Big Data for
Better Outcomes

PRACTICAL AND METHODOLOGICAL **ROADMAP** TO REALISE THE MISSION AND VISION OF THE BIG DATA FOR BETTER OUTCOMES PROGRAMME

2018-04-27

Table of contents

Acronyms and abbreviations

BBMRI-ERIC	Biobanking and BioMolecular resources Research Infrastructure - European Research Infrastructure Consortium
BD4BO	Big Data for Better Outcomes
BigData@Heart	Big Data for Better Hearts
CDM	Common Data Model
EHDEN	European Health Data and Evidence Network (formerly EHDN)
EHR4CR	European Health Records for Clinical Research
ELIXIR	Intergovernmental organisation that brings together life science resources from across Europe.
EMA	European Medicines Agency
EMIF	European Medical Information Framework
EUnetHTA	European Network for Health Technology Assessment
GDPR	General Data Protection Regulation
HARMONY	Healthcare Alliance for Resourceful Medicines Offensive against Neoplasm in Hematology
HMA	Heads of Medicines Agencies
HTA	Health Technology Assessment
MHMD	My Health My Data
OECD	Organisation for Economic Co-operation and Development
PAES	Post-Authorisation Efficacy/Effectiveness Study
PASS	Post-Authorisation Safety Study
PIONEER	Prostate Cancer Diagnosis and Treatment Enhancement through the Power of Big Data in Europe
PRIME	EMA's Priority Medicines scheme
ROADMAP	Real World Outcomes Across the AD Spectrum for Better Care
RWD	Real World Data
RWE	Real World Evidence

Acknowledgement

This report was compiled by a team from the Swedish Institute for Health Economics (IHE), the London School of Economics and Political Science (LSE) and Imperial College London (ICL).

The report was reviewed by members of the IMI2 DO-IT consortium, including NICE, TLV, GSK, Bayer, TMF UCB and Pfizer as well as representatives of BD4BO projects ROADMAP, HARMONY, BigData@Heart and PIONEER. DO-IT international advisory board also provided comments.

Disclaimer

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.

1. Executive summary

- The Big Data for Better Outcomes (BD4BO) programme's **mission** is to maximise the potential of Big Data to transform healthcare across the entire pathway for every stakeholder, with the ultimate outcome of improved patient care and value. The programme is ambitious and seeks to gather health data for 100 million people over the next five years. In the mid-term, it aspires to become the "go-to" source for pan-European outcomes data and a trusted partner for patients, researchers, healthcare professionals, and decision-makers alike. It requires collaboration between different stakeholder groups.
- This roadmap is intended to provide **guidance and direction** to the continued development of the BD4BO programme. It identifies five key challenges and three opportunities for the programme, and seeks to outline potential activities and tactics to mitigate these risks and explore the opportunities. We have chosen to focus on what we believe to be the most important ones where BD4BO can make a difference. We have strived to be as forward looking as possible, therefore omitting many of the activities that are already underway within the programme.
- There is a lack of alignment on the **value of Big Data** among decision makers. BD4BO should develop evidence to guide decision makers. BD4BO should develop novel methods to help assess the value of new technology utilising big data and provide guidance on how these can be used to complement more traditional methods of evidence generation.
- A key objective of the BD4BO programme is to address the lack of **interoperability**. Using common data models (CDMs) BD4BO will make it possible to access data from multiple data sources. BD4BO should promote outcome measure standardisation and the inclusion of minimum baseline data to adjust for risk. BD4BO should also promote the use of the European Network for Health Technology Assessment's (EUnetHTA's) quality standards for registries and implement a data validation and quality management infrastructure.
- There is a lack of **data sharing culture** today. Much data is collected and regulated through monopolistic structures. BD4BO strives to promote better understanding between data custodians, users and patients who today is sceptical to sharing their health data in many parts of Europe. To foster trust, BD4BO should implement a governance structure where all key stakeholders are represented and have an equal say. BD4BO should also consider developing principles for ethical data sharing beyond the legal framework currently being developed through BD4BO informed consent templates. BD4BO also need to ramp up its communication efforts to demonstrate the value of big data, and showcase the technical and structural safeguards that are implemented to ensure safe and ethical reuse of data. Finally, BD4BO should implement a business model that is well aligned to data custodians and user's needs.
- Despite the ambitions of the General Data Protection Regulation (GDPR) to provide a harmonised legal framework for the protection of personal data, significant **legal barriers** remain for researchers who want access to health data. All current BD4BO projects have

work packages dedicated to addressing legal, ethical and data privacy issues. Moving forward, BD4BO should systematically collect evidence on remaining barriers to efficient, ethical and safe use of health data. Any insights relevant to the implementation of GDPR should be presented to European policy makers. Finally, BD4BO should establish a permanent body to coordinate requests for legal advice from BD4BO-projects, share best practice and monitor GDPR jurisprudence.

- As with many EU projects, the **sustainability** of results and systems represents a key challenge. All current BD4BO projects have established knowledge repositories to ensure that the results generated remain accessible after the project finishes. But this is not enough. To deliver on its mission and vision, BD4BO need to implement a business model to ensure that the BD4BO system, i.e. infrastructure, governance organisation and operations, becomes financially self-sustainable on a medium-long term. Moving forward, BD4BO should also streamline some of its infrastructure such as ethical review boards, advisory boards, data repositories and communications activities.
- There is an increased interest in the **opportunities** Big Data provides. Many key organisations are looking at how they can harness its potential. BD4BO should explore this interest and wherever possible seek to develop long-term relationships and, over time, refine its value proposition to these stakeholders.
- There is an increased demand for Big Data in the **regulatory setting**. BD4BO could support EMA's and HMA's joint Big Data Task Force around mapping of data sources, validation of the value of Big Data and providing open source data analysing tools. In return, the task force could provide BD4BO with important input on regulatory requirements, insights on future data needs generated through its early dialogue processes and help prioritise and align data requirements for regulatory and HTA decision making.
- There is also increased demand for RWE generated though Big Data in the **HTA setting**. BD4BO could support EUnetHTA's workstream on a lifecycle approach to improve evidence generation. Short term, BD4BO could support the use of quality standards for registries developed by EUnetHTA. Mid-term, BD4BO could potentially support EUnetHTA pilots through its use case studies.
- BD4BO should explore how it can support OECD's agenda on **improved efficiency & reduced waste** to achieve sustainable health care. Several of the areas discussed, around unwarranted practice variation and identification of low-value care could benefit from an outcome focus and evidence generation through Big Data solutions. BD4BO could contribute in this space.
- BD4BO should adopt a **collaborative** approach and liaise with other projects active in the area. New technologies are being explored which may help us to overcome some of the barriers identified in this paper. By keeping an open mind to new ideas, BD4BO should be able to realise our mission and vision faster than if we work in isolation and avoid duplication.

2. 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 drug discovery process, through clinical development and to the final use by providers and patients. 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 [1], Big Data in health care is described as:

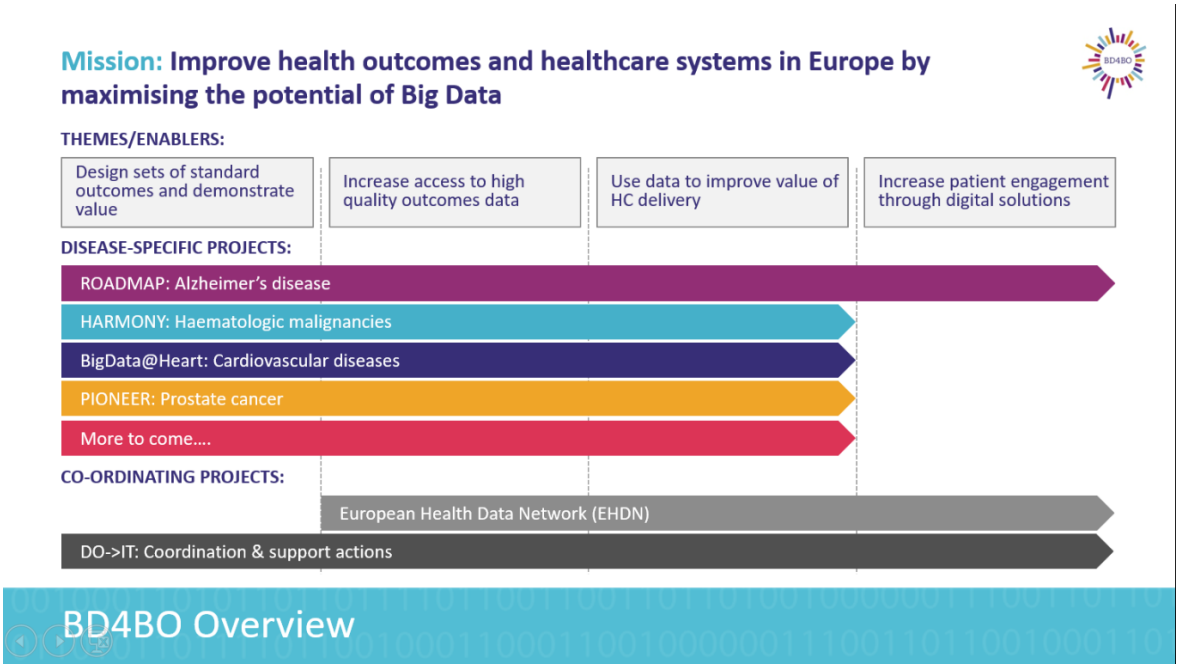
> *"Big Data in Health refers to large routinely or automatically collected datasets, which are electronically captured and stored. It is reusable in the sense of multipurpose data and comprises the fusion and connection of existing databases for the purpose of improving health and health system performance.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

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 two additional Vs are added; Veracity (trust of data) and Value (return on investment) [2].

The Big Data for Better Outcomes (BD4BO) programme, launched in 2016, is a part of the public-private Innovative Medicines Initiative-2 (IMI-2). Its mission is to maximise the potential of Big Data to transform healthcare across the entire pathway for every stakeholder, with the ultimate outcome of improved patient care and value. The programme is ambitious and seeks to link health data for 100 million people over the next five years through the use of common data models. In the mid-term, it aspires to become the "go-to" source for pan-European outcomes data and a trusted partner for patients, researchers, healthcare professionals, and decision-makers alike. Patients are recognised to be central in the undertaking. BD4BO provides a structure for empowering patients as data providers and partners in research and recognises the ultimate right of patients to their personal data. As of this writing, the programme has focused on cardiovascular diseases, haematological malignancies, Alzheimer's disease and prostate cancer.

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 defining the of overall programme strategy for BD4BO, managing and sharing knowledge from and between the different disease specific projects, acting as a central point of collaboration and communication across BD4BO projects, and communicating minimum

personal data protection standards as applicable for informed consent forms and supporting materials.



3. Aim

This roadmap is intended to provide guidance and direction to the continued development of the BD4BO programme. It identifies some key challenges and opportunities for the programme and seeks to outline potential activities to mitigate these risks and explore the opportunities. The identified challenges and opportunities are clearly not exhaustive. There are many more that must be overcome to realise BD4BO's mission and vision. We have chosen to focus on what we believe to be the most important ones where BD4BO can make a difference. We have strived to be as forward-looking as possible, therefore omitting many of the activities that are already underway within the programme.

The roadmap was developed without an exact time horizon for implementation. However, most of the recommendations should be possible to address within the next five years. Some of the recommendations, around implementation of a business model and alignment of BD4BO bodies and infrastructure may have to wait until the early launched disease specific projects concludes.

4. Method

This roadmap is largely based on a grey literature review, a questionnaire and structured interviews with key stakeholders which were originally conducted for the development of BD4BO's interim strategic guidance document. The roadmap has been developed by a team of DO-IT partners from both public and private sector. Representatives across all BD4BO projects were given the opportunity to provide input. Finally, it was reviewed by DO-IT's international advisory board,

which include representatives of all key stakeholder groups (patients, regulators, HTA-bodies, payers, industry, government).

5. Key Challenges

There are many challenges ahead to fully realise the BD4BO mission and vision. In the BD4BO's interim strategic guidance document we identified eight barriers and success factors for Big Data projects. In this roadmap we've chosen to focus on five challenges: the lack of a uniform understanding of the value of Big Data; lack of interoperability; lack of data sharing culture; legal barriers and privacy issues; sustainability of results and systems. These are the ones which we consider to be the greatest barriers for the programme to deliver on its mission and where BD4BO can best effectuate change.

5.1. Value of Big Data

Real world evidence (RWE) generated from Big Data gets increased attention for the benefit it can provide to researchers, patients, industry and decision makers. Researchers want to mine large data sets, combining observational data from multiple sources and omics to find new avenues to prevent, treat and cure disease. Patients play an increasingly important role in Europe's health care system, as both data providers and users. Informed, empowered and engaged, patients have a key role to play to manage their health and help maintain universal health coverage sustainably. Industry want to tap the potential of Big Data to inform their research & development strategies and use RWE on health outcomes to underpin new market access models, allowing new, stratified medicines to reach the market faster and supporting new approaches to risk sharing between the private and public sector through advanced managed entry agreements. RWE is central to enable a shift towards pay-for-performance models.

Among decision makers the views vary regarding the potential of Big Data. Figure 1 shows an image from a paper examining the current use of RWE to support pricing and reimbursement decision making across Europe and attitudes towards the future potential for RWE [3]. This illustrates the need for better understanding and alignment on of the value of Big Data among decision makers.

BD4BO's research can contribute to this through validating and demonstrating the value of Big Data. Novel methods and technologies as well as existing analytical techniques from other disciplines should be explored and employed for Big Data analysis to show the potential of Big Data to complement randomised control trials as well as for undertaking pragmatic trials. This will

help realise the potential of Big Data and improve the acceptability of RWE in decision making.

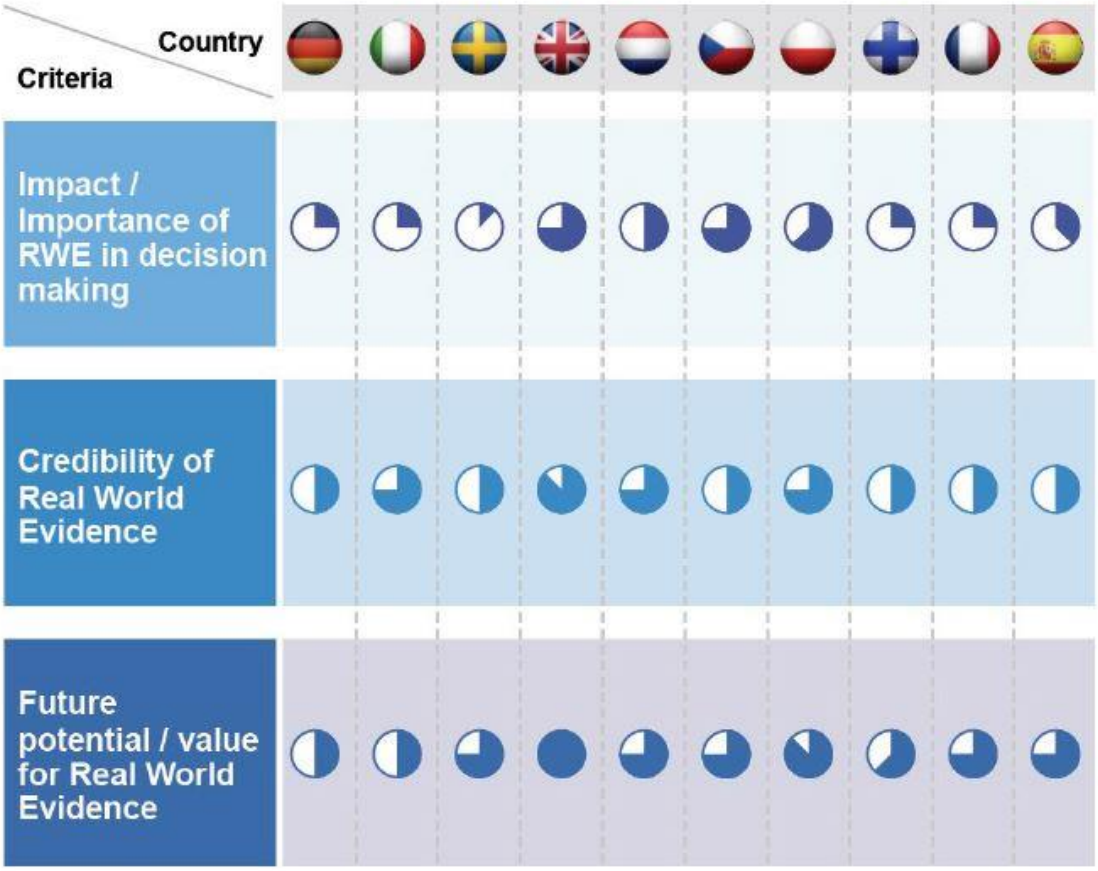


Figure 1: Opinions of RWE use across ten European countries. Gill, J.L, et. al. 2016 [3]

Machine learning can contribute to the realisation of Big Data's potential by allowing precise and reliable prediction using data with very large numbers of variables. It can also create predictive classification models and test their reliability and accuracy without needing data-generative models or coefficient estimation, and by enabling causal inference without data being experimental [4]. Using causal discovery algorithms, discussed further elsewhere, within a machine learning framework can help identify strategies for preventative intervention and can be used to better understand, predict and ultimately improve health outcomes. Using new technologies such as machine learning can help utilise Big Data efficiently and is an area where BD4BO can show if predictive classification models can be applied to the disease areas focused on by BD4BO programmes [5].

Big data can be used both to identify causal associations and for pragmatic trials. As data can be reused for multiple purposes, big data can be used to complement traditional randomised controlled trials (RCT) as well to explore efficiency and modifiers. Methods such as regression discontinuity design, interrupted time series analysis and propensity score matching have been validated through comparison with RCT results [6-9]. Confounding is a problem in non-randomised data and techniques to deal with it vary depending on whether it is observable or not. Techniques to

overcome confounding can be applied to the study design and data analysis. Regarding the study design, the study cohort can be restricted to make the analytical sample more homogenous. This can also aid to overcoming other issues such as spurious correlation due to large sample size. The sample can be matched based on common confounders such as age, gender and ethnicity. Once matched variables have an equal distribution in both the treatment and control groups, they cannot have an effect as confounders. Most observational data do not contain variables to measure and adjust for all potential confounding factors. As it is not possible to check for residual confounding by unobserved factors, instrumental variable estimation can be used to provide unbiased estimates of causal effects in non-randomised studies by mimicking random assignment of patients into groups of different likelihood for treatment [10].

The utilisation of Big Data in health care could be encouraged by validation studies using causal methods on Big Data to confirm findings from RCTs. Techniques such as regression discontinuity design, interrupted time series analysis, difference-in-difference, Mendelian randomisation could be show cased to convey their potential in this field. This could help overcome issues typical of traditional RCT such as the lack of external comparator and, the ecological validity and intrinsically high costs. Big Data also allows for sub group analysis and the effects of modifiers to be explored. Techniques such as multilevel modelling can be used to overcome clustering (such as within hospitals geographic areas) in Big Data. Utilising these methods in Big Data has the capacity to alleviate some of the issues with conducting RCT such as budget constraints, unfeasibility due to small patient populations in rare diseases and limited evidence of efficacy in wider populations. BD4BO should undertake research using these methods, showing how issues using Big Data can be overcome and showing its potential to decision makers.

RCTs are set to remain gold standard, but efficiency, utility of information generated (PROs) and generalisability could be enhanced through pragmatic trial approaches which big data could be used for. BD4BO can show the value of Big Data in this area through undertaking effectiveness studies, disseminating their findings and showcasing how these can be useful for decision makers. RCTs are limited by a lack of generalizability. Results from pragmatic trials are particularly useful for informing decision making in routine practice and to understand how the treatment is associated with outcomes in a patient population which may be broader (such as having comorbidities) than that included in the RCT and where clinician and patient behaviour (such as compliance) may impact on treatment [11]. Using pragmatic trials in combination with RCTs adds to our understanding of illness and treatment by using a variety of perspectives and methodologies to study it [12]. External validity may be greater in pragmatic trials but internal validity may be weaker.

5.2. Lack of Interoperability

Interoperability is essential to realise the full value of Big Data. Consistent use of "core outcome sets" is key for comparison of interventions, comparison across studies and for the linking of data sources. Core outcome sets are sets of outcomes agreed between stakeholders on what should be measured and reported in all trials in a specific area as a minimum. If these are routinely collected

from various sources, it allows outcomes to be consistently monitored across data sources, across countries and over time. Most core outcome sets do not specify how to measure the outcomes in question. The health condition in question may influence how transferable outcomes are across different data sources and geographies. For instance, for mortality or cause-specific causality as an outcome, the international classification of diseases (ICD-10) is widely used to classify causes of death. For other outcomes, common use of diagnostic tools such as magnetic resonance imaging for strokes can contribute to consistency across geographic settings and data sources. On the other hand, quality of life and other patient reported outcomes are difficult to measure consistently across settings. There is a potential for machine learning to contribute to standardising measures across datasets and reduce how labour intensive this process is [13], however it is still dependent on how substantively interchangeable the measures in question are.

In addition to variation in outcomes and their measurement across data sets, there is variation in the collection of additional information such as demographics, health behaviours and co-morbidities. Measures such as these are important in pragmatic trials to adjust for confounding in a way that may not be necessary in traditional RCTs. The inclusion of minimum baseline data to adjust for risk (as advocated by International Consortium for Health Outcomes Measurement (ICHOM) standard sets) in addition to core outcome sets may help promote interoperability. Agreeing on measurement is also factor for minimum datasets. Measurement may vary across data sources depending on the practice setting and geographic setting data are captured in.

Developing quality standards for Big Data could be a means of improving interoperability. The principles being applied to quality standards for registries by EUnetHTA could be adapted and developed for other routinely collected data comprising Big Data in healthcare. Shared data quality standards would improve interoperability as well data quality through effective governance, consistency in data entry and identification of coding errors. BD4BO should promote the use of core outcome sets and quality standards in routine data sources to allow for interoperability.

5.3. Lack of data sharing culture

A key challenge for the BD4BO programme is to build trusted and sustained relationships between all stakeholders, to achieve a wide societal acceptance for the concept of ethical data sharing to advance science and improve health outcomes. Today, there is a lack of a data sharing culture. Much data is collected and regulated through monopolistic structures. In addition, citizens in many parts of Europe are still unwilling to share their health data for research purposes. In 2016, only one in five Europeans would be willing to give anonymised data to public authorities or public-sector companies for medical research purposes (21%), and even fewer would give data to private sector companies for medical research purposes (14%). Only one in twenty would be willing to give their anonymised data to private sector companies for commercial purposes. Almost one quarter would not be willing to give access to their personal health and wellbeing data under any circumstances [14].

Looking forward, there are several opportunities that BD4BO should explore to help improve the data sharing culture:

Stakeholders involvement is a key factor for the legitimacy and implementation of data sharing systems. The BD4BO programme's multi-stakeholder approach should be further explored. Beyond the technical and legal solutions and standardisation of data, there are several dimensions needed for the creation of a fruitful data sharing environment. BD4BO should consider developing principles for ethical data sharing to help foster trust, providing guidance beyond the legal framework currently being developed through the informed consent templates.

All key stakeholders should be represented and have an equal say in the governance structure of BD4BO data network.

The European Health Data and Evidence Network (EHDEN) is expected to establish application domains and to develop supporting infrastructure, services and analytics in key areas such as health service efficiency and individual patient care [15]. This should help further the understanding between data custodians and data users to find pragmatic solutions to improve access to the evidence needed to support expanding value-based pricing and outcomes-focused healthcare delivery in Europe.

An upfront agreed- upon and transparent business model is needed for an ideal data sharing environment [16]. The EHDEN should develop its business model taking into account the multiple interests of different actors, not the least data holders. Hence, data should be considered both an asset and a currency. Several value propositions for the data access may co-exist: 1) providing access to data with the aim to improve health care efficiency 2) enlarging access to data that may lead to publications with higher impact; 3) financial compensation for data and expertise; 4) speed in analyzing data using standardized scripts; 5) creation of capacity by rotation of roles and participation in study teams and ability for peer consultation. Through clever design an ecosystem should be created that caters to various needs.

Stakeholders involvement and willingness to participate in data sharing hinges on their ability to recognize the value of Big Data and collaboration and that citizens have confidence that their integrity is ensured. BD4BO should therefore increase its efforts to communicate about the value of Big Data, being transparent with results generated through data sharing and the various measures it implements to ensure ethical data sharing and the safeguards put in place to protect the integrity of individuals, in particular around goverance structure and technical security measures. This action will interlink with the progress made in the other areas analysed in this paper.

5.4. Legal barriers and privacy issues

The questionnaires conducted by DO-IT have shown, that legal uncertainty especially regarding data protection requirements are one of the most important factors in hindering data sharing. The legal fragmentation within and across Europe, expected to remain even after May 2018 when GDPR comes into force, and the resulting lack of clarity on data privacy issues is a central challenge facing Big Data initiatives. In particular, derogation clauses in the GDPR and resulting national particularities regarding data retention periods, restrictions on the secondary processing of data without consent and restrictions on linking data were key legal issues identified in the current

Big Data landscape. All current BD4BO projects have work packages dedicated to address legal, ethical and data privacy issues. Though the approach and activities vary somewhat between BD4BO projects over short-medium term, BD4BO envisages an alignment of practices and structures over time:

- A European-wide health data network built around a federated data model where data access is based on local contracts and ethical review
- A uniform consent management approach, drawing on experience from other Big Data related projects and infrastructures such as BBMRI-ERIC, ELIXIR, EMIF and EHR4CR.
- Integration of best practices in data de-identification to protect patient data privacy across disease areas.

The Legal Think Tank set up between the disease projects and DO-IT should be formalised under DO-IT to develop BD4BO positions on common data protection topics and foster the harmonisation process across the BD4BO programme. This is especially important to align already ongoing activities and to avoid a duplication of work and a fragmentation of results. The Think Tank should also consider whether anonymization approaches could be a possible solution to enable clinical data sharing where the demands of GDPR cannot be met and consider the inclusion of dynamic consent strategies into the programme.

Whereas the EHDEN and the disease specific programmes should remain responsible for providing practical solutions to manage legal and ethical barriers for data sharing, -their experience should be systematically analysed within DO-IT, discussed with the legal/ethical work packages of the respective projects and turned into common standpoints and solutions to support the development of best practice guidance across the programme as well as provide important feedback to European policy makers. In the long term it may also generate recommendations to policy makers and other decision makers to support the continued development of EU Big Data eco-system. Some of the foundations for this work is already in place. For instance, 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. This work will further feed the Legal Think Tank discussion and thus lead to aligned approaches, providing support and advice for future BD4BO projects.

5.5. Sustainability of results and systems

The sustainability of results and systems to harness the value of Big Data arising from the BD4BO project represents a key challenge. While all current BD4BO projects have established knowledge repositories to ensure that the knowledge generated through the projects remains accessible after the project finishes, it is often a challenge to find someone willing to take on the maintenance of the repository long- term. But sustainability of results is not enough if we are to realise the mission and vision of BD4BO programme. Without ensuring the long-term access to the BD4BO network data and its infrastructure, there is little hope that project results will be adopted in health system practice. For instance, European regulators and HTA-bodies are unlikely to want to build their real world data (RWD) access plans on a data provider which may be gone in five years. It is crucial therefore to ensure that the BD4BO system, i.e. infrastructure, governance organisation and operations becomes financially self-sustainable on a medium-long term. This has been recognised in the EHDEN call.

This is a challenging task. The business model should be built based on insights around the key drivers & barriers for data sharing and an understanding of the value of the services offered to a wide variety of potential customers. See section 5.3 for more detail on this. Though challenging, there are examples of EU-projects who have successfully transitioned to a self-sustainable funding model. The European lead factory for instance adopted a model based on fee for service with several payment options to cater to different customer preferences [17].

As the number of projects under the BD4BO increases, some key stakeholders may find it burdensome to engage with each project individually. Rather than having multiple ethical review boards, advisory boards etc, these should be developed into common resources, to streamline work flow, reduce duplication and improve programme efficiency. This should also be true for other infrastructure such as data repositories and communications activities. Moving forward, we recommend that future disease specific projects are advised to use common resources where they exist, rather than developing their own work packages.

6 Key opportunities

As outlined so far, there are many challenges on our road ahead. But there are also significant opportunities. The greatest opportunity is possibly the increasing number of stakeholders who see the potential of Big Data. Many have stated that “Data is the new Oil”, e.g. the Economist in May last year [18]. Many key organisations are increasingly looking at how they can harness this potential. BD4BO should explore this interest and wherever possible seek to develop long-term relationships and, over time, refine its value proposition to these stakeholders.

6.1 Increased demand for Big Data in the regulatory setting

Personalised medicines are changing the regulatory pathways in Europe. Though the mission of European regulatory authorities remains the same, i.e. to ensure patient safety and assess the risk benefit of medicines, the processes to generate the knowledge needed to answer questions around safety, quality, efficacy and relative effectiveness are evolving. Several tools have been developed

over the years to support continued knowledge generation after authorisation approval (PASS, PAES).

Big Data has the potential to further add to how the risk-benefit balance of medicinal products is assessed. In the US, there is an increased recognition that there is a need for new ways to collect and utilise patient data [19]. In Europe, EMA and ten national regulatory agencies established The Big Data Task Force in March 2017. One of their key deliverables is to develop their own roadmap for Big Data. There is a significant opportunity for a mutually beneficial dialogue between BD4BO and the Task Force. As the BD4BO programme matures and the catalogue of data sources available through the BD4BO network grows, it should be able to provide input to the Task Force in several key areas:

- Mapping of data sources in several key disease areas
- Validating the usefulness of Big Data for regulatory decision-making (through pilots/use case studies)
- Providing an open source data analysing tools and systems needed to handle Big Data (including quality management and data governance)
- A network of expertise that the regulatory agencies could consult around specific issues.

In return, the Task Force could provide BD4BO important input on:

- Regulatory requirements through its recommendations
- Insights on future data needs generated thorough early dialogue processes (such as PRIME)
- Aligning and prioritising data requirements for regulatory and HTA-decision making

6.2 Increased demand for Big Data to support HTA-decision making

Demand for RWE generated through Big Data is on the rise also among HTA bodies and payers. Acceptability of RWE to HTA bodies varies across HTA bodies and depending on the context of its use. HTAs do not use RWE in most of cases, but it has the potential to confirm the generalizability of RCTs, to extend the findings to different outcome measures more relevant to HTA, to justify utility estimates and to confirm the durability of results. Improving quality standards in RWD to make it a source of more acceptable evidence and show casing how RWD can be used effectively are two ways of to increase the demand for RWE to support HTA decision making. This could also help alleviate concerns over the validity of data from other countries and HTA studies carried out by other HTA-bodies. Concerns exist regarding non-randomisation, incomplete and missing data, biases undermining quality in RWD and variation in this quality. However, the characteristics of RWD mean that it can support HTA decision making in a way which RCTs cannot. RWD can be utilised to support evidence in rare diseases, to show the effects of an intervention in the wider patient population, whether there is variation in the treatment's

effectiveness, long- term outcomes and cost effectiveness. RWE can also provide insights into comparative effectiveness.

RWE could provide HTA with evidence regarding rare diseases if RCT are not possible or to provide evidence where there has been conditional licencing. As effect estimates would be necessary in such cases, high quality complete data would be required and the utilisation of causal methods. This could be important where there are too few individuals in a single country.

Where efficacy has been established by RCT, RWE can provide knowledge on the long-term safety and effectiveness of recently introduced treatments, and on factors influencing the level of effectiveness. RWE conveys information on long-term clinical benefits and rare adverse events while additionally providing insights into health-care resource utilization and the impact of a given intervention on complementary outcomes such as treatment satisfaction.

RWE can be used to support HTA decision-making: when traditional trials do not show a well-established efficacy/safety ratio, RWE can provide a means of assessing long term safety outcomes and actively monitoring adverse events. Ultimately, it should be possible to track all patients that receive a specific drug or treatment.

RWE can be used for comparative effectiveness research and assess the benefits or harms of alternative interventions on outcomes in routine clinical practice. Biases can be identified and adjusted for. Comparative effectiveness studies can show the most efficient procedures to optimise outcomes and ensure healthcare is value based.

BD4BO should explore how it can support HTA-bodies and payers both at a national and European level. For instance, BD4BO could support EUnetHTA's workstream on a lifecycle approach to improve evidence generation. Short term, BD4BO could support the use of quality standards for registries developed by EUnetHTA. Mid-term, BD4BO could potentially support EUnetHTA pilots through its use case studies.

6.3 Focus on improved efficiency & reduced waste to achieve sustainable health care

It is increasingly recognised that not all activities in the health care sector bring value to patients. A recent OECD report [20] distinguished between three types of wasteful spending: (i) wasteful clinical care, (ii) operational waste, and (iii) governance related waste.

Wasteful clinical care is defined as care that either harms patients or fails to deliver benefits. According to the OECD report it is estimated that more than 10% of hospital expenditure is related to the correction of preventable medical harm. Care that does not deliver benefits include unnecessary diagnostic testing, for example the use of imaging services for low back pain, and procedures for which there is little evidence of benefits either in general or for specific patient groups, for example the inappropriate use of antimicrobials.

Operational waste covers instances where the same level of effectiveness of care could be reached with fewer resources. The prime example is the use of branded medicines where generic

alternatives exist. Inappropriate admissions to hospital, for example in the case of hospital so-called ambulatory care sensitive conditions that are assumed to be preventable through increased primary care provision, is another example of waste arising from the operation of the health care sector.

Governance related waste relates to unnecessary administrative procedures and fraudulent behaviour. Large variation in administrative costs across countries suggest the scope for reducing administrative costs for example through digitalisation of administrative procedures and prescriptions. It is difficult to assess the level of fraudulent behaviour across countries, but one third of citizens in the OECD countries believe that the health sector is corrupt or very corrupt.

The OECD report identified better information on the extent and causes of wasteful spending as being of key importance in addressing the problem. BD4BO thus has a considerable opportunity to contribute to the identification and reduction in wasteful spending through the collection and analysis of Big Data.

In general, the collection of data on practice and outcome data facilitates between and within country comparisons that can be used to identify unwarranted variation in treatment, spending and outcomes. BD4BO can thus potentially play a key role in identifying causes and consequence of wasteful spending which can be addressed in the three areas identified as drivers of wasteful spending by the OECD.

In low value care, specifically, the collection and analysis of Big Data through BD4BO, can be used to identify providers with unusually high rates of harm with the intention of addressing these. In areas where benefits of treatment vary between patient groups, Big Data can be used to identify the patients most likely to benefit from treatment and prioritise treatment to these groups. Likewise, the development of better information sharing systems could prevent the repetition of diagnostic testing where this occurs due to lack of information sharing. Health Technology Assessments, an area where analysis of Big Data has already been identified to potentially contribute (see section 6.2), has been identified also by the OECD as potentially playing a key role in evaluating the value of new and existing treatments.

7. Conclusions

We live in many ways in an era of exploration. To quote Herbert Hoover, 31st President of the United States, *"New discoveries in science will continue to create a thousand new frontiers for those who still would adventure."* Big Data and the digitalisation of our society is a new frontier and the BD4BO-programme's journey has just begun. This roadmap help should guide and focus the BD4BO efforts over the coming years.

BD4BO is not alone. There are many other projects currently ongoing looking to explore Big Data in the field of health. Some focus more on the technical aspects of building an efficient infrastructure to manage and safeguard the increasing amount of health data and make it accessible for researchers, thereby maximising the value of publicly funding research (ELIXIR). Others focus on new technologies such as blockchain, new encryption solutions and smart contracting to create

an information market place where citizens are the ultimate owners of their health data (MHMD). Several more are being set up by the European Commission (DG RTD, DG CONNECT and DG SANTE) around HTA, AI, evidence-based healthcare, personalised medicines and more. Exploration is about embracing the unknown and realising that the journey has its own reward. Time will tell which of these will flourish and likely the solution will be an amalgam of several projects. BD4BO need to keep track of these projects and engage when necessary to stay at the forefront of health data digitalisation.

In the meantime, many awaits the outcome of BD4BO's journey with high expectations. Life-science is currently developing fast with many new treatments coming to market every year. Drugs are also being used in increasingly complex combinations (e.g. in oncology). To be able to accurately measure outcomes in a real-world setting should help improve the lives for patients, help researchers push the boundary of science to find the next generation of treatments in areas of unmet medical need, and help decision makers steer resources and maximise the impact of our health systems.

With a collaborative approach, working with other projects active in digital health, should we be able to realise our mission and vision faster than if we work in isolation.

Big Data for Better Outcomes

Mission	Improve health outcomes and healthcare systems in Europe by maximising the potential of Big Data				
Challenges	**Value of Big Data**	**Interoperability**	**Data sharing culture**	**Legal barriers & privacy issues**	**Sustainability**
Recommendations BD4BO Activities	Development of new methods to assess the value of new technology utilising big data.	Promote outcome measure standardisation (outcomes selection tool kit) and the inclusion of minimum baseline data to adjust for risk. (EHDEN)	Develop and implement principles for ethical data sharing, beyond legal requirements (EHDEN).	Systematically gather evidence on barriers to an efficient, ethical and safe use of health data.	Develop and Implement a business model for EHDEN to make BD4BO self-sustainable over medium-long term. (EHDEN)
	Provide guidance when different methods should be utilised.	Promote the use of quality standards for registries developed by EUnetHTA.	Implement a multi-stakeholder governance organisation for the BD4BO data networks. (EHDEN)	Provide feedback to policy makers on GDPR implementation.	Alignment of BD4BO infrastructure and processes.
		Implement data validation and quality management infrastructure (EHDEN, Disease Projects)	Develop of a business model based on data custodians' and data users' needs. (EHDEN)	Provide permanent body to coordinate requests for legal advise and monitor GDPR jurisprudence.	
			Increased comms on the value of big data and BD4BO security measures (governance & technical). (EHDEN, Disease Projects)		

Opportunities	**Big Data in the regulatory setting**	**Big Data in HTA-decision making**	**Sustainable health care (improved efficiency & reduced waste)**
Recommendations BD4BO Activities	Establish strategic partnerships with key stakeholders , including EUnetHTA, EMA and OECD. Explore opportunities for collaboration. Define and refine BD4BO value proposition to key customers over time.		
	Support EMA Big Data Task Force with mapping of data sources, validation of value of big data, providing open source data analysing tools to handle big data. (EHDEN, DP)	Promote the use of quality standards for registries developed by EUnetHTA. Support EUnetHTA pilot programme on mid-term through use case studies. (EHDEN, Disease Projects)	

BD4BO Roadmap recommendations

1

8. References

1. European-Commission, *Study on Big Data in Public Health,Telemedicine and Healthcare.* 2016.
2. ABPI, *Big data road map.* http://www.abpi.org.uk/about-us/resources/publications-library/big-data-road-map, 2013.
3. Gill, J., et al., *The use of Real World Evidence in the European context: An analysis of key expert opinion.* 2016.
4. Saxe, G.N., et al., *Machine learning methods to predict child posttraumatic stress: a proof of concept study.* BMC psychiatry, 2017. **17**(1): p. 223.
5. Malinsky, D. and D. Danks, *Causal discovery algorithms: A practical guide.* Philosophy Compass, 2018. **13**(1): p. e12470.
6. Maas, I.L., et al., *The regression discontinuity design showed to be a valid alternative to a randomized controlled trial for estimating treatment effects.* Journal of clinical epidemiology, 2017. **82**: p. 94-102.
7. van Leeuwen, N., et al., *Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes.* Epidemiology, 2016. **27**(4): p. 503-511.
8. Fretheim, A., et al., *Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result.* Journal of clinical epidemiology, 2013. **66**(8): p. 883-887.
9. Mayne, S.L., B.K. Lee, and A.H. Auchincloss, *Evaluating propensity score methods in a quasi-experimental study of the impact of menu-labeling.* PloS one, 2015. **10**(12): p. e0144962.
10. Takač, M.J.-M. *ENCePP Guide on Methodological Standards in Pharmacoepidemiology*. in *3rd PharmSciFair, Pharmaceutical Sciences for the Future of Medicines*. 2011.
11. Berger, M.L., et al., *Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report.* Value in Health, 2012. **15**(2): p. 217-230.
12. Berger, M.L., et al., *Good practices for real‐world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR‐ISPE Special Task Force on real‐world evidence in health care decision making.* Value in Health, 2017. **20**(8): p. 1003-1008.
13. Rajkomar, A., et al., *Scalable and accurate deep learning for electronic health records.* arXiv preprint arXiv:1801.07860, 2018.
14. Eurobarometer, *Attitudes towards the impact of digitisation and automation on daily life.* 2017:460, 2017.
15. Innovative-Medicines-Initiative-2, *12th call for proposals* http://ec.europa.eu/research/participants/data/ref/h2020/other/call_fiches/jtis/h2020-call-fiche17-12-imi2-ju_en.pdf 2017. **ref: IMI2/INT/2017-02169**.
16. Innovative-Medicines-Initiative-2, *Topic: European Health Data Network (EHDN).* https://www.imi.europa.eu/sites/default/files/archive/uploads/documents/Future_Topics/IndicativeTopic_EHDN.pdf 2017. **Research-and-Innovation-action**
17. European-Lead-Factory. *European Lead Factory- Concept*. 2017 [cited https://www.europeanleadfactory.eu/about/concept. .

18. The-Economist, *The world's most valuable resource is no longer oil, but data*. 2017, 6 May https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.
19. Brennan, Z., *FDA's Janet Woodcock: the clinical trials system is 'broken'*, in *Endpoints News*. 2017: https://endpts.com/fdas-janet-woodcock-the-clinical-trials-system-is-broken/.
20. OECD, *Tackling Wasteful Spending on Health Highlights revised.* OECD report, 2017. **https://www.oecd.org/els/health-systems/Tackling-Wasteful-Spending-on-Health-Highlights-revised.pdf**