatory framework, because then you can't exchange or collect data."*

However, the existing situation is one of extensive and variable legal fragmentation between and within countries, both regarding personal data protection and scientific research (biomedical research, clinical trials, etc.) e.g. great uncertainty regarding appropriate retention periods, how to carry out secondary processing for research purposes without consent, restrictions on data linkage among different researchers; how to apply the "data minimization principle", and differing legal treatment of electronic and other data.

Several respondents mentioned the lack of clarity around legal aspects of big data projects as a key challenge in big data projects. Trust and buy-in, especially with respect to addressing personal data protection, was also mentioned as one of the most important challenges that need to be addressed in the near future regarding big data in healthcare. A multi-stakeholder networking organisation wrote on security:

> "Providing assurance to hospitals of the robust governance and security measures in place, developing shared understandings between Pharma and academia."

Technology provides a set of opportunities for informing data subjects, but also presents a challenge

in maintaining public trust as the potential uses of data expand. The dilemma for policymakers is to balance the demand for big data with the need to protect individuals. An international association elaborated on privacy and confidentiality:

> *"One of the most important challenges in relation to big data is to convince the general public that privacy of personal data is respected. It is important to find a balance between using big data in healthcare to increase effectiveness and cost-effectiveness of care on the one hand and guaranteeing that confidentiality of personal data is respected."*

Thus, it is clear that as well as law and regulation, governance, consultation, and communication mechanisms will be vital to maintaining public confidence in the use and re-use of personal health data. This is a major challenge, which should involve all stakeholders.

The benefits of harmonising legal and regulatory practices and illustrating key principles relating to the use of personal health data for research purposes is consistently recognized in the reports identified in the grey literature review, as the need for reform. There is also implicit support for the need for a distinct legal and regulatory regime for personal health data (including ethical aspects) that encompasses research and the delivery of care.

The EU's General Data Protection Regulation (hereinafter GDPR) could contribute to reducing fragmentation in some of these areas. At the same time, the GDPR comes at a time when pragmatic solutions permissive of data access have been developed in Member States and there may be concern that the implementation may undermine these arrangements. The GDPR will be applicable from May 2018.

However, a risk that was taken into observation in the literature was that the worry and concerns regarding personal data protection and patient data could lead to too much restriction limiting access and reducing the possibility to use big data. Instead it was recommended in a consultancy report that the collective mind-set would move from "protect" to "share, with protection." [4]

All citizens have the right to personal data protection and thus countries may need support to evaluate and ensure adequate protection regarding personal data and protection of health data. In the literature it was also proposed that a risk classification of data would be developed so that data that is more sensitive as regards personal data protection easily would be identified and addressed [1, 5] .

There is a recommendation of the OECD council on health data governance [6]. Among the key health data governance recommendations two referred to legal aspects on personal data protection:

- The processing of personal data for public health, research and statistical purposes are permitted, subject to conditions specified in the legislative framework for data protection.
- Best practices in data de-identification are applied to protect patient personal data.

3.2 Structures for data control

The control of data is related to legal and regulatory aspects. The structures for control of real-world data (RWD), which can be seen as part of big data, vary based on the source of data [7]. For example, general practice records are normally managed by GPs/national health systems, EHRs and point of care records are owned by hospitals. Similarly, pharmacy records will be embedded in pharmacies while health insurer databases will be managed either privately or publicly depending on the existing health insurance system in the country. The data ownership of patient registries may vary, with some owned by patients (e.g. International Niemann-Pick Disease Registry) while others may be owned by the registries themselves (e.g. Pharmachild) [8].

Based on the online questionnaires, the data source that was most often used in big data projects were EHR (80 %), followed by disease specific patient registries (57 %) and administrative systems/claims data (47%). Stakeholders also used other sources of data for their projects including pharmacy data, published non-randomised studies and clinical trial data. Social media and wearables were in this questionnaire only mentioned in less than 5% of the responses.

In a OECD study on health data governance from 2015 countries that have concentrated the collection and processing of national health datasets were identified [9]. These countries have distinct advantages in further developing these data for health outcomes research and health care system performance monitoring by undertaking data linkage studies and improving secure access to data for external researchers. Among the EU5 countries (Germany, France, Italy, Spain, UK), the UK's devolved healthcare system in Scotland has the highest concentration of national datasets, where 78% of national datasets are concentrated within one data controller. Beyond the EU5 countries, the US and Sweden also have highly concentrated datasets, with 73% and 70% of national datasets concentrated within one data controller respectively [9].

The concentration of the data linkage process also varies between countries. For example, Scotland and Wales regularly link over ten key national datasets within one organisation [9]. Three organisations are involved in regularly linking national datasets in Sweden and the UK, two in Italy, and over four organisations in the US and Spain. The key advantage for countries having a concentrated control of national datasets is that data linkage projects can be conducted without the need to get into negotiations or data sharing agreements with multiple stakeholders. As a result, these countries are more likely to have regular programmes that monitor health care quality and performance based on data following the entire pathway of care [9]. In addition, these countries are more likely to have the resources to develop efficient data processing, enabling them to provide high quality services to external data users. While concentration may pose risks to personal data protection, implementing a good governance mechanism can help manage these risks. Finally, some of these countries only enable the access to these public data sets to the researchers who work for the public sector but not to researchers of the private sector.

A few countries are making efforts to concentrate data processing and access through the implementation of an accreditation process. In England, NHS Digital is recognised in law as a safe haven for the collection, processing, analysis and dissemination of data about the health and social care system in England. NHS Digital are involved in developing standards for electronic health records (EHR) and running a data linkage service to provide more information about a patient's health and social care journey [9]. In Scotland, there are five health-related accredited safe havens that provide a safe environment to process health related data for research purposes. In addition, these safe havens provide data linkage services and access to anonymised health microdata to approved researchers. The Secure Anonymised Information Linkage project (SAIL) in Wales is another safe haven created to enable routinely collected data from health care provision and other government service delivery to be anonymised and linked by a trusted National Health Service organisation.

The interviews that were conducted support the notion that the structures for control of RWD vary based on the source of data. Interviewees stated that the structures need to be clear and that it affects how and to what extent stakeholders can collaborate when it comes to sharing data. For many of the projects included the data is owned by the sponsor (e.g. RCT data from pharmaceutical companies) or provider (e.g. health care regions) of the data. The common way to proceed is that the project is using de-identified data and is acting as the data processor/technical facilitator conducting and providing data analyses. One person involved in previous big data initiatives explained the situation as follows:

"The ownership of collected data is regulated most of the time in a monopolistic way, which means that it's very difficult to reuse data that was previously collected in the context of a project or a patient registry…"

One solution could be that laws, contracts, informed consents, convections and collaboration agreements are adapted to the new reality where data is needed and the reuse of previously collected data, by a third party may be foreseen.

3.3 Ensure data interoperability and sharing

As presented above, the structure for data control and management of the databases within a country will affect and/or define the nature of data sharing and collaboration. As a result, the generation, analysis and sustainability of RWD will rely on the multiple stakeholders who own, curate and manage the data and the extent to which they interact and collaborate with each other [7].

The key to improving data quality and linking is to support the development of standards. Current technical standards and best practices should be adopted and core definitions must be established for data governance and usability, evidence and value, and analytical protocols [1]. For example, developing standards for the interoperability of clinical data (e.g. EHRs) will incentivise the pooling of data for system level research. Adopting standards will lead to the creation of structured data to facilitate more effective health care outcomes research.

However, the value of high quality linked data will only be realised if a culture and environment of data sharing is created [10]. To achieve this, a structured set of incentives needs to be created to incentivise partnerships between the public and private sectors.

A recent OECD survey on data governance revealed that most countries already had the technical pre-requisites needed to effectively utilise and share RWD but the key issues revolved around governance, legislation and culture [3]. The World Health Organisation eHealth survey conducted in 2015 [11] found that most countries surveyed, including Italy, Spain, Sweden, the UK and the US have no policies governing the use of big data in the health sector. In addition, the survey only identified Italy and Spain out of the countries under consideration in this review, as countries with a national EHR system. However, most countries, including Italy, Spain, Sweden, the UK and the US, did have policies or legislation governing data sharing between health professionals in other health services within the country. Notably, only Italy and Spain had policies in place governing data sharing between health professionals in the health services of other countries.

This points to an overall lack of governance models for managing and sharing data, both within and across countries [1]. Currently, there is only limited standardisation of data access rules in the EU [7]. As a result, there is no clear pathway for researchers to access data and the type of data available is therefore highly dependent on the kind of interactions that are developed between stakeholders that own and process the data and the underlying governance structure shaping these interactions. This has resulted in activities related to health such as hospital services, research, clinical activities, education and administrative services operating in silos with each of them having their own organisational data and information infrastructure [12].

For example, the absence of standardised platforms for sharing information is a critical problem limiting international data sharing and linking in dementia research [10]. Dementia patients showing signs and symptoms are usually diagnosed based on crude end points, and this makes the lack of a shared platform for metrics of biomarkers and dementia more problematic [10]. Therefore, for diseases like this, which has multiple causes, developing a governance framework to internationally collaborate and share data is crucial.

Inconsistent data quality is another factor that has to be addressed through effective governance [7].

The usability of data is limited by inconsistent data entry, coding errors and discontinuities in data collection. Therefore, harmonised strategies to develop data quality standards are crucial to address these limitations. One example of an initiative working towards improving data quality is the international working group that has been formed to improve the quality procedures for the European Cystic Fibrosis Society Patient Registry [7]. The group is responsible for developing a standardised procedure document for data quality checks.

Another key component needed to support the standardisation and sharing of core sets of information is the definition and use of minimum data sets [5]. The minimum dataset is a defined set of data that could be shared among clinicians treating the same patients. England maintains a minimum dataset known as the "national summary record" for sharing patient information electronically to support unscheduled emergency care. One fourth of patients have a summary record and transaction standards have been agreed upon for this dataset. Scotland has one minimum dataset to support emergency care that has been agreed upon for all patients. Spain has specified a minimum dataset and it has been incorporated within the EHRs for an estimated 27% of patients. France, Germany and the US have not specified minimum datasets.

The information from the grey literature review that most countries already have the technical pre-requisites needed to effectively utilise and share RWD but that often the key issues concern governance and data sharing is also reflected in the questionnaires and the interviews. Respondents in the questionnaires cited both data access and linkage as success factors for their projects. Respondents from academia however highlighted the importance of technical aspects more than others, especially regarding data linkage and data access, whereas respondents from the pharmaceutical industry, HTAs and multi stakeholders networking organisations emphasized the importance of good collaboration and communication.

Results from the grey literature review shows that poor interoperability limits the combination of multimodal data which prevents health systems from taking advantage of existing synergies between data [13]. This issue is particularly relevant when there are disparities between national and international health data standards and this variation within and across countries affect the interoperability of datasets. For example, the inconsistent coding of dementia in routinely collected health care data and the inconsistent diagnosis or case-finding protocols limit the interoperability and linking of datasets related to dementia [10].

Also, one of the key challenges encountered in the projects as mentioned in the on-line questionnaires was a poor interoperability between the data sets (e.g. the coding and terminology between data sets were not compatible and difficult to link). In the interviews, the importance of interoperability was also highlighted.

Improving the interoperability of datasets is challenging because the stakeholders involved in developing, inputting and using datasets are primarily motivated by different incentives [3]. For example, EHR systems are often designed to have proprietary features that make it difficult to integrate outside data sources like patient generated health data (PGHD) into the EHR systems [13]. Industry standards organisations like Health Level Seven (HL7) are developing standard methods for capturing, recording and making PGHD interoperable.

To ensure the interoperability of datasets, countries must adopt common IT standards and terminologies that take account of the needs of the various stakeholders (patients, providers, EHR vendors, applications developers, etc.) [13]. However, the new technology must be flexible enough to adapt to constantly evolving health care recommendations, standards and policies.

3.4 Align healthcare and research

A prerequisite for a big data project in healthcare is to have access to data. As was described earlier, big data projects often use EHR, which are often owned by hospitals. Thus, hospitals and healthcare providers are essential partners in big data projects. The hospitals have their way of working, which must be taken into account when designing a big data project, which was described by an interviewee:

> *"From a research point of view, you have your own way of working, your scientific workflow. Sometimes it is clinical trial workflow in the mindset of the researcher. But if you want to capture data from the real world you have to align the care processes, which is not a clinical trial process. So, you have to find some alignment between these two different processes."*

In another interview, it was stressed that it is important to reflect on the reasons for a hospital to participate in big data projects and how participation will impact the hospitals. One interviewee described that a less successful attitude would be:

> *'We like data and we can do cool things with it for the greater good', as the greater good doesn't impact them in the next five revenue cycles."*

Instead, it was thought to be important to try to identify the value in participation and motivate the hospitals to participate and get involved in the project.

3.5 Early co-operation between stakeholders in big data projects

Due to the complexity and regulations in the healthcare sector, the use of big data requires knowledge and the involvement of many functions and disciplines: legal and regulatory, policy, business, healthcare process, IT infrastructure etc.

Of the 47 responders to the online questionnaire the largest group of responders represented the pharmaceutical industry (28%), followed by other, such as international organisations (21%), academia (13%), user/patient group (13%), HTA organisation (11%), regulatory agency (9%), healthcare provider, e.g. hospitals (4%) and health care payer organisations (2%). This broad involvement is also illustrated in the responses to a question on the types of stakeholder groups included in big data projects. Most of the projects that respondents were involved in were collaborative projects, where academia (25 responders or 78%) and healthcare providers (72%) were the most common groups included across projects. This is followed by pharmaceutical industry (50%), regulatory agencies (44%), healthcare payer organization (31%), user/patient groups (31%) and HTA organisations (25%).

However, getting the right stakeholders involved can be a challenge. Moreover, stakeholders often have different interests and perspectives. Coordination and culture between different organisations was often emphasised as a challenge and mentioned as a general theme to consider in future big data projects. According to interviews the technological challenges were less problematic and could be resolved in a fairly straightforward way. This was stressed in several questionnaires and interviews.

> "It is not about technology but it is about people."

> *"There is an abundance of technology and platforms and partners to work with, but the real issue is human and nothing to do with data."*

> *"It is always more challenging than anticipated, with the human factor being more of a challenge than data/IT."*

"Coordination and culture between different organisations is a challenge."

At the same time, commonly cited success factors included multi-stakeholder collaboration and cooperation including trust and buy-in. One responder to the online questionnaire wrote on success factors:

"Collaborators with a common vision, trust, institutional support, funding, research projects that address a range of issues, involvement of patients."

Co-operation provides an opportunity to do work in a multi-stakeholder way, which leads to a lot of learning between the different stakeholders. Involving experts from a very wide range of stakeholders was mentioned as an important success factor for big data projects, as this provides a lot of different skillsets within the consortium and a lot of different possibilities and enabling factors for insightful and detailed research. Many problems in healthcare require multidisciplinary solutions. Different disciplines such could have tackled a problem from a different lens and a different angle.

"Nobody has solved the problem, so we have been taking the approach of trying to bring these different disciplines together and having statisticians who work with epidemiologists, who work with medical informatics, [and this] has brought new and different solutions and ways to think about the problem than we would have got if we consider this an epidemiology problem or treated it only as a statistics problem."

Another benefit from involving all stakeholders from the start that was mentioned was that it increased the chance that the results from the big data project will be of use and implemented:

"So, if you don't have everyone involved in creating a solution, it's very unlikely that everyone will implement that solution."

The importance of involving a wide variety of stakeholders was also found in the scoping review of grey literature. The majority of the included reports drew upon input from a wide range of stakeholders including academia, authorities, regulators, industry, insurers and other payers, health professionals and patients. A central recommendation from the grey literature being made is the need for collaboration across disciplines and the importance of identifying all relevant stakeholders. Several rationales for this are given: it allows for identifying and addressing needs across stakeholder groups [1, 7]; ensure that the design and process for outcomes measurement is optimized [14] and to improve collaboration and develop networks across different sectors [15]. In order to optimise the process when designing infrastructure and data capture models the International Consortium for Health Outcomes Measurement (ICHOM)[2] pointed out the inclusion of clinical expertise, business and IT, while in the report from NICE the participants emphasized the involvement of regulators, HTA agencies and payers to clarify the requirements of data in regulatory and reimbursement processes. One way to increase knowledge that was exemplified was to start measuring outcomes in small scale using existing databases and then scale up initiatives [14, 15] .

3.5.1 Involve all stakeholders early and set the framework

It is important to get different stakeholders on board in big data projects, such as people responsible for legal and regulatory framework, people at IT and technical level as well as hospital managers and

[2] The International Consortium for Health Outcomes Measurement (ICHOM) is a non-profit organization with the purpose to transform health care systems worldwide by measuring and reporting patient outcomes in a standardized way.

researchers. Several interviewees mentioned that patience is important and the need to have enough time to really talk to each other/having a dialogue and sharing perspectives in order to get to a common understanding of the project. As a part of this it is noted important to engage and value all the stakeholders equally. As stressed by a patient group in the online questionnaire:

> *"Big data need big cooperation."*

> *"Don't wait so long to interact with stakeholders, misperception on how information might be useful for other stakeholders stands in the way of efficient use of existing data sources."*

An important factor as to get everybody onboard which was raised during interviews is to establish a common vocabulary to get different people to talk the same language. Involving stakeholders early gives a possibility to set a common framework, which is of paramount importance for the future success and implementation of the results. A key factor to ensuring success is engaging and valuing all the stakeholders equally as this will help create recommendations and tools that make a difference.

> *"If a stakeholder is not included in creating a solution, it's hard to make that solution come through. If you involve all the stakeholders, you also start understanding incentives for maintaining non-value-added activities and your solutions are mindful of these."*

Another interviewee emphasised that an important success factor was to have a common agreement or consensus amongst the wide constituency of multi stakeholders as to the value behind the research question or the general aim of the project. The challenges of co-operation between stakeholders varied in different parts of big data projects.

> *"We didn't experience problems in, for example, interpreting the data or interpreting the results. Because if you follow scientific procedure then it's very easy to make the interpretation: it's objective…. So the struggle was more on agreeing on the data research plan and agreeing beforehand on what we want to do and why. "*

> *"But being able to have a very clear ask that speaks in the language of both the legal team and the IT team, when you want access to the data, will go a long way."*

A problem that was mentioned, was that in statistical analysis different disciplines, such as epidemiology, statistics, health economics and informatics, would solve the exact same problem with the exact same mathematics, but they would have labelled it with a different word. It was stressed that this process took time and required patience. Also, it is not about producing documents but it is about meeting, interacting and discussing.

> *"But really establishing a common ground is not about producing some document that nobody reads, it is about getting everybody to buy in, support, acknowledge and agree to a common approach of moving forward. And that takes real effort to align on, not even opinions on how to do science, aligning on the language we are going to use to describe the same idea."*

3.5.2 A transparent agenda to build trust

Another factor, linked to collaboration, which was stated in the interviews as an important challenge, is the issue of trust among stakeholders. It is important to have a transparent agenda to ensure alignment between partners already in the early stage of the project. The rationale behind the project needs to be clear and questions such as: why collecting data, what is done with the data, how is data governed, what are the risks, or what happens if data is misinterpreted, need to be answered. Here, optimal project management is a critical factor as it could be a challenge to develop a management

approach that is effective across different stakeholder cultures (e.g., academia, industry).

> *"Create a safe environment where people are comfortable enough to speak their mind."*

3.5.3 Balancing expectations

Further, it is important to manage expectations to align the expectations of different parties interested in the data. Based on the interviews some projects try to deal with these issues through organising workshops, conferences and meeting and some projects also found that they should have focused more on such activities. These expectations came from both within the projects and from outside the projects.

One example of how to handle expectations from within the project would be to be clear in terms of commitment, e.g. having a "description of work" where the deliverables are specified, as that helps stakeholders to know what their specific tasks were beforehand.

Expectations from outside of the project could involve how to communicate information on what data and results to expect and when to expect it. It was mentioned that many people are very enthusiastic about health data and its way of working.

> *"You have to start being more careful with what you promise… You probably have to have some timing and be somewhat more strategic...and realise that we're not quite finished yet."*

> *"It is very interesting to see that when the government announces that health data start with a new project, the day after you receive questions from the industry for instance to reuse the data that is not already collected, because it was just announced that we will start with this data collection."*

3.6 Patients at the heart of Big Data

Patients and patient groups are central stakeholders in health care projects in general as the end-goal is improved patient outcomes. Benefits can take several forms, such as patient empowerment, putting patients at the centre of managing their health, to bring care closer to them, and to connect them with the right information, services and institutions at the right time. Big data can help patients get the right treatment and that it will help them matching with the right provider to achieve the best outcome [3, 4].

> *"It was worth all the effort and it was worth every moment because it was an eye opener for us as well internally to see how different the patient voice might be and its brought up a lot of discussion internally on how we view quality of life data and how we would want to incorporate patient preferences within our processes in a more systematic manner."*

One view that was expressed was that there can sometimes be an uncertainty whether the representative from a patient organisation represents a large body of members or if they present their own personal views. Therefore, it is also important to determine what role the patient group has within the big data project.

> "Ultimately, the most important consideration is having a clear idea about what the patient input is for. This will ultimately dictate how the patient perspective is incorporated."

It was mentioned that an umbrella organisation could be a good start as they can lead to the specific indication-base or disease-based patient network that later is needed within a specific project This will however depend on the type of big project on health data that is considered and what specific

needs there are for the project.

Patients are not only the end receivers of the potential benefits of big data solutions but are also crucial in proactively driving data capture and providing informative patient perspective [8]. Based on the grey literature review, one of the identified barriers that restricts the use of real-world data in general is personal data protection concerns related to patients [7]. Involving patients in a discussion about the benefits and risk of data use was explicitly mentioned as one of the factors that would support access to data with the purpose of improving cure and care [10]. Indeed, the consent or full anonymization are not the sole possibilities to address the personal data protection concerns of patients, as it has been recognized in the GDPR. In particular, there is need for national authorities to put in place sound data governance frameworks that can ensure the integrity of patients and citizens while enabling the secondary use of data to improve health systems and support research and development. A need to educate patients and citizens on their rights and the societal benefits of reuse of health data has been identified [9]. One way of doing this is to report back to patients on the findings of studies that they have been involved in for them to understand the value of sharing their data [15].

Similar indications were found in the online questionnaire that user/patient groups were included in less than a third of the projects. This was also noted in the question on whether the project incorporated a patient perspective, in which 40 percent answered that it did include a patient perspective. When asked how the patient perspective was included a few answered that it was addressed through involvement by patient organisations, whilst two responded that is was addressed through the ethical approval process. Thus, from these answers it seems as if patients were not actively involved in the design of the project.

> *"One final recommendation is to take them seriously, because patient organisations, I think, are very susceptible to scepticism from other stakeholders and rightly so. Because their aims may be different to the aims of other, let's say, more common stakeholders within consortia, such as the IMI consortia. It maybe that the other stakeholders lose interest quickly in trying to incorporate patient organisations in their work. I think that culture needs to change a bit."*

Inclusion of the patient perspective can be solved in different ways, but an important question is to reflect on which role the patient group will have. Some solutions mentioned in the online questionnaires and the interviews were to appoint patient representatives to the Steering Committee, the Executive Committee, and to establish a patient-leadership council.

Clinical Trials Transformation Initiative (CTTI)[3] has created guidelines and recommendations on patient engagement in clinical trials. The CTTI project Patient Groups in Clinical Trials works to provide guidance and structure for patient groups and sponsors to be able to better engage with each other across the research and development continuum. Another example of incorporating the patient perspective is the Health e-Research Centre Citizens' Jury [4]. In this model, members of the public, i.e. the jury, are paid for their time to participate for a discussion related to a particular topic with experts. Following this discussion, questions will be put forward to the jury to debate and make a decision on.

Related to this is the group sometimes referred to as the general public, which could include previous,

[3] Clinical Trials Transformation Initiative (CTTI) is a public-private partnership, co-founded by Duke University and the FDA in 2007, with a current mission to develop and drive adoption of practices that will increase the quality and efficiency of clinical trials.

[4] The Health eResearch Centre is a world-leading digital health network that unites and provides data-intensive research and education across Northern England and beyond. https://www.herc.ac.uk/get-involved/citizens-jury/

current and future patients. Patients and the general public have two roles as they are both the source of generating data and at the same time may have the power to refuse its use by not giving consent [8]. Increased awareness among the general public and patient groups could increase the understanding of and the willingness to give consent to sharing data, thus facilitating and ensuring the availability of data. Rising awareness could include both informing the public/patients on their rights to their own data, consent, and on the societal benefits from re-using health data [9, 10]. Increased awareness could be reached by communication strategies, campaigns and presenting real examples[1, 7].

3.7 Sustainability of results and systems

Sustainable provision of project results could be interpreted as continued access to data or infrastructure after the end of the project, so that it is not dependent on project funding. The scoping review found that little attention was paid to best practices for the critical phase when moving from the project phase into something more permanent. However, making the outputs available for others and disseminating the results and learnings is important in order to ensure sustainability. One of the interviewees responded to the question on how to ensure a sustainable provision of project results so that others can benefit from earlier work as follows:

> *"…. creating awareness and a good outward communication policy, presence within international meetings with other multiple stakeholders"*

Based on the scoping review several actors pointed to the need for sustainable data ecosystems [2, 13]. To facilitate this there is a need for strategic partnerships to align interests of stakeholders: payers, providers, professionals and patients [7]. One way to operationalize this may be through the establishment of centres of excellence [10]. In this context, data access and use should be a common good [16]. There is a need to promote the openness of government data as well-as non-proprietary private data [1]. One option to create incentives for providers to contribute data is tying collection and sharing of data to reimbursement [17].

Public-private and public-public partnerships have been highlighted as one feasible approach to funding big data initiatives [1, 10]. One of the recommendations from the European Commission report is to use initial multi-source financing to spread the risk and financial burden. The public has a clear role in financing big data system development and audit, shared data handling infrastructure, continuous partnerships between various stakeholders, enhanced infrastructure for public health department to receive and process EHR data, and coherent strategies that incorporate disparate pilot programs [1]. Another suggestion was that collection and sharing of data would be a condition for reimbursement for pharmaceuticals and medical devices and that that a bigger part of the reimbursement would be contingent on how well it contributes to better outcomes.

Based on the online questionnaire the issue of sustainable provision of project results such as continued access to outcomes data or infrastructure after the end of the project (not dependent on the project funding) seem to be delicate. The majority of the projects did not have or were not aware of a solution to sustainable provision of project results. For the projects that had a solution for sustainability there was no single general solution given. Some of the respondents meant that *by developing new projects* the sustainable provision of results was ensured. Another respondent meant that there was *no particular action needed,* while some meant that their solution was to *develop a continued funding consortium for the project platform and a hospital network of those willing to continue participating.*

The need for sustainability of results and systems is also emphasized in the interviews and is defined

as an important challenge. Factors that came up during the interviews were the importance of utilising what has been developed in other projects and linking this to other initiatives as to not reinvent the wheel. To have a more open-science and open-source collaborative approach was mentioned as to be beneficial for the sustainability of a project. Another factor mentioned for ensuring sustainability of outputs was to utilize an infrastructure that is possible to manage centrally.

It is important to ensure awareness of the research also to a broader public for example at international meetings with other stakeholders and to ensure that input also comes from outside stakeholders. However, to get people using the work is meant to be more of a challenge as it requires strategic thinking into the possible future trends of research. One respondent expressed the difficulty as:

> *"It's a question of identifying current needs in the healthcare community, identifying possible future needs, monitoring, so gearing your deliverables towards that, and then re-reflecting and re-adapting that process throughout your project, part of which also must stress on inputs from outside stakeholders."*

It summarizes in that the work process in a project is adaptive but also responsive.

> *"I think that's the whole philosophy of science to climb on the shoulders of the others."*

Recommendation on the above areas was presented in the report published by the European Commission and validated by experts, stakeholders and private companies [1]. The recommendations covered in this area were:

- Develop and implement a communication strategy to raise awareness of the benefits of big data in healthcare
- Reinforce education and training to increase human capital for maximising the potential of big data in health
- Secure that funding and financial resources lead to purposeful investments that guarantee cost-effectiveness and sustainability

In projects that have generated recommendations it is important to spread these and implement them. This can only be done if they are available and easy accessible, i.e. that the recommendations are packaged correctly. If poorly packaged, extra work will be created for people to adopt and implement recommendations, and fewer will use and implement them.

> *"If you have recommendations, the packaging of that matters a lot: Why does it matter? Why do people have to pay attention to the recommendations? What will it do? Who is going to be the people that are going to implement this change? Who are the change agents? - these aspects are vital to drive adoption."*

3.8 Big data to develop value-based and outcomes-focused healthcare systems

Value-based and outcomes focused healthcare is becoming more common and can also be related to the increased focus on patient-centred healthcare. Value-based healthcare can be described as healthcare that improves patient outcomes in a cost-effective way. This requires measurement of healthcare outcomes.

In the online questionnaire, it was found that the pharmaceutical manufacturers were nearly unanimous in their view that big data are critical to the foundation for outcome-based healthcare systems, the research and development of precision medicines, and the rational use of medicines. Other stakeholders expressed broad agreement with this view, although several respondents

(including some HTA bodies) offered somewhat more tempered endorsement, acknowledging the potential impact of big data in these areas while observing that the possibilities are yet to be realized and it is too soon to tell how impactful these data actually will be. A respondent from a medical society suggested that more work is needed to demonstrate that big data can generate accurate evidence before a proper assessment of utility and impact can be made. Both HTA bodies and regulatory agency respondents suggested that they view evidence derived from big data as a supplement to, and not a substitute for, RCT data – a point consistent with statements made by such stakeholders in other fora. One payer stated that the data are indeed critical but need to be developed and managed internally – implying that evidence derived from the payer's own data will be preferred over evidence generated from other data sources. Finally, it was noted that the data and its applications should benefit the patient, i.e., patient-centred and capturing patient-relevant outcomes and not just used at the health system level.

> *"Big data has a fundamental role in developing value-based and outcomes-focused healthcare systems and should be considered at all stages of developing and evaluating health interventions and services."*

> *"[W]e are obliged to find ways to better collect & analyse data to better understand the impact of interventions on health outcomes and facilitate decision-making."*

The same opinions were presented in the interviews, including the large opportunities that big data can have and but also stressing that there needs to be a development of methods and evaluation of them in order to establish a best practice.

> *"It is a very complex matter, and certainly very difficult to have a simple forecast of success in the big data projects for the near future. So, some projects will certainly be successful, but others sometimes lack societal approval in order to have authorisation to start the project."*

> *"Observational data can fill a critical gap in providing evidence to inform medical decisions. That it is never going to be filled by clinical trials or other sources that are out there. And yet, the challenge that we have is that when we generate evidence from observational data, we currently don't know whether or not we could trust it and whether or not what we are generating is actually signal instead of noise. So, we feel as if the need is for us to first establish some confidence in what we can learn from observational data, and what we cannot learn, to understand how reliable that evidence is going to be when we generate it."*

4. Activities of the current BD4BO projects

Summary
This chapter discusses the current activities of the BD4BO projects and how they relate to the challenges and success factors identified in Chapter 3. There are currently three projects operating within the BD4BO programme – HARMONY, ROADMAP and BigData@Heart. These BD4BO projects will be operating to advance the broader BD4BO programme objective of accelerating the transition towards value-based and outcomes-focused healthcare systems in Europe. More specifically, the BD4BO projects have dedicated work packages to address legal and ethical issues, technical and governance challenges, stakeholder collaboration and project sustainability.
The BD4BO projects plan to address the legal and ethical challenges related to managing patient data by analysing the legal landscape on issues pertaining to collecting, using and sharing patient data with the objective of developing guidelines and protocols that incorporate the best practices for sharing, linking and aggregating sensitive medical data. These activities will be supported by external ethics/governance advisory boards. The BD4BO projects will also take several steps to deal with the technical and governance challenges related to data harmonisation. They will build upon previously developed open-source informatics infrastructures like the OMOP common data model to harmonise data from multiple sources. Novel computational and methodological tools will be developed to analyse large-scale observational data and some of these tools will be validated through pilot studies.
The BD4BO projects will also engage with decision making bodies to identify guiding principles for the use of real-world evidence through feedback forums and advisory groups. These platforms will be used to discuss and refine these guiding principles to ensure the generalisability and applicability of the project outputs to support decision-making in terms of regulatory approval, HTA recommendations, reimbursement approval and patient access. The importance of patient engagement is also recognised and the BD4BO projects will rely on focus groups to collect patient attitudes towards data integration and sharing. Finally, the BD4BO projects will take active measures to ensure the sustainability of key project outputs by incorporating sustainability factors into their governance frameworks and adopting best practices from previous projects.

This chapter will discuss how the current BD4BO projects HARMONY, ROADMAP and BigData@Heart are planning to address the identified success factors and challenges related to big data initiatives. The planned activities for the projects were extracted from the DOAs of each project.

4.1 BD4BO programme objective

There is an increased wealth of digital information generated in European health care systems. The increased supply of health-related data from multiple sources ("big data") has the potential to transform European health care systems in terms of clinical operations, research and development. The "Big Data for Better Outcomes" (BD4BO) programme creates research platforms and big data networks for various disease areas with the objective of *accelerating the transition towards value-based and outcome-focused healthcare systems in Europe, maximising the health outcomes for patients and society by improved clinical pathways and optimised health care expenditure*.

Four key enablers have been identified in the text for IMI 2 Call 7, topic 2 to support a data driven transition:

- agree on definition of outcome metrics
- develop protocols, processes and tools to access high quality data
- improve methodologies and analytics to drive improvements
- implement digital and other solutions that increase patient engagement.

To reach the programme's objective, it seeks to exploit the opportunities offered by big and deep data sources in a few representative disease areas, to put together a methodological framework to guide big data research and to invite a wide range of stakeholders to discuss the future of health systems shaped by big data (insert reference to Eurohealth article). So far, three disease-specific BD4BO projects have been launched within IMI2: Alzheimer's disease, haematological malignancies and heart diseases.

The first project, "Real world outcomes across the Alzheimer's disease spectrum for better care: multi-modal data access platform" (ROADMAP) will align outcomes and methods to develop an approach within existing data systems to efficiently enable initiation, maintenance, and evaluation of the right treatment to the right patient at the right time in health care systems [18]. The project also seeks to work with national authorities to investigate the needs for future prospective data collection to ensure that the second phase of the project is relevant to access and reimbursement decision making.

The second project, "Healthcare Alliance for Resourceful Medicine Offensive against Neoplasms in Hematology" (HARMONY) aims to increase access to relevant outcomes data which can help improve the care of patients with certain haematological malignancies [19]. It plans to collect, integrate and analyse anonymous patient data from a number of high quality sources to help define clinical endpoints and select a core set of outcomes for these diseases that are recognised by all key stakeholders.

The third project "Big Data for Better Hearts" (BigData@Heart) is expected to improve understanding of the risks of serious outcomes in these patients compared to the general population [20]. The existing knowledge should be further improved on how these patients are treated in the real world and what affects outcomes with more efficient surveillance of safety and effectiveness in real world settings.

A fourth call for proposals on Prostate cancer is currently underway. All disease-specific projects attempt to leverage the potential of big data, but with a somewhat different emphasis on the various enablers.

As a complement to the disease-specific project, a call to set up a European Health Data Network (EHDN) has recently been launched (July 2017). The purpose of the EHDN project is to map a large number of European data sets to the OMOP CDM, building on tools and methodologies developed in other IMI projects (including EMIF). The EHDN will be able to service current and future disease-specific BD4BO projects with access to data.

To maximise the impact of the BD4BO programme, a coordination and support action DO→IT has been set up to support the disease specific projects on common issues, to identify areas of unmet data need, to provide overall strategic direction and communications for the programme, to integrate and synthesise the key learnings and act as a centre of excellence to exchange knowledge.

The potential applications for the output that is generated from the BD4BO programme are many and it has the opportunity to drive value-based outcomes-focused healthcare in Europe.

Big data will be an important component of a “learning” health care system that generates insights about the effectiveness of treatments in clinical practice. Some possible applications of big data to support this process may include obtaining estimates about treatment effectiveness from routine practice data, identification of relevant patient groups to support the research and development process to develop targeted treatments and generating evidence to support regulatory and reimbursement decision-making.

The BD4BO programme is expected to contribute to building up knowledge about big data in health care and develop tools that will be helpful for current and future big data initiatives. Specifically, the programme has the potential to have an impact on three areas that are crucial to realise the potential of big data to drive value-based outcomes focused health care systems: a) improving data access b) addressing data privacy issues and c) creating strong stakeholder alignment.

The BD4BO programme is expected to build a foundation for improving data access in Europe. By identifying multi-jurisdiction data sources and creating catalogues of harmonised data, the BD4BO programme can provide a basis for the more efficient use of data in Europe. The data management tools that are developed will give researchers access to linked high-quality data sources, which would enable health organisations to understand and improve patient management and the delivery of services.

The BD4BO programme could also raise awareness and provide solutions to issues pertaining to data privacy. The programme can facilitate clear agreement at the EU level and provide guidance to data controllers for conducting effective research without compromising patient privacy. There are dedicated work packages and working groups within the BD4BO projects working to address legal considerations related to data anonymization and sharing. This gives BD4BO the opportunity to develop a common understanding between researchers and patients on sharing patient-level data (whether anonymised or pseudonymised) while ensuring compliance with data protection regulations. The BD4BO programme also has the opportunity to provide guidance on the long-term implications of the European General Data Protection Regulation (GDPR). In addition, by working towards developing standardised informed consent templates, BD4BO can facilitate a uniform solution to consent management, improving access to patient-level data.
The BD4BO programme is expected to create strong partnerships between public and private sector organisations and develop a common understanding among stakeholders on the role of big data in outcomes focused health care systems, and also increase the awareness of the benefits of outcomes-

based approaches for patients and healthcare systems. This would help ensure stakeholder alignment, which is necessary to agree on priority outcomes, design data management tools, create unified solutions to consent management and ensure the uptake of these outputs at the health care system level.

4.2 Legal and ethical issues

The extensive legal fragmentation within and across Europe and the resulting lack of clarity on data privacy issues is a central challenge facing big data initiatives. In particular, complex laws regarding data retention periods, restrictions on the secondary processing of data without consent and restrictions on linking data were key legal issues that were identified in the current big data landscape. All three current BD4BO projects have work packages dedicated to address legal, ethical and data privacy issues.

HARMONY will provide a framework to deal with the legal, ethical and governance issues related to managing patient-level data. The legal landscape of the countries participating in HARMONY will be analysed to produce guidelines, focusing on issues concerning data protection and research and fundamental rights issues concerning the secondary use of medical data[21]. The analysis will be conducted by reviewing publications on the secondary use of data and seeking advice from other big data related projects such as BBMRI, ELIXIR, EMIF and EHR4CR. Further, relevant cases from the European Court of Human Rights will be analysed and input from the patient community as well as participating physicians and industry partners will be incorporated into the guidelines on the secondary use of data. The guidelines produced on the secondary use of data will be tested in a proof-of-principle study, where a protocol for the sharing of data will be developed in collaboration with other work packages. For secondary use of prospective data collection, HARMONY will develop a template of an informed consent form in collaboration with the local Ethics Committees of the participating countries. HARMONY will also conduct a literature review to produce guidelines on how to deal with the ethical and legal implications of incidental findings (clinically relevant findings outside the scope of the original clinical trial paradigm). To seek advice on on-going legal issues throughout the life time of the project, HARMONY will establish a team of legal and ethical experts.

ROADMAP also has a dedicated work package to deal with the ethical, legal and social implications of using patient-level data[22]. ROADMAP recognises the ethical challenges and opportunities that arise when patient-level data is pooled and re-purposed for research. In particular, ROADMAP expects to encounter ethical and legal challenges in two primary areas - re-purposing or reusing existing multi-modal data and priority setting in terms of relevant measures and outcomes for each type of modelling. In recognition of these challenges, ROADMAP will develop an ELSI (ethical, legal and social implications) framework for a real-world evidence platform for Alzheimer's disease(AD). To develop the ELSI framework, ROADMAP will conduct a systematic review on best practices for sharing, linking and aggregating sensitive medical data across all data types – trials, cohorts, patient-reported outcomes, national registries and EHR databases. The identified ELSI framework requirement will be drafted to guide the data integration efforts within ROADMAP and to provide recommendations for an EU-wide RWE platform for AD. The framework will be designed to supplement the local ethics and governance procedures managed by local ethics bodies, with the emphasis being on providing recommendations for issues arising in a multi-national RWE platform.

BigData@Heart will also address legal and ethical issues through a dedicated work package. Specifically, they will focus on developing an appropriate governance framework for the data infrastructure of the project [23]. In order to do this, the project will review the literature on models for data collection and sharing. In addition, the conditions under which the data to be used in the project

was originally collected will be assessed – e.g., informed consent, data protection measures used such as anonymization or pseudonymisation. The lessons learned from this exercise will be used to map and design the decision-making procedures that need to be considered when integrating data. These decision-making procedures will provide the foundation for BigData@Heart to develop a governance framework for data-intensive health research.

In recognition of the implementation of the GDPR, HARMONY has designated a task to review the relevant data protection laws after the application of the GDPR, taking account of the differences across the participating countries. This review will give HARMONY an extensive and up-to-date picture of the legal landscape concerning data sharing following the implementation of the GDPR, which would enable the project to adapt data acquisition and transfer protocols accordingly. In a similar vein, ROADMAP has also recognised the likely implications of the implementation of the GDPR and will address the related ethics and privacy challenges in a systematic review. BigData@Heart will also consider the implications of the GDPR when designing the decision-making procedures for data integration and sharing.

A key recommendation made in the OECD governance principles related to big data is the use of best practices in data de-identification to protect patient data privacy. HARMONY will produce guidelines on the anonymization and pseudonymization of patient data to minimise the risk of re-identification of subjects. Particular attention will be given to handling genetic data that carry a higher risk of re-identification. In order to produce the guidelines, HARMONY will seek advice from other similar projects such as BBMRI as well as the patient community. HARMONY will use the limited dataset from a proof-of-principle study. The existing informed consent forms (ICONs) for pilot study data will be checked by a specialised legal counsel. Based on this feedback, a proper, scalable and defendable method balancing the benefits of the project for the health of the community and the protection of personal privacy will be proposed.

In addition to the activities of all three BD4BO projects with regards to questions towards the legal framework a Think Tank was set up, consisting also of members from the DO→IT Work Package 4, working on aligned positions among all relevant data protection topics to foster the harmonisation process within the framework and to collate constraints/requirements of the BD4BO programme. On a long term view this Think Tank may also provide opinions to policy law makers and other decision makers to support the big data landscape and follow up on the principle on stakeholder feedback.

In order to monitor ethics issues, HARMONY has nominated an international external Ethics Advisory Board (EAB), where the members will be independent of the HARMONY consortium. The EAB will issue periodic ethics reports to the IMI and will be part of HARMONY's reporting documents. ROADMAP will also establish an Ethics Advisory Board (EAB) to provide input on the data integration tools and methodologies developed within the project. The EAB will also provide guidance on the ELSI framework developed by ROADMAP. BigData@Heart will establish a governance advisory committee to evaluate the work done within the project with regard to ethical and legal issues. The advisory committee will consist of members representing all stakeholders within the consortium. There will also be an independent Ethics Advisor in this committee.

4.3 Technical and governance challenges

The legal issues related to data privacy represent one challenge to data sharing. Another impediment to data sharing relates to the technical and governance challenges, particularly with regard to data linking and interoperability. Common issues that were identified in this study within the current big data landscape were the lack of data standards, limited standardisation of data access rules as well as problems with data quality -inconsistent data entry, coding errors and discontinuities in data

collection.

The HARMONY Big Data platform will support data harmonisation by developing a common data model. The common data model will provide a unified data model to acquire data from multiple data sources which will make the data comparable – e.g., comparisons between treatments. The common data model will also support the future data collection from existing data sources. In work package 3, HARMONY will focus on three aspects of developing the Big Data platform – the data intake and collection process, findings process and the harmonisation process. Issues related to data quality will be dealt with at the intake and collection stage through ETL (Extract-Transform- Load) software. During the findings process, relationships between variables will be created and the Common Data Model will standardise the format and content of the observational data. At the harmonisation stage, advanced technologies such as rule engines and intelligent algorithms will be applied to harmonise and categorise the findings and ultimately create an organised system with observational methods. HARMONY will also implement a data governance framework with consideration to data privacy, rights of access and security features. For example, there will be two levels of firewall to prevent unauthorised access to the Platform. Hardware and software will be designed and configured to define two separate security zones to control or limit access to the servers and network resources. In addition, HARMONY aims to reach consensus on common minimum data sets for hematologic malignancies.

ROADMAP will also have a work package dedicated to addressing challenges related to data harmonisation. The work package will focus on identifying, mapping and integrating patient-level data for AD in order to address the challenges related to extracting, harmonising and analysing data from RWE sources relevant to AD. The potential challenges include technical issues related to accessing datasets with varying modality, structure and provenance, as well as governance and data pooling issues. Solutions for these challenges will be explored through pilot data integration and harmonisation exercises in several countries. The pilot exercises will utilize the existing informatics infrastructures developed by the European Medical Information Framework (EMIF) and Dementia Platform UK (DPUK). These pilot studies will be used to inform and validate tools and methodologies for real-world data handing and to allow stakeholders to assess the suitability of existing types of data for outcome measurement, disease progression modelling and pharmaco-economic analyses.

BigData@Heart will work towards consolidating biomedical knowledge, phenotyping algorithms and developing informatics tools for exploring and analysing data from multiple sources. The project is planning to achieve this through an open-access informatics platform that will bridge knowledge across multiple sources under a diverse set of formats with varying degrees of accessibility. BigData@Heart will do this by creating a high-level biomedical knowledge data resource, developing novel computational and informatics approaches for enabling discovery and replication in large-scale clinical information systems such as electronic health records, and facilitating smart remote monitoring data collection. Similar to HARMONY and ROADMAP, BigData@Heart will also rely on the OMOP Common Data Model and other tools developed by the OHDSI community to integrate various data sources. In doing so, the BD4BO projects are planning to utilise available resources in an efficient manner rather than opting to reinvent the wheel.

4.4 Stakeholder collaboration and inclusion of patients

Ensuring effective stakeholder collaboration was an important challenge unanimously identified by previous and current big data initiatives. Including a wide-variety of stakeholders while ensuring the right mix was highlighted as a critical success factor for big data initiatives. Stakeholder engagement is also crucial to ensure the adoption and acceptance of evidence derived from big data, which in turn,

would lead to value-based and outcomes focused healthcare systems. The online survey found that HTA bodies view evidence generated from big data as supplementary and that more work needs to be done to demonstrate the accuracy of the evidence generated from big data. In addition, including patients and patient groups as key stakeholders was emphasised as an important success factor. However, patient involvement has been generally viewed as an area that requires further engagement.

HARMONY aims to identify and address evidence gaps that lead to delays in decision making by regulatory agencies, HTA bodies and payers, potentially delaying patient access to innovative treatments for hematologic malignancies. HARMONY will engage with key stakeholders through the Policy Health Stakeholders Feedback Forum (PHSFF). The PHSFF will collect viewpoints from patient organisations, haematologists/clinicians, the pharmaceutical industry, regulators and HTA bodies. The objective of these engagements is to discuss barriers, gaps and needs to get a better understanding of the current situation on market access decision criteria and evidence requirements for hematologic malignancy treatments. These engagements would potentially lead to reaching consensus on innovative solutions based on understanding enablers and barriers, opportunities and constraints, and feedback on resolving conflicting opinions.

The stakeholder clusters involved in the PHSFF are intended to be [21]

- Patient organisations
- Haematologists/ clinicians
- Medicines authorities
- HTA bodies
- Payers
- Pharmaceutical industry

HARMONY recognises that the key to successful and effective stakeholder integration is the generation of relevant evidence to address uncertainties and support decision-making in terms of regulatory approval, HTA recommendations, reimbursement approval and patient access. With this view in mind, HARMONY will develop an Access Evidence Framework focused on outcome acceptability for centralised regulatory approval as well as national/regional HTA to support the decision-making process for market access of innovative medicines for hematologic malignancies. Further, HARMONY will develop a Clinical Value Framework to assess the magnitude of clinical benefit/added therapeutic value offered by a new medicine in relation to standard treatment offered for hematologic malignancies. The PHSFF will serve as a platform to discuss and refine these frameworks once all stakeholder views are taken on-board. This would ensure the generalisability and applicability of these frameworks to the decision-making bodies within the EU member states.

Similarly, ROADMAP will take steps to engage with regulators, HTA agencies and payers. The objective of such activities will be to enable an understanding of potential regulatory and HTA implications on the use of real-world data to develop evidence on the value of interventions in AD. In order to do this, ROADMAP will establish a Regulatory-HTA-Payer Expert Advisory Group (EXAG) to inform and evaluate the applicability of work package plans and outputs in the regulatory and HTA/payer context. The EXAG will consist of representatives from regulatory agencies, HTA bodies and payer organisations from different European countries representing the various payer archetypes. The EXAG will also have one patient representative. The EXAG will engage through virtual and face-to-face meetings. The EXAG in partnership with EFPIA and CBG-MEB will collect and collate guiding principles for the use of real-world data based on the development of three Use Cases of RWE in a regulatory context, HTA context and payer context. The identified guiding principles will be used to develop recommendations for the use of RWE in the regulatory and HTA context. These actions taken to engage with HTA bodies and regulators will increase the likelihood of project outputs being adopted

at a health system policy level, which is a key step in transforming heath care systems into value-based outcomes-focused health systems.
The stakeholder engagement efforts through the PHSFF will be supported by the HARMONY communication activities, the latter being co-ordinated by the European Haematology Association (EHA). Outputs and policies that are agreed upon will be disseminated via papers/policy briefings, congresses, white papers, conferences and round tables. ROADMAP will adopt a similar communications strategy. In particular, ROADMAP will partner with Alzheimer Europe to disseminate and communicate project results to patients and the wider dementia community through 36 national Alzheimer associations. BigData@Heart will also adopt a similar communication and dissemination strategy, utilising the European Society of Cardiology (representing over 80, 000 cardiology professional) and the European Heart Network (representing 30 national heart foundations and patient organisations).
Recognising the importance of including the patient perspective, ROADMAP will establish focus groups to collect patient attitudes towards the proposed RWE platform and data integration. The focus groups will include members of the European Working Group with Dementia and representatives of national Alzheimer Europe associations. In addition, empirical studies on AD patients and carer attitudes towards RWE, data sharing and re-purposing will be analysed through a systematic review, which will be fed into the ELSI framework. BigData@Heart will also take steps to include the patient perspective by constructing a model for public and patient involvement in the governance of big data health research. Once the requirements for meaningful patient and public involvement and alignment have been defined, the model will be integrated into the governance framework.

4.5 Project sustainability

Ensuring the sustainability of project outputs is a key success factor for big data initiatives. Having a clear plan to ensure continued access to the project data and infrastructure and the sustainable provision of project results was seen as critical factors to ensure adoption of outputs at a health system policy level. BigData@Heart will be taking several steps to ensure the sustainability of project outputs. Sustainability will be one factor considered in developing the governance framework, where potential future developments in areas such as the implementation of learning healthcare systems will be addressed. In addition, the project plans to use best practices from other IMI projects, for example, adopting the IMI code of practice for secondary use of medical data for scientific research projects, to ensure that the governance platform remains harmonised with other IMI initiatives. BigData@Heart will also attempt to commercialise some outputs and deliverables to build on the investment made by the EU and other partners, and ensure that the science developed within the project can be continued to meet the needs of the future.
HARMONY will also take several measures to ensure project sustainability. This sustainability plan will include – i) the use of the HARMONY database for future analyses by the European health systems, scientists, regulators, and industry; ii) potentially including additional diseases, such as chronic myelogenous leukemia (CML) and myelofibrosis or Hodgkin lymphoma; iii) transfer of the HARMONY “Proof-of-Principle” findings along with the HARMONY tools and algorithms to other non-hematologic malignancies and rare diseases, as well as to non-EU countries.

5. Development of minimum data privacy standards for informed consent forms

Summary
The introduction of two new regulations (The Clinical Trials Regulation EU No 536/2014/536 (CTR) and the General Data Protection Regulation 2016/679 (GDPR) is shifting the legal framework for data protection in medical research. Many core questions have been left open for interpretation or are still subject to national legislation; thus, pan-European harmonisation may continue to present a challenge for innovative medicines research activities.
Within DO→IT work package 4 the focus therefore is first the development and continuous alignment of minimum data personal data protection standards for informed consent forms (ICFs) and the generation of supporting materials to provide and to enable the use of patient health data and human biological samples for both R&D and policy development purposes in the future. Activities include developing content structure recommendations for ICFs, the latter including considering new approaches such as charts and images. The work package also deals with interpretative issues of GDPR. In order to monitor ethics issues, two external advisory committees where members will be independent of the DO→IT work package 4 consortium has been created.
The GDPR comes in force in force in 25 May 2018, and it is therefore premature to issue recommendations at this point. Aligned positions, recommendations and explanatory documents are being developed during the DO→IT project duration. The final data protection section will be made available to the BD4BO projects as soon as possible.

While there is a legitimate request from researchers, health care systems and other stakeholders to

collect, process and link data within the field of innovative medicines, including clinical trials, the undergoing legal framework is shifting due to two new regulations: The Clinical Trials Regulation EU No 2014/536 (CTR) and the General Data Protection Regulation 2016/679 (GDPR). While it is true that these regulations aim to harmonize the European data protection framework, many core questions have been left open for interpretation or are still subject to national legislation; thus, pan-European harmonisation may continue to present a challenge for innovative medicines research activities.

The BD4BO programme offers the great opportunity to establish a dialogue between patients, researchers and public authorities and review boards in order to develop aligned solutions for balancing privacy and research needs on the one hand, and trust in personal data protection personal data protection on the other hand among multiple stakeholders. The GDPR rules on the *ex lege* compatibility test of the scientific research, the scientific research exemption as well as the need to have broader consents in this context (where consent is required) should be applied.

Within DO→IT work package 4 the focus therefore is first the development and continuous alignment of minimum data personal data protection standards for ICFs and the generation of supporting materials to provide and to enable the use of patient health data and human biological samples for both R&D and policy development purposes in the future.

In the first phase of this project an industry partners´ draft for informed consents and structure recommendations was shared with public partners. This draft had been generated by comparing EFPIA partner´s practice regarding their ICF templates and thorough discussions about legal, research and ethics prerequisites for using patient health data and biological samples. Comments where received by public partners and a list of critical issues, both for industry and public partners, was set up. Overall, it was common understanding, that the EFPIA draft addresses essential points in connection with the minimum standards of personal data protection for ICFs and recommended structure. There is an agreement that this EFPIA draft can serve as good starting point for Work Package 4 for further development and alignment during the project duration.

Also new structure approaches for ICFs like charts, images were identified as helpful and can provide an easier and clearer understanding for participants. New technologies to change the informed consent process in a more practical way in the future are also taken into account. The use of quizzes and videos in computer-supported informed consent procedures can also enhance the understanding of participants about what they are consenting to. Internet-connected devices, such as smart phones and tablets, can also change the research process in a more fundamental way, including recruitment of participants through the internet. On the other hand, it can be stated, that new technologies must integrate with mandatory legal requirements on informed consents like procedural requirements. Work Package 4 has identified the need to develop an approach to tackle this topic.

As a main aspect of DO→IT work package 4, the GDPR which applies from May 25, 2018, sets the scene for data protection topics in this work package 4 and for all BD4BO projects taking the relevant projects time into account. It can be said that the emphasis of DO→IT work package 4's efforts lies on interpretative issues regarding the GDPR and therefore is focused on prospective data collections. The following none- exhaustive list shows some of the main issues to be addressed:

- What is personal data and what could be considered "anonymous" data for the purposes of scientific research?
- What are the general requirements for informed consent under the GDPR for scientific research according to recital 33? In particular, how to address new research methods and scientific research purposes?
- Which role must both the scientific research exemption (Art. 9.2j)) to the consent and the compatibility test (Art. 5.1.b) of the GDPR play in the ICF?

- How is genetic data regulated in the Member States?
- In case of consent withdrawal: Which data must be deleted if any or which processing activity should be discontinued? Could patient level data? Article be anonymized? Art. 7 para. 3 GDPR
- What has to be considered in an ICF with regards to participants rights, Article 12 seq GDPR in conjunction with the CTR?

Further, it has to be recognized that data usage scenarios may often rely on pseudonymized codified data sets rather than anonymized by operation of the clinical laws. So also technical-legal questions on pseudonymization approaches are tackled by the ongoing consultations of work package 4 in view of art. 89 of the GDPR and the scientific research exemption.

In order to monitor ethics issues, work package 4 has nominated two external advisory committees where members will be independent of the DO→IT work package 4 consortium. In particular these committees will be set up as an operational and a regulatory committee (DPEC-O and DPEC-R), the first representing research and industry stakeholders, the second consisting of data protection authorities and Ethics Committees throughout the EU. This gives the great opportunity for alignment as mentioned above. The once established dialogue is planned to be maintained even after the project's end and can be used for future consultations, since constant dialogue seems to be a key factor for the success of using big data for healthcare related research and quality assessment

Conclusion

The GDPR becomes in force in 25 May 2018, current and future research must take into account that there are key issues and questions with regards to data protection under the GDPR and research which are subject to ongoing alignment among the different stakeholders. Aligned positions, recommendations and explanatory documents are being developed during the DO→IT project duration taking into account opinions from Data Protection Ethic Boards, Data Protection Authorities, Article Working Part 29, OECD, the literature and stakeholder feedback. The final data protection section will be made available to the BD4BO projects as soon as possible. At this stage, it would have been too premature to bring forward a final proposal on this topic rather a defined guidance on the issues listed above.

6. Guidance for BD4BO projects

Summary

Data management and collection

Do→IT has developed a set of operational and technology recommendations that encapsulate the lessons learnt from earlier IMI Data and Knowledge activities that have taken place within the IMI Framework.

A key consideration for a consortium in making secondary use of the data is how they plan to pool pseudonymised/anonymised individual patient data (IPD) for the purposes of enabling the most flexible data analysis approaches. An approach that has been implemented with success is to keep access to IPD only at the data controller level and "bring the analysis to the data" instead of pooling data together in a data warehouse to subsequently perform the analysis on the pooled data. This federated model with only data controllers processing IPD and other consortium members processing summarised data has been demonstrated as the generally best approach to facilitate analysis across distributed network of data providers, by balancing the risk of data misuse against the need of technical innovation This model addresses a lot of the key concerns around sharing of IPD but might be incapable of addressing certain research questions (e.g. when requiring pooled data sets for statistical analysis). In these situation, a physical pooling of pseudonymised/anonymised IPD is the only viable approach. A central third party (the data coordinator) can assume here responsibility for data harmonization and provide pooled data sets to the community of analysts.

The recommended working practices have implications for the type of Data Management Technology that is selected and deployed. DO→IT provides a primer for the selection of technology to ensure compatibility with existing technology investments of the IMI. Briefly, this implies using OMOP/OHDSI for clinical data using a distributed model, TransSMART for omics-data and RADAR Tools for continuous remote biosensors.

It is recommended that the discussion of data management is driven in partnership with data controllers of background sources. These data controllers will need allocation of sufficient resources.

Summary
Branding and communication summary
DO→IT has developed a graphic identity for the BD4BO programme that is intended to be used along with the distinct graphic identity of the individual BD4BO projects. The graphic identity is accompanied by guidelines on how to properly use the logos, fonts, colours and other visual elements.
DO→IT has also developed a communication plan and an outreach plan which aims to raise the profile of the BD4BO programme and expand the dissemination activities to audiences beyond the BD4BO projects' target groups. The communication plan covers the communication activities to be implemented by DO-->IT on behalf of the BD4BO programme and these activities will complement the individual and independent communication activities of the BD4BO projects. The outreach activities as defined by the outreach plan will identify and prioritise health care system stakeholders and will set objectives and milestones for engaging with external stakeholders. The effectiveness of the programme's communication and outreach activities will be assessed through pre-defined key performance indicators.

Summary
Stakeholder engagement summary
Effectively and efficiently engaging with all relevant stakeholders is a critical success factor for a BD4BO project. Three key recommendations are given: 1) undertake a stakeholder mapping exercise, which should identify institutions and existing initiatives that are relevant to the activities of the BD4BO project; (2) assemble an external advisory board comprising of representatives from regulatory bodies, patient groups, health insurance funds, health ministries, academic institutions and health care providers to provide strategic input to the BD4BO project; (3) ensure that resources are used efficiently by reaching out to other BD4BO projects to identify overlaps and opportunities for synergies.

Summary
IP management for the BD4BO programme
The IMI Intellectual Property (IP) provisions/rules govern the IP regime of all projects supported by IMI and apply equally to all partners in the projects. The IP provisions are designed to promote the creation and exploitation of knowledge and to reward innovation, while ensuring that the assets and interests of the project partners are respected.
The IMI Programme Office offers impartial advice to all partners during negotiations on IP and ensures that the resulting agreement is in line with the IMI IP provisions. The neutral role adopted by the IMI along with the flexibility of the IP provisions have allowed IMI project partners to share resources and knowledge in unprecedented ways. IP issues are agreed before the launch of the project ensuring that the knowledge developed and shared within the project will be used appropriately.
The base of all IMI2 IP regulations is the legal framework of the HORIZON2020 (H2020) programme which covers all H2020 research and innovation actions. Due to the special nature of IMI2, the H2020 legal framework was adapted in specific areas, one particular area being IP. The IP provisions for all IMI 2 projects are set out in Articles 23 to 31 of the IMI 2 Model Grant Agreement.

Summary
IP management for the BD4BO programme
The IMI Intellectual Property (IP) provisions/rules govern the IP regime of all projects supported by IMI and apply equally to all partners in the projects. The IP provisions are designed to promote the creation and exploitation of knowledge and to reward innovation, while ensuring that the assets and interests of the project partners are respected.
The IMI Programme Office offers impartial advice to all partners during negotiations on IP and ensures that the resulting agreement is in line with the IMI IP provisions. The neutral role adopted by the IMI along with the flexibility of the IP provisions have allowed IMI project partners to share resources and knowledge in unprecedented ways. IP issues are agreed before the launch of the project ensuring that the knowledge developed and shared within the project will be used appropriately.
The base of all IMI2 IP regulations is the legal framework of the HORIZON2020 (H2020) programme which covers all H2020 research and innovation actions. Due to the special nature of IMI2, the H2020 legal framework was adapted in specific areas, one particular area being IP. The IP provisions for all IMI 2 projects are set out in Articles 23 to 31 of the IMI 2 Model Grant Agreement.

Summary
Project management
An IMI2 funded project will have a work package dedicated to the overall management of the project. Due to the complexity of the projects and the large number of partners involved, project management resources will be needed at the task, deliverable and work package level in addition to the resources allocated for overall project management. Three key documents will be used throughout the project: 1) the Grant Agreement; (2) the Consortium Agreement; and (3) the Description of Action. It is also recommended that every project develop a project management handbook containing information on project management procedures relating to governance, internal communications, administration and finance. This handbook should be distributed among the project partners. It is recommended that key performance indicators are developed to measure the projects' performance overtime. Finally, it is suggested that all project coordinators and project managers devote time to understand the financial implications of running an IMI2 project. IMI has produced webinars and presentations to aid in this task.

The four work packages[5] within the DO→IT consortium operate with the objective of supporting and broadening the impact of the individual projects within the BD4BO programme. The objectives of these work packages include providing operational and strategic support, providing methodological guidance for the collection and management of data in real-world settings, giving the BD4BO projects an effective means of communication and collaboration with health care system stakeholders and developing minimum data privacy standards for informed consent forms and providing explanatory information on the use of such documents.
Based on the work done by the DO→IT work packages, this section provides guidance for current and future BD4BO projects on issues pertaining to issues relating to data management and collection, branding and communication, stakeholder engagement, IP management, and project management.[6]

6.1 Data management and collection

6.1.1 Introduction

The BD4BO initiative is a second-generation IMI program that aims to leverage “big data” to support the advancement of health outcomes research. To ensure that IMI resources committed to this program are spent appropriately and not wasted on duplicative work, it is critical that BD4BO projects re-use existing technological or governance solutions to the maximum extent possible, acknowledging this is fast paced field where technological advancements can and should be incorporated as appropriate. To that end, this subchapter summarises a set of operational and technology recommendations that encapsulate the lessons learnt from earlier IMI Data and Knowledge activities that have taken place within the IMI Framework.
New BD4BO projects should adhere to these recommendations as part of the development of their technology development and where deviating from them, justify their decisions in the development of the full project proposal.
There are three principles that underlie these recommendations:

- Data Management Practices in large research consortia are not a purely technical concern. They require new working practices between data providers and data consumers. Applying existing best working practices is critical to ensure that mistakes from earlier projects are not repeated.

- Data Management Technology has received considerable funding support from both EFPIA and Public funders in IMI, FP7 and H2020. To ensure best value for the European tax payer it is preferred that where technology is available and open for reuse, it is reused. Applying new technology for achieving the same purpose is discouraged.

[5] WP1: Programme Strategy and Coordination, WP2: Knowledge Integration and Management, WP3: Communication and Collaboration, WP4: Minimum Data Privacy Standards for IFC:s and Supporting Materials
[6] The guidance notes were submitted by the DO→IT work packages.

- The BD4BO program has been developed as a strategic program that will enhance the overall competitiveness of health outcomes research. As such individual projects within the program that develop data management solutions that do not add to the overall value of the BD4BO will not contribute to this transformative strategic objective.

In combination, these three principles create a strong argument for new BD4BO projects to use existing technologies and governance solutions and not build new solutions, unless absolutely required by the outcomes questions that form part of the study.

6.1.2 Data management in research consortia

From a data management perspective, it is helpful to categorise the stakeholders and the data access scenarios that are being used.
With respect to a given data set, the categories of functional participants are (note that these categories are not mutually exclusive):[7]

- Data Subject: The individual to which the data relates. (note that the term data subject here refers to the functional definition and not necessarily to the definition as per privacy regulation i.e. for this further discussion data related to a particular data subject might be fully de-identified or not)

- **Data controller** – The entity that manages a dataset and controls access and content of the dataset. They may be a clinical site, a research laboratory or other data producing facility. Data owner and data controller might be one and the same organization while in other cases (e.g. regional health databases or primary care databases) it will be different organizations or individuals

- **Data co-ordinators** – These are member of the consortium who are typically data management professionals that seek to enable the consortium by supporting the harmonisation and integration of the study data on behalf of the consortium.

- **Data Consumers** – These are all members (including associate members) of the consortium who would want to use the data either individually or part of an integrated consortium wide data set to answer the research questions that are the objectives of the consortium.

With respect to data access scenario categories – these fall into:

- **Use of data collected during the project** - This scenario occurs where the consortium sponsors a new research study to collect real world data. In the context of the IMI project data generated during the time of the project is considered foreground IP. The patient consent and ethical review are consistent across the project and data management tools are all generated

[7] *Advice received from Covington:* From a data protection law perspective, study participants would qualify as "data subjects." The other categories of participants would most likely qualify as "data controllers" depending on the types of data they process and the functions they perform. "Data controller" is a designation defined by applicable privacy regulation. The definition in GDPR Art. 4.7 is as follows: "the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by EU or Member State laws, the controller may be designated by those laws."

by the consortium. For example, a novel biomarker discovery cohort study (IMI Examples include U-BIOPRED & Oncotrack)

- **Use of data collected prior (or outside) the project** – The scenario occurs where the consortium aims to pool a number of background datasets provided by consortium members for either a secondary use to its initial purpose or an extended use to in initial purpose. In the context of and IMI project this data is considered background IP and owned by the data owner who is responsible to ensure its use is compliant with both the patient consent and ethical review boards who approve its use. Examples include EMIF and Predict-TB)

With respect to data types, three main types can be distinguished:

- Directly identifiable data: the study participant can easily be identified by means of name, address, contacts details, or other direct identifier.
- Pseudonymised data: identification of the study participant is not possible without additional information which is held separately from the study data (e.g., key coded data). Identification is possible using reasonable means.
- Anonymous data: the data cannot reasonably be associated to a particular identifiable study participant.

Data protection laws and the GDPR only apply to the first two categories of data and not to anonymous data. Where research objectives can be met with anonymous data, anonymous data should be used. If not, studies should use pseudonymised data.

6.1.3 Logical and physical data flow

The logical data flow describes how and by which organizational data are transformed from their 'raw' or source format into analysis results. A physical flow indicates how data are physically transferred and where data are stored.

In relation to the physical data flow, the implementation is dependent upon the specific use case in the respective project but the following guidelines can be applied:

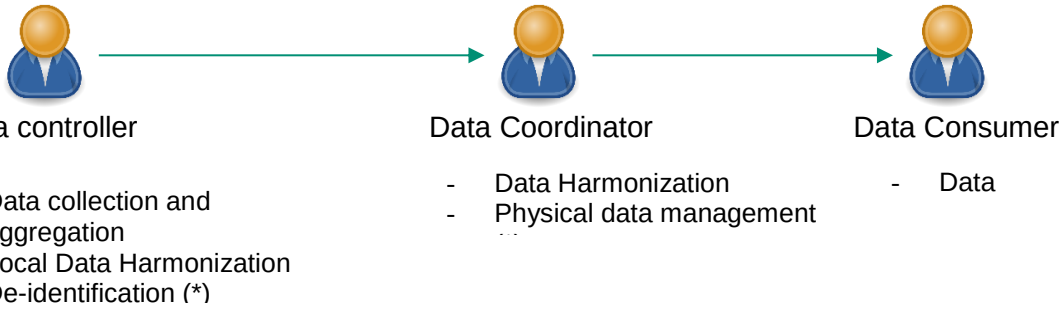


Figure 2: Logical flow of data across different stakeholders

1) Disclosure of individual patient level data (IPD) should be kept to an absolute minimum, especially if it relates to directly identifiable data
2) Pseudonymisation/anonymisation should occur as close as possible to the original source – to the extent that can be accommodated by the research questions
3) Data controllers should keep control over the access to their data - especially for IPD – irrespective of the place where these data are stored

6.1.4 Working practice recommendation

In the context of the BD4BO collection, it could be anticipated that a significant portion of the data used will initially come by accessing data from existing real-world data resources, such as electronic health records, case management systems or patient registries.

A useful reference in this scenario is the “Code of Practice for Secondary Use of Medical Research Data” [21, 24].

A key consideration for a consortium in making secondary use of the data is how they plan to pool pseudonymised/anonymised individual patient data (IPD) for the purposes of enabling the most flexible data analysis approaches.

An approach that has been implemented with success is to keep access to IPD only at the data controller level and “bring the analysis to the data” instead of pooling data together in a data warehouse to subsequently perform the analysis on the pooled data. To achieve the necessary scalability, this approach consists of the following steps:

1. A consortium will agree on a common data model
2. Each data controller will transform their data into a local version of the common data model
3. Research questions will be developed and approved by the respective Internal Review Board (IRB) and an “analysis script” - a computer program that assumes the common data model - is developed.
4. The IPD element of the analysis script is sent to each data controller who then runs this on their common data model. Summarised patient data are returned from each controller to the data co-ordinator who can then pool these data sets for continued analysis

This federated model with only data controllers processing IPD and other consortium members processing summarised data has been demonstrated as the generally best approach to facilitate analysis across distributed network of data providers, by balancing the risk of data misuse against the need of technical innovation. It has been implemented by the European and US based projects: FP7 EU-ADR, IMI EMIF, OMOP, OHDSI and the FDA Sentinel project.

This model addresses a lot of the key concerns around sharing of IPD but might be incapable of addressing certain research questions (e.g. when requiring pooled data sets for statistical analysis). In these situation, a physical pooling of pseudonymised/anonymised IPD is the only viable approach. A central third party (the data coordinator) can assume here responsibility for data harmonization and provide pooled data sets to the community of analysts (statisticians, informaticians and machine learners).

In this case, the data harmonisation is the responsibility of the network of data controllers within the consortium. Each controller will need to ensure that the processing of their data is compliant with:

- Local privacy regulations and is suitably pseudonymised/anonymised for scientific purposes (i.e., context-specific small risk of re-identification)
- The ethical governance and patient consent where required. Each use of the data may need to be reviewed by an IRB or data access committee
- The security practices and management of the third-party data processing capability are sufficiently mature that they can be audited

6.1.5 Preferred system recommendations

The recommended working practices have implications for the type of Data Management Technology that is selected and deployed:

- It should support the recommended operating model
- It should support the common harmonisation standards that are required to deliver outcomes research

- It should support the strategic aims of the BD4BO program by being aligned to the other technology in this space

A useful introduction on considerations about data standards is the eTRIKS Standards Starter Pack [25]. This has been compiled by both clinical standards authority (CDISC) and the biomedical standards community (ISA Tab organisation)

With these conditions in mind the following overview should be used to help identify technology platforms complimentary to existing IMI investments and practice

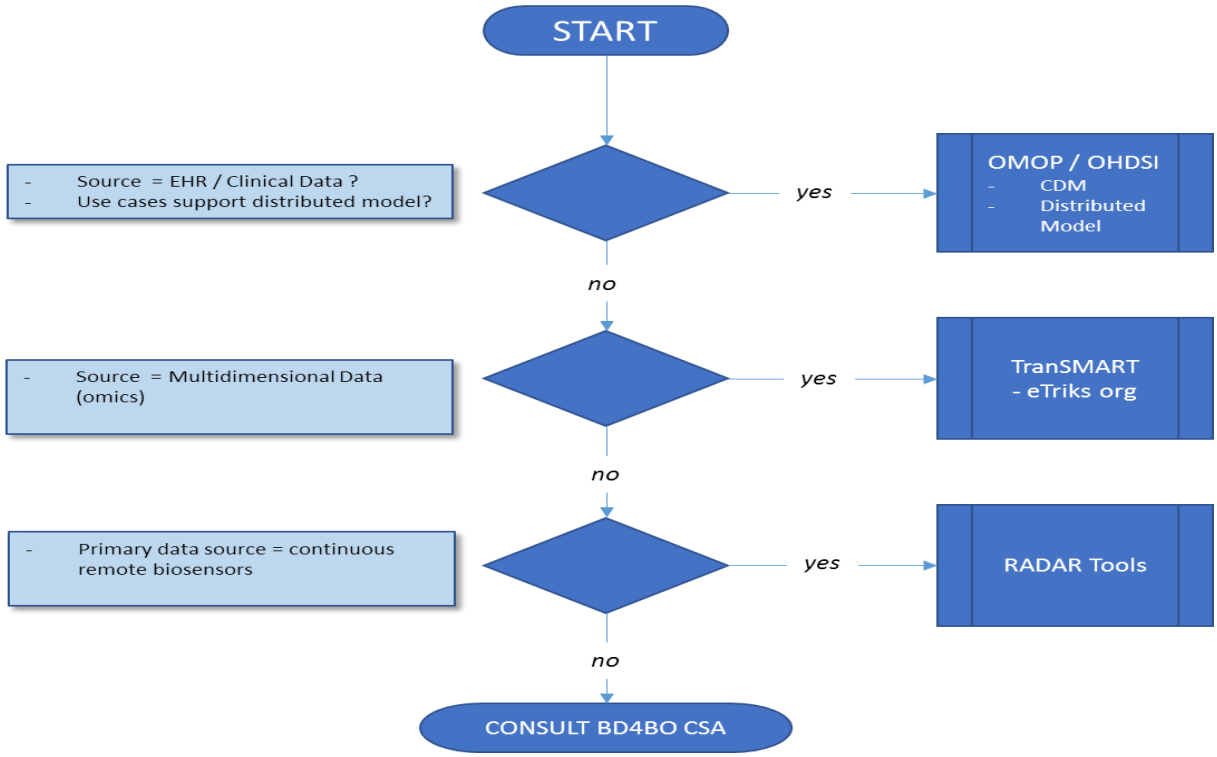


Figure 3: Technology selection overview to ensure compatibility with existing technology investments of the IMI. Depending on the architecture of the project is may be justifiable to select a combination of methods

OHDSI CDM & Tools

The OHDSI Project [26] is an approach and series of tools built for using healthcare data to support outcomes research. It is a successor to the FNIH OMOP project where the definition of the Common Data Model was defined for healthcare research that enabled the approach to processing IPD at a controller's site, but enable analytics to be run over all controller's in the network.

OHDSI tools enable a common technology standard that has been adapted by both EMIF and EU-ADR projects as the basis of enabling larger scale research networks than previously thought.

OHDSI tools are open source so do not require software licensing fees and EU SME's provide technology support to help projects leverage these tools.

eTRIKS/TranSMART

The TranSMART project [27] is an open source technology for providing knowledge management tools that has been most successful in prospective biomarker cohorts. It is a patient centric repository that captures patient data.

IMI eTRIKS project [28] has supported the use of tranSMART in over 30 research consortia. As a consortium they are willing to advice any IMI project on all aspects of data management technology that can be applied to translation research. From advice to training in how to deploy and use

tranSMART in a project to sustainable support hosting translational research project data via the Luxembourg Elixir Node.

RADAR Tools

The IMI program RADAR is a sister to the BD4BO program that aims to better develop and validate the science of using continuous remote measuring technology such as wearable devices in Depression, Multiple Sclerosis and Epilepsy with other topics such as Alzheimer's planned in the future.

The RADAR consortium leverages technology such as purple robot [29], openSMILE [30] and tranSMART for knowledge management in this emerging area of science. They can be contacted via the IMI project office.

6.1.6 Considerations for projects

As mentioned in the introduction, these recommendations have been made on the basis of practical experience in health outcomes research in European consortia. They aim to balance the risk to data controllers with the ability to drive innovative research, to ensure that investments are complimentary and competitive, and to ensure a transformative future for health outcomes research in Europe across all disease area.

It is recommended that the discussion of data management is driven in partnership with data controllers of background sources. This will increase significantly the chance of user acceptance and generally accommodate a better use than defining the data management approach in isolation.

Data controllers will need allocation of sufficient resources in a project to take on the following responsibilities (who may be the projects data coordinators):

- Anonymise and transform their data for analysis, in preference by the common data model outlined
- If IPD is processed locally – resource to run local analytical tasks when received from the project
- If IPD can be pooled by a third-party data co-ordination site to audit and validate the technology environment

This may require projects to reflect these recommendations in their scope / budgets

6.1.7 Future evolutions

It is recognized that the field of big data and outcomes research is in full development. The technico-legal aspects therefore need to be monitored closely when deploying solutions for new initiatives. Two relevant evolutions include:

- EU General Data Protection Regulation [31]: The upcoming change in data protection regulation will come into effect in May, 2018 and relates to the processing of personal data. Although – to the extent known- the above text is written with the current understanding of the upcoming GDPR regulation, specific elements
- Blockchain technology is a distributed ledger technology and is mostly known for its association with the digital currency Bitcoin but it has potential to be applied as well in healthcare [32]. It may pose some issues from a GDPR viewpoint since no data are deleted.

6.1.8 Summary

The above recommendations should be the basis for the default strategy for BD4BO initiatives. Where possible the proposed technologies should be applied - primarily to ensure IMI investments in BD4B0 are focused on defining health outcomes research and not on development of novel underlying technology. At the same time, technological evolutions, the emergence of novel data types or analysis

methods or changes in the legal context might necessitate either adoption of the proposed tools or application of new technologies. Such should be done, following a conscious evaluation of alternatives and in consultation with other BD4BO / IMI projects and DO→IT.
See appendix 6 for a checklist on data management for reviewers of BD4BO projects to consider.

6.2 Toolkit for BD4BO projects

Many of the individual BD4BO projects are working to develop data access platforms, enabling data to be accessed centrally from different sources, in order to be able to pool resources across different jurisdictions and get more information from more people and ultimately get better estimates of a drug's effectiveness. Part of this work involves defining and agreeing a common dataset and outcome set that should be prospectively be collected in order to improve interoperability between sources. This ensures that the outcomes which are collected are 'fit for purpose' for all stakeholders throughout the development process and ultimately ensure patients' lives are improved through access to effective medicines.

The purpose of this toolkit is to aid current and future BD4BO projects in the identification, selection, and measurement of outcomes in real-world settings by combining strong methodological background (obtained through systematic synthesis of available literature) and practical considerations in BD4BO projects on one side, and considerations of outcomes usage by relevant stakeholders on the other side to ensure maximum and sustained impact beyond the initial 2-year duration of the CSA. The toolkit will be specifically for BD4BO projects that are working to establish and agree the minimum outcomes (the 'core outcome set') that should be collected and measured in a disease area. However, this tool will also be useful for other organisations in Europe or internationally intending to initiate data collection. It may also be useful for regulators, health technology assessment (HTA) agencies and payers to better understand differing levels of acceptability of outcomes and the factors contributing to this and to consider using developing methods to specify outcomes of interest.

The toolkit will combine both methodological considerations on developing these core outcome sets with practical considerations including the acceptability of outcomes for different stakeholders such as the decision-makers who determine access to medicines or opportunities for collecting outcomes not normally considered in trials.

For the toolkit, the focus will be on clinical and patient-reported health outcomes which are used to demonstrate the effect of a medicine. However, it is clearly important to collect patient, disease and intervention-related data, as well, to contextualise the data as part of the dataset. Whilst this is important for both randomised controlled trials (RCTs) and observational studies, it is particularly important for observational data which is typically more subject to bias than RCTs. Knowing these characteristics can enable a better understanding of where these biases exist and may enable more analyses to be conducted on this data to attempt to minimise these biases.

6.3 Branding and communication

6.3.1 Branding and logos

DO→IT has developed a graphic identity for the BD4BO projects which is intended to be used alongside their own distinct graphic identity.

The logo, figure 4, is composed of the visual representation of data that is shared between 28 countries. All together, they represent a brighter future for healthcare development. The shape of the sun is an illustration of the "brighter outcome", while the sun rays are an illustration of a bar chart "representing data", see also appendix 7.

Figure 4: BD4BO Logos

The graphic identity is accompanied by guidelines to make proper use of the logos, fonts, colours etc. These guidelines will be shared with the designers in charge of developing the identities of the various BD4BO projects in order to ensure their compatibility with the overall BD4BO identity. More details are available in the document "Big Data for Better Outcomes Brand Guidelines".

Together with the branding, different communicative support tools such as templates for letters, PowerPoint presentations and deliverables have been developed and are available for the BD4BO projects to use and adapt to their own branding identities.

6.3.2 Communication plan

Communication is one of the key areas of the BD4BO programme to achieve its objectives. The communication strategy aims to raise the profile of the BD4BO programme and expand the dissemination activities to audiences beyond the BD4BO projects' target groups. This will ensure that the outputs generated from the BD4BO projects reach a wider group of health care sector stakeholders, which will include policy makers, health care professionals, HTA bodies, payers and patients. In addition, it will facilitate the alignment of the individual BD4BO projects' communication activities on a sustainable basis also beyond the timeline of DO→IT. The communication strategy should also ensure the visibility and transparency of the BD4BO programme's activities.
The main objectives of communication are:

- to raise awareness about the programme among stakeholders,
- to facilitate effective dissemination of the programme's aims and objectives, as well as results and outcomes,
- to ensure participation at conferences, workshops, and other events relevant for the programme,

- to gather knowledge, experience, and best practices for the successful and effective implementation of the programme
- to provide effective means for BD4BO partners in order to support their communication activities and stakeholder engagement
- to facilitate engagement with key external stakeholders, foster policy discussion and help build consensus with them,

With this in mind, DO→IT has developed a communication plan and an outreach plan for the programme. The communication plan covers the communication activities to be implemented by the DO→IT project on behalf of the BD4BO programme. The communication activities of DO→IT support and complement the individual and independent communication activities of the BD4BO projects. Meanwhile, the outreach activities will aim to identify and prioritise health care systems stakeholders, and will set objectives and milestones for engaging with external stakeholders, i.e. stakeholders beyond the BD4BO consortia.

Besides the support that the general communications and outreach plans will provide to the BD4BO projects, each project is expected to develop its own communication strategy to reach their disease specific target audience. To that end, BD4BO projects can use the Horizon 2020 communication strategy checklist (see appendix 8) to help them develop a communications plan. It is also recommended to follow the communication guide to IMI projects, also in appendix 8. This is not an exhaustive list of resources, but they can serve as a first step to prepare the dissemination and communication strategy of individual BD4BO projects.

The main outcome measures of the dissemination and communication relate to the awareness and support for the project from the core stakeholder groups. The effectiveness and success of the programme's communication and outreach activities are measured by pre-defined key performance indicators (KPIs), both quantitative (e.g. no. of website visits, no. of participants at programme workshops) and qualitative (e.g. awareness among stakeholders). In order to ensure effective BD4BO project communication, the use of key performance indicators is encouraged. A non-exhaustive list of indicators is found in appendix 9.

6.4 Stakeholder engagement

Succeeding in gathering and engaging all relevant stakeholders is a key area to succeed in the implementation of any BD4BO project. To do so, it is recommended that a careful stakeholder mapping of the relevant actors is developed and a comprehensive outreach plan implemented. It is also important to keep in mind that many overlaps may happen at the project level with other BD4BO projects. Therefore, potential synergies in stakeholder engagement efforts should be recognised. In addition, a devoted work package or an appropriate number of deliverables to address stakeholder engagement is highly recommended. For example, DO→IT regularly facilitates coordination calls on communication and outreach between BD4BO projects to ensure there is a coordinated approach to these activities across the BD4BO programme.

6.4.1 Stakeholder mapping and outreach plan

As a first step to successfully engaging with all the relevant stakeholders of the project you should undertake a stakeholder mapping exercise. The stakeholder map should contain institutions but also

existing initiatives that may have developed their work in the area. Besides contacts the mapping exercise should identify prior level of knowledge and engagement. An initial list of stakeholders could include: Patient organisations, healthcare professionals, trade organisations, research and academia, HTA Agencies, Regulators, Payers, Disease registries, industry and others.
The development of an outreach plan is the next step. It should make the best possible use of the resources the project has allocated. Measure the possible impact your activities could have and decide on the basis of what would be more effective for your project. The outreach plan such as the stakeholder mapping will be an evolving document that could change depending on arising opportunities and events.
In order to be able to measure the effectiveness of the stakeholder mapping exercise it would be important to carry out interviews, develop questionnaires and the use other tools to measure the knowledge and engagement of these institutions with the project. This would help measure the effectiveness of your project at different moments and evaluate the need to change the strategy to reach these groups of interest. The stakeholder map is an evolving document that should be modified throughout the lifespan of the project.

6.4.2 Advisory boards

A multi-stakeholder consortium like the one of the BD4BO projects covers a lot of expertise in the complex field of adoption and use of big data in healthcare. It is nevertheless essential to complement the existing consortium with strategic input from a broader group of stakeholders within and outside the EU. Therefore, building an external advisory board will be very valuable for the project. This group of advisors can include experts from ministries of health, health insurance funds, regulatory bodies, patient representatives, healthcare providers, academic institutions, and international organisations and initiatives. Besides this group of experts, it can be necessary to include specific expert input in the form of a Data Protection and Ethics Committee.

6.4.3 Efficient use of resources

Finally, it is recommended that every project reaches out to other BD4BO projects to identify overlaps and to avoid repeating work that has been already developed by other projects or institutions. This could save time and resources that could be better used in other areas of the project or for engagement activities with other stakeholders.

6.5 IP management for the BD4BO programme

The IMI Intellectual Property (IP) provisions/rules govern the IP regime of all projects supported by IMI and apply equally to all partners in the projects. The guiding principle behind provisions is IMI's objective of making a very practical contribution to improving the efficiency of drug development. The IP provisions are therefore designed to promote the creation and exploitation of knowledge generated and reward innovation, while respecting the assets and interests of all project partners.
An important aspect of IMI's IP provisions is their flexibility, which allows them to be adapted to the needs of the individual projects and their partners. Of significance, here is the neutral role played by the IMI Programme Office, which offers impartial advice to all partners during negotiations on IP, and ensures that the resulting agreement is in line with the IMI IP provisions and does not leave some project partners at a disadvantage. The flexibility of the provisions, coupled with IMI's neutral role in negotiations, have allowed IMI project partners to share resources and knowledge in unprecedented

ways and deliver results that would not have been possible otherwise.
Another key element of successful negotiations is that IP issues are agreed before the launch of the project. Project partners can therefore be confident that knowledge developed and shared within the project will be used appropriately.
The base of all IMI2 IP regulations is the legal framework of the HORIZON2020 (H2020) programme which covers all H2020 research and innovation actions. Due to the special nature of IMI2, the H2020 legal framework was adapted in specific areas, one particular area being IP. The IP provisions for all IMI 2 projects are set out in Articles 23 to 31 of the IMI 2 Model Grant Agreement (IMI2 IP Policy).
The principles of this policy address multiple interests (support to industry, freedom of access and compensation for IP) and have already allowed unprecedented levels of sharing as demonstrated under IMI1 by companies pooling legacy toxicity data, the European platform for antibiotic development, companies pooling & sharing old trial data and the European Lead Factory compound collection. This extensive sharing of data also allows project partners to validate each other's findings which is to promote the level of quality of European R&D activities significantly.
An important aspect of the IMI2 IP policy is the Background regulations. Partners can contribute selected assets as Background which can be used by all partners to carry out the project work, i.e. for the implementation of the Action. The rights to the Background will remain with the original owner and the rights to results which are obtained based on the Background (Foreground) can be pre-assigned to that the participant who contributed them to the Action. These regulations were designed to motivate partners to contribute sensitive assets with appropriate protection against unintended use by other partners/third parties.
In addition to Background and Results, the IMI2 IP policy also contains the concept of Sideground, which is generated during the Action but outside of its objectives and not needed for implementation or Research Use.
As a principle, the Results/Sideground of the project belong to the beneficiary who generated it. However, the regulations allow the transfer of ownership rights to individual partners in the Consortium Agreement. Different rules apply for single vs joint ownership of Results.
Results and Sideground can be transferred within the consortium to affiliates and purchasers without prior notification in case of individual ownership, whereas the individual use jointly owned Results and Sideground requires a prior notice and fair & reasonable compensation to the other joint owners.
It is mandatory for beneficiaries receiving EU funding (public partners) to ensure adequate and effective protection of the Results in line with the relevant (national) legal provisions, Action peculiarities and other legitimate interests. If valuable Results are left unprotected, the owners need to discuss this within the consortium.
Each beneficiary has the obligation to disseminate its own results as soon as reasonably practicable and the open access to publications is mandatory. It is also mandatory to mention the IMI support & Partners' in-kind contribution in patent applications and all other communications.
A very important aspect of the IMI2 Policy are the access rights to the Results, Background and Sideground which can be summarized as in table 1.

Access rights granted by a beneficiary to/on	**Background** (necessary and identified)	**Results**	**Sideground**
Beneficiaries for completion of the action	Royalty-free	Royalty-free	N.A.
Beneficiaries and affiliates for Research Use	Fair & reasonable terms (including royalty-free) for background needed for using the results	Fair & reasonable terms (including royalty-free)	N.A.

Third Parties for Research Use after the action	Appropriate conditions for background needed for using the results	Appropriate conditions	N.A.
Beneficiaries and affiliates or Third Parties for Direct Exploitation	To be negotiated	To be negotiated	N.A.

Table 1: Access rights

It is important to note the modalities of granting potential access to Results or Background for the use of Results for third parties:

- Time-limits for requesting access can be agreed in the Consortium Agreement of each project
- Almost all ongoing IMI projects agreed that access rights to background are granted without any additional administrative step

The basis for the cooperation and implementation of the DO→IT Description of Action (DoA) is the DO→IT Consortium Agreement which regulates the governance and IP in DO→IT.

A group of EFPIA companies did develop a draft Informed Consent Template which will be further developed by DO→IT. In addition, NICE defined some contributions as Background which will be subject to the above stipulations of the IMI2 IP policy.

The DO→IT partners agreed the access to Results and Background for the implementation of the action and for research use as royalty-free.

With regard to use for Direct Exploitation, beneficiaries are not required to grant Access Rights for Direct Exploitation to their Results. However, such Access Rights to the Results for Direct Exploitation may be negotiated between the Beneficiary owning such Results and the Beneficiary wishing to perform Direct Exploitation.

Third parties can submit a written request directly to the Beneficiary owning the Results or the Background concerned. Such Access Rights will be granted subject to an agreement between the Third Party and the Beneficiary owning the Results or Background on the conditions of such access for the agreed usage.

If the cooperation with an advisor generates results with intellectual property rights the advisors shall promptly disclose any Results to the Project Leader in writing. All rights, title and interest in any Results will be owned exclusively by the Participants in equal shares, and Advisor shall assign (or cause to be assigned).

6.6 Project management

This section on project management highlights some basic issues BD4BO projects should consider during the implementation of the projects.

The objectives of project management for any IMI2 project should ensure:

- A smooth execution of the project's work plan
- The work plan is adapted and corrected as the result of unforeseen circumstances
- An efficient communication within the consortium and the engagement of external stakeholders
- The coordination of all activities within the consortium
- Reporting to the IMI and the administrative and financial management of the project take place in a timely manner

In general terms, any IMI2 funded project will have a work package dedicated to the management of the overall activity. Due to the complexity of these projects and the big number of partners involved, project management resources will be needed at the task, deliverable and work package level on top

of the ones allocated to the overall management

6.6.1 Key documents

This is the list of key documents that will be used throughout the execution of the project:

1) Grant Agreement the contract concluded between the Innovative Medicines Initiative 2 Joint Undertaking and the beneficiaries under which the parties receive the rights and obligations [33].

2) Consortium Agreement: the internal agreement signed between the participants of the consortium establishing their rights and obligations with respect to the implementation of the action in compliance with the grant agreement. Ensure the correct execution of the project's work plan including its adaptation to necessities that may occur during the project

3) The Description of Action (Annex 1, part A of the Grant Agreement) is also a key document compiling a specific description of the tasks that will be carried out along the project and the expected results, deliverables and milestones to be obtained.

6.6.2 Project management handbook

It is recommended that every project includes as part of the deliverables the development of a project management handbook to be distributed among the partners. This document should include all the information concerning project management procedures and should provide answers to all the main questions regarding governance, internal communications, administrative and financial procedures as well. A document "Project management manual" has been developed as part of the Do ->IT project, the Coordination and Support Action of the Big Data for Better Outcomes Programme. The document provides practical guidelines, procedures and support documents that will help the project implementation. It will be kept up to date as needed throughout the project lifecycle.

6.6.3 Key performance indicators

Key performance indicators (KPIs) are a set of quantifiable measures that a project should develop to measure its performance over time. These measurements need to be presented concisely and in a user-friendly manner to make their interpretation easy.

The suggested approach would be to identify a few KPIs that can easily be read and interpreted. As such, the best strategy to utilize is to identify and leverage a few pertinent KPIs that will convey the information to any participant or stakeholder. For project management purposes, it will allow the consortium to identify potential problems and the project manager to address the issues ahead of time and take steps to mitigate it. Beware of the burden that excessive KPIs could put on the project. Too many KPIs can create confusion and 'noise' instead of shedding light.

In order to be effective, KPIs should cover the following areas:

- Timeliness
- Budget:
- Quality
- Effectiveness

See Appendix 10 for example of KPIs.

6.6.4 Financial management

It is recommended that all project coordinators and project managers devote time to understand what

the financial implications of running an IMI 2 project are. In order to facilitate this task, IMI has produced a presentation on financial Management rules of IMI 2 projects [34]. This presentation is the result of a training webinar that IMI ran between 2015 and 2016. In the past, IMI has organized webinars for IMI project participants and it is recommended that BD4BO project managers attend any such webinars in the future.

7. Conclusion

Big data has an important and increasing role in healthcare, where big data can be used for different purposes, such as improving patient outcomes, quality, and efficiency of healthcare. Projects on big data are of varying types, have different objectives and include different stakeholder groups. This means that there is *no one solution* that is applicable to all kinds of big data projects. Still, there are some learnings that can be drawn and lessons to be learnt for future projects.

The study shows that many of the identified success factors are at the same time challenges. Also, the different aspects identified are clearly interrelated. *A clear legal and regulatory framework* is a prerequisite for a well-functioning big data project, where it is important to have a balance between facilitating data use and the protection of sensitive patient data. It is also a challenge to harmonise regulation between sectors and countries, though the EU's General Protection regulation (GDPR) could be as a step forward. The concentration of *data control* affects how and to what degree stakeholders interact and collaborate regarding data sharing. Standardisation of data and platforms, as well as *interoperability* also affect the possibility to share and link data, and further development of this could improve. The work from this study shows that data ownership that allows re-use of data and ensuring data interoperability facilitate big data projects, without sacrificing the aspects of privacy and confidentiality. *Collaboration between stakeholders* is a key issue in the management of big data projects. An *early involvement* of stakeholders, including the patients, setting the framework, building trust and handle expectations increases the chances for a successful project. However, the study did not find clear evidence on *how* to involve stakeholders and to *what degree* they should be involved, though this of course is dependent on the individual project and its objectives.

This study shows similar results regarding identified success factors and challenges from the scoping review, online questionnaire and interviews. Based on this, even though big data projects have varying objectives and scope, it would be valuable to map out best practice for big data projects and to further investigate best practice, including hands-on information on for example how to operationalise *stakeholder collaboration*, including patients. Other aspects that would merit further work are how to *agree on core outcomes sets* in order to facilitate sharing and collaboration on data and how to ensure a *sustainable transition (of data and infrastructure) from project to continuous operation.*

The current BD4BO projects (HARMONY, ROADMAP, BigData@Heart) have taken into account most of the challenges and success factors identified in this study. All three projects have dedicated work packages to address the legal and ethical issues arising from integrating and sharing patient-level data. All projects will review and identify best practices and laws related to data sharing and protecting patient privacy in order to develop guidelines for the use of big data. Importantly, the three projects will assess the implications of the GDPR on the legal landscape in Europe. They will be given guidance and advice throughout the project through advisory committees comprising of legal and ethics experts. Further, the projects will seek advice from previous initiatives on issues related to data anonymization and pseudonymization.

In terms of addressing the technical and governance challenges related to integrating and sharing data, the BD4BO projects will build upon previously developed infrastructure (e.g., OMOP Common Data model) rather than reinventing the wheel. This will be supplemented by developing data governance frameworks that will take account of data privacy, rights of access and security.

All three BD4BO projects have emphasised the importance of engaging with decision making bodies. HARMONY and ROADMAP will establish feedback forums and advisory groups to identified guiding principles for the use of RWE in the regulatory and HTA context. These platforms will be used to discuss and refine these guiding principles to ensure the generalisability and applicability of the project outputs to support decision-making in terms of HTA recommendations, reimbursement approval, regulatory approval and patient access.

The importance of patient engagement is also recognised in the current BD4BO projects. This will be done through focus groups that will collect patient attitudes towards data integration and sharing. BigData@Heart will integrate a model for patient involvement into their data governance framework. These engagement efforts will be supported by communications activities involving disease associations and patient societies.

BigData@Heart will take measures to ensure the sustainability of key project outputs by incorporating sustainability factors into their governance framework and adopting best practices from previous projects. Further, BigData@Heart will attempt to commercialise some project outputs and deliverables to build on the investment made by the EU and other partners. HARMONY will also take measures to ensure project sustainability by providing access to the HARMONY database for future analyses by the European health systems, scientists, regulators and industry. In addition, HARMONY may potentially expand to other disease areas like chronic myelogenous leukemia (CML) and Hodgkin lymphoma. The current BD4BO projects plan to address most of the critical success factors and challenges identified in this study. It is hoped that the guidelines laid out in this document in the areas of BD4BO programme priorities, data management and collection, branding and communication, stakeholder engagement, intellectual property management and project management will help current and future BD4BO projects to effectively execute their plans and ensure that they are run in a well-functioning manner, allowing them to achieve their objectives. Further, following the suggested recommendations would ensure that all outputs produced by future BD4BO projects remain harmonised with other BD4BO projects, creating synergies and improving the sustainability of outputs.

References

1. European Commission, *Study on Big Data in Public Health, Telemedicine and Healthcare*. 2016.

2. ABPI, *Big data road map*. 2013.

3. Paris, V., et al., *OECD New Health Technologies: managing access, value and sustainability*. 2017, OECD.

4. Groves, P., et al., *The "big data" revolution in healthcare. Accelerating value and innovation* 2013, McKinsey&Company.

5. OECD, *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities, and data Privacy Protection Challenges* 2013.

6. OECD, *Recommendation of the OECD Council on Health Data Governance*. 2017.

7. Miani, C., et al., *Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe*. 2014, RAND Europe.

8. European Medicines Agency, *Patient Registries Workshop, 28 October 2016. Observations and recommendations arising from the workshop*. 2017.

9. Oderkirk, J., et al., *Health Data Governance. Privacy, Monitoring and Research*. 2015, OECD.

10. Anderson, G. and J. Oderkirk, *OECD Dementia Research and Care – Can Big Data Help?* 2015, OECD.

11. WHO, *Atlas of eHealth country profiles. The use of ehealth in support of universal health coverage.* 2016.

12. European Commission, *The Use of Big Data in Public Health Policy and Research*. 2014.

13. Big Data Value Association, *Big Data Technologies in Healthcare - Needs, opportunities and challenges*. 2016.

14. ICHOM, *What matters most. Patient outcomes and the Transformation of health care* 2015.

15. NICE, *Data Science for Health and Care Excellence. Harnessing the UK opportunities for new research and decision-making paradigms.* 2016.

16. European Medicines Agency, *Identifying opportunities for Big Data in medicines development and regulatory science*. 2017.

17. Soderlund, N., et al., *Progress Toward Value-Based Health Care. Lessons from 12 Countries.* 2012, The Boston Consulting Group,.

18. http://roadmap-alzheimer.org/.

19. https://www.imi.europa.eu/content/harmony.

20. http://www.bigdata-heart.eu/.

21. IMI, *Healthcare Alliance for Resourceful Medicines Offensive against Neoplasms in Hematology(HARMONY).* 2016(1).

22. IMI, *Real world Outcomes across the AD spectrum for better care : Multi-modal data Access Platform.* 2016. **0**.

23. IMI, *BigData @ Heart.* 2016. **1**.

24. https://www.etriks.org/code-of-practice.

25. https://www.etriks.org/standards-starter-pack/.

26. https://www.ohdsi.org.

27. http://transmartfoundation.org.

28. https://www.etriks.org.

29. https://tech.cbits.northwestern.edu/purple-robot.

30. http://audeering.com/technology/opensmile.

31. http://www.eugdpr.org.

32. Halamka, J., A. Lippman, and E. Ekblaw, *The potential for blockchain to transform electronic health records*. 2017.

33. http://ec.europa.eu/research/participants/data/ref/h2020/other/mga/jtis/h2020-mga-imi_en.pdf.

34. *IMI2 Financial Management Webinar*.

Appendices

Appendix 1: List of documents in scoping review of grey literature

International

- OECD
  - Dementia Research and Care**.** Can Big Data Help? (2015) [9]
  - New Health Technologies: Managing access, value and sustainability (2017) [2]
  - Health Data Governance: Privacy, Monitoring and Research (2015) [8]
  - Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges (2013) [4]
- WHO Europe
  - Global eHealth survey (2015) [10]
- EC
  - Study on Big Data in Public Health, Telemedicine and Healthcare (2016) [1]
  - The Use of Big Data in Public Health Policy and Research. Background information document (2014) [11]
- BDVA (Big Data Value Association)
  - Big Data Technologies in Healthcare (2016) [12]
- EFPIA: to be released in April 2017: report on the Value of Health data written by Rand Europe and commissioned by EFPIA
- ICHOM
  - What Matters Most: Patient Outcomes and the Transformation of Health Care (2015) [13]
- EMA
  - Final report on the adaptive pathways pilot (2016)
  - Workshop report from the EMA Registry Initiative – Patient Registries Workshop, 28 October 2016: observations and recommendations arising from the workshop (2016) [7]
  - EMA – Identifying opportunities for 'big data' in medicines development and regulatory science – this was off the back of an EMA held workshop on Big Data (2016)[16]

National

UK

- NICE
  - Data Science for Health Care Excellence. Harnessing the UK opportunities for new research and decision-making paradigm. Summary of a meeting (2016) [14]
- ABPI
  - Big data road map (2013) [15]

US

- RAND
  - Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe (2014) [6]
- AHRQ Agency for Healthcare Research and Quality
  - User's Guide to Registries Evaluating Patient Outcomes: Summary (2007) [21]
- FDA
  - Sentinel initiative report (2010) [22]

Commercial

- BCG
  - Progress towards value-based healthcare. Lessons from 12 countries (2012) [17]
- McKinsey
  - The Big Data Revolution in Healthcare. Accelerating Value and Innovation (2013) [3]
- IMS
  - White paper: Why pharma needs to work differently with payers and IDNs on RW (2015)
- IBM
  - Predictive Analytics in Value-Based Healthcare: Forecasting Risk, Utilization, and Outcomes (2017) 8 pages
  - The Future of Health is cognitive (2016) 12 pages

Appendix 2: Online questionnaire - Questions in IMI2 Survey on Big Data in Healthcare

Background questions * Mandatory

1. Name *
2. Title *
3. Name of company/organisation *
4. Which group of **stakeholders** do you represent? *

   □ Health care provider (e.g. hospital)
   □ Pharmaceutical industry
   □ Health care payer organisation
   □ HTA organisation
   □ Regulatory agency
   □ User/patient group
   □ Academia
   □ Don't know/prefer not to answer
   □ Other, please specify

5. Have you/your organisation participated in any projects involving external partners on collecting, analysing and/or utilising the analysis of big data related to healthcare in the last ten years? *
   Definition of big data: For the purpose of this survey, we are applying the following definition as stated by the European Commission in the report "Study on Big Data in Public Health, Telemedicine and Healthcare" (2016): *"Big data in health refers to largely routinely or automatically collected datasets, which are electronically captured and stored. It is reusable in the sense of multipurpose data and compromises the fusion and connection of existing databases for the purpose of improving health and health system performance. It does not refer to data collected for a specific study."* For the purpose of our exercise, we have specified that to qualify as big data it needs to involve the linkage of two or more data sets. Data sets could include registries.

   □ Yes-> Question 6 and onwards
   □ No-> Questions 22 and onwards
   □ Don't know/prefer not to answer -> Questions 22 and onwards

6. How many projects on big data in healthcare have you/your organisation participated in? *
   □ 1
   □ 2-5
   □ 6-10
   □ >10□ Don't know/prefer not to answer

Specific project with external partners on big data in healthcare

Please note that in the following questions in this section refer to experiences only from the **last completed project** on big data in healthcare that also included **external stakeholders.**

7. What was the **name of the *last* completed project** with external partners on big data that you/your company participated in? *
8. Who **initiated** that project?
9. What **stakeholder groups** were included in that project? Please tick one or more boxes. *

   □ Health care provider (e.g. hospital)
   □ Pharmaceutical industry
   □ Health care payer organisation
   □ HTA organisation
   □ Regulatory agency
   □ User/patient group
   □ Academia
   □ Don't know/prefer not to answer
   □ Other, please specify

10. What was **your/your organisation's role** in that project? *
    - □ Overall coordinating partner
    - □ Participating partner coordinating subproject
    - □ Participating partner with no coordination role
    - □ Don't know/prefer not to answer
    - □ Other role, please specify

11. In your opinion, what were **your/your organisation's key contributions** to that project in terms of expertise? Please tick one or more boxes. *
    - □ Setting up or maintaining hardware infrastructure
    - □ Governance and project management, including conducting workshops and engaging stakeholders
    - □ Data access and collection
    - □ Clinical expertise
    - □ Dissemination and communication
    - □ Analytic capacities
    - □ Legal experience with big data
    - □ Don't know/prefer not to answer
    - □ Other, please specify

12. Describe the **strategic vision** (the long-term aim) of the project in a few sentences:

13. What was the **objective**(s) of the project? Please tick one or more boxes. *
    - □ Improve access to outcomes data
    - □ Early detection or prevention of diseases (e.g. predicating epidemics, disease monitoring)
    - □ Improve quality of care (e.g. improve outcomes)
    - □ Improve patient safety and pharmacovigilance
    - □ Reduce uncertainty of the value or effectiveness of a treatment to support HTA or other decision making linked to pricing and reimbursement
    - □ Improve efficiency for the healthcare system
    - □ Facilitate outcomes-based payment models
    - □ Improve the effectiveness of drug development
    - □ Improve academic research
    - □ Don't know/prefer not to answer
    - □ Other, please specify

14. Thinking about your last completed big data project, imagine you can send your past self a message highlighting issues to put more effort on: what would it say, i.e. what were the **key learnings**?

15. Please describe in a few sentences the **key success factors** of the project?

16. What **key challenges** were encountered in the project? *

    Scale for each alternative: Fully agree- partly agree- partly disagree- disagree - don't know/prefer not to answer
    - □ Relevant data sets were not available for the purpose, e.g. no registries on the disease
    - □ Relevant data sets were difficult to access when it came to ownership, standards of sharing and collaboration with other researchers etc.
    - □ The interoperability between data sets was a hinder, e.g. the coding and terminology between data sets were not compatible and difficult to link.
    - □ Relevant outcomes measures were not available in the data set.
    - □ Methods and analysis of data were difficult
    - □ Legal aspects were unclear as to e.g. privacy issues
    - □ Other, please specify

17. Please elaborate in a few sentences the **key challenges**

18. Summarise in a few sentences how the project dealt with the **challenges** encountered?

19. Did the project incorporate **patient perspectives** on big data? *
    - □ Yes -> If yes: How was the patient perspective addressed?
    - □ No

□ Don't know/prefer not to answer

20. Was there a solution to **sustainable provision** of project results (such as e.g. continued access to outcomes data or infrastructure) after the end of the project, i.e. not dependent on the project funding and organization? *
    □ Yes -> If yes: How was the sustainable provision of project results organized?
    □ No
    □ Don't know/prefer not to answer

21. What categories of **datasources** were used in the project? Please tick one or more boxes. *
    □social media
    □wearables
    □electronic health records
    □disease specific patient registries
    □administrative systems/claims data
    □don't know/prefer not to answer
    □other, please specify

General views on big data in healthcare

22. In your opinion, what are the three most important **challenges** that need to be addressed in the near future on a national and/or international level regarding big data in healthcare?

23. In your opinion, what are the three most important **success factors** that could benefit projects on big data in health care?

24. In your opinion, who would be the most **appropriate institution** to address the challenges you identified in the previous question? E.g. national authorities, European commission, OECD, EFPIA.

25. Do you believe such initiatives should be done on a **voluntary** basis or on a **binding legal** basis?

26. In your opinion, what role does big data have in developing **value-based** and **outcomes-focused** healthcare systems?

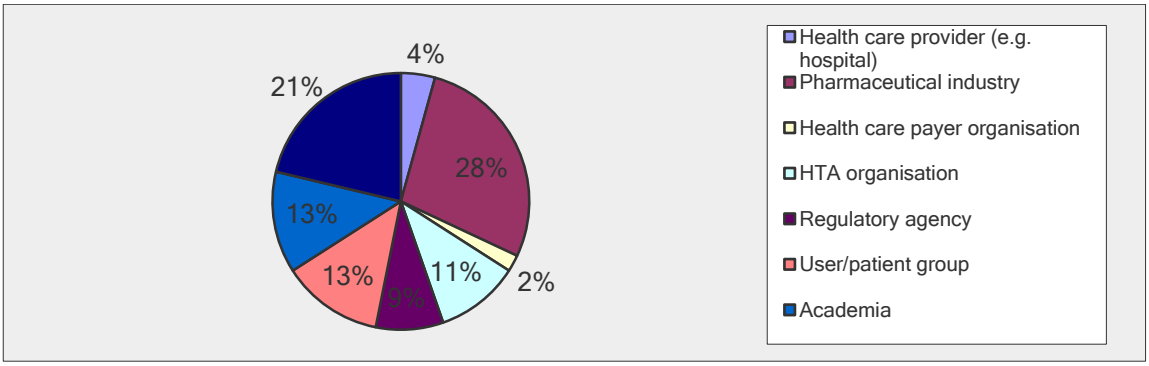
Appendix 3: Results from online questionnaire

The following summary of questionnaire data, collected from a range of stakeholder groups, provides an overview of key points from each question followed by tabulation/figure of the data.[8] Responses are divided into two sections, questions 4-23 reflecting experience of stakeholders with recent big data projects and questions 24-28 reflecting stakeholder opinion on factors for success and challenges of using big data given the current external data, regulatory and payer environment.

BD4BO Survey Analysis: Questions 4-23

Question 4: Which group of stakeholders do you represent?

This survey captured a diverse range of stakeholders including pharma, academia, regulatory and payer organisations. The figure below captures the distribution of respondents. The *other* category itself captured diverse stakeholders including government institutes, multinational organisations and service providers.



Respondent Breakdown [# of Responses]

HCPs [2]
Pharma [13]
HTAs/Payers [6]
Regulatory agencies [4]
Patient groups [6]
Academia [6]
Other [10]:

- International association [2]
- European institution [1]
- Government institution [1]
- Research database service [1]
- Multi stakeholder networking organisation [2]
- Open science collaborative with industry/government/academia/health systems all represented [1]
- Medical society [1]
- Association of local authorities and regions [1]

Question 5: Have you/your organisation participated in any projects involving external partners on collecting, analysing and/or utilising the analysis of big data related to healthcare in the last ten years?

It is of note that 4 out of 13 Pharmaceutical manufacturers did not participate in any projects involving big data related to healthcare in the last 10 years or did not know about big data projects in their organisation. Of the other stakeholder groups surveyed all, but the user/patient groups, had a higher proportion of respondents with experience of big data projects.

Stakeholders Breakdown [# of Responses]:

Stakeholder	Yes	No	Don't know
Health care provider (e.g. hospital)	2		
Pharmaceutical industry	9	2	2
Health care payer organisation	1		
HTA organisation	4	1	
Regulatory agency	4		
User/patient group	4	2	

[8] This section has been written by GSK.

Academia	5	1	
Don't know/prefer not to answer	0	0	
Other, please specify	8	1	1

Question 6: How many projects on big data in healthcare have you/your organisation participated in?
Pharmaceutical manufacturers had the greatest experience with big data in healthcare; 7 out of 9 Pharmaceutical manufacturers have participated in more than 10 projects. Most of the other stakeholders have participated in 2-5 projects on big data related to healthcare.
Stakeholders Breakdown [# of Responses:

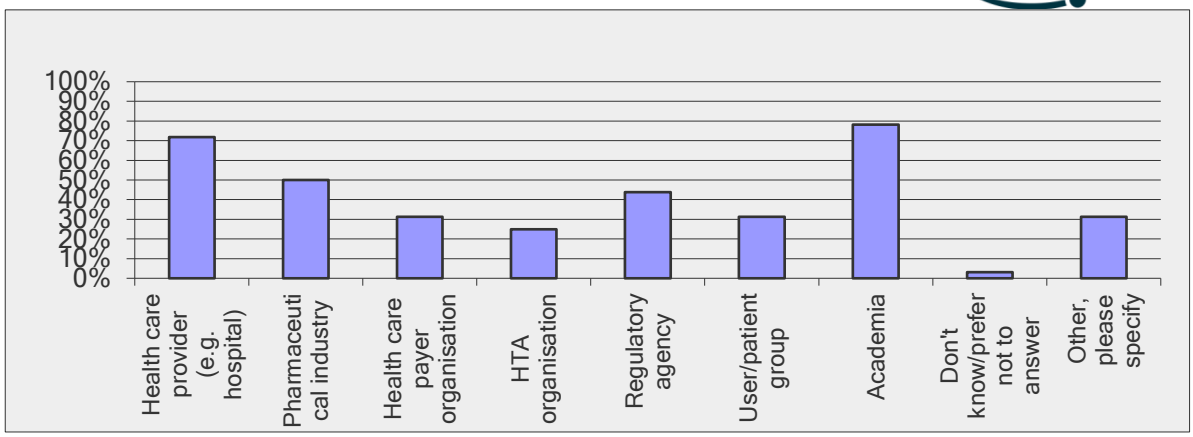
Stakeholder	1	2-5	6-10	>10	Don't know
Health care provider (e.g. hospital)		1			
Pharmaceutical industry		1		7	1
Health care payer organisation		1			
HTA organisation		4			
Regulatory agency	1	2		1	
User/patient group		3			1
Academia	1	2		1	2
Don't know/prefer not to answer					
Other, please specify		4		4	
Total	**2**	**18**	**0**	**13**	**4**

Question 7: What was the name of the last completed project with external partners on big data that you/your company participated in?
Respondents provided the name of very specific projects. IMI funded projects were frequently given as the example project (EHR4CR and IMI GetReal n=5).

Question 8: Who initiated that project?
The majority of big data initiatives were started by the stakeholders themselves. The exception was for regulatory agencies, though the number of respondents was small (n=3).

Respondent Breakdown [# of Responses]
Pharma: 6 of 7 responses indicated pharma had initiated the project, the exception was IMI initiated.
Regulatory agencies: the 3 responses showed diversity: regulatory agency, IMI and academia all quoted as initiators.
HTAs/Payers: The 5 responses indicated all projects were initiated by payor or government organisations with the one exception being IMI.
HCPs: Only two responses were received with both projects initiated by academic/public health partners.
Patient groups: 3 responses were received with no pattern observed: initiators included academic, a patient group and one unknown answer.
Academia: Of 5 responses, 3 projects were initiated by the academics with a further two initiated by research council or health fund.
Other: Of the 5 responses from a range of stakeholders not captured by the categories above, projects were initiated by the institute themselves, IMI or other collaborative initiatives (e.g. OMOP).

Question 9: What stakeholder groups were included in that project? Please tick one or more boxes.
Most of the projects were collaborative projects with Academia and HCPs the most common groups included across projects. It is also notable that User/Patient groups were only included in less than a third of the projects.



Question 10: What was your/your organisation's role in that project?

Pharmaceutical manufacturers were rarely co-ordinating partners. Academia or 'other' organisations were more likely to take the co-ordinating role. Pharmaceutical manufacturers were more likely to coordinate sub-projects.

Stakeholders Breakdown [# of Responses

Stakeholder	Coord Partner	Particating partner co-ord subproject	Participating partner no co-ord	Don't know	Other
Health care provider (e.g. hospital)	1				
Pharmaceutical industry	1	4	1		2
Health care payer organisation		1			
HTA organisation	1	1	1	1	
Regulatory agency	1		2		
User/patient group		2	1		
Academia	4		1		1
Don't know/prefer not to answer					
Other, please specify	6				
Total	**14**	**8**	**6**	**1**	**3**

Question 11: In your opinion, what were you/your organisation's key contributions to that project in terms of expertise? Please tick one or more boxes.

It is notable that Academia most commonly rated contributions to analytics as key while Pharmaceutical manufacturers considered their key contributions to be to clinical expertise. Both academia and pharmaceutical manufacturers were strong contributors to governance, while HTA organisations scored highly with dissemination of the project.

Stakeholders Breakdown [# of Responses]:

Stakeholder	Set up	Governance	Data Access	Clinical	Dissemination	Analytics	Legal	Don't know	Other
Health care provider		1	1	1	1	1	1		
Pharmaceutical industry		4		4	3	2	2	3	1
Health care payer organisation		1			1				
HTA organisation	1	3	1	1	4	1	2		1
Regulatory agency				2	1	1			2
User/patient group		2	3		3				
Academia	2	4	2	1	2	5	1		
Other	2	6	4	3	5	4	4		
Total	**5**	**21**	**11**	**12**	**20**	**14**	**10**	**3**	**4**

Question 12: Please describe the strategic vision (the long-term aim) of the project in a few sentences

Responses showed high variability between stakeholders. Some of the projects and stakeholders had very specific research vision, for example:

Cardiovascular disease population-based outcomes assessment

Study the safety and effect of pandemic flu vaccines

To describe methodology, interim baseline, and longitudinal magnetic resonance imaging (MRI) acquisition parameter

characteristics of the multiple sclerosis clinical outcome and MRI in the United States
While other projects had much broader aims (e.g. ensuring patient safety, supporting clinical trials design, assessing patient outcomes).
Ensuring safe use of medicines
making electronic health records accessible for clinical trial recruitment
To show how robust new methods of RWE collection and synthesis could be adopted earlier in pharmaceutical R&D and the healthcare decision making process.
Overall the projects involving regulatory agencies and HTAs had broader objectives.

Question 13: What was the objective(s) of the project? Please tick one or more boxes.
As expected, the HTAs were more interested in the outcomes based payment models than other stakeholders. The Pharmaceutical manufacturers indicated prioritisation of patient safety and drug development effectiveness, whereas Academics championed improving academic research.
Patient safety and improving data access scored quite consistently across all the stakeholders. Stakeholders Breakdown [# of Responses]:
Improve access: improve data access

Stakeholder	Improve access	Early detection	Quality of care	Patient safety	Reduce uncertainty	Increase efficiency	Facitliate outcomes models	Increase DD effectiveness	Improve academic research	Don't know	Other
Health care provider (e.g.	1	1	1			1	1		1		
Pharmaceutical industry	2		1	3	3	2	1	3	2	1	2
Health care payer organisation	1	1	1	1		1		1	1		
HTA organisation	3		1	1	4	2	4	1	1		
Regulatory agency	2		3	3	2		1	2	2		
User/patient group	2	2	2	2	1	1		1	2		1
Academia			1			2			5	1	1
Don't know											
Other, please specify	3	2	4	5	1	2		3	4		1
Total	14	6	14	15	11	11	7	11	18	2	5

DD is drug development

Question 14: Thinking about your last completed big data project, imagine you can send your past self a message highlighting issues
Respondents cited a broad range of issues with the following general themes emerging across all stakeholder groups:

- Complexity and quality of data (widely cited across stakeholders especially Pharma, Academia and Regulatory Agency)
- Governance/legal/compliance issues (of particular concern for Academia, Governmental institutes, Healthcare Payor Organisation)
- Stakeholders management and coordination (widely described across stakeholders especially Pharma and HTAs)

The responses from Pharmaceutical placed more emphasis on challenges with co-ordination across stakeholders whereas Academia placed more emphasis on data complexity and governance. It is notable that the two regulatory agency responses were very different with a positive comment on big data for hypothesis testing in one hand and a more sceptical comment on data validity in the other hand.

Question 15: Please describe in a few sentences the key success factors of the project.
Respondents cited a broad range of successes with general themes emerging around:

- Data access and linkage
- Communication, engagement and communication across partners

Academics highlighted the importance of technical aspects more than others, especially regarding data linkage and data access, whereas Pharmaceutical manufacturers, HTAs and multi stakeholders networking organisations emphasised the importance of good collaboration and communication with a clear vision of the project. Testing feasibility early in the process was also a factor for success highlighted by the Pharmaceutical manufacturers.

Question 16: What key challenges were encountered in the project?
The respondents highlighted that data availability was a key challenge. This response was consistent across Pharmaceutical manufacturers, HTAs, Regulatory agencies and Academia.
Lack of clarity around legal aspects of big data projects were the key issue for 'other' stakeholder, though this was also raised by several Pharma and HTA respondents.
Stakeholders Breakdown [# of Responses]:

Stakeholder	Data sets not available	Data access difficult	Poor interoperability	Outcomes N/A	Methods difficult	Legal unclear
Pharmaceutical industry	6	5	4	3	3	3
Health care payer organisation	1	1	1	1		1
HTA organisation	3	4	2	1	1	3
Regulatory agency	3	2	1	1	1	1
User/patient group	1	1	1	2	2	1
Academia	4	4	3	3	1	1
Don't know/prefer not to answer						
Other, please specify	4	4	4	3	2	6
Total	**22**	**21**	**16**	**14**	**10**	**16**

Question 17: Please elaborate in a few sentences the key challenges.

Respondents cited a variety of key challenges that could be grouped into the following themes:

- Data issues regarding quality, harmonisation and linkage (widely cited by pharmaceutical manufacturers, Academia, HTAs)
- Contract issues (of particular concern of Academia)
- Motivation of data provider
- Sustainability of the project
- Agreements and governance/ data protection/ shared understanding and political consensus (Multi stakeholder networking organisation from 'other' stakeholder group)

Question 19: Did the project incorporate patient perspectives on big data?

About 60% of the respondents did not have patient perspectives incorporated in their project or did not know about it. It is particularly notable for most of the Pharmaceutical manufacturers (6 out of 7 respondents) did not include patient perspective in their project. A similar response proportion was observed for Academia.

Stakeholders Breakdown [# of Responses]:

Stakeholder	Yes	No	Don't know
Health care provider (e.g. hospital)		1	
Pharmaceutical industry	1	6	
Health care payer organisation	1		
HTA organisation	2	2	
Regulatory agency	2	1	
User/patient group	2		1
Academia	1	3	1
Don't know/prefer not to answer			
Other, please specify	3	3	
Total	**12**	**16**	**2**

Question 20: How was the patient perspective addressed?

For the respondents who incorporated patient perspectives in their project (n=12), this question was not clearly addressed and in reality patients did not seem to have been actively involved in the design of the proposal. For example, one HTA and Healthcare payor organisation only addressed patient perspective through ethics committees. Other respondents had involved patient groups/advocates on project advisory boards or steering committees or had included patient groups as key project partners.

Question 21: Was there a solution to sustainable provision of project results (such as e.g. continued access to outcomes data or infrastructure) after the end of the project, i.e. not dependent on the project funding and organization?

More than 50% of the Pharmaceutical manufacturers had no knowledge of a sustainable solution.

None of the Academia partners had a solution to sustain provision of project results after the end of the project. The *other* category of stakeholders seem better organised with respect to knowledge of sustainable solutions.

Stakeholders Breakdown [# of Responses]:

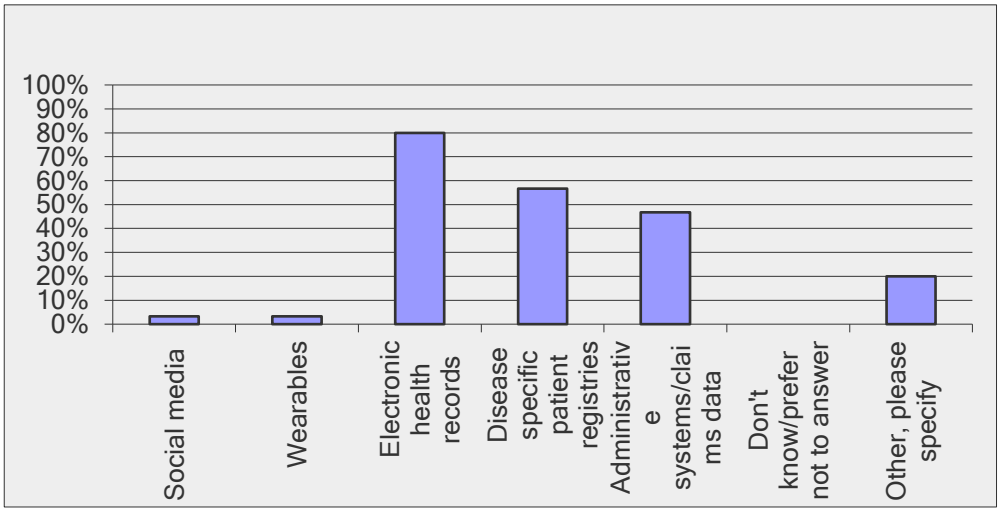
Stakeholder	Yes	No	Don't know
Health care provider (e.g. hospital)	1		
Pharmaceutical industry	3		4
Health care payer organisation	1		
HTA organisation	2	1	1
Regulatory agency			3
User/patient group	1	1	1
Academia		3	2
Don't know/prefer not to answer			
Other, please specify	5		1
Total	**13**	**5**	**12**

Question 22: How was the sustainable provision of project results organized?

The responses were very variable. Only 12 stakeholders responded to this question with no specific theme identified for ensuring the sustainable provision of project results.

Question 23: What categories of data sources were used in the project? Please tick one or more boxes

It is notable that Electronic health records were the main source of data used by stakeholders, followed by disease specific patient registries and administrative systems/ claims data, whereas social media and wearable technologies were rarely used. Stakeholders also used other sources of data for their projects including pharmacy data, published non-randomised studies and clinical trial data.



BD4BO Survey Analysis: Questions 24-28

Question 24: In your opinion, what are the three most important challenges that need to be addressed in the near future on a national and/or international level regarding big data in healthcare?

Respondents cited a broad range of challenges with key themes emerging around:

- data governance and privacy concerns (widely cited by patients, pharmaceutical manufacturers, HTA bodies, government institutions, academics, and associations, among others);
- data standards and quality;
- interoperability;
- access to and sharing of data (of particular concern to pharmaceutical companies and academics);
- capacity building (technical/infrastructure, human/skills); and
- agreement on validated outcome measures.

The need for political support and public investment also received multiple mentions.

Perhaps expectedly, the three regulatory agency respondents emphasized data quality, validity, and representativeness. By contrast, the six HTA/payer respondents raised a highly diverse range of challenges and did not coalesce around particular themes.

Respondent Breakdown [# of Responses]

Pharma [10]
Regulatory agencies [3]
HTAs/Payers [6]
HCPs [2]
Patient groups [3]
Academia [5]
International association [1]
Government institution [1]
Research database service [1]
Multi stakeholder networking organisation [1]
Open science collaborative with industry/government/academia/health systems all represented
Medical society [1]
Association of local authorities and regions [1]

Question 25: In your opinion, what are the three most important success factors that could benefit projects on big data in healthcare?

It was apparent that some respondents interpreted the question to ask about the benefits of using big data rather than factors contributing to the success of big data projects. Among those responsive to the latter, a range of factors were identified. As to be expected, these factors were broadly aligned with and responsive to the challenges identified in the answers to Question 24. Commonly cited success factors included multi-stakeholder collaboration and cooperation, trust/buy-in (especially with respect to addressing privacy concerns), data quality and reliability, data standards/interoperability, methodological standards, and the need to share data and results. Pharmaceutical manufacturers, in particular, emphasized the importance of multi-stakeholder collaboration (such as through public-private partnerships) and alignment on shared goals and outcomes. A multi-stakeholder networking organization highlighted the importance of taking an end-user approach, i.e., ascertaining the needs of those who will use data-derived evidence for decision-making purposes and curating data and analysis to meet these needs. Other noted factors included the use of open-source standards and the importance of political support and management support for HTAs and HCPs, respectively.

Respondent Breakdown [# of Responses]

Pharma [9]
Regulatory agencies [3]
HTAs/Payers [6]
HCPs [2]
Patient groups [2]
Academia [4]
International association [1]
Research database service [1]
Multi stakeholder networking organisation [1]
Open science collaborative with industry/government/academia/health systems all represented [1]
Medical society [1]
Association of local authorities and regions [1]

Question 26: In your opinion, who would be the most appropriate institution to address the challenges you

identified in the previous question? E.g. national authorities, European commission, OECD, EFPIA.
The European Commission (or an EU institution), national authorities, or both, were the most common responses to this question, gaining favour among at least some respondents in all stakeholder categories. A few respondents suggested an organization more global in scope (WHO) or with membership stretching beyond the EU's borders (OECD). Numerous respondents opined that no single institution can successfully address the broad and complex challenges and thus suggested a multi-stakeholder/consortia approach (e.g., public-private partnerships). These responses are not necessarily mutually exclusive, with the European Commission or a trusted, independent non-profit organization (either existing or established specifically for the purpose) possibly serving in a convening/secretariat role in this regard.

Respondent Breakdown [# of Responses]
Pharma [9]
Regulatory agencies [2]
HTAs/Payers [6]
HCPs [2]
Patient groups [3]
Academia [5]
International association [1]
Government institution [1]
Multi stakeholder networking organisation [1]
Open science collaborative with industry/government/academia/health systems all represented [1]
Medical society [1]
Association of local authorities and regions [1]

Question 27: Do you believe such initiatives should be done on a voluntary basis or on a binding legal basis?
Of those expressing a clear opinion, 11 responded "voluntary," 9 responded "legally binding," and several suggested a hybrid approach with both voluntary and binding aspects. Others offered conditional responses ("depends on the initiative"), indicated that the situation is more complicated than the question allows ("not a black and white question – cannot answer this way"), or responded that they did not understand the question. No clear pattern emerged within defined stakeholder groups, with split opinion in all groups. The absence of any definable pattern within the groups, along with numerous responses explicitly or implicitly indicating uncertainty or an inability to respond, suggests that the question lacked sufficient specificity, was subject to multiple interpretations (e.g., which aspects of the initiatives – participation? sharing data? confidentiality requirements?), and thus may not be a reliable source of perspectives on this topic.

Respondent Breakdown [# of Responses]
Pharma [11]
Regulatory agencies [2]
HTAs/Payers [5]
HCPs [2]
Patient groups [3]
Academia [4]
International association [1]
Government institution [1]
Multi stakeholder networking organisation [1]
Open science collaborative with industry/government/academia/health systems all represented [1]
Medical society [1]
Association of local authorities and regions [1]

Question 28: In your opinion, what role does big data have in developing value-based and outcomes-focused healthcare systems?
Pharmaceutical manufacturers were nearly unanimous in their view that big data are critical to the foundation for outcome-based healthcare systems, the research and development of precision medicines, and the rational use of medicines. Other stakeholders expressed broad agreement with this view, although several respondents (including some HTA bodies) offered somewhat more tempered endorsement, acknowledging the potential impact of big data in these areas while observing that the possibilities are yet to be realized and it is too soon to tell how impactful these data actually will be; a respondent from a medical society suggested that more work is needed to demonstrate that big data can generate accurate evidence before a proper assessment of utility and impact can be made. Both HTA bodies and regulatory agency respondents suggested that they view evidence derived from big data as a supplement to, and not a substitute for, RCT data – a point consistent with statements made by such stakeholders in other fora. One payer stated that the data are indeed critical but need to be developed and managed internally – implying that evidence derived from the payer's own data will be preferred over evidence generated from other data sources. Finally, it was noted that the data and its

applications should benefit the patient, i.e., patient-centered and capturing patient-relevant outcomes and not just used at the health system level.

Respondent Breakdown [# of Responses]
Pharma [10]
Regulatory agencies [1]
HTAs/Payers [6]
HCPs [2]
Patient groups [3]
Academia [5]
International association [1]
Government institution [1]
Research database service [1]
Multi stakeholder networking organisation [1]
Open science collaborative with industry/government/academia/health systems all represented [1]
Medical society [1]
Association of local authorities and regions [1]

Appendix 4: Template for semi-structured telephone interviews

Alternative 1: IF INTERVIEWEE HAS RESPONDED TO ONLINE QUESTIONNAIRE:
The interview will focus on the project xxx, that you also described in the online questionnaire.

1) **Data sets and outcome measures**

Could you please elaborate on data sets and outcomes measures? Did the project manage data? What data sets where used? Did your project focus on outcomes? What outcomes measures where used?

2) **Success factors**

In your response to the online questionnaire you wrote about success factors and...
What are the basis for success factors for successful application of results on big data in healthcare? What components are needed in a project to make this possible? It could be on a national or international level, and it could regard government level, companies, project teams etc.

2) Challenges

In your response, you also wrote about challenges in big data in healthcare…
In your opinion, what are the most important challenges that need to be addressed, either from your experience in a specific project or on a general level. It could be on a national or international level, and it could regard government level, companies, project teams etc.

3) Intellectual property, data privacy and governance

Did the project address intellectual property aspects? How did you manage IP rights, data access and governance? How was data security and data privacy ensured? How was data transfer and data sharing issues solved? *(Comment: In projects that involve several regions and/or countries different rules and ethics might apply when it comes to data sharing and transfer.*

4) Sustainable provision

Many projects on big data in health care are started, but it can be difficult to secure future access to results, outcomes data, knowledge and program infrastructure after a project is ended?
What is your general view on this? In an ideal world how could we ensure a sustainable provision of project results so that others can benefit from earlier work?

5) **Patient involvement** is often emphasised regarding big data.

What are your general views on that? Is there any experience in what type of patient organisation preferably should be involved?
Do you think there is a “best practice” on how to involve patients in a more structured way? How should that be done? Have you seen any examples, good or bad on this regarding big data?

6) Is there **anything else** you would like to highlight regarding big data in healthcare based on your experience?

Alternative 2: IF INTERVIEWEE HAS <u>NOT</u> RESPONDED TO ONLINE QUESTIONNAIRE:

1) Background

Could you please give some background information on the project xxx

a. General description
b. Purpose
c. What was your/your organisation’s role in the project.
d. Did the project manage data? What data sets where used? Did your project focus on outcomes? What outcomes measures where used?

2) Success factors

Could you please describe the success factors of the project? What are the basis for success factors for successful application of results on big data in healthcare? What components are needed in a project to make this possible? It could be on a national or international level, and it could regard government level, companies, project teams etc.

3) **Challenges**

Could you please describe the challenges of the project? In your opinion, what are the most important challenges that need to be addressed, either from you experience in a specific project or on a general level. It could be on a national or international level, and it could regard government level, companies, project teams etc.

4) **Intellectual property, data privacy and governance**

Did the project address intellectual property aspects? How did you manage IP rights, data access and governance? How was data security and data privacy ensured? How was data transfer and data sharing issues solved? *(Comment: In projects that involve several regions and/or countries different rules and ethics might apply when it comes to data sharing and transfer.*

5) **Sustainable provision**

Many projects on big data in health care are started, but it can be difficult to secure future access to results, outcomes data, knowledge and program infrastructure after a project is ended?
Was there a solution to sustainable provision of project results (such as e.g. continued access to outcomes data or infrastructure) after the project ended, i.e. not dependent on project funding and organisation) addressed in the project and if so, how?
In an ideal world how could we ensure a sustainable provision of project results so that others can benefit from earlier work?

6) **Patient involvement** is often emphasised regarding big data.

Was the patient perspective addressed in the project and if so, how? Is there any experience in what type of patient organisation preferably should be involved?
Do you think there is a “best practice” on how to involve patients in a more structured way? How should that be done? Have you seen any examples, good or bad on this regarding big data?

7) Is there **anything else** you would like to highlight regarding big data in healthcare based on your experience?

Appendix 5: List of interviewees

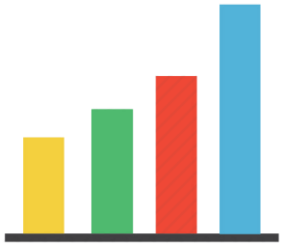
Clinical Trials Transformation Initiative (CTTI) - Pamela Tenaerts & Zachary Hallinan
European Medical Information Framework (EMIF) - Nigel Hughes
Health Data BE - Johan van Bussel
International Consortium for Health Outcomes Measurement (ICHOM) - Andre Dias
IMI GetReal project - Amr Makady (ZIN, Nederland)
Observational Health Data Sciences and Informatics (OHDSI) - Patrick Ryan
Sveus, Sweden - Charlotte Karbassi

Appendix 6: Data management and collection – points for reviewers to consider

This appendix provides a series of points that we recommend reviewers of BD4BO projects consider critically when reviewing proposals to ensure that sufficient attention has been considered by the complexities of data management in consortium working.

- Has the project specified that data will be collected prospectively (Foreground IP) or retrospectively (Background IP)?
- Prospective Data
    - Has the mechanism for collected data from patients in clinic or elsewhere been described?
    - Is their sufficient budget to set up the data processing logistics?
    - Has a technology solution that matches the systems proposed in the document been selected?
    - If no, is a meaningful justification why not given?
- Background Data
    - Is a clear definition in the proposal about how data controllers will be expected to provide data and how others will access the data controller's data?
    - Is there evidence that data controllers are comfortable with mechanism for accessing their data?
        - If IPD will be transferred to a third party is there explicit support that data controllers are committed to the approach (e.g. explicit MoU for each controller)
        - If IPD will be transferred to a third part is there sufficient resource allocated from each data controller to audit the 3rd party for compliance to their institutional standards of privacy, security and ethic and consent governance.
    - Is there sufficient resource allocated to data controllers to prepare, anonymise and transform their data ready for it to be used?
- Technical Platform
    - Has one of the technical platforms highlighted in this document been selected?
    - If not has suitable justification been given why not?
    - Has the justification included as how it will help outcomes research in Europe more generally rather than just in the context of the project?

Appendix 7: Branding and communication



Representation of data

Brighter outcome

Sky Blue	Indigo Blue	Mulberry Purple	Ruby Red	Gold Yellow

Big Data for Better Outcomes

BigData@Heart
Big Data for Better Hearts

ROADMAP
Real world Outcomes across the AD spectrum for better care

HARMONY
Healthcare Alliance for Resourceful Medicines Offensive against Neoplasms in Hematology

Appendix 8: Horizon 2020 communication strategy checklist and the IMI project communication guide

Horizon 2020 communication strategy checklist

1) Ensure good management

a) Have resources been allocated (time and money)?

A word of warning
When working with external professionals, the costs of hiring them need to be justifiable economically and in terms of effectiveness. Shopping around among several service providers can assure best value for money.
We are aware that from time to time participants in projects funded under the framework programmes are contacted — often by telephone — by organisations seeking **payment in return for publishing** information on the work being done in their projects. As with 'cold calling' in general, the claims and assertions made should be treated with appropriate caution before deciding on the best course of action. Contrary to some of the 'sales pitches' used, **these publications and their services have not been endorsed by the Commission**. Common tactics to secure business include vague references to high-level contributions from decision-makers, or making project participants believe that their activities have been singled out on account of special merit, which may not be the case.

- Does your proposal include a work package on communication?
- Have you prepared a communication strategy and timeline?
- Does the communication element of the project involve all consortium partners (and their respective staff, including researchers)?
- Is there awareness that communication is a continuous process, not a one-time effort when the project ends?
- Are you ready for the unexpected? Have you thought about how to respond effectively to such things as publication in high-ranking journals or a sudden new event related to the project's theme?

b) Are professional communicators involved?

- Have resources been allocated to professional assistance with the drafting of press releases, graphic design, maintenance of the website and other communication tasks? Larger institutions usually have an in- house capacity for this.
- Have you considered taking any training in the field of communication or including a communication expert in your team?

c) Is continuity ensured?

- Are there any arrangements to ensure that information will not be lost once the project comes to an end?
- Does the project provide for any feedback loops back to the European Commission that can help with amplifying the message, for example by notifying an event, or before publishing a press release?

2) Define your goals and objectives

a) Are there any goals and objectives?

- Have the final and intermediate communication aims of the project been specified, what impact is intended, what reaction or change is expected from the target audience? For example:
  - Receiving feedback or engaging in dialogue
  - Influencing the attitudes of decision-makers
  - Having people make a decision or take action
  - Ensuring that the project outcomes will be taken into production

b) Are your goals and objectives neither too ambitious nor too weak?

- Is there a deadline by which the goals should be achieved, taking into account different stages of the research and possible intermediary outcomes?
- Are the objectives specific and measurable, rather than vague? Does the project envisage ways of measuring its communication efforts and impact? For example:
  - Evidence of debates in the media
  - Evidence of new funders for your area
  - Evidence of transfer of research and innovation into practice (patents, prototypes, licenses)

- Number and turnover of new products, practices or procedures developed, based on your research outcomes
- Number of articles in the press
- Number of people asking for feedback or more information
- Number of references in scientific publications
- Participation in project events and seminars
- Speaker evaluations from conference presentations
- Survey of end-users
- Trends in website visits

3) Pick your audience

a) Is your audience well defined?

- Is each target audience a relatively homogenous group of people (not: 'the public at large' or 'all stakeholders')?
- Can the indicated audiences be further specified? For example: from 'the general public' to 'female citizens commuting by train to work in one of the EU-10 countries' or from 'decision- makers' to 'Euro-parliamentarians involved in the design of the new transport policy 2013'.

4) Does it include all relevant target groups?

- Can your audience help you reach your objectives?
  - Who has an interest in your research?
  - Who can contribute to your work?
  - Who would be interested in learning about the project's findings?
  - Who could or will be affected directly by the outcomes of the research?
  - Who are not directly involved, but could have influence elsewhere?
- Does the project aim to address both a direct audience and intermediaries to reach more people?
- What about the possibility of audiences at local, regional, national and European level?
- Is the audience external (not restricted to consortium partners)?

5) Choose your message

a) Is it news?

- Why do we need to know? What will change? What solutions are you offering? What makes the issue urgent? What are the consequences if no action is taken?

Tell a story, don't just list facts

A story is an effective way to make people remember your message. Why not tell one to disseminate your results?

Which stories work best? A good story consists of a succession of events with a beginning, a middle and an end, a scene setter and a plot, a climax and a conclusion, all of this in a rich context. It is hence more than a list of results achieved. A good story is one with which others can identify, with the project content as a basis, and focused on a person (for example: the researcher). Such stories also allow your message to be conveyed through shared values that will touch people's hearts and provoke emotion, and the promise of a better future.
You have forgotten how to tell a story? There are plenty of resources on the internet to help you. Just search for 'storytelling'.

- Have you tried to stir your audience's imagination and emotions?
- How does your work relate to everyday life? Does it link to any broader societal issue? Rather than focusing only on the provision of factual information, is your project research positioned within a broader socio-economic and policy context, so that it will be easier to explain the results and their relevance to policymakers and citizens?

b) Are you connecting to what your audience wants to know? See through your audience's eyes:

- What do they already know about the topic?
- What do they think about it?
- Do they need information and/or persuasion?
- Have you tested your message?

c) Are you connecting to your own communication objectives?

6) Use the right medium and means

a) Do they reach the audience?

- Are you working at the right level (local, regional, national)?
- Are you using dissemination partners and multipliers? Dissemination partners can help amplify and multiply a message. Rather than aiming to build an audience from scratch, the project should indicate which partners to use and how.

b) Do they go beyond the obvious?

- If input or contributions are needed, are there mechanisms in place to make communication interactive so as to obtain responses?
- Are you taking into account the different ways to communicate

Examples of interpersonal , two-way communication	**Examples of mass media, one-way communication**
- Dialogues, face-to-face conversation	- Newspapers and magazines
- Group discussions	- Press releases
- Conferences	- Newsletters
- Brokerage events	- Manuals
- School visits	- Brochures, booklets, flyers
- Tours	- Letters
- Round tables	- Radio
- Exhibitions	- Television
- Meetings	- Video
- Workshops	- Posters
Smaller audience, lower costs, more effort (more effect) Interactive, good for acquiring input Flexible (easy to change tone, strategy and content)	Potentially large audience Uses the credibility of the mass media

7) Evaluate your efforts

- Go back to your goals and objectives. Have they been reached? What lessons have you learn

In addition to taking into consideration the H2020 programme communication strategy checklist, the following structure is recommended for the communication plan:

Communication guide to IMI projects

I. Introduction

1. Executive Summary
   a. Situation Analysis
   b. Short Introduction of BD4BO and DO-IT
2. Mission, Vision, Values
3. Aim of dissemination/communication activities
4. Objectives

a. General
b. Specific
c. Operational

5. Principles of Dissemination
   a. Time
   b. Language
   c. Requirements
   d. Authorship
   e. Cooperation

II. Compulsory Elements

1. What a communication document should involve
2. Checklist
   a. Formal acknowledgements of IMI support
   b. Link to IMI website
   c. Link to the BD4BO website
   d. Logos
   e. Disclaimers

III. Messages

1. Type of content
2. Key messages

IV. Channels and Tools

1. External/Internal
2. Typology of Regular Deliverables

V. Target Groups

1. Mapping: identification, profile, prioritization
2. Classification: primary/secondary

VI. Resources

1. Budget
2. Responsibilities
   a. Inside Consortium
   b. Outside Stakeholders
   c. Project Governance
   d. Management and supervision
   e. Responsibility Matrix
   f. Organisation and Governance Structure Matrix

VII. Action Plan

1. Milestones and deliverables
2. Action Matrix
3. Timing

VIII. Risk assessment and analysis

1. Crisis communication plan
2. Monitoring

IX. Evaluation & Impact Assessment

1. Achievements/ project results
   a. for the research community
   b. for policy makers
   c. for public agencies
   d. for HTA bodies
   e. for the industry
   f. for patient organisations
   g. for professional organisations
   h. for payers
   i. for civil society and civilians
2. Measurement and evaluation of the communication activities
3. Key Foregrounds
4. Reporting on dissemination: publishable summary in periodical reports and final indicators
5. KPI

Appendix 9: Key Performance Indicators to Monitor Communication

Primary outcome Metrics – impact on key stakeholders

Activity	KPI	Output/Deliverables	Measurement (6 months)
Impact on Stakeholders	External communication with key stakeholders raises profile of BD4BO programme and value of big data	Core Stakeholder baseline survey (repeated at the end of year one and year two). Report on focus groups with key stakeholders to measure awareness in (M9-12) and (M 21-24) Pre-launch, then quarterly report including analysis of online conversation, sentiment analysis, key influencer, geographic analysis and impact in core stakeholder groups. [Budget dependent - €20K]	% (TBD) increase in awareness and support on baseline survey % Change in awareness and support in core stakeholder groups % Increase in online awareness and support % Increase in awareness and support for BD4BO in target countries

Communication output metrics - impact on key stakeholders

Activity	KPI	Output/Deliverables	Measurement (6 months)
Impact on Stakeholders	External communication with key stakeholders raises profile of the project	Establishing and maintaining outreach channels (website, social media etc...) Project brand Communications/Editorial Plan Articles, blogs collateral	Number of presentations of outputs at external events Number of times the project is quoted in the media (google statistics and press clipping) Attendance at dissemination events/workshops hosted by the project Number of articles published by the project Number of people reached by the articles published Altmetrics results of reports and papers Unique website visits Downloads of publically available project outputs (papers, toolkits etc) Total number of core-stakeholders contacted, broken down by key stakeholder group

Appendix 10: Key Performance Indicators to Monitor Project Implementation

A set of quality tests/ key performance indicators were needed in order to monitor the implementation of the project. The following list of KPIs was set to measure the overall project performance. The consortium agreed basing them on four key impact areas:

1. Effectiveness of CSA coordination and management
2. Impact on key stakeholders (academic, scientific, policymakers, BD4BO programmes)
3. Impact on IMI in terms of future planning
4. Sustainability of CSA activities

These KPIs will be measured quarterly and will be part of the regular internal report submissions, with the first measurement taking place during M3. Following that first measurement, final changes will be made to the KPIs. The baseline and any relevant targets for objectives will then be set after M6 KPI measurements.

Effectiveness of the CSA coordination areas

Activity	KPI	Output/Deliverables	Measurement (6 months)
Effectiveness of CSA coordination (focus area 4)	• Project Management and reporting tools are effective (i.e. project activities and deliverables on track)	• GANNT chart • Quarterly internal reports • Portal platform • Annual reports to IMI • Meetings	• % of quarterly reports uploaded to portal on time. • List of views of documents on SharePoint • % tasks running behind schedule • Number of requested revisions from IMI to reports submitted • Proportion of public organisations representatives at BD4BO meetings as a percentage of total invited participants. • Proportion of private organisations representatives at BD4BO meetings as a percentage of total invited participants • Frequency of and attendance at Coordinating team Committee Meetings
	• Internal communication within BD4BO effective (i.e. streamlined communication and knowledge transfer enhances and informs BD4BO activities)	• Internal communications mechanisms/guidelines • Internal communication monitored • BD4BO group meetings • Workshop for BD4BO program • Strategic guidance document • Newsletter • Repository	• Instances noted of BD4BO programmes or WPs not adhering to guidelines resulting in problems or complaints. • % of partners opening the internal newsletter + number of click throughs (separate figures for BD4BO and CSA engagement if possible) • Frequency of contact with vertical programmes (number of emails/calls/meeting attendance etc) • Vertical programme attendance at the BD4BO group meetings • Number of documents from vertical programmes uploaded to repository (measured 6 monthly). • Number of downloads of the documents shared on the repository. • Number of BD4BO partners taking part in the dissemination workshops • Number strategic guidance document downloads • Number of joint publications with the BD4BO projects

Impact on Key Stakeholders

- Academic stakeholders
- Scientific stakeholders
- Policy makers

Activity	KPI	Output/Deliverables	Measurement (6 months)
Impact on	• External	• Establishing	• Number of presentations of

Stakeholders	communication with key stakeholders raises profile of BD4BO programme and value of big data	and maintaining outreach channels (website, social media etc...) • Dissemination Workshops (WP2) • BD4BO Narrative • Communications/Editorial Plan • Policy briefs • Engagement with T2.2.4 • Toolkit • Advisory Board engagement (IAB and DPECs)	BD4BO outputs at external events • Number of times the programme is quoted in the media (google statistics and press clipping) • Attendance at dissemination events/workshops hosted by CSA • Number of articles and other non-scientific reports published by CSA WPs • Altmetrics results of reports and papers • Unique website visits • Downloads of publically available CSA outputs (papers, toolkits etc) • Social media metrics • Total number of stakeholders contacted, broken down by key stakeholder group • Focus groups/ Survey with key stakeholders to measure awareness in (M9-12) and (M 21-24) • IAB/DPEC attendance rates at meetings/response rates for requested inputs
	• Effectiveness of the CSA in developing documents	• ICF • Guidelines/Explanatory Information	• Number of access to the different downloads of ICF documents as well as guidelines / explanatory information •

Engagement of IMI with future priorities / new research areas proposed.

Activity	KPI	Output/Deliverables	Measurement (6 months)
New Scientific Priorities, unmet 'big data' needs (Focus Area 5)	• CSA effectively informs future IMI activities	• Overview of existing unmet big data needs • List of scientific priorities for IMI • Strategic guidance	• List of new scientific priorities for IMI2 acknowledged by relevant IMI committees or EFPIA • Number of projects launched as call topics • Number of ideas that enter topic development stage • IMI JU supports projects from list for new IMI projects

Sustainability of CSA activities

Activity	KPI	Output/Deliverables	Measurement (6 months)
Sustainability	• The CSA effectively reaches sustainability	• Repository • Sustainability planning across CSA	• Number of CSA deliverables accessible beyond project lifetime. • Evidence of continued engagement with CSA/BD4BO platforms built into future BD4BO activities/programmes/calls.