Study on Big Data in Public Health, Telemedicine and Healthcare

Final Report

Written by Gesundheit Österreich Forschungs- und Planungs GmbH
Submitted: 4.11.2016

Gesundheit Österreich
Forschungs- und Planungs GmbH

SOGETI

Study on Big Data in Public Health, Telemedicine and Healthcare

Final Report

Authors

Claudia Habl (GÖ FP)
Anna-Theresa Renner (GÖ FP)
Julia Bobek (GÖ FP)
Anja Laschkolnig (GÖ FP)

Supported by:

Laurent Jacquet (SOGETI)
Katharina Habimana (GÖ FP)

Project assistants

Brigitte Marsteurer, Romana Landauer (GÖ FP)

Contract number:	Specific Contract N°20157305 under Framework contract № EAHC/2013/Health/01 Lot 1 'Health reports'
Duration of the project:	8 months
Starting date:	11/01/2016
Ending date:	31/10/2016

Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

Abstract - English

The aim of the study on Big Data in Public Health, Telemedicine and Healthcare is to identify applicable examples of the use of Big Data in Health and develop recommendations for their implementation in the European Union.

Examples of Big Data in Health were identified by a systematic literature review, after which the added value of twenty selected examples was evaluated. Based on the assessment of the added value and the quality of the evidence, ten priority examples were selected. Furthermore, potential policy actions for the implementation of Big Data in Health were identified in the literature, and a SWOT analysis was conducted to check the feasibility of the proposed actions. Based on this analysis, and with the help of renowned experts, the study team developed ten policy recommendations in the field. These recommendations were validated through public consultations at three relevant conferences in Europe and were again reviewed by the Expert Group.

The recommendations aim to benefit European citizens and patients in terms of strengthening their health and improving the performance of Member State's health systems. They should be seen as suggestions for the European Union and its Member States on how to utilise the strengths and exploit the opportunities of Big Data for Public Health without threatening privacy or safety of citizens.

Recommendations were developed for ten relevant fields: awareness raising, education and training, data sources, open data and data sharing, applications and purposes, data analysis, governance of data access and use, standards, funding and financial resources, as well as legal aspects and privacy regulation.

Abstract - French

L'objectif de l'étude des *Big Data* dans le domaine de la santé publique, de la télémédecine et des soins médicaux est d'identifier des exemples applicables des *Big Data* de la Santé et de développer des recommandations d'usage au niveau de l'Union Européenne.

Une étude documentaire systématique a permis l'identification d'exemples de *Big Data* de la Santé. Vingt exemples ont été choisis après une évaluation de leur valeur ajoutée, dont dix prioritaires sélectionnés sur la qualité des informations recueillies. De plus, les mesures potentielles pour l'application des *Big Data* de la Santé ont été identifiées dans la documentation et une analyse SWOT a été réalisée pour vérifier la faisabilité des actions proposées. L'équipe de recherche a élaboré à partir des résultats de l'analyse et avec l'aide d'experts renommés, dix recommandations stratégiques dans le domaine. Celles-ci ont été validées par des consultations publiques lors de trois conférences pertinentes en Europe, puis ont été réexaminées par un groupe d'experts.

Le but de ces recommandations est d'aider les citoyens et patients européens à améliorer leur santé, et de renforcer les services proposés par les systèmes de santé des États membres de l'Union Européenne. Ceux-ci devraient les considérer comme des suggestions sur l'exploitation efficace des atouts et des opportunités des *Big Data* pour la santé publique sans menacer la vie privée et la sécurité des citoyens.

Des recommandations ont été élaborées pour dix domaines pertinents : sensibilisation, éducation et formation, sources d'information, ouverture et échange de données, applications et objectifs, analyse de données, gouvernance de l'accès et de l'utilisation des données, normes techniques, financement et ressources financières, aspects légaux et protection de la vie privée.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A & E	Accidents and emergencies
AEGLE	Analytics framework for integrated and personalized healthcare services in Europe
CEO	Chief Executive Officer
CHAFEA	Consumers, Health, Agriculture and Food Executive Agency
CEPHOS-LINK	Comparative Effectiveness Research on Psychiatric Hospitalisation by Record Linkage of Large Administrative Data Sets
Comet K	Competence Centre for Excellent Technologies
CPRD	Clinical Practice Research Datalink
DEXHELPP	Decision Support for Health Policy and Planning
DG COMP	Directorate-General Competition
DG CONNECT	Directorate-General for Communications Networks, Content & Technology
DG GROW	Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs
DG RTD	Directorate-General for Research and Innovation
DG SANTE	Directorate-General for Health and Food Safety
DoPHER	Database of promoting health effectiveness reviews
EFPIA	European Federation of Pharmaceutical Industries and Associations
EHR	Electronic health record
EU	European Union
EUROSTAT	Statistical office of the European Union
EXPAND	Expanding Health Data Interoperability Services
FDA	Food and Drug Association
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
GÖ FP	Gesundheit Österreich Forschungs- und Planungs GmbH
HES	Hospital Episode Statistics
IBBL	Integrated BioBank of Luxembourg
IHEC	International Human Epigenome Consortium
IHE.net	Integrating the Healthcare Enterprise
LHU	Local Health Units
MeSH	Medical Subject Headings
MS	Member States
NHS	National Health Service
NIH	National Institutes of Health

NIS	Network and Information Systems
OECD	Organisation for Economic Co-operation and Development
PASSI	Progressi delle Aziende Sanitarie
RD	Rare disease
SEMCARE	Semantic Data Platform for Healthcare
SpainRDR	Spanish Rare Diseases Registries Research Network
SSNAP	Sentinel Stroke National Audit Programme
SWOT	Strengths, Weaknesses, Opportunities, Threats
UN	United Nations
YODA	Yale University open data access
WHO	World Health Organisation

Country Codes

AT	Austria
BE	Belgium
DE	Germany
DK	Denmark
EE	Estonia
GR	Greece
ES	Spain
FI	Finland
FR	France
IT	Italy
LU	Luxembourg
NL	The Netherlands
NO	Norway
PT	Portugal
RO	Romania
SE	Sweden
SI	Slovenia
UK	United Kingdom
US	United States

Executive Summary – English

Background

There are various common definitions of Big Data [1-3] in place, but none of them specifically focuses on Health, Telemedicine and Healthcare. However, recurring characteristics of datasets to be considered as Big Data are the 3 Vs: **V**olume, **V**elocity, and **V**ariety [4]. Other researchers even add a fourth V, which some refer to as **V**alue [1, 4] and some as **V**eracity [5]. For the purpose of this study the authors together with an Expert Group (cf. section 2.2.1) developed the following definition.

"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."

The analysis of Big Data, whether it is structured or unstructured, usually requires significant logistic efforts and computing power. Independent of the actual definition and the field of application (e.g. public health), the value chain of Big Data consists of generating and collecting data, storing and processing, and, finally of the distribution and analysis of the relevant data [4]. Applications may either be prospective data monitoring or retrospective data analysis and may contribute to [6-10]

- increasing the **effectiveness and quality of treatments** by e.g.:
    - earlier disease intervention,
    - reduced probability of adverse reactions to medicines,
    - less medical errors,
    - determination of causalities, understanding of co-morbidity,
    - cross-linkage of health care providers and professionals,
    - intensification of research networks, and
    - fusion of different networks such as social networks, disease networks or medicine networks,
- widening possibilities for the **prevention of diseases** by identification of risk factors for disease at population, subpopulation, and individual levels, and by improving the effectiveness of interventions to help people achieve healthier behaviours in healthier environments,
- the improvement of pharmacovigilance and **patient safety** through the ability to make more informed medical decisions based on **directly delivered information** to the patients,
- **prediction of outcomes**, e.g. containment and improvement of chronic diseases, global infectious disease surveillance through evolving risk maps and better **understanding of demographic challenges and trends as well disease transmission pathways,**
- **knowledge dissemination,** e.g. help physicians to stay current with the latest evidence guiding clinical practice, and
- **reduction in inefficiency** and waste, improvement in cost containment.

Study objectives

In its recently adopted conclusion on open, data-intensive and networked research as a driver for faster and wider innovation, the European Council calls for action regarding the identification of sectorial priorities for research and innovation with the greatest potential for social and economic benefits in the data economy [11]. The European

Council also emphasises the importance of digital economy and recognizes its high potential as well as the need for a strong data value chain in Europe. Both, the European Council and Member States (MS), are willing to set the necessary steps to enhance data innovation, especially in the light of exponential increase of data, highlighting that making data discoverable, accessible, assessable, reusable and interoperable is the key to innovation. This study commissioned by CHAFEA / DG SANTE explores the use of Big Data in Health in general to support an improvement in citizen's health and health outcomes.

The study aims to identify applicable examples of use of Big Data in Health and develop recommendations for their implementation. The **specific objectives** are:

- To provide a list of examples of Big Data in Public Health, Telemedicine and Healthcare already being used for possible implementation in the EU MS.
- To propose at least the 10 most important priorities related to Big Data in relation with the practice of public health, telemedicine and healthcare, where a specific action could be developed, in particular at EU level. The added value as well as the support to sustainability of health systems, improving quality and effectiveness of treatment, combating chronic disease and support of healthy lifestyle will be the main factors for the selection of these priorities.
- To develop a list of at least 10 policy recommendations with the aim to provide guidelines for the development of a Big Data value chain in the EU.

Examples and priorities of Big Data in Health

The research team performed a systematic literature review to identify examples of Big Data in Health. After identifying relevant examples, the added value of twenty selected examples was evaluated. The literature databases Medline, Cochrane Databases, Embase and Scopus where searched systematically, and a hand search of different websites (EU institutions, OECD, WHO, Google Scholar etc.) was conducted additionally. In order to maximize the number of relevant publications, reference tracking was applied.

The following twenty examples of use of Big Data in Health were identified and selected for further analysis:

- Comet K-Project DEXHELPP – AT
- The Shared Care Platform – DK
- E-Estonia – National Identity Scheme – EE
- AEGLE (An analytics framework for integrated and personalized healthcare services in Europe) – UK, IT, GR, SE, BE, NL, PT, FR
- The Business Intelligence database system – GR
- PASSI (Progressi delle Aziende Sanitarie) – IT
- Arno Observatory – IT
- The Swedish Big Data Analytic Network – SE
- Clinical Practice Research Datalink (CPRD) – UK
- Sentinel Stroke National Audit Programme (SSNAP) – UK
- Hospital Episode Statistics (HES) – UK (England)
- The YODA Project (Yale University open data access) – US
- FDA Adverse Event Network Analyser - US
- CEPHOS-LINK – FI, AT, RO, NO, SI, IT
- Twitter (Adverse drug reactions and public health) – International
- Flatiron – US
- UK Biobank – UK
- Semantic Data Platform for Healthcare (SEMCARE) – DE, NL, AT, UK, ES
- Integrated BioBank of Luxembourg (IBBL) – LU
- Spanish Rare Diseases Registries Research Network (SpainRDR) – ES

The examples were attributed to one or more of the following fields of application: Health system and service research, Epidemiology, Surveillance (pharmaceutical / public health), Clinical research.

The added value of the examples was assessed from four different perspectives: 1.) Patient, 2.) Provider, 3.) Policy and 4.) Research. Based on these results, the added value in terms of quality and effectiveness of treatment, sustainability of health systems, combating chronic disease and/or supporting healthy lifestyles was evaluated. This served as the basis for the selection of 10 priority examples.

Based on the assessment of the added value and the quality of the supporting evidence, the following examples were identified as priorities:

- Comet K-Project DEXHELPP
- The Shared Care Platform
- E-Estonia
- ARNO observatory
- PASSI
- Health Episode Statistics
- The YODA Project
- CEPHOS-LINK
- Flatiron
- Spanish Rare Diseases Registries Research Network (SpainRDR)

The aim was to select a balanced set of examples regarding the field of application and added value to health. Furthermore, different technological approaches to cover the whole range of possibilities of Big Data were included. Examples where no or only low-quality information sources could be found were not regarded in the priority selection.

Policy actions for Big Data in Health

The identification of potential policy actions was based on the literature and included recent reports [12]. The identified policy actions for Big Data in Health were clustered around twelve fields:

1. Legal aspects
2. Stakeholders
3. Privacy and data protection
4. Open data and data sharing
5. Standards and protocols
6. Technological development
7. Data sources
8. Data analysis
9. Applications
10. Communication
11. Human capital
12. Funding

The distinction between these fields of policy actions is not always perfectly precise, and other categorizations (broader or more narrow) are conceivable. However, based on the literature, these were the fields that became apparent when clustering the single policy actions identified.

To reflect on the feasibility of implementing the proposed policy actions, a SWOT[1] analysis was conducted using the following framework:

Table 1: Framework for SWOT analysis

	Positive aspects	Negative aspects
Internal factors	What are internal strengths of the actions with respect to the added value they provide for DG SANTE's health policy?	What are internal weaknesses of the actions with respect to the added value they provide for DG SANTE's health policy?
External factors	What are opportunities that can promote the implementation of actions on behalf of DG SANTE?	What are threats that can jeopardise the implementation of actions on behalf of DG SANTE?

Source: GÖ FP

The SWOT analysis aimed to reveal the most important internal strengths and weaknesses as well as external opportunities and threats for each field of policy action and to relate these to the identified examples of use of Big Data.

The policy actions were validated during a workshop by experts (cf. section 6.3) in the following fields:

- Health issues (health systems, telemedicine, chronic diseases and healthy lifestyle) and health policy working in national public administrations
- Stakeholders bringing in the views of patients/citizens, health professionals, service providers and healthcare payers
- Health Information issues, Big Data and telemedicine from the industry or multi-stakeholder associations.

Policy recommendations for Big Data in Health

Based on the SWOT analysis of the policy actions and the results of the expert workshop, ten policy recommendations were developed and validated by public consultations at three relevant conferences in Europe as well as a final feedback round with the identified experts. The recommendations aim to benefit European citizens and patients in terms of strengthening their health and improving the performance of MS's health systems. Therefore, the recommendations are explicitly written from a public health perspective.

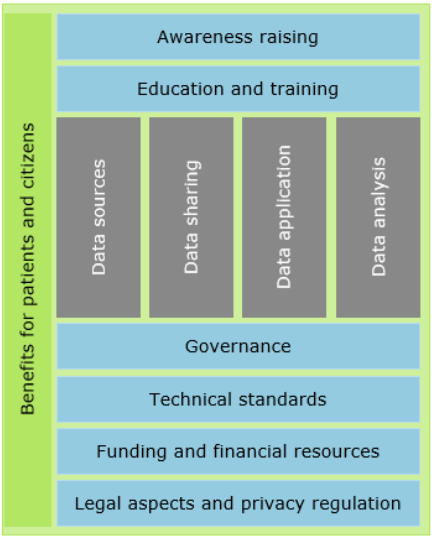
The following general notions should be considered for all ten policy recommendations. First, the scope of the recommendations is to give suggestions for the European Union (EU) and its Member States (MS) on how to utilize the strengths and exploit the opportunities of Big Data for Public Health without threatening privacy or safety of citizens. Second, Big Data in Health should not be seen as a goal in itself, but as a tool to reach certain purposes that benefit the patient or citizen. Third, current ethical standards must not be weakened or compromised for potential benefits of Big Data. Fourth, stakeholders need to be included in the implementation of the proposed recommendations and in the production of future recommendations on Big Data in Health. Especially patients (represented by their advocacy groups), who ultimately have to consent to the use of Big Data in Health, have to be involved in the process of producing and implementing recommendations. Despite the importance of patient involvement, all other stakeholders that are part of the data value chain (health professionals, data scientists, health research, industry, public administrations, etc.) should be considered. Fifth, issues related

[1] Strengths-Weaknesses-Opportunities-Threats

to Big Data in Health that are covered in existing regulatory frameworks (e.g. data protection, informed consent, quality, safety and reliability) are only included if distinctively important for the use of Big Data in Health.

Figure 6 depicts the fields of policy recommendations and their respective dimension (vertical, horizontal, and overarching), and the following list summarizes the main messages of the ten fields.

Figure 1: Overview of fields of policy recommendations



Source: GÖ FP

- Recommendation 1 on Awareness Raising:
  Develop and implement a communication strategy to increase the awareness of the added value of Big Data in Health and encourage a positive public mind set towards Big Data in Health
- Recommendation 2 on Education and Training
  Strengthen human capital with respect to the increasing need for a workforce that can utilize the potential of Big Data in Health
- Recommendation 3 on Data Sources:
  Expand existing and explore new sources of Big Data in Health and secure their quality and safety
- Recommendation 4 on Open Data and Data Sharing:
  Promote open use and sharing of Big Data in Health without compromising patients' rights to privacy and confidentiality
- Recommendation 5 on Applications and Purposes:
  Increase target-oriented application of Big Data analysis in health based on the needs and interests of stakeholders including patients
- Recommendation 6 on Data Analysis:
  Identify the potentials of Big Data analysis, improve analytical methods and facilitate the use of new and innovative analytical methods
- Recommendation 7 on Governance of Data Access and Use:
  Implement governance mechanisms to ensure secure and fair access and use of Big Data for research in health
- Recommendation 8 on Standards:
  Develop standards for Big Data in Health to enhance and simplify its application and improve interoperability

- Recommendation 9 on Funding and Financial Resources:
  Ensure purposeful investment steered by the European Commission to warrant cost-effectiveness and sustainability
- Recommendation 10 on Legal Aspects and Privacy Regulations:
  Clarify and align existing legal and privacy regulation of Big Data in Health

Executive Summary – French

Contexte

Il existe plusieurs définitions courantes de *Big Data* [1-3], mais aucune d'entre elles ne s'applique spécifiquement aux domaines de la santé, de la télémédecine et des soins médicaux. Pourtant, on trouve fréquemment parmi les caractéristiques des ensembles de données qui entrent dans la catégorie « *Big Data* », ce qu'on appelle les 3 V : **V**olume, **V**élocité et **V**ariété.[4]. D'autres chercheurs vont même jusqu'à ajouter un quatrième V qui, pour certains, représente la **V**aleur [1, 4] et, pour d'autres, la **V**éracité. Pour les besoins de ce rapport, les auteurs et les membres d'un groupe d'experts (voir section 2.2.1) ont développé la définition suivante.

« On appelle Big Data de la Santé des ensembles de données volumineux recueillis de manière régulière ou automatique, qui ont été enregistrés et stockés électroniquement. On peut réutiliser ce concept dans le sens de données polyvalentes et y inclure la fusion et la connexion de bases de données existantes dans le but d'améliorer la santé ainsi que la performance du système de santé. Il ne s'agit pas de données recueillies pour une étude spécifique. »

L'analyse du *Big Data*, qu'il soit structuré ou non, nécessite des efforts logistiques et une puissance informatique importants. Indépendamment de la vraie définition et du champ d'application (par exemple, la santé publique), la chaîne de valeur du *Big Data* consiste en la création et la collecte de données, dans le stockage, le traitement et finalement, la distribution et l'analyse des données pertinentes [4]. Il peut être appliqué soit pour un suivi prospectif de données, soit pour une analyse rétrospective de données, ce qui peut contribuer à [6-10] :

- augmenter **l'efficacité et la qualité des traitements** grâce à, par exemple :
  - des interventions médicales plus rapides,
  - une probabilité réduite dans l'apparition d'effets indésirables liés aux médicaments,
  - une baisse des erreurs médicales,
  - une meilleure détermination des causes et une meilleure compréhension de la comorbidité,
  - une mise en relation entre les prestataires de soins de santé et les professionnels,
  - une intensification de réseaux de recherche et
  - une fusion de différents réseaux (réseaux sociaux, réseaux dédiés à des maladies, réseaux de médecins…)
- élargir les possibilités pour la **prévention de maladies** en identifiant les facteurs de risques de maladie au niveau des populations, sous-populations et des individus, et en améliorant l'efficacité des interventions pour aider les gens à adopter des comportements plus sains dans des environnements plus sains,
- améliorer la pharmacovigilance et la sécurité des patients en rendant possible des prises de décisions plus informées et basées sur des **informations directement transmises** aux patients ;
- **Prédiction des résultats** avec par exemple, des cartes d'évolution des risques, une **meilleure compréhension des défis et tendances démographiques** ainsi que des **voies de transmission des maladies** afin d'endiguer et d'améliorer la surveillance des maladies chroniques et des maladies transmissibles à l'échelle mondiale ;
- **disséminer la connaissance** – par exemple, aider les praticiens à rester informés sur les éléments scientifiques disponibles pour les pratiques cliniques et
- **réduire l'inefficacité** et le gaspillage, améliorer la maîtrise des coûts.

Objectifs de l'étude

Le Conseil européen a récemment conclu qu'une recherche ouverte, en réseau et axée sur les données contribuait activement au renforcement et à l'accélération de l'innovation, et appelle donc à identifier les priorités sectorielles en matière de recherche et innovation qui sont les plus susceptibles de bénéficier à l'économie de données sur le plan social et économique [11]. Le Conseil européen souligne également l'importance de l'économie numérique et reconnaît son potentiel élevé ainsi que la nécessité d'une chaîne de valeur forte dans la gestion de données en Europe. Le Conseil Européen aussi bien que les états membres (EM) sont disposés à prendre les mesures nécessaires pour renforcer l'innovation en matière de données, compte tenu notamment de l'augmentation exponentielle des données, et soulignent que la clé de l'innovation réside dans la possibilité de rendre les données décelables, accessibles, évaluables, réutilisables et interopérables. Cette étude commandée par CHAFEA / DG SANTE étudie en profondeur l'utilisation du *Big Data* de la Santé en général afin de favoriser une amélioration dans la Santé des citoyens et des résultats sanitaires.

Le but de cette étude est d'identifier des exemples applicables d'utilisation des *Big Data* de la Santé et de développer des recommandations d'usage. Ces **objectifs spécifiques** sont :

- De fournir une liste d'exemples de *Big Data* déjà utilisés en vue d'une application possible dans les états membres de l'UE.
- De proposer au moins 10 priorités essentielles dans la mise en relation entre les *Big Data* et les pratiques exercées dans la santé publique, la télémédecine et les soins médicaux, qui nécessitent une action spécifique, en particulier à l'échelle de l'UE. La sélection de ces priorités reposera principalement sur les facteurs suivants : la valeur ajoutée et la contribution à la durabilité des systèmes de santé, l'amélioration de la qualité et de l'efficacité des traitements, l'engagement en faveur de la lutte contre les maladies chroniques et de la promotion d'un mode de vie sain.
- De développer une liste d'au moins 10 recommandations stratégiques en vue de fournir des directives pour le développement d'une chaîne de valeur du *Big Data* au niveau de l'UE.

Exemples et priorités de *Big Data* de la Santé.

L'équipe de recherche a effectué une étude documentaire systématique pour identifier des exemples de *Big Data* de la Santé. Après avoir identifié des exemples pertinents, la valeur ajoutée de vingt exemples sélectionnés a été évaluée. Les bases de données documentaires Medline, Cochrane Databases, Embase and Scopus ont fait l'objet d'une recherche systématique complétée par une recherche manuelle de différents sites web (institutions européennes, OCDE, OMS, Google Scholar etc.). Des outils de gestion bibliographiques (suivi de références) ont été utilisés afin d'optimiser le nombre de publications pertinentes.

Les vingt exemples d'utilisation de *Big Data* de la Santé ci-dessous ont été identifiés et sélectionnés pour une analyse plus détaillée :

- Comet K-Project Dexhelpp – AT
- The Shared Care Platform – DK
- E-Estonia – National Identity Scheme – EE
- AEGLE (Un cadre d'analyse des services de soins intégrés et personnalisés en Europe) – UK, IT, GR, SW, BE, NL, PT, FR
- The Business Intelligence database system – GR
- PASSI – IT
- Arno Observatory – IT
- The Swedish Big Data Analytic Network – SW
- Clinical Practice Research Datalink (CPRD) – UK

- Sentinel Stroke National Audit Programme (SSNAP) – UK
- Hospital Episode Statistics (HES) – UK (England)
- The YODA Project (Yale University open data access) – US
- FDA Adverse Event Network Analyser - US
- CEPHOS-LINK – FI, AT, RO, NO, SI, IT
- Twitter (Informations sur les effets indésirables liés à l'usage de médicaments et sur la santé publique) – International
- Flatiron – USA
- UK Biobank – UK
- Semantic Data Platform for Healthcare (SEMCARE) – DE, NL, AT, UK, ES
- Integrated BioBank of Luxembourg (IBBL) – LU
- Spanish Rare Diseases Registries Research Network (SpainRDR) – ES

Ces exemples ont été attribués à un ou plusieurs des champs d'application suivants : recherche sur les services et systèmes de santé, épidémiologie, surveillance (pharmaceutique/santé publique), recherche clinique.

La valeur ajoutée de ces exemples a été évaluée sous quatre angles différents : 1) Patient, 2) Fournisseur, 3) Politique et 4) Recherche. Ces résultats ont permis d'évaluer la valeur ajoutée en ce qui concerne la qualité et l'efficacité des traitements, la durabilité des systèmes de santé, la lutte contre les maladies chroniques et/ou la promotion d'habitudes de vie saines, et ont servi de base à la sélection des 10 exemples de priorités.

Les exemples ci-dessous ont été identifiés comme priorités à partir de l'évaluation de la valeur ajoutée et de la qualité des informations recueillies :
- Comet K-Project DEXHELPP
- The Shared Care Platform
- E-Estonia
- ARNO observatory
- PASSI
- Health Episode Statistics
- The YODA Project
- CEPHOS-LINK
- Flatiron
- Spanish Rare Diseases Registries Research Network (SpainRDR)

L'objectif était de sélectionner un ensemble objectif d'exemples de champs d'application et de valeurs ajoutées au domaine de la santé. Il inclut également une gamme complète de possibilités qu'offre le *Big Data* d'un point de vue technologique. La sélection des priorités a exclu les exemples n'ayant pas de source d'information ou provenant d'une source médiocre.

Mesures pour le *Big Data* de la Santé

L'identification des mesures stratégiques potentielles s'est basée sur la documentation qui comprenait des rapports récents [12]. Les stratégies identifiées pour le *Big Data* de la Santé ont été regroupées autour de douze domaines :
1. Aspects légaux
2. Parties prenantes
3. Protection des données personnelles et de la vie privée
4. Ouverture et partage des données
5. Normes et protocoles
6. Développement technologiques
7. Sources de données
8. Analyse de données
9. Applications

10. Communication
11. Capital humain
12. Financement

Il n'y a pas toujours de distinction précise entre ces domaines stratégiques et il est possible de concevoir d'autres catégorisations (plus larges ou plus restreintes). Cependant, les champs d'action listés ci-dessus sont ceux qui sont ressortis du regroupement des exemples identifiés lors de l'étude documentaire.

Une analyse SWOT[2] a été effectuée pour servir de base de réflexion sur la faisabilité des actions politiques proposées :

Table 2: Cadre de l'analyse SWOT

	Aspects positifs	**Aspects négatifs**
Facteurs internes	Quelles sont les forces internes des actions par rapport à la valeur ajoutée qu'elles apportent à la politique de santé de DG SANTE ?	Quelles sont les faiblesses internes des actions par rapport à la valeur ajoutée qu'elles apportent à la politique de santé de DG SANTE ?
Facteurs externes	Quelles sont les opportunités qui peuvent favoriser l'application de ces actions au nom de DG SANTE ?	Quelles sont les menaces qui peuvent compromettre l'application de ces actions au nom de DG SANTE ?

Source: GÖ FP

Le but de l'analyse SWOT a été de révéler les forces et les faiblesses les plus importantes ainsi que les opportunités et les menaces externes pour chaque domaine stratégique, et de les relier avec les exemples identifiés d'utilisation du *Big Data*.

Les stratégies ont été validées lors d'un atelier d'experts (voir Section 6.3) dans les domaines suivants :

- Problèmes liés à la santé (systèmes de santé, télémédecine, maladies chroniques et mode de vie sain) et politique sanitaire dans les administrations nationales publiques.
- Les parties prenantes apportant les avis de patients/citoyens, professionnels de la santé et organismes payeurs.
- Problèmes liés à l'information, le *Big Data* et la télémédecine fournie par l'industrie ou les associations multipartites.

Recommandations pour l'utilisation du *Big Data* de la Santé

Dix recommandations stratégiques ont été développées à partir des analyses SWOT et des résultats des ateliers d'experts, puis ont été validées par des consultations publiques lors de trois conférences pertinentes en Europe. Le but de ces recommandations est d'aider les citoyens et patients européens à améliorer leur santé, et de renforcer les services proposés par les systèmes de santé des états membres. Par conséquent, ces recommandations ont été rédigées de manière explicite du point de vue des fonctionnaires de la santé publique.

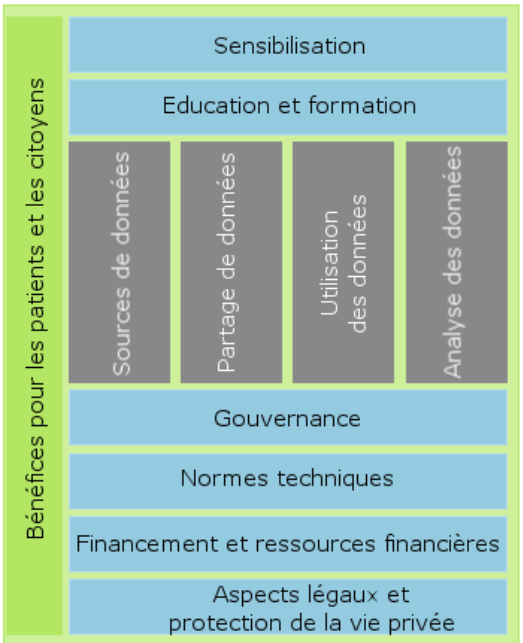
Il est indispensable de prendre en considération les notions générales suivantes pour les dix recommandations stratégiques. Premièrement, l'objectif des recommandations est de présenter des suggestions pour l'Union Européenne (UE) et ses états membres

[2] Strengths-Weaknesses-Opportunities-Threats (ou FFOM – Forces, Faiblesses, Opportunités, Menaces)

(EM) sur l'exploitation efficace des atouts et des opportunités des *Big Data* pour la Santé publique sans menacer la vie privée et la sécurité des citoyens. Deuxièmement, le *Big Data* de la Santé ne doit pas être perçu comme un but en soi, mais comme un outil permettant d'atteindre certains objectifs qui bénéficieront au patient. Troisièmement, les normes éthiques actuelles ne doivent pas être affaiblies ou compromises au profit des avantages potentiels du *Big Data*. Quatrièmement, les parties prenantes ne doivent pas être exclues de l'application des recommandations proposées ainsi que dans la création de futures recommandations pour l'utilisation de *Big Data* de la Santé. Ceci inclut surtout les patients (représentés par leurs groupes de défense) qui en fin de compte, doivent pouvoir consentir à l'utilisation du *Big Data* de la Santé, et doivent pouvoir participer au processus de création et d'application de recommandations. Malgré l'importance de l'implication du patient, il est nécessaire de prendre en compte tous les autres participants de la chaîne de valeur des données (professionnels de la santé, spécialistes des données, instituts de recherche médicale, industries, administrations publiques…). Cinquièmement, les problèmes liés au *Big Data* de la Santé abordés dans les cadres réglementaires existants (par exemple, la protection des données, le consentement informé, la qualité, la sécurité et la fiabilité) ne sont inclus que s'ils se rapportent spécifiquement à l'utilisation du *Big Data* de la Santé.

Le schéma 5 représente les domaines d'application des recommandations stratégiques et leur dimension respective (verticale, horizontale et globale), et la liste suivante résume les principaux messages des dix domaines.

Figure 2: Vue d'ensemble des domaines des recommandations stratégiques



Source: GÖ FP

- Recommandation 1 sur la sensibilisation:
  Développer et mettre en œuvre une stratégie de communication pour sensibiliser à la valeur ajoutée du *Big Data* de la Santé et promouvoir une image positive du *Big Data* de la Santé.
- Recommandation 2 sur l'éducation et la formation :
  Renforcer le capital humain par rapport au besoin croissant d'une main d'œuvre qui peut tirer profit du potentiel du *Big Data* de la Santé.

- Recommandation 3 sur les sources de données :
  Développer et diversifier les sources existantes de *Big Data* de la Santé et en découvrir de nouvelles, garantir leur qualité et assurer leur protection.
- Recommandation 4 sur l'ouverture et le partage des données:
  Promouvoir l'accès libre et le partage des *Big Data* sans compromettre le droit des patients à la protection de leur vie privée et à la confidentialité.
- Recommandation 5 sur les applications et les objectifs :
  Appliquer de manière ciblée les résultats des analyses de *Big Data* en se basant sur les besoins et les intérêts des parties prenantes, y compris des patients.
- Recommandation 6 sur l'analyse des données :
  Identifier les potentiels de l'analyse des *Big Data*, améliorer les méthodes d'analyse et promouvoir l'utilisation de méthodes novatrices.
- Recommandation 7 sur la gouvernance de l'accès et de l'utilisation des données:
  Mettre en œuvre des mécanismes de gouvernance qui assurent un accès et une utilisation équitable et en toute sécurité des *Big Data* pour la recherche médicale.
- Recommandation 8 sur les normes techniques:
  Développer des normes techniques pour les *Big Data* de la Santé afin de simplifier leur application et améliorer leur interopérabilité.
- Recommandation 9 sur le Financement et les Ressources Financières:
  Assurer un investissement réfléchi guidé par la Commission Européenne afin de garantir rentabilité et durabilité.
- Recommandation 10 sur les aspects légaux et la protection de la vie privée:
  Clarifier et aligner les règlementations existantes des *Big Data* de la Santé concernant la protection de la vie privée.

1 Introduction

A defining characteristic of today's data-rich society is the collection, storage, processing and analysis of immense amounts of data. Due to its characteristics, commonly known as the 3 Vs: **V**olume, **V**elocity, and **V**ariety [4], this kind of data is called Big Data. Other researchers even add a fourth V, which some refer to as **V**alue [1, 4] and some as **V**eracity [5]. Big Data is generated from an increasing plurality of sources including internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks or business transactions such as sales queries and purchases. In addition, genomics, health care, engineering, operations management, the industrial internet and finance all add to the Big Data pervasiveness. Looking in the literature [1-3], one finds various common definitions of Big Data, and it was even pointed out, that "the use of the term is quite nebulous" (Philip Ashlock, [2]).Big Data is defined in various ways. However, none of them specifically focuses on Big Data in the relation to Health, Telemedicine and Healthcare. For the purpose of this study the authors together with an Expert Group (cf. **Fehler! Verweisquelle konnte nicht gefunden werden.** 2.2.1) developed the following definition.

"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[3]."

1.1 Background and context

The analysis of Big Data, whether it is structured or unstructured, usually requires significant logistic efforts and computing power. Independent of the actual definition and the field of application (e.g. public health), the value chain of Big Data consists of generating and collecting data, storing and processing, and, finally of the distribution and analysis of the relevant data [4].

There are many terms, including e-health, m-health, digital health, health information technology, health 2.0, e-medicine, telemedicine, e-health that are linked to the collection, analysis and application of Big Data in Health, which is an emerging field of action in recent years. For a long time Big Data in Health only played a major role in medical and clinical research. Translation into the practice of public health was not a distinct objective of the collection of Big Data. As data gets more available, financial resources are more and more limited and the technical progress increases, stakeholders in public health as well as the scientific community are opening up to the opportunities offered by applications of Big Data not only for the health of the individual but also for the health of the whole population. [7, 13].

It is now important to take further and coordinated action in absorbing the full potential of Big Data in Health as a driver for faster and wider innovation as recently stressed by the European Council in its Conclusions on open, data-intensive and networked research [11]. Through the use of Big Data it might be possible to improve health of individual persons (personalised medicine) as well as to improve the performance and outcomes of health care systems.

[3] Primary data collected for (clinical) study purposes only without the purpose of data sharing outside the scope of the study

The current approach regarding the use of data on individual health is the collection of data during diagnosis and monitoring. This makes it 1) difficult to gather high frequency longitudinal data and 2) necessary to rely on retrospective recollection which may be inaccurate [14]. For the implementation of personalised medicine not only individual genomic data but also population data are highly relevant for estimating a posteriori probabilities. The combination of population-level information with individual-level measurements provide exciting opportunities for the implementation of personalized medicine [15].

Big Data in Health is already being generated and ready for use from various different sources as listed in the tender specifications:
Health care records and patient summaries

- Social media
- Genomic data
- Pharmaceutical data
- Insurance claims
- Telemedicine, mobile apps and sensors
- Other sources (income statistics, environmental databases etc.)

Possible data sources for Big Data are continuously evolving, therefore, the presented list cannot be regarded as exhaustive. Additionally the combination of data sets generates another level of complexity, yet also creates new possibilities.

This is why the concept of data fusion is gaining significance [16] and the connection of existing Big Data and Big Data research in platforms or tools has become more and more important over the last few years. Moreover, these data sets require the use of powerful computational techniques to unveil trends and patterns within and between the datasets [17]. It is crucial to find ways of systematic approaches to manage, integrate, analyse, and interpret such large complex data sets [18].

High level experts in a conference meeting on Personalised Medicine found that in order to intensify action on this topic a stronger cooperation between the European Union's (EU's) Member States (MS) a cross-sectoral approach is needed. Furthermore, a need for flexible methods to evaluate the added value of use of Big Data in Health, for the empowerment of patients and for a move to an adaptive approach of data collection were expressed during the conference [19].

Big Data use in health care will gain importance quickly. Applications may either be prospective data monitoring or retrospective data analysis and may contribute to [6-10]

- increasing the **effectiveness and quality of treatments** by e.g.:
    - earlier disease intervention,
    - reduced probability of adverse reactions to medicines,
    - less medical errors,
    - determination of causalities, understanding of co-morbidity,
    - cross-linkage of health care providers and professionals,
    - intensification of research networks, and
    - fusion of different networks such as social networks, disease networks or medicine networks,
- widening possibilities for **prevention of diseases** by identification of risk factors for disease at population, subpopulation, and individual levels, and by improving the effectiveness of interventions to help people achieve healthier behaviours in healthier environments,

- improvement of pharmacovigilance and **patient safety** through the ability to make more informed medical decisions based on **directly delivered information** to the patients,
- **prediction of outcomes**, e.g. containment and improvement of chronic-diseases, global infectious disease surveillance through evolving risk maps and better **understanding of demographic challenges and trends as well as disease transmission pathways,**
- **knowledge dissemination,** e.g. help physicians to stay current with the latest evidence guiding clinical practice, and
- **reduction in inefficiency** and waste, improvement in cost-containment.

1.2 Rationale and objectives

In its recently adopted conclusion on open, data-intensive and networked research as a driver for faster and wider innovation, the European Council calls for action regarding the identification of sectorial priorities for research and innovation with the greatest potential for social and economic benefits in the data economy [11]. The European Council also emphasises the importance of digital economy and recognizes its high potential as well as the need for a strong data value chain in Europe. Accordingly, both the European Council and MS are willing to set the necessary steps to enhance data innovation especially in the light of exponential increase of data, highlighting that making data discoverable, accessible, assessable, reusable and interoperable is the key to innovation. This study commissioned by CHAFEA / DG SANTE explores the use of Big Data in Health in general to support an improvement in citizen's health and health outcomes.

The study aims to identify applicable examples of use of Big Data in Health and develop recommendations for their implementation.

The **specific objectives** are, as stated in the Request for Specific Services N° CHAFEA/2015/Health 21 for the implementation of Framework Contract N° EAHC/2013/Health/01 – lot 1- health reports – for the provision of a study on Big Data in public health, telemedicine and healthcare:

- To provide a list of examples of Big Data already being used for possible implementation in the EU MS.
- To propose at least 10 most important priorities related to Big Data in relation with the practice of public health, telemedicine and healthcare, where a specific action could be developed, in particular at EU level. The added value as well as the support to sustainability of health systems, improving quality and effectiveness of treatment, combating chronic disease and support of healthy lifestyle will be the main factors for the selection of these priorities.
- To develop a list of at least 10 policy recommendations with the aim to provide guidelines for the development of a Big Data value chain in the EU.

2 Methodology

The study at hand was based on a systematic literature review as well as consultation(s) of experts. Details on the methodology applied are presented in the following sections 2.1 and 2.2.

2.1 Systematic literature review

The aim of the systematic literature review was to identify relevant literature on Big Data in Health based on which examples of the use of Big Data in the practice of public health, telemedicine and healthcare were identified, their added value analysed and potential policy actions derived.

The following research questions were addressed:

- According to the literature, what are examples for the use of Big Data in the practice of public health, telemedicine and healthcare?
- What added value in terms of sustainability of health systems, improving quality and effectiveness of treatment, combating chronic disease and support of healthy lifestyle do the examples identified bring?
- What are potential policy actions DG SANTE could launch to develop and support the use of Big Data?

To answer the above mentioned research questions, the following search strategy was applied (simplified presentation; see Annex 1 in section 6.1 for the comprehensive search strategies): *(Big Data OR large public health data OR large health care data) AND (Public health)*. For the systematic literature search, the following databases were used: Medline, Cochrane Databases, Embase and Scopus. Search terms were searched for in title, abstract and descriptor fields in English and covered a period of 15 years (2000-2015). Free-text truncation (e.g. truncation like data* for dataset/s, database/s, etc.) and subject headings (e.g. Medical Subject headings (MeSH)) were used when appropriate.

In order to identify relevant grey literature, the team complemented the systematic literature search by a thorough hand search, including the websites of international organisations and networks:

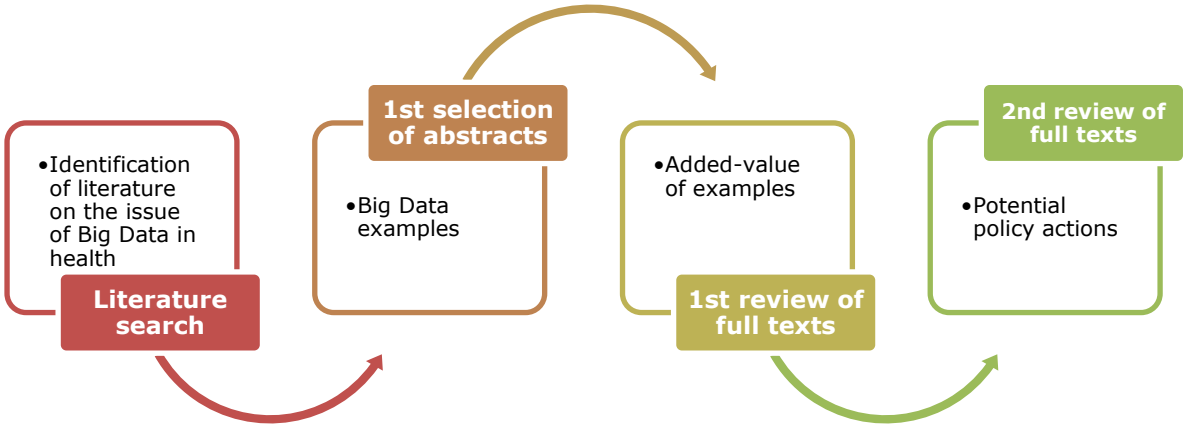
- EU (European Commission, particularly of DG SANTE, DG GROW, DG COMP, DG CONNECT, DG RTD, European Parliament, Council of the EU, CHAFEA project database, EUROSTAT)
- OECD
- WHO publications
- World Bank publications
- UN Statistics Division
- Google Scholar
- National governmental homepages (e.g. National Institutes of Health (NIH))
- National statistics institutes (e.g. Statistics Netherlands, Federal Statistical Office Germany)
- DoPHER

The websites have been searched for publications, including grey literature, on the issue of Big Data use in health care. In order to maximize the number of relevant publications, reference tracking was applied, too.

All references identified were first collected in an Endnote® file and analysed. The challenge of a systematic literature search is to maximize the amount of relevant literature while keeping the number of unrelated papers as small as possible. Therefore, the authors' selection of the literature for the proposed study was subdivided into three selection stages (cf. Figure 3):

- 1st abstract selection, to identify Big Data examples
- 1st review of full texts, to identify the added-value of Big Data in Health
- 2nd review of full texts, to derive potential policy actions.

Figure 3: Working process based on systematic literature review



Source: GÖ FP

2.1.1 Identification of examples on the use of Big Data in Health

For the selection of the examples on the use of Big Data in the practice of public health, telemedicine and healthcare, the abstracts and titles of the identified literature were screened in the first selection stage based on the following pre-defined inclusion and exclusion criteria.

A study was considered as relevant, if the following criteria (inclusion criteria I 1 – I 5) were met:

- I 1: The primary investigated subject is Big Data
- I 2: Actions to develop or support the use of Big Data are addressed in the studies.
- I 3: A connection to public health, telemedicine and/or health care is evident
- I 4: The studied practice/policy/action/strategy is implemented on a large-scale and the data collection is not limited to a small subgroup of the population.
- I 5: Outcomes such as quality and effectiveness of treatment, sustainability of health systems, combating chronic disease and/or supporting healthy lifestyle were considered in the study

A study was excluded, if the following criteria (exclusions criteria E 1 – E 7) were met:

- E 1: Big data in public health, telemedicine or health care is not the primary subject of the study
- E 2: The data collection is limited to a small subgroup of the population
- E 3: Implementation was only done at regional level
- E 4: The study was published in a language other than English or German
- E 5: Duplicate
- E 6: Publication date before 2000
- E 7: No clear description of the context

2.1.2 Identification of the added-value of Big Data in Health practice

The added value was identified by a first full text review of the abstracts, which were included in the first selection. Studies considering the following outcomes were regarded as relevant in the review:

- quality and effectiveness of treatment,
- sustainability of health systems,
- combating chronic disease
- supporting healthy lifestyle

The systematic search was complemented by an extensive grey literature search. As information gathered in that manner varied greatly with regard to structure and comprehensiveness, standardized quality assessment tools (e.g. quality criteria checklists for systematic literature reviews) were unsuitable to assess the risk of bias. Therefore, the authors followed a more pragmatic approach and decided to use a reduced set of general criteria derived from classic assessment tools (cf. tables in Annex 2 in section 6.2).

The examples were ascribed to one or more of the following fields of application:

- Health system and service research
- Epidemiology
- Surveillance (pharmaceutical / public health)
- Clinical research

The added-value of the examples was assessed from four different perspectives: 1.) Patient, 2.) Provider, 3.) Policy and 4.) Research. Based on these results, a final statement regarding the added-value of the examples in terms of quality and effectiveness of treatment, sustainability of health systems, combating chronic disease and/or supporting healthy lifestyles could be given.

2.1.3 Derivation of policy actions for Big Data in Health

Policy actions for Big Data in Health were identified by a second full text reviews of the abstracts included in the systematic literature research. Additionally, the systematic search was complemented by an extensive hand search in order to cover the most recent literature [12].

An iterative clustering process was use to derive policy a first set of policy actions. In a first round, the policy actions identified during the second selection of full texts were clustered around the dimensions of the data value chain (i.e. generation and collection, storage and processing, distribution and analytics) including a dimension on privacy and safety and an overarching dimension. In a second round, a more precise clustering was done based on specific and frequently occurring topics.

For a more concise set of policy actions, a SWOT[4] analysis was conducted. The SWOT analysis served for internal critical reflection of the feasibility of implementing the policy actions identified. In general, the concept of SWOT analysis has its roots in strategic management and originally aimed to facilitate a firm's strategy setting by analysing internal strengths and weaknesses as well as external opportunities and threats [20, 21]. For the purpose of this study, the concept was adapted to policy actions (cf. Table 1). The SWOT analysis for each policy action aimed to give an overview of the most important internal strengths and weaknesses as well as external opportunities and threats and served as foundation for drafting policy recommendations in the further course of the study.

Table 3: Framework for SWOT analysis

	Positive aspects	Negative aspects
Internal factors	What are internal strengths of the actions with respect to the added value they provide for DG SANTE's health policy?	What are internal weaknesses of the actions with respect to the added value they provide for DG SANTE's health policy?
External factors	What are opportunities that can promote the implementation of actions on behalf of DG SANTE?	What are threats that can jeopardise the implementation of actions on behalf of DG SANTE?

Source: GÖ FP

2.2 Expert consultation

In order to validate the research results and the derived policy actions (and recommendations), experts, stakeholders and private companies dealing with the topic of Big Data in Health were consulted. The validation of results comprised three kinds of consultations: 1.) consultation of the study's Expert Group, 2.) public consultation and 3.) internal feedback loops on behalf of the European Commission.

2.2.1 Expert Group

The study was accompanied by 16 experts (cf. Annex 3 in section 6.2), ensuring a well-balanced mix of Member States and different health system types. The Expert Group included:

- Experts in Health issues (health systems, telemedicine, chronic diseases and healthy lifestyle) and in health policy working for the government of 5 different member States with a geographical and health system types (Beveridge, Bismarck, National health system, Private) balance
- Stakeholders bringing the views of patients/citizens, health professionals, providers and healthcare payers
- Experts of Health Information issues, Big Data and telemedicine from the industry or multi-stakeholder associations.

The Expert Group was involved in a partly electronic Delphi panel including an Expert Workshop.

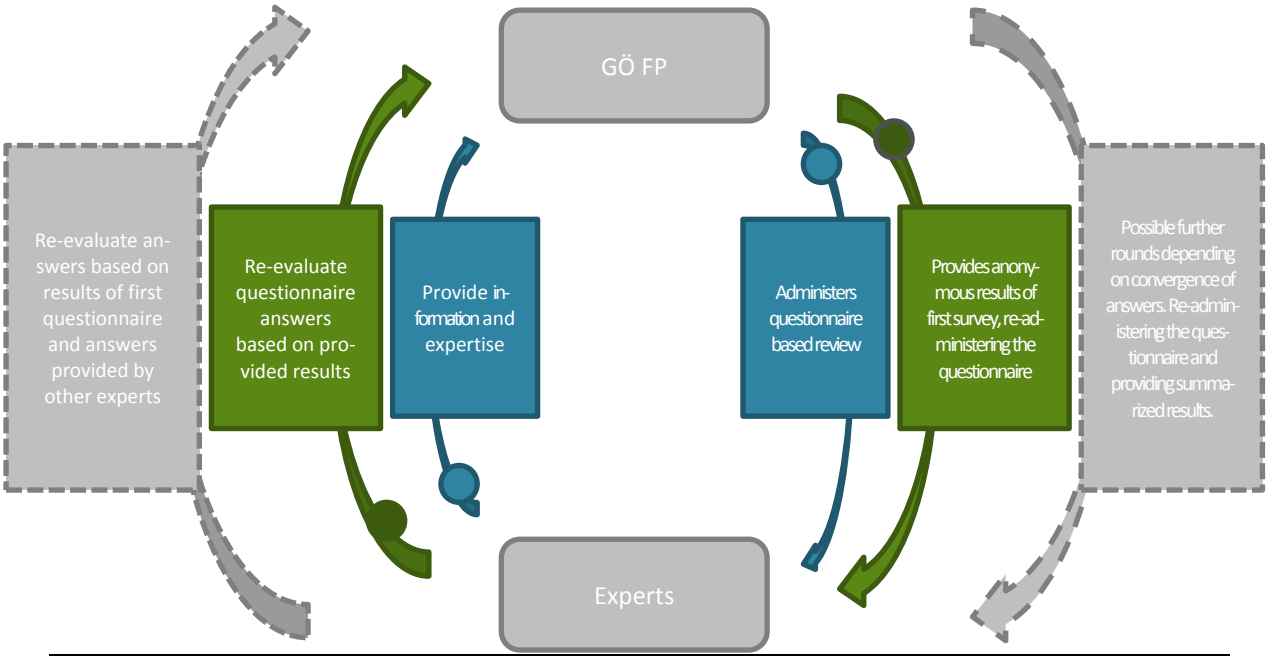
[4] Strengths-Weaknesses-Opportunities-Threats

2.2.2 Delphi consultation

The structured interview process was based on the methodological approach of a Delphi consultation. The aim was to validate and refine the derived policy actions and to develop policy recommendations for the use of Big Data in Health accordingly.

The iterative process of the Delphi method allows expert opinion to converge towards common sense. Moreover, studies have shown that averaging individual responses is inferior to the averages produced by group decision processes, such as the Delphi methodology [23]. During a Delphi process answers of interviewees remain anonymous, encouraging free expression of opinion and controlling for influence of reputation or authority or certain experts. If the level of consensus reached by the Expert Group after the second round is unsatisfactory, or unsatisfactory for a subset of questions, further rounds can be conducted also for specific parts of the questionnaire until an adequate level of consensus is reached (cf. Figure 4).

Figure 4: Expert consultation based on Delphi methodology



Source: GÖ FP

The Delphi consultation was built upon the results of the literature review and consisted of two rounds. The first Delphi round was conducted face-to-face during an expert workshop (organised in EC premises in Brussels) aiming to discuss the policy actions derived by the literature review and the related results of the feasibility study (i.e. the results of the SWOT analysis). Based on this, main elements to be included in the policy recommendations were identified by the Expert Group during the workshop. The second Delphi round was done electronically and aimed at validating the policy recommendations that were drafted based on the results of the Expert Group workshop. By the end of the project period the Expert Group had the chance to have a final look at the recommendations before sharing them with the eHealth Network.

2.2.3 Public consultation

The policy recommendations that were framed through literature review and the consultation of experts during the workshop, were refined through an additional public consultation loop. This was achieved by presenting the recommendations to a wider audience of experts and stakeholders at public conferences and/or meetings in the field of public health, eHealth and related fields.

The draft recommendations were presented for comments at three public conferences and meetings respectively, which are known to attract experts and stakeholders in the relevant fields. Table 4 provides an overview of conferences/meetings in 2016 at which the draft policy recommendations were presented and discussed.

Table 4: Conferences/meetings for public consultation

Conference / meeting	Place	Date (2016)	Web link
eHealth Week	Amsterdam, The Netherlands	June 8 – 10	http://www.ehealthweek.org/
2nd International Health Congress	London, UK	June 20 - 22	http://www.globalhealthcongress.org/
7th World Congress on Healthcare & Technologies	London, UK	September 26 - 27	http://healthcare.global-summit.com/europe/registration.php

Source: GÖ FP

The comments from the participants and main discussion points from the three conferences/meetings were recorded in written by the authors. Based on the received inputs the authors prepared the final set of recommendations.

In addition to the public consultation, the policy recommendations were validated internally by several feedback loops within the European Commission (including DG SANTE, DG CONNECT, DG RTD).

3 Results

3.1 Added-value of Big Data use in Health

The systematic literature search in Medline, Cochrane Databases, Embase and Scopus yielded a total of 588 abstracts (duplicates were excluded). The hand search delivered in total 64 publications.

3.1.1 Shortlist of examples on the use of Big Data in Health

After the selection of the relevant abstracts and the hand search, a shortlist of 20 examples for the use of Big Data in Health was defined and agreed upon, incorporating the feedback received by DG SANTE and DG CONNECT:

- Comet K-Project DEXHELPP – AT
- The Shared Care Platform – DK
- E-Estonia – National Identity Scheme – EE
- AEGLE (An analytics framework for integrated and personalized healthcare services in Europe) – UK, IT, GR, SE, BE, NL, PT, FR
- The Business Intelligence database system – GR
- PASSI (Progressi delle Aziende Sanitarie) – IT
- Arno Observatory – IT
- The Swedish Big Data Analytic Network – SE
- Clinical Practice Research Datalink (CPRD) – UK
- Sentinel Stroke National Audit Programme (SSNAP) – UK
- Hospital Episode Statistics (HES) – UK (England)
- The YODA Project (Yale University open data access) – US
- FDA Adverse Event Network Analyser - US
- CEPHOS-LINK – FI, AT, RO, NO, SI, IT
- Twitter (Adverse drug reactions and public health) – International
- Flatiron – US
- UK Biobank – UK
- Semantic Data Platform for Healthcare (SEMCARE) – DE, NL, AT, UK, ES
- Integrated BioBank of Luxembourg (IBBL) – LU
- Spanish Rare Diseases Registries Research Network (SpainRDR) – ES

3.1.2 Priorities of Big Data in Health and their added-value

In order to select ten priorities of practical use of Big Data in Health, the contractor has identified the added-value of the shortlisted examples on Big Data use in Health by the literature search described in section 2.1. Based on the added-value identified, ten priorities have been selected. An overview of results is presented in Table 5. Detailed results for all shortlisted examples can be derived from Annex 4 (section 6.4).

Table 5: Overview of initiative's added-value

	Sustainability of health systems	Quality and effectiveness of treatment	Chronic disease	Healthy lifestyle	Field(s) of application	Quality of identified sources
Comet K-Project DEXHELPP	x				▪ Health system and service research	High quality
The Shared Care Platform		x	x		▪ Health system and service research	High quality
E-Estonia – National Identity Scheme	x				▪ Health system and service research ▪ Epidemiology	High quality
AEGLE		x			▪ Health system and service research	High quality
Greek e-Prescription System	x				▪ Health system and service research ▪ Surveillance (Pharmaceutical) ▪ Epidemiology	Moderate quality
PASSI				x	▪ Surveillance (Public health) ▪ Epidemiology	Moderate quality
Arno Observatory	x	x	x		▪ Health system and service research ▪ Epidemiology ▪ Surveillance (Public health)	Low quality
The Swedish Big Data Analytic Network	x				▪ Health system and service research ▪ Epidemiology	Low quality
Clinical Practice Research Data link (CPRD)	x	x			▪ Clinical research ▪ Health system and service research	High quality
Sentinel Stroke National Audit Programme		x			▪ Clinical practice ▪ Health system and service research	Low quality
Health Episode Statistics	x	x			▪ Health system and service research	Moderate quality
The YODA Project		x	x		▪ Clinical research	Unclear quality
FDA Adverse Event Network Analyser		x			▪ Surveillance (pharmaceutical)	

	Sustainability of health systems	Quality and effectiveness of treatment	Chronic disease	Healthy lifestyle	Field(s) of application	Quality of identified sources
CEPHOS-LINK	x	x			▪ Health System and Service Research ▪ Clinical Practice	High quality
Twitter (Adverse drug reactions and Public Health)		x			▪ Surveillance (Pharmaceutical) Surveillance (Public health)	Moderate quality
Flatiron		x			▪ Health system and service research and clinical research	High quality
UK Biobank	x	x			▪ Surveillance (Public health), Clinical and Epidemiological research	High quality
SEMCARE		x			▪ health service research, epidemiology and clinical research	High quality
Integrated BioBank of Luxembourg (IBBL)		x			▪ clinical research	Moderate quality
Spanish Rare Diseases Registries Research Network (SpainRDR)	x	x			▪ Health system and service research and Epidemiology	High quality

Source: GÖ FP

Based on the evaluation results of the added-value and the quality assessment of the underlying sources, the following examples of Big Data use in Health were selected as priorities:

- Comet K-Project DEXHELPP
- The Shared Care Platform
- E-Estonia
- ARNO observatory
- PASSI
- Health Episode Statistics
- The YODA Project
- CEPHOS-LINK
- Flatiron
- Spanish Rare Diseases Registries Research Network (SpainRDR)

The added-value reflects the contribution of the 10 priorities in supporting the sustainability of health systems, in improving the quality and effectiveness of treatment, in combating chronic disease and in supporting healthy lifestyles. In the selection, the authors aimed for a balanced set of examples regarding the field of application and added value to health. Furthermore, it was tried to include different technological approaches to cover the whole range of possibilities of Big Data in Health. Examples where no or only low quality information sources could be found, were not regarded in the priority selection.

Comet K-Project DEXHELPP - AT

The initiative's field of application falls within *health system and service research* and deals with different routinely collected data sources ranging from epidemiological to cost data.

"Decision Support for Health Policy and Planning: Methods, Models and Technologies based on Existing Health Care Data" (DEXHELPP) is a project of the Austrian COMET-K (Competence Centre for Excellent Technologies) supported by two Austrian ministries and the City of Vienna. The Vienna University of Technology coordinates DEXHELPP which is conducted in collaboration with ten Austrian partners. The project aims to develop new methods, models and technologies in order to analyze the status-quo of the health care system, forecast future developments and compare scenarios based on different interventions within the health care sector. The overall objective of all these subprojects is to support health policy and planning. Within the project, existing data sets as well as new data sources are used and data will be enhanced and linked [22]. DEXHELPP focuses on a variety of relevant interdisciplinary topics with questions ranging from data security and data management, statistical methods, causal inference, mathematical and decision-analytic modelling and simulation to visualization of data and public health [23]. One main application project is the development of a research server incorporating different sources of routine data, which aims to facilitate the exchange and storage of data in a secure way, provide a safe environment to test developed methods and to improve research [24].

Table 6: The added value of Dexhelpp

Patient	Provider	Policy	Research
▪ not applicable for this example	▪ not applicable for this example	▪ target-oriented healthcare planning ▪ anticipation of future developments in order to set timely actions	▪ linkage of different data sources ▪ combines different perspectives through collaboration of various partners
Overall added value: Sustainability of health systems			

Source: GÖ-FP based on DEXHELPP [22, 24]

The Shared Care Platform - DK

The initiative's field of application falls within *health system and service research* and deals with routinely collected data from the health and social care providers' individual IT systems.

The Shared Care Platform is an internet IT platform that supports a cross sector collaboration within healthcare by facilitating the coordination between the general practitioner, the municipality and the hospital [25]. The project was funded by the national fund for chronic diseases and developed by the Region of Southern Denmark in cooperation with IBM. At the moment the Shared Care Platform focuses on patients with chronic illnesses but it is planned to broaden its application [25, 26].

The platform collects data from the health and social care providers' individual IT systems, which represents the basis for a common treatment plan for the patient. Moreover, patients have access to their own data on their computer, tablet or smartphone and can add additional data to the system e.g. by answering questionnaires or sending their vital monitoring information collected at home. Data entered into the system by the patient is collected and the origin of the data and the responsible author can be traced [26, 27].

Data stored in the Shared Care Platform can be printed and analysed [26]. Therefore, the health care resources can be focused and used for patients where a disease is not treated accordingly. With the Shared Care Platform health and social care providers are enabled to offer a coherent course of treatment for patients.

Table 7: The added value of The Shared Care Platform

Patient	Provider	Policy	Research
▪ a coherent course of treatment ▪ better informed patient ▪ home monitoring of patients	▪ comprehensive picture of all medical records ▪ well-coordinated treatment facilitated by collaboration of various health care providers and the municipalities	▪ efficient use of resources	▪ not applicable for this example
Overall added value: Combating chronic disease and improving quality and effectiveness of treatment			

Source: GÖ-FP based on the Shared Care Platform [25-27]

E-Estonia National Identity Scheme - EE

E-Estonia's field of application falls within *health system and service research* as well as e*pidemiology* and deals among other things with routinely collected patient record data.

In 2005, the Estonian eHealth Foundation[5] was established with the aim to promote and develop national e-solutions within the health care system. During the period 2005-2008 e-health projects like the e-Health Record, digital registration, digital image, and e-Prescription have been developed. The common aim of these projects was to decrease the bureaucracy in physician's work flows and allow for a more efficient distribution of work time, to make medical information accessible for physicians and to make health care services more patient-friendly. The projects have been funded between the years 2005-2008 by the EU Structural Funds [28-30]. In 2010, e-prescription and e-Health Record was implemented as part of E-Estonia[6] throughout Estonia. The e-Health record is a nationwide system that integrates data from different health care providers into a common patient record. The e-Health Record comprises information on diagnoses, physician visits, tests (including image files), inpatient treatments as well as medication prescribed. Also, data is compiled for national statistics, in order to measure health trends, track epidemics and to ensure wise spending of resources. Patients can access their own records through an online patient portal. E-Prescription is a centralized system for issuing and handling medical prescriptions. All hospitals and pharmacies are connected to the system. Prescriptions are filled by presenting an ID card. Routine refills can be issued without a prior visit of the physician, but via e-mail, Skype or phone. Further, state medical subsidies, patients are entitled to, are discounted automatically, as the system draws on data of the national health insurance fund [31-33].

The backbone of all Estonian e-services including the e-health services is the so called X-Road. It is an environment which allows the nation's various e-services databases, both in the public and private sector, to link up and operate in harmony. Thus, it is not a centralized national database, but retrieves data from various providers using different systems, and presents it in a standardized format. As a consequence, there is no single owner or controller, and every government agency or business is free to choose an IT solution which fits their requirements best [34].

Table 8: The added value of Estonian E-Health

Patient	Provider	Policy	Research
▪ better informed patients through comprehensive overview of medical records	▪ less paperwork ▪ improved quality and efficiency of health services	▪ improved health care planning and administration facilitated by comprehensive and accurate medical statistics	▪ improvement of medical statistics
Overall added value: Sustainability of health systems			

Source: GÖ-FP based on E-Estonia [35]

[5] The Estonian eHealth Foundation was established on October 18 of 2005 by: Ministry of Social Affairs of Estonia, North Estonia Medical Centre, Tartu University Hospital Foundation, East Tallinn Central Hospital, Estonian Hospitals Association, The Estonian Society of Family Doctors, Union of Estonian Emergency Medical Services [28]

[6] E-Estonia is the umbrella term for the Estonian government's efforts to facilitate citizen interactions with the state through the use of a wide range of electronic solutions. Services included under the initiative are: e-Police, e-Residency, e-School, e-Tax. e-Voting, etc. These services can be accessed either by the ID card which almost 90 per cent of Estonian citizens owned in 2012 or by mobile phone [31].

ARNO Observatory - IT

The ARNO Observatory's field of application falls within *Health System and Service Research, Surveillance (Public health)* as well as *Epidemiology* and deals with routinely collected clinical and administrative patient data.

The ARNO Observatory – a network system for the epidemiological and economic surveillance – was implemented in 1987 by CINECA, a non-profit Consortium made up of 70 Italian universities, four Italian research institutions and the Italian Ministry of Education [36]. The database has been designed to combine and aggregate huge masses of administrative patient data: pharmaceutical prescriptions, hospital discharges, medical home services, diagnostic examinations, laboratory analyses. This information is linked to other data flows from different databases (e.g. the GP's registry, population registry, pharmacies' registry, National Vital Statistics, National Drugs Formularies). Through the epidemiological orientation of the ARNO Observatory, great emphasis is given to data quality. By cross-checking of the original data bases coming from the different areas a population oriented database could be created. This ensures not only a high degree of confidence for general analyses but also for stratified analyses. Further, by linking data from different sources, the ARNO Observatory is able to build comparable epidemiological and economic indicators. Thus, it can provide the Italian Local Health Units with homogeneous data derived from different geographical areas. In 2007, the ARNO Observatory involved seven Italian regions (i.e. Venetia, Liguria, Tuscany, Lazio, Abruzzo, Marches and Campania) comprising 30 Local Health Units and almost 11 million inhabitants. The ARNO Working Group is composed of Local Health Authorities, the Italian Society of hospital Pharmacy and the Mario Negri Institute for Pharmacological Research. Thus, it is a data warehouse [37]. According to the literature identified, one field that utilizes data of the ARNO Observatory is diabetes research [38, 39].

Table 9: The added value of the ARNO Observatory

Patient	Provider	Policy	Research
not applicable for this example	▪ benchmarking with other providers ▪ evaluation of clinical practice	▪ monitoring and verifying the impact of interventions on quality and costs ▪ target-oriented healthcare planning	▪ wide range of quantitative research in the field of health system, health service ▪ identification of cohorts of population
Overall added-value: Improving sustainability of health systems, quality and effectiveness of treatment, and potentially combating chronic diseases			

Source: GÖ-FP based on ARNO Observatory [37-39]

PASSI (Progressi delle Aziende Sanitarie) - IT

The initiative's field of application falls within *surveillance (public health)* as well as *epidemiology* and deals with collected data about adult behavioral risk factors and preventive measures.

PASSI is the surveillance system for behavioral risk factors in Italy and in place since 2006. The main objective of PASSI is to estimate the frequency and evolution of behavioral risk factors for health and the diffusion of preventative measures, over time. Production and dissemination of information relevant to public health professionals and communities are PASSI's priorities. This information can be used for designing, implementing and assessing public health actions. The system is based on an ongoing nationwide collection of data using a standardized questionnaire and is coordinated on a national level by the National Institute of Public Health. Participating local health units (LHU) collect data on a monthly basis via telephone interviews of a random sample of

resident adults aged 18 to 69 years [40]. In 2012, 93 % of LHUs participated which means that 90 % of the adult Italian population were covered [41]. The questionnaire comprises a variety of topics related to health behavior and prevention, which are all stated as priorities in the Italian National Health Plan. Particular attention is given to subjective aspects, such as the respondents' perceptions, opinions, knowledge, and attitudes about health behaviors and whether their doctors provide them with appropriate medical advice. Many questions are provided only to specific population subgroups. Data is transmitted to a national coordinating center, where it is cleaned, managed, and made available for local, regional, and national analysis. Moreover, data quality is routinely monitored [41, 42].

Table 10: The added value of PASSI

Patient	Provider	Policy	Research
▪ specific public health actions for individual risk behaviour	▪ Not applicable for this example	▪ targeted planning of public health actions	▪ Data base for research about behavioural risk factors and the effectiveness of preventive measures ▪ analyses for population subgroups, different geographic aggregation levels
Overall added value: Contributing to healthy life styles			

Source: GÖ-FP based on PASSI [40-42]

Hospital Episode Statistics (HES) – UK

The Hospital Episode Statistics' field of application falls within h*ealth system and service research* and consists of routinely collected patient record data in hospitals.

HES was designed in 1987 as data warehouse containing detailed information on admissions, outpatient appointments and accident and emergency (A&E) attendances for secondary non-clinical purpose use, including the basis for hospital payment. Thus, HES is a system based on patient records. All National Health Service (NHS) trusts in England are covered by the system, including acute hospitals, primary care trusts and mental health trusts [43]. Also, private patients treated in NHS hospitals and patients with residency outside of England who received treatment are covered by the system. Each patient record includes data on: clinical information (i.e. about diagnosis and operations), patient information (i.e. age, gender, and ethnicity), administrative information (i.e. waiting time, dates, and methods of admission and discharge) and geographical information (i.e. treatment place and area of residency). HES provides admitted patient care data from 1989, outpatient attendance data from 2003, and A&E data from 2007 onwards [44]. In terms of data security and patient confidentiality, all data is stored in a secure data warehouse and strict statistical disclosure control is applied to all published HES data, so that patients cannot identify themselves or others. Since 2006, HES is the responsibility of the Secondary Uses Service, run by the Health and Social Care Information Centre and the National Programme for IT, who publish a number of standard analyses on a regular basis [43].

Table 11: The added value of the Hospital Episode Statistics

Patient	Provider	Policy	Research
▪ not applicable for this example	▪ benchmarking with other NHS hospitals in England ▪ assessment of effective delivery of care	▪ development, monitor and evaluate government policy ▪ development of national clinical quality indicators ▪ support of local and national service planning	▪ analysis of health trends over time ▪ monitoring trends in NHS hospital activity ▪ revealing of health trends over time

Overall added value: Sustainability of health systems and improving quality and effectiveness of treatment

Source: GÖ-FP based on Hospital Episode Statistics [43, 44]

Yale University Open Data Access (YODA) Project – USA

The YODA Project's field of application falls within *clinical research* and deals with participant-level clinical research data and/or comprehensive clinical research reports.

The YODA Project was initiated with the aim of facilitating the access to participant-level clinical research data and/or clinical research reports, which are more detailed than in journal publications, in order to promote scientific research. By these means the YODA project addresses the problem that clinical evidence is often not at all, selectively, deferred or incompletely published. The project is run by a group of academically-based clinical researchers who partner with data holders (e.g. Medtronic, Inc. and Johnson & Johnson) to access their clinical trial programme's data. Thus, the YODA Project is the mediating authority between third party data owners and the general research society [45, 46] and yields the potential to advance science or improve individual and public health as well as health care delivery. The provided clinical research data is diverse, ranging from infectious diseases (e.g. HIV, Tuberculosis) and cancer to chronic diseases (e.g. diabetes). Until now about 125 clinical trials can be accessed via the YODA project (February 2016).

Table 12: The added value of the YODA Project

Patient	Provider	Policy	Research
▪ not applicable for this example	▪ not applicable for this example	▪ not applicable for this example	▪ sharing of clinical research data ▪ transparency of clinical research

Overall added value: Improving quality and effectiveness of treatment and potential to combat chronic disease. Limitation of added value due to the dependency of clinical research data provision by data holders.

Source: GÖ-FP based on the YODA Project

CEPHOS-LINK – FI, AT, RO, NO, SI, IT

CEPHOS-LINK's field of application falls within h*ealth system and service research* as well as *clinical practice* and deals with routinely collected patient record data.

CEPHOS-LINK (Comparative Effectiveness Research on Psychiatric Hospitalisation by Record Linkage of Large Administrative Data Sets) is a research project investigating psychiatric services in six European countries (i.e. Finland, Austria, Romania, Norway,

Slovenia and Italy) between April 2014 and March 2017. In many EU countries, psychiatric hospital admissions and re-admissions are high and the reasons for variations in re-admissions are not clear. As frequent and unplanned re-admissions might be an indication for poor or inefficient psychiatric care, CEPHOS-LINK aims to investigate psychiatric re-hospitalisations by comparing different types of interventions focusing on differences in rehospitalisation outcomes of adult psychiatric patients. Specifically, it explores the relationship between factors such as patient, service and health system on the re-hospitalisation of discharged psychiatric inpatients. Further, it compares the outcome of service use patterns. In order to do so, it uses patient record linkage methods for large data sets of administrative electronic healthcare databases of six EU countries. The project is led by the Finnish National Institute for Health and Welfare and funded by the EU's Seventh Framework Programme [47-49].

Table 13: The added value of CEPHOS-LINK

Patient	Provider	Policy	Research
▪ *not applicable for this example*	▪ benchmarking with other providers in different countries	▪ harmonization of psychiatric services across Europe ▪ facilitating the development of psychiatric practice guidelines	▪ availability of robust and linked data
Overall added value: Improving quality and effectiveness of treatment and sustainability of health systems			

Source: GÖ-FP based on CEPHOS-LINK [47-49]

Flatiron – US

The initiative's field of application falls within *health system and service research and clinical research* and deals with routinely collected data from electronic medical record systems.

Flatiron is a start-up company based in New York, which was founded in 2012 with the objective to build up the world's largest cancer database by collecting a huge amount of clinical data. The company has already developed a cancer-focused data analytics platform (disruptive software platform), which is called OncologyCloud platform. It aggregates relevant data from electronic health record (EHR) systems, standardizes it and organizes it in order to gain knowledge about which treatments work best in treating cancers. In this way, EMR data becomes usable for research and analytics. The difficulty with EMR data is that it usually comes from many different sources. For a single patient data can come from internists, oncologists, radiologists, surgeons, laboratory and pathology reports and more. Even if this data is digitalized it is often in an unstructured format. [50].

The OncologyCloud software suite was designed by a team of leading oncologists and software engineers and supports the entire scope of clinical workflow and delivery of patient care. Around 200 cancer centers across the US are connected to the software platform helping to support the treatment of nearly one million active cancer patients [51]. The software suite includes "OncoEMR" which is an advanced EMR system with several features [52], the "OncoBilling", which provides an integrated system for filing and managing claims with insurance companies within "OncoEMR", and the "OncoAnalytics", which helps providers to gain detailed clinical insights from their EMR and practice management systems in near real-time [53]. For interactions with patients the SeeYourChart was developed, which is a cloud-based communication portal for patients. In that way, clincial documents, laboratory results, an appointment calendar and educational materials can be shared with the patient [54].

Flatiron cooperates with various clinical and life science companies such as "Guardant health", which is a company that developed a cancer blood test that provides a cheaper and less painful way of accessing genomic information about a tumor compared to the standard way of repeated biopsies. Results from these tests are assimilated into the OncologyCloud [55]. Through a collaboration with <Foundation Medicine, Inc.> they integrated the Flatiron Health OncologyCloud platform with Foundation Medicine's comprehensive genomic profiling capabilities. In the future, life science companies will be able to utilize this cloud-based platform to enable better selection of molecular candidates, more efficient clinical trial design and faster patient recruitment into clinical trials. [56]

In 2014 Google invested into the company with more than $100 million via Google Ventures and combined with money from other investors they raised around $138 million [50].

Table 14: The added value of Flatiron

Patient	Provider	Policy	Research
▪ access to clinical documents, lab results, appointment calendar and educational material	▪ sharing of clinical documents, lab results, appointment calendar and educational material directly with patient	▪ Not applicable for this example	▪ simplified identification of suitable candidates for clinical studies
Overall added value: Improving the quality and effectiveness of treatment			

Source: GÖ-FP based on [50-56]

Spanish Rare Diseases Registries Research Network (SpainRDR) – ES

https://spainrdr.isciii.es/en/Pages/About.aspxThe SpainRDR's field of application falls within *health system and service research* as well as *epidemiology* and deals with data from various registries including epidemiological and clinical data.

For developing clinical research with focus on rare diseases (RD), it is essential to have registries as they facilitate the recruitment of suitable patients for the launch of studies in order to explore disease etiology, pathogenesis, diagnosis or therapy. Moreover, also existing information from registries and other sources is often sealed in different places and difficult to access, complicating research in the area of rare diseases [57, 58].

This is why the SpainRDR was established within the scope of the Spanish call for the international rare diseases research consortium. The development of the SpainRDR was financed by the Institute of Health Carlos III with 2.4 million Euro for the years 2012 to 2014. [57]

SpainRDR aims at providing a central platform with access to information and data for health policy making and clinical research. Data included in this platform is coming from two types of sources. The first type is patient registries, which were built up for a group of diseases or one specific disease in order to conduct patient outcome research. The second type is population-based registries, which were established to conduct epidemiologic research and social-health planning. These registries were already set up by the autonomous regions in Spain. Also mortality registries, health insurance card databases and electronic hospital records and other sources will be integrated. Data included in these different registries will ultimately be harmonized and combined into one comprehensive platform. Moreover, the national rare diseases registry is linked to the national biobank of rare diseases. [57, 59]

This project involves a variety of partners including all Health Departments of the Autonomous Communities (regions) of Spain, the Spanish Ministry of Health and the Spanish Centre of Reference of People and Families affected by RD (CREER). Moreover, six Spanish Medical Societies, four research networks, pharmaceutical and biotechnological organizations (ASEBIO and FARMAINDUSTRIA), the Spanish Federation of RD (FEDER) and its foundation (FEDER TELETHON FOUNDATION) are cooperating. The project is coordinated and led by the Institute of Rare Diseases Research (IIER). [57]

With this new comprehensive registry prevention, diagnosis, prognosis, treatment and quality of life for RD patients can be improved due to high quality information. This will facilitate the implementation of RD-oriented health and social policies and promote transnational research.

Table 15: The added value of SpainRDR

Patient	Provider	Policy	Research
▪ faster diagnosis of rare diseases	not applicable for this example	▪ facilitates RD-oriented health policy making ▪ facilitates RD-oriented social policies	▪ simplifies identification of suitable candidates for clinical studies ▪ comprehensive data for research on rare diseases
Overall added value: Improving the quality and effectiveness of treatment and sustainability of treatment			

Source: GÖ-FP based on SpainRDR [57-59]

3.2 Fields of policy actions for Big Data in Health

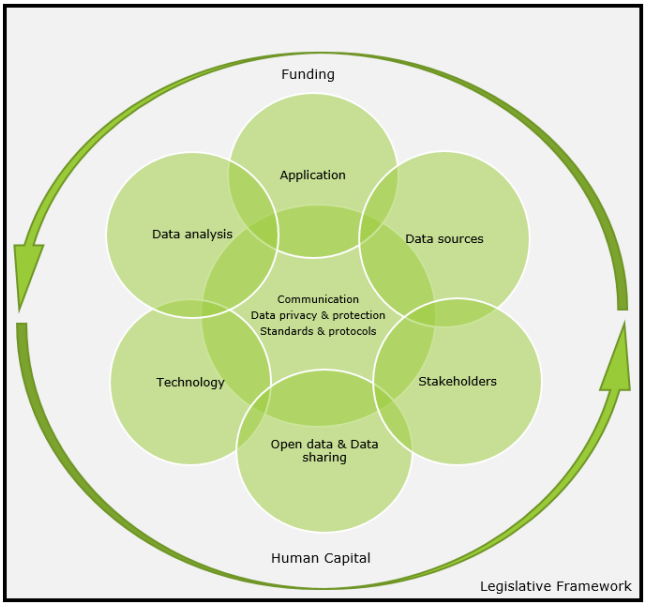
The literature research described in section 2.1 not only aimed at selecting the prioritized examples of use of Big Data in Health but also served as the basis for the identification of policy actions. The identified policy actions for Big Data in Health were clustered around twelve fields:

13. Legal aspects
14. Stakeholders
15. Privacy and data protection
16. Open data and data sharing
17. Standards and protocols
18. Technological development
19. Data sources
20. Data analysis
21. Applications
22. Communication
23. Human capital
24. Funding

The authors are aware of the fact that the distinction between these fields of policy actions is not always perfectly precise, and that other distinctions (broader of more narrow) are conceivable. However, based on the literature, these were the fields that gradually became apparent when clustering the identified single policy actions.

The following Figure 5 gives an overview of the relevant fields and their potential intersections.

Figure 5: Fields of policy actions and their interaction



Source: GÖ FP

For internal critical reflection of the feasibility of implementing the policy actions, a SWOT analysis was performed. The results of the SWOT analysis also served as basis for the drafting of policy recommendations in the further course of the study. Detailed results of the SWOT analysis are presented in Annex 5 (section 6.5).

For external critical reflection and validation of the policy actions, the study's expert panel was consulted. In an expert workshop, the proposed policy actions were discussed and approved by the study's experts (cf. section 3.2). Detailed results of the expert workshop are provided in Annex 6 (section 6.6)

Field 1: Legal aspects

Secure data generation and sharing, accessibility of data for patients and health care providers, and patient confidentiality concerns needs to be addressed in a legislative framework. Legal frameworks and internal policies should be aligned and clarified, specifically regarind:

- privacy and data-ownership,
- secondary use of health data,
- cloud services and
- institutions hosting and managing Electronic Health Records (EHR)

Furthermore, patient confidentiality concerns might be addressed by the EC through amendments to existing data protection directives and regulations, especially the General Data Protection Regulation[7] provides a sound framework and tackles some of the above mentioned topics. [60-63]

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Field 2: Stakeholders

Stakeholders' roles and responsibilities, the engagement of and with individuals as well as partnerships in the field of Big Data should be addressed.

The establishment of expert teams and/or decision making bodies both at EU level and at user level is recommended in order to oversee health data quality initiatives (EU level) and for defining and prioritizing key data needs (user level). Further, roles and responsibilities across the data value chain need to be clear and designed in an adaptable and flexible way avoiding single actor responsibilities (e.g. for collection, use, security, data quality).

Roles and responsibilities regarding data governance, which refers to the overall management of data regarding availability, usability, integrity and security, need to be addressed. Specifically, EU should have a clearly defined role in health data governance and management. Stakeholder interests should be considered and individuals should be engaged in planning, executing and improving Big Data efforts in order to ensure their success.

Partnerships, such as public-private-partnerships and partnerships between physicians and Big Data behavioural scientists, need to be supported for accelerating Big Data use in health practice. [4, 60-62, 64-66]

Field 3: Privacy and data protection

Privacy and data protection policies should address personal data protection policy, a protective design of big data systems and incentives for enhancing data privacy.

Specifically, consensus on privacy breaches should be established and gaps in legal protection potentially causing harm need to be identified. Furthermore, a comprehensive and coherent policy on personal data protection and security needs to be drafted addressing:

- standards for data ownership and control (i.e. opt-out clauses),
- purpose specification and use limitation,
- confidentiality,
- data access (i.e. for patients and health professionals),
- cloud services,
- storage and processing, including archiving durations of data, and
- re-use and cross-border flow of data.

Thereby, emphasis should be placed on new and innovative data security solutions. Highest ethical standards have to be adopted (e.g. a code for responsible analytics), and multipurpose consent models addressing patients and data holders should be developed, especially in biomedical and genomic research, in order to meet ethical and legal requirements. Also, the collective – currently rather critical – mind-set about patient data should be shifted towards a sharing of data with protection, which can be supported by transparent information about the use of health information and incentives for privacy-enhancing technologies and privacy-protecting technical architectures. [4, 60, 62, 63, 65, 67]

Field 4: Open data and data sharing

To enhance, facilitate and promote secure data sharing and open use of data policy actions are needed both at national and at international level. Openness and transparency of government data (including health data) as well as non-proprietary private data

should be promoted and the safe access to public-sector data as well as its re-use should be improved. Data sharing between health care providers (e.g. clinician practices, hospitals, imaging centres) and other settings in which care and services are delivered should be supported through public and private sector incentives and resources, which will enable key data holders to participate in data sharing.

In order to facilitate data sharing

- common protocols for users and resources,
- data architecture,
- governance models for managing and sharing data, and
- mutual recognition procedures

should be established. These issues can be summarized in a common framework for responsible data sharing. [4, 60, 61, 64, 65, 68]

Field 5: Standards and protocols

Standards and protocols that aim to enhance interoperability of different data sets are a highly relevant field for policy action. The adoption of current technical standards[8], policies and best practices should be promoted, and new common ground and core definitions should be established (e.g. for data government and usability, evidence and value, and analytical protocols). Specifically, (international) standards for the interoperability of clinical data, such as EHR or for genome experiment data, should be developed, promoted, and incentivized to allow for pooling of data and comparison of system-level research. Furthermore, quality and outcomes-based protocols are needed in order to align the definition of "what is working" and what constitutes a "better outcome". [60-62, 64-66, 68, 69]

Field 6: Technological development

The development of software solutions that enable smooth data collection and storage as well as data linkage in order to facilitate the extraction of relevant data, its analysis and the communication of findings to relevant parties should be supported. This includes supporting technological solutions that simplify data sharing (for example within organizations), enable faster data transmission as well as the development of innovative software to process EHR (taking into account the issue/integration of text data).

Best practices in converting data analytics into a useful presentation of data for clinicians and consumer need to be shared and encouraged. Moreover, the development of privacy-enhancing technologies should be intensified and its spread should be encouraged. This includes the use of data de-identification and pseudonymisation methods as well as an objective evaluation of statistical methodologies regarding the chances of re-identification of individuals. Furthermore, new data analytic engines, which allow for a performance comparison between providers and networks and the sharing of performance data, need to be developed. Technological developments should be observed and if possible channelled by coordination entities at EU level. [4, 60, 64, 65, 67-69]

[8] Such as file transfer protocols

Field 7: Data sources

Existing sources such as EHR, patient registries and hospital information systems as well as new sources for data collection should be a policy focus. Specifically, EHR systems and their sources of information should be adapted and expanded in order to include necessary information not yet captured. Further, other sources than EHR should be considered for gathering personal health data including:

- data from search engines and web browsing,
- data supplied by individuals in (health-focused) social networks (e.g. Twitter),
- genomic or biomedical data,
- environmental data (e.g. air quality),
- socio economic data,
- individual data supplied by smart and embedded medical devices (including smart phones), and
- remote monitoring applications (e.g. sensors, wearable devices).

Health data of various sources need to be linked in order to provide a comprehensive picture of patients' entire care pathway and care history and statistical methods to use these new sources of data need to be developed. [4, 65, 68, 69]

Field 8: Data analysis

Opportunities of Big Data and the potentials of analysis should be identified, analytical methods improved, and the use of new and innovative analytical methods facilitated.

Specifically, data and analytics should be applied in new, innovative and sophisticated ways. In the development of new methodologies (such as data mining, living laboratories, rich open data repositories) focus should be set on their predictive power and ability to integrate clinical data (e.g. biomedical, genomic data) with contextual, real-world data. In order to foster the use of new Big Data methodologies transparency needs to be increased by spreading awareness, and by understanding the demand for cutting-edge data and analytical tools and techniques. [12, 61, 64-66]

Field 9: Applications

Measured and adequate use and application of Big Data in Health, which implies specific conditions and requirements for data analysis and interpretation, should be explicitly addressed. This is specifically important to reduce the risk of bias, and implies a careful assessment of data accuracy and quality as well as the adequate estimation of error rates before analysing and interpreting Big Data.

Another aspect regarding the application of Big Data in Health is the expansion of application fields which include but are not limited to the following:

- *Health system management analysis* helps inform decision makers regarding workforce and infrastructure planning, fluctuating demand, and the assessment and management of expensive drugs and technologies.
- *Care coordination* can be improved by use of comprehensive and integrated health data
- *Infection surveillance* may give detailed and timely information about disease patterns and epidemiology (e.g. antimicrobial resistance)
- *Performance monitoring* of healthcare products and services can combat fraud
- *Biomedical research* can benefit from the use of Big Data by scaling up scientific enquiry, creating a broader and more reliable evidence base and complementing existing methods in a relatively inexpensive and rapid way.

- *Economic analysis* can help contain expenditures by investigating the cost-effectiveness of new drugs and treatments and evaluate trends related to various cost drivers. [4, 62, 69, 70]

Field 10: Communication

Policies in the field of (public) communication and public relations should promote a strategy of open and transparent exchange of information regarding Big Data and its implications for data privacy and safety. The public as well as healthcare providers, technology vendors and other stakeholders should be informed, about:

- Added value of the use of Big Data in Health
- legal protection and its limits,
- trade-offs between benefits and costs of Big Data,
- algorithms that are used (e.g. what data is included, how is the data collected), and
- the security of health information collection and storage in general.

Using effective communication (1) awareness regarding data privacy and protection mechanisms can be raised, (2) possible public concerns about the misuse of collected data can be addressed and (3) public and patient engagement may be encouraged.

Communication should also help to anchor the idea among all groups of the population that health data is a core asset that can leverage, amplify and generate valuable knowledge. [60-62, 67]

Field 11: Human capital

New demands should be addressed by expanding the capabilities of the workforce by training

- behavioural and health scientists in the use of open-source tools for data analysis,
- (more) computational social scientists as well as clinicians and managers, and
- data scientists in general.

This can be achieved by adapting existing or developing new education and training programmes for data and system management in particular by focussing on technological skills and advanced dynamic/statistical modelling capabilities. Furthermore, Big Data analysis could be integrated in the curricula of medical schools to ensure a minimum degree of skill development in statistics and programming for health care professions. [12, 60, 61, 64, 65]

Field 12: Funding

Funding policies for Big Data in Health should aim to ensure financial sustainability and cost containment. This can be achieved by using multi-source financing and public-private partnerships. In particular (public) investments in

- 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

should be made. Furthermore, the administrative burden of funding and grant support should be kept at a minimum level. [60-62, 65, 68, 69]

3.3 Policy recommendations

Building on the policy actions and the input of the study's Expert Group, the policy recommendations were drafted (face-to-face Delphi) and validated electronically (electronic Delphi). In addition, the authors had the recommendations further validated during three public consultations and several feedback loops with the European Commission (i.e. DG SANTE, DG CONNECT, DG RTD). By considering the comments received during these diverse consultations, the final set of policy recommendations was developed and shared for final commenting with the Expert Group.

The recommendations at hand aim to provide guidelines for the development of an EU Big Data value chain on behalf of European and national decision makers in this field. As already pointed out in the introductory section 0, in this specific context 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.

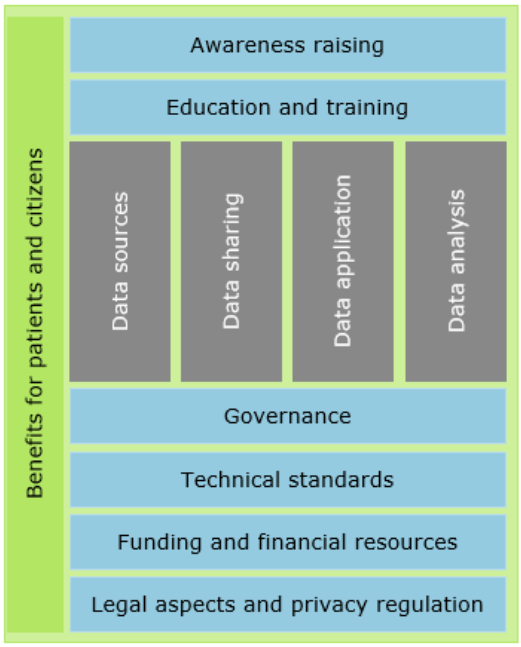
Considering this definition, the recommendations presented in the following aim to benefit European citizens and patients in terms of strengthening their health and improving the performance of MS's health systems. Therefore, the recommendations are explicitly written from a public health perspective.

The following general notions need to be considered for the policy recommendations presented below.

First, the scope of recommendations is to give suggestions for the European Union (EU) and its Member States (MS) on how to utilize the strengths and exploit the opportunities of Big Data for Public Health without compromising privacy or safety of citizens. Second, Big Data in Health should not be seen as a goal in itself, but as a tool to reach certain purposes that benefit the patient and the public. Third, current ethical standards must not be weakened or compromised for potential benefits of Big Data. Fourth, stakeholders need to be included in the implementation of the proposed recommendations and in the production of future recommendations on Big Data in Health. Especially patients (represented by their advocacy groups), who ultimately have to consent to the use of Big Data in Health, have to be involved in the process of producing and implementing recommendations. Despite the importance of patient involvement, all other stakeholders that are part of the data value chain (health professionals, data scientists, health research, industry, public administration, etc.) need to be considered. Fifth, issues related to Big Data in Health that are covered in existing regulatory frameworks (e.g. data protection, informed consent, quality, safety and reliability) are only addressed in the recommendations if distinctively important for the use of Big Data in Health.

Figure 6 depicts the fields of policy recommendations and their respective dimension (vertical, horizontal, and overarching).

Figure 6: Overview of fields of policy recommendations



Source: GÖ FP

Recommendation 1 on Awareness Raising:
Develop and implement a communication strategy to increase the awareness of the added value of Big Data in Health and encourage a positive public mind set towards Big Data in Health

It is of high importance to raise awareness of the practical use of Big Data in Health and its benefits to make it more tangible and understandable for the public and concerned citizens. It is therefore necessary to encourage a positive public mind set towards Big Data in Health by strengthening both the dialogue between the stakeholders in the field and the fact-based information towards the European citizens and patients. Communication and discourse with and within the society, patients, health care providers, technology vendors and other stakeholders should aim for reducing potentially unjustified reservations against the use of Big Data in Health.

It is recommended to develop a sound communication strategy that highlights the scientifically proven added value of Big Data in Public Health and Healthcare, but also addresses possible concerns about data protection and misuse of collected data. Therefore it is essential to:

1) map and align all current national and EU-efforts on communication and awareness raising between the various actors and stakeholders in the field and to the patients and citizens,
2) generate scientific evidence of the added value of Big Data in Health, and
3) anchor the idea that Big Data in Health can leverage, amplify and generate valuable knowledge, which potentially leads to higher quality and efficiency of healthcare for European citizens by giving concrete examples of good practice.

The communication strategy has to take into account all relevant stakeholders, and inform them about the specific benefits that can be realised and the risks that have to be avoided when utilising Big Data in Health. Involving patient organisations in the development of a communication strategy is crucial to (re)establish the trust and confidence of patients in the application of Big Data in Health. Health professionals, decision

makers in health policy, and the industry have to be involved to identify questions that are relevant to them, and to find out if and to what extent the application of Big Data can help answering these questions. It is further recommended that target groups within the main stakeholder groups are identified (e.g. less technophiliac persons) and purposefully addressed.

In order to ensure that the communication strategy reaches citizens and patients, it should be developed bottom-up by each MS accounting for their specific national circumstances. In the preparation of national communication strategies, all relevant stakeholders, especially patients' representatives, need to be involved, in order to safeguard wide commitment of the public. The European Commission can support these national processes by providing best practice examples and toolkits for communication strategies and by mapping existing communication activities. The aim of this should be to stimulate a continuous, open dialogue with all stakeholders and patient groups, which could be fostered further, for instance, by setting up a European platform to exchange experiences.

Recommendation 2 on Education and Training:
Strengthen human capital with respect to the increasing need for a workforce that can utilize the potential of Big Data in Health

Digital health literacy of healthcare professionals and allied health professionals (e.g., managers) should be increased through information and education. To achieve this, existing training and education programmes for public health or healthcare should integrate data handling in the curricula to ensure the development of the necessary skills and competencies.

Increasing numbers of available, accessible and useable sources of Big Data, lead to an increasing demand for human capital. We recognise that in some regions new working fields such as health data analysts are already evolving. Nevertheless, we recommend increasing the training resources in the field of data and system management, advanced statistical analysis and information technologies for scientists using Big Data.

Adapting and augmenting the educational system to strengthen the human capital for Big Data in Health has to be realised on the national level, but can be facilitated by cooperation at the EU level. The European Commission can facilitate the accumulation and knowledge sharing of Big Data in Health including the integration of data handling into the educational curricula of health professionals, by enabling and funding relevant initiatives.

Recommendation 3 on Data Sources:
Expand existing and explore new sources of Big Data in Health and secure their quality and safety

To enhance scientific analysis and relevant applications of Big Data in Health it is recommended to adapt and expand existing Big Data sources (e.g. data repositories in hospitals) in order to include necessary information not yet captured (e.g. biomedical data) and to complement them with newly explored sources. Priority should be given to the assessment of available data sources and their infrastructure to ensure a targeted exploration of new data sources (e.g. wearable health devices and social networks). The outpatient and primary care sector as well as other sectors than health care (e.g. social, labour, environment) should be encompassed in the expansion and exploration of new data sources. The linkage of various health data sources needs to be assured within and between MSs.

The expansion of existing data sources and the exploration of new data sources should be done at MS level. In order to safeguard data quality across the EU, it is important to set standards ensuring quality and reliability of Big Data sets in order to yield reliable results. The dissemination and monitoring of the adherence to these quality standards in the MSs should lie within the responsibility of the appointed health data officers (cf. Recommendation on Governance).

Recommendation 4 on Open Data and Data Sharing:
Promote open use and sharing of Big Data without compromising patients' rights to privacy and confidentiality

Access to complementary sources of Big Data enables improved analytical insights and facilitates data analysis. To utilize this asset, it is recommended to support secure open use and sharing of government data, non-proprietary private data, and data of different healthcare providers for research in public interest on a national and international level.

To support open data and data sharing, a common framework including guidelines for operative processes should be established. This should be done not only at EU-, but on a global level, e.g. considering the Global Alliance for Genomics and Health (GA4GH) white paper on responsible sharing of genomic and clinical data[9]. All relevant elements of usage of Big Data, such as access to data, data safety, data quality, reliability and completeness of data, format and standards (also linked to accreditation), exchange processes and the exploration of possibilities for automated extraction, need to be included in this framework. Furthermore, it has to address the interests of all relevant stakeholders and recognise different cultures regarding data protection in different countries in order to ensure broad support. Data security, data stewardship and data ownership should explicitly be addressed and technological requirements have to be in place to guarantee safe data sharing.

Recommendation 5 on Applications and Purposes:
Target-oriented application of Big Data analysis in health based on the needs and interests of stakeholders including patients

The production of reliable information is important for addressing the concerns that are related to the application of Big Data in Health. Therefore, Big Data in Health needs to be applied target-oriented, where personalised medicine is a good example for the added value it generates. For achieving this, it is recommended to identify the stakeholders that could benefit from Big Data in Health and consider their needs. As stakeholders' interests vary, it is recommended to expand and open up new application fields accordingly (e.g. infection surveillance, biomedical research).

Areas for Big Data applications would be developed in accordance to stakeholders' needs and expert advice at EU level, e.g. by the means of a platform for open dialogue. Particular emphasis should be given to patient benefits subject to ethical considerations. The implementation at MS level can be facilitated by reserving subsidies especially for Big Data application in health.

[9] GA4GH White Paper: https://genomicsandhealth.org/files/public/White%20Paper%20June%203%20final_0.pdf

Recommendation 6 on Data Analysis:
Identify the potentials of Big Data analysis, improve analytical methods and facilitate the use of new and innovative analytical methods

The predictive power of Big Data and the ability to integrate clinical data (e.g. biomedical, genomic data) with contextual, real-world data provide the possibility for high-quality analysis to produce reliable and valid health information. To fully exploit this potential, it is recommended to constantly improve and update existing analytical methods and tools. Furthermore, their development and use (e.g. data mining, living laboratories, and rich open data repositories) should be facilitated while keeping in mind the practical use of Big Data. In cases where new analytical approaches are insufficiently covered by existing ethical framework (e.g. the In Silico method in clinical trials for 'personalised' medicine or treatment) governance structures an independent research committee and/or an independent review board should be established and authorised to oversee research processes.

New analytical methods and tools should be developed by experts of (Big) Data analysis at MS level, and their potential should be shown in pilot studies. Nationally appointed health data officers (cf. Recommendation on Governance) as well as the European Commission can facilitate this development by increasing awareness about the necessity of up-to-date analytical approaches through an adequate communication strategy and by providing financial resources for methodological research.

Recommendation 7 on Governance of Data Access and Use:
Implement governance mechanisms to ensure secure and fair access and use of Big Data for research in health

To utilize the full potential of Big Data in Health, the processing and secondary use of data (facilitated by the General Data Protection Regulation (GDPR)) should be permitted for health research and statistical purposes. However, it is necessary to oversee the purposeful use and the quality of Big Data in Health. We suggest to define roles and responsibilities regarding the access and processing of Big Data in Health as well as to develop core definitions for Big Data in Health governance (e.g., secondary use, data donor cards) in the form of a glossary to encourage a common language. Also, so-called health data officers should be appointed in each MS to oversee and coordinate national activities from a public health perspective. These health data officers should be established within the existing national frameworks and should collaborate with data protection authorities and ethics committees that are already in place at national level. One of their tasks would be, for instance, to make sure that health apps need to strictly be evidence based and not only purely designed in the interest of market entities. A platform for regular dialogue between these health data officers should be established on European level to tackle such questions.

These governance mechanisms need to be transparent and fair. The processing and linkage of data sources as well as access and secondary use should be approved by independent reviewers to ensure non-discriminatory and adequate data access and use. To facilitate this, the appointed national health data officers should coordinate the governance process and foster the cooperation between different database owners and stakeholders. In order to streamline the governance of different data sources in a safe way, a platform for securely linking and accessing data from different sources should be made available on a national level.

On EU-level, the implementation of national governance mechanisms for Big Data in Health can be supported by giving guidance on the process of data access approval and the technical implementation of data platforms, e.g. by providing information on models of good practice for good data governance at research level such as the International

Human Epigenome Consortium (IHEC)[10]. EU-wide harmonized guidelines and operative processes ensure that the rules for data access are similar in different countries and that the interoperability of the technical components (e.g. a platform for linking and accessing data) is possible. Furthermore, the European Commission should enable the continuous exchange of experiences and knowledge between the national health data officers, and inform MSs about new potential security risks and how to prevent them.

Recommendation 8 on Standards:
Develop standards for Big Data in Health to enhance and simplify its application and improve interoperability

By setting common standards across the Big Data value chain in Health, pooling, exchanging and analysing data will become more efficient. It is therefore recommended to adopt, or to develop where non-existent, standards with global scope addressing the issues of interoperability (cf. ICT Standardisation Priorities for the Digital Single Market[11]) e.g. in areas related to patient consent in the use of Big Data in Health or nomenclature of genotyping or ethics to name only a few.

As diverse data representation and formats hamper the easy combination of data sets originating from different sources, introducing standards can improve the interoperability of different data formats, data sets and data warehouses (e.g. following the example of ELIXIR and eTRIKS[12]). Another crucial aspect of improving interoperability is to standardise and harmonise the core content and structure of patient consent forms. Different models of consent (e.g. dynamic consent, enlarged consent) and digitalised consent forms should be considered in order to facilitate secondary use of data and data sharing.

The abovementioned development of standards should build on already existing (international) work, such as the IHE.net initiative[13] or the EXPAND[14] network, wherever possible rather than developing new ones. Consequently, these existing standards need to be aligned before being adopted by the MS national legislation (especially regarding patient consent, cf. Recommendation 7), taking into account the differences in technological development across MSs. Relevant experts and stakeholders should be included in the formulation of these EU level standards. Furthermore, regular updates of standards are important in order to keep up with the rapidly changing technological environment.

Recommendation 9 on Funding and Financial Resources:
Ensure purposeful investment steered by the European Commission to warrant cost-effectiveness and sustainability

Investments in Big Data in Health should eventually yield sustainable social, economic and budgetary returns to all members of society. To spread the initial financial burden and the risk of the investment, multi-source financing (public-private or public-public partnerships) is recommended. In particular, EU level organizations should invest in Big

[10] International Human Epigenome Consortium (IHEC): http://epigenomesportal.ca/ihec/about.html

[11] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. ICT Standardisation Priorities for the Digital Single Market: https://ec.europa.eu/digital-single-market/en/news/communication-ict-standardisation-priorities-digital-single-market

[12] eTRIKS: https://www.etriks.org/wp-content/uploads/2015/11/Code-of-Practice-on-Secondary-Use-of-Medical-Data-with-recognition.pdf

[13] IHE.net (Integrating the Healthcare Enterprise): https://www.ihe.net/About_IHE/

[14] http://www.expandproject.eu

Data system development, a shared data handling infrastructure and a communication strategy (cf. Recommendation 3). National investments in enhanced infrastructure for processing health data and continuous partnerships between various stakeholders should be incentivized.

The EU institutions can support purposeful investment on the national level by informing about existing national Big Data initiatives in the MSs and by issuing guidelines or best-practice examples for an efficient and outcome-enhancing use of Big Data in Health that will benefit all citizens. To steer and monitor these activities it is recommended to provide resources for the establishment of a coordinating body at EU level.

Recommendation 10 on Legal Aspects and Privacy Regulations:
Clarify and align existing legal and privacy regulation of Big Data in Health

A clearly defined and consistent legal framework for Big Data in Health is important in order to provide security when generating, accessing and sharing Big Data and to ensure privacy rights. It is therefore recommended to align the existing legal frameworks and internal policies, especially regarding the aspects of data-ownership, confidentiality of data and patient consent. Further legal aspects that should be addressed are the secondary use of health data (e.g. by introducing data donor cards), storage (e.g. cloud computing) and processing of data as well as legal foundations for re-use and cross-border flow of data.

Privacy and data protection are addressed by clear and coherent regulations regarding data management and control on an EU level. In April 2016, the European Parliament and the council adopted the GDPR, Regulation 2016/679[15], which aims at strengthening the rights of natural persons and to harmonize national laws. This regulation is regarded as the foundation for EU data protection rules and has direct impact on all issues related to Big Data in Health. Moreover, the Directive on security of network and information systems (NIS Directive)[16] was adopted by the European Parliament in July 2016 providing legal measures to increase the level of cybersecurity and the cooperation of MSs regarding this issue. These new stipulations are currently put into place, but their practical implementation needs to be closely followed and monitored by MSs and the European Commission to ensure legal interoperability. Support for a coherent approach in the practical application of this new legislation should be provided by EC Services, e.g. by convening discussions amongst national data protection authorities and healthcare stakeholders on the interpretation before the Regulation comes into force in May 2018.

[15] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[16] Directive 2013/0027 (COD) of the European Parliament and of the Council concerning measures to ensure a high common level of network and information security across the Union

4 Conclusions

The study provides directions for the further development of an EU Big Data value chain on behalf of European and national decision makers in the health domain, among them the members of the eHealth Network.

The mapping exercise of priority activities in the field of Big Data in Public Health, Telemedicine and Healthcare clearly showed the broad variety of initiatives and undertakings in the area. It ranged from single, but complex genomic detection tools to efforts for defining a minimal data set for cross-country exchange of patient data to a still growing industry of health and wellbeing mobile applications. So basically, policy makers are faced with similar challenges for Big Data in Health, as for Big Data in other fields, for instance the need to adopt existing regulations and frameworks to new technologies but also the mind-set of people.

But, as health is a special good, and ethical standards are especially high in this field, a careful and sensitive, though innovation-friendly approach is necessary. It is also crucial to involve patient groups in ongoing processes, as many patients and citizens have concerns regarding the privacy and security of their own health data. Only by their involvement, the chances for shifting the public mind set towards Big Data in Health can be ensured.

The most important lesson learned in this process is therefore that awareness raising regarding the added value of Big Data in Health is needed quite urgently, as pointed out in Recommendation 1, **a communication strategy to encourage a positive public mind set towards Big Data in Health is needed.** The aim of this should be to stimulate a continuous, open dialogue with all stakeholders and patient groups, which could be supported by setting up a European platform to exchange experiences and discuss how to best address current and future challenges.

One task of such a platform could be a discussion on 'low hanging fruits', i.e. areas which could benefit quickest and easiest from Big Data in Health; whether these are chronic diseases, nutrition, cancer diagnosis or treatment or rehabilitation to name only a few.

Another important finding of the study was that a number of experts called MSs and the European Commission to closely follow the practical implementation of the General Data Protection Regulation to ensure legal interoperability. If necessary, support for a coherent approach in the practical application of this new legislation should be provided by European Commission Services. Also, further developments and initiatives such as the European Science Cloud and the Free Flow of Data initiative shall be considered in this field.

Concluding, it is obvious that the public discussion around the 'major' potential benefits and challenges of Big Data in Health is in full flow and will continue for the next years. The ten developed recommendations offer guidance for decision makers and the citizens of the European Union alike on how to best to utilise Big Data in Health for the ultimate benefit of strengthening their health and improving the performance of MS's health systems.

5 References

[1] De Mauro A, Greco M, Grimaldi M. What is big data? A consensual definition and a review of key research topics. 2015;1644: 97-104.

[2] Dutcher J. What is Big Data? https://datascience.berkeley.edu/what-is-big-data/2014, accessed 01.02.2016.

[3] James Manyika MC, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big data: The next frontier for innovation, competition, and productivity. 2011.

[4] OECD. Exploring data-driven innovaiton as a new source of growth: Mapping the policy issues raised by "big data". OECD Publishing, 2013.

[5] IBM. The four V's of Big Data. In: Hub IBDaA, editor. Infographics & Animations: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

[6] European Commission. Report on the use of -omic technologies in the development of personalised medicine Brussels: European Commission, 2013.

[7] Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data. 2013;1(3): 168-75.

[8] Hay SI, George DB, Moyes CL, Brownstein JS. Big Data Opportunities for Global Infectious Disease Surveillance. PLoS Medicine. 2013;10(4): e1001413.

[9] Murdoch TB, Detsky AS. THe inevitable application of big data to health care. JAMA. 2013;309(13): 1351-2.

[10] Klimek P. Big Data, Complexity, und Gesundheit in Österreich. n.d.

[11] Council of the European Union. Council conclusions on open, data-intensive and networked research as a driver for faster and wider innovation. Brussels; 2015.

[12] WHO. FROM INNOVATION TO IMPLEMENTATION: eHealth in the WHO European Region. Copenhagen: WHO Regional Office for Europe, 2016.

[13] Kumar S, Nilsen W, Pavel M, Srivastava M. Mobile health: revolutionizing healthcare through transdisciplinary research. Computer. 2013(1): 28-35.

[14] Onnela J-P. Using Big Data to Improve Individual Health and Health Care Systems. 2015.

[15] Altman RB. Personal Genomic Measurements: The Opportunity for Information Integration. Clinical pharmacology and therapeutics. 2013;93(1): 21-3.

[16] Bernstein S. The US big data report and fully utilising big data within healthcare. eHealth Law and Policy. 2014;1(4).

[17] George G, Haas MR, Pentland A. Big data and management. Academy of Management Journal. 2014;57(2): 321-6.

[18] Khoury MJ, Lam TK, Ioannidis JPA, Hartge P, Spitz MR, Buring JE, et al. Transforming Epidemiology for 21st Century Medicine and Public Health. Cancer epidemiology, biomarkers & prevention. 2013;22(4): 508-16.

[19] Ministry of Health Luxembourg. Conference Report. High-level Conference Making Access to Personalised Medicine a Reality for Patients; Luxembourg, Cercle Cité: Ministry of Health Luxembourg; 2015.

[20] Kessler EH. Encyclopedia of management theory: Sage Publications; 2013.

[21] Thompson A, Strickland AJ. Strategic Management: Concepts and Cases; US: McGraw-Hill; 2003.

[22] DEXHELPP. K-Projekt DEXHELPP. http://www.dexhelpp.at/?q=de/homen.d.

[23] IMEHPS.research. Austrian COMET k-project (Competence Centre for Excellent Technologies) http://www.imehps.at/index.htmln.d.

[24] DEXHELPP. DEXHELPP Research Server. http://www.dex-helpp.at/?q=de/node/93n.d.

[25] News Medical. Southern Denmark to use IBM's Shared Care Platform to improve care for patients with chronic illness,. http://www.news-medical.net/news/20130515/Southern-Denmark-to-use-IBMs-Shared-Care-Platform-to-improve-care-for-patients-with-chronic-illness.aspx2013

[26] Bruun-Rasmussen M, Lehrmann T, Nasser A, editors. A shared care platform for chronic disease management. International Congress on Health Informatics; 2013; Havana. shared care.

[27] European Innovation Partnership on Active and Healthy Ageing. Good Practice Region of Southern Denmark n.d.

[28] Estonian e-Health Foundation. The Foundation. n.d. Available from: http://www.etervis.ee/index.php/en/2012-07-22-13-35-31/organization, accessed 22.02.2016.

[29] Estonian e-Health Foundation. Services. n.d. Available from: http://www.etervis.ee/index.php/en/health-information-system/services, accessed 22.02.2016.

[30] Widén S, Haseltine W. Case Study: The Estonian eHealth and eGovernance System. New York: Access Health Internaitonal, 2015.

[31] e-Estonia.com. Components for Digital Society. n.d. Available from: https://e-estonia.com/components/, accessed 22.02.2016.

[32] Estonia.eu. e-Estonia. n.d. Available from: http://estonia.eu/about-estonia/economy-a-it/e-estonia.html, accessed 22.02.2016.

[33] Rannala R. An Electronic Health Record for every citizen: a global first Estonia. Brussels, 2007.

[34] E-Estonia. X-Road. n.d. Available from: https://e-estonia.com/component/x-road/, accessed 22.02.2016.

[35] Estonian e-Health Foundation. Electronic Health Record. n.d. Available from: http://www.etervis.ee/index.php/en/health-information-system/electronic-health-record, accessed 22.02.2016.

[36] CINEA. About us. n.d. Available from: http://www.cineca.it/en/content/about-us, accessed 15.02.2016.

[37] ARNO Observatory. Network System for the Epidemiological and Economic Surveillance. Introduction. n.d. Available from: https://osservatorioarno.cineca.org/arnoeng.htm, accessed 15.02.2016.

[38] Marchesini G, Bernardi D, Miccoli R, Rossi E, Vaccaro O, De Rosa M, et al. Undertreatment of migrants with diabetes in a universalistic health care system: the ARNO observatory. Nutrition, Metabolism and Cardiovascular Diseases. 2014;24(4): 393-9.

[39] Marchesini G, Forlani G, Rossi E, Berti A, De Rosa M. The direct economic cost of pharmacologically-treated diabetes in Italy-2006. The ARNO observatory. Nutr Metab Cardiovasc Dis. 2011;21(5): 339-46. Epub 2010/02/16.

[40] Baldissera S, Campostrini S, Binkin N, Minardi V, Minelli G, Ferrante G, et al. Features and initial assessment of the Italian Behavioral Risk Factor Surveillance System (PASSI), 2007-2008. Prev Chronic Dis. 2011;8(1): A24. Epub 2010/12/17.

[41] Quarchioni Eea. Socio-economic status and behavioral risk factor in the adult population. Data from the Italian behavioral surveillance system. Italy, 2010-2012. European Journal of Public Health. 2014;Vol. 24, Supplement 2/2014.

[42] EpiCentro. La sorveglianza Passi - Organization. http://www.epicentro.iss.it/passi/en/organization.asp2015

[43] Health and Social Care Information Centre. Hospital Episode Statistics. n.d. Available from: http://www.hscic.gov.uk/hes, accessed 23.02.2016.

[44] Health and Social Care Information Centre. What HES data are available? n.d. Available from: http://www.hscic.gov.uk/hesdata, accessed 24.02.2016.

[45] Krumholz HM, Ross JS, Gross CP, Emanuel EJ, Hodshon B, Ritchie JD, et al. A Historic Moment for Open Science: The Yale University Open Data Access Project and Medtronic. Annals of Internal Medicine. 2013;158(12): 910-1.

[46] The YODA Project. Policies & Procedures to Guide External Investigator Access to Clinical Trial Data. n.d. Available from: http://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data, accessed 23.02.2016.

[47] National Institute for Health and Welfare. Project Background. 2015 Available from: https://www.thl.fi/en/web/thlfi-en/research-and-expertwork/projects-and-programmes/comparative-effectiveness-research-on-psychiatric-hospitalisation/project-background, accessed 22.02.2016.

[48] National Institute for Health and Welfare. CEPHOS-LINK in a nutshell. 2015 Available from: https://www.thl.fi/en/web/thlfi-en/research-and-expertwork/projects-and-programmes/comparative-effectiveness-research-on-psychiatric-hospitalisation/cephos-link-in-a-nutshell, accessed 22.02.2016.

[49] National Institute for Health and Welfare. CEPHOS-LINK. 2015 Available from: https://www.thl.fi/en/web/thlfi-en/research-and-expertwork/projects-and-programmes/comparative-effectiveness-research-on-psychiatric-hospitalisation, accessed 22.02.2016.

[50] Fortune. Can Big Data cure cancer? http://fortune.com/2014/07/24/can-big-data-cure-cancer/2014, accessed 08.03.2016.

[51] Fortune. Flatiron Health's bold proposition to fight cancer with big data. http://fortune.com/2014/06/12/flatiron-healths-bold-proposition-to-fight-cancer-with-big-data/2014, accessed 08.03.2016.

[52] Flatiron. OncoEMR. http://pages.flatiron.com/rs/801-FPC-090/images/Onco-EMR%20%281%29.pdf?mkt_tok=3RkMMJWWfF9wsRonvqzNZKX-onjHpfsX56egrXqW%2BlMI%2F0ER3fOvrPUfGjI4ATspnI%2BSLDwEYGJlv6SgFQrDEMaZw2bgNUhE%3Dn.d.

[53] Flatiron. OncoAnalytics. http://pages.flatiron.com/rs/801-FPC-090/images/OncoAnalyt-ics%20%283%29.pdf?mkt_tok=3RkMMJWWfF9wsRonvqzNZKX-onjHpfsX56egrXqW%2BlMI%2F0ER3fOvrPUfGjI4ATspnI%2BSLDwEYGJlv6SgFQrDEMaZw2bgNUhE%3Dn.d., accessed 08.03.2016.

[54] Flatiron. SeeYourChart. http://www.flatiron.com/products/see_your_chart2015

[55] techcrunch. Guardant Health And Flatiron Health Team Up To Cure Cancer With Big Data. http://techcrunch.com/2015/08/19/guardant-health-and-flatiron-health-team-up-to-cure-cancer-with-big-data/2015

[56] Foundation Medicine I, editor. Foundation Medicine and Flatiron Health Collaborate to Develop First-in-Class Data Platform to Accelerate Precision Medicine for Cancer. 2014; http://investors.foundationmedicine.com/releasedetail.cfm?releaseid=885539. flatiron.

[57] Research IoRD. Executive Summary n.d.

[58] International rare disease research consortium. SpainRDR aims to build the National Rare Diseases Registry in Spain – Q&A with Dr Manuel Posada. http://www.irdirc.org/spainrdr-aims-to-build-the-national-rare-diseases-registry-in-spain-qa-with-dr-manuel-posada/2013 Available from: http://www.irdirc.org/spainrdr-aims-to-build-the-national-rare-diseases-registry-in-spain-qa-with-dr-manuel-posada/

[59] Alonso Vea. National rare diseases registry in Spain: pilot sutdy of the Spanish Rare Diseases Registries Research Network (SpainRDR). Orphanet Journal of Rare Diseases; 2014.

[60] OECD. Data-Driven Innovation, Big Data for Growth and Well-Being. OECD Publishing, 2015.

[61] Pentland A, Reid TG, Heibeck T, editors. Big Data and Health: Revolutionizing medicine and Public Health; Report of the Big Data and Health Working Group 2013. World Innvation Summit for Health (WISH); 2013; DOHA.

[62] Salas-Vega, S., Haimann A, Mossialos E. Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU. Health Systems & Reform. 2015.

[63] European Commission. Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth services, Final report and recommendations. Brussels, 2014.

[64] Bipartisan Policy Center (BPC), editor. A Policy Forum on the Use of Big Data in Health Care. 2013; Washington.

[65] Groves P, Kayyali B, Knott D, Van Kuiken S. The 'big data' revolution in healthcare - Accelerating value and innovation. Center for US Health Sytem Reform - Business Technology Office: McKinsey&Company, 2013.

[66] Yiu C. The Big Data Opportunity: Making government faster, smarter and more personal. London: Policy Exchange, 2012.

[67] Health IT Policy Committee; Privacy and Security Workgroup. Health Big Data Recommendations. USA, 2015.

[68] Strategic Policy Forum on Digital Entrepreneurship. Accelerating the uptake of big data and establishing competitive digital platforms in Europe, draft,. European Commission. Brussels, 2016.

[69] Hoffman S, Podgurski A. Big bad data: law, public health, and biomedical databases. Journal of Law, Medicine & Ethics. 2013;41 Suppl 1: 56-60.

[70] Directorate for employment lasahc. Preperations for the health ministerial meeting and policy forum: bureau, participation and issues notes Organisation for Economic Co-operation and Development. Paris, 2015.

[71] AEGLE. Pioneering Healthcare Future. http://www.aegle-uhealth.eu/en/2015

[72] Soudris Dea. AEGLE: A Big Bio-Data Analytics Framework for Integrated Health-Care Services. 2015.

[73] AEGLE. AEGLE's Partners. http://www.aegle-uhealth.eu/en/a.htmln.d.

[74] Pangalos G, Asimakopoulos D. The New Greek National E-Prescription System: A Powerful Ehealth Tool for Improving Services to Citizens and Administrative Control. DATABASE. n.d.;7: 237-8.

[75] Liatis S, Dafoulas G, Tentolouris N, Kokkinos A, Arvanitis M, Diakoumopoulou E, et al. Nation-wide estimated prevalence of diabetes in Greece: Identification of patients with prescribed pharmacological treatment among 94.5% of the total Greek population. Diabetologia. 2015;58(1): S176-S7.

[76] Sfikakis P, Dafoulas G, Boumpas D, Drosos A, Kitas G, Liossis SN, et al. SAT0361 Large, Nation-Wide Data Analysis-Derived Estimated Prevalence of Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE), and Systemic Sclerosis (SSC) in Caucasians: Insights from the Identification of Patients with Prescribed Pharmacological Treatment Among 7.742.629 Greek Citizens. Annals of the Rheumatic Diseases. 2015;74(Suppl 2): 790.

[77] Pangalos G, Asimakopoulos D, Pagkalos I. The new Greek national e-prescription system: an effective tool for improving quality of care and containing medication costs. Studies in Health Technology & Informatics. 2013;190: 13-7. Epub 2013/07/05.

[78] Pangalos G, Sfyroeras V, Pagkalos I. E-prescription as a tool for improving services and the financial viability of healthcare systems: the case of the Greek national e-prescription system. Int J Electron Healthc. 2014;7(4): 301-14. Epub 2014/08/28.

[79] SEMCARE. The Project. http://semcare.eu/the-project-2/the-project/2016

[80] averbis. Neues EU-Projekt bringt Patienten und klinische Forschung näher zusammen https://averbis.com/presse/neues-eu-projekt-bringt-patienten-und-klinische-forschung-naeher-zusammen/2014

[81] SEMCARE Consortium. Semantic Data Platform for Healthcare, Report on the use and dissemination of foreground,. 2016.

[82] Horizon 2020. IBBL (Integrated Biobank of Luxembourg). https://www.b2match.eu/behealthparis/participants/752016

[83] IBBL. About IBBL. http://www.ibbl.lu/about-ibbl/2016

[84] IBBL. PERSONALISED MEDICINE CONSORTIUM. http://www.ibbl.lu/research/personalised-medicine-consortium/2016

[85] Ginn R, Pimpalkhute P, Nikfarjam A, Patki A, O'Connor K, Sarker A, et al., editors. Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark. Proceedings of the fourth workshop on building and evaluating resources for health and biomedical text processing; 2014: Citeseer.

[86] O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on Twitter? Mining Tweets for Adverse Drug Reactions. AMIA Annual Symposium Proceedings. 2014;2014: 924-33.

[87] Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf. 2014;37(5): 343-50. Epub 2014/04/30.

[88] Broniatowski DA, Paul MJ, Dredze M. National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic. PLoS ONE. 2013;8(12): e83672.

[89] Paul MJ, Dredze M, Broniatowski DA, Generous N, editors. Worldwide influenza surveillance through Twitter. AAAI Workshop on the World Wide Web and Public Health Intelligence; 2015.

[90] Paul MJ, Dredze M. You are what you Tweet: Analyzing Twitter for public health. ICWSM. 2011;20: 265-72.

[91] Royal College of Physicians. Sentinel Stroke National Audit Programme (SSNAP). 2015 Available from: https://www.rcplondon.ac.uk/projects/outputs/sentinel-stroke-national-audit-programme-ssnap, accessed 15.02.2016.

[92] Kavanagh M. Bringing Big Data to Quality Improvement in the Sentinel Stroke National Audit Programme (SSNAP). 2015.

[93] Görnerup O, Gillblad D, Holst A, Bjurling B. Big Data Analytics - A Research and Innovation Agenda for Sweden. In: ICT SS, editor.

[94] CPRD - more dimensions to data. The Clinical Practice Research Datalink,. https://www.cprd.com/intro.asp2016

[95] Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3): 827-36. Epub 2015/06/08.

[96] Botsis T, Scott J, Goud R, Toman P, Sutherland A, Ball R. Novel algorithms for improved pattern recognition using the US FDA Adverse Event Network Analyzer. Studies in Health Technology & Informatics. 2014;205: 1178-82.

6 Annexes

6.1 Annex 1: Search strategies

6.2 Annex 2: Extraction and quality evaluation

6.3 Annex 3: List of participating experts

6.4 Annex 4: Added-value of shortlisted examples of Big Data use in Health

6.5 Annex 5: Feasibility study of policy actions

6.6 Annex 6: Minutes of the Expert Workshop

6.1 Annex 1: Search strategies

Search strategy Medline, Cochrane via OVID

Search date: 27th January 2016

Databases:

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present,

EBM Reviews - Cochrane Database of Systematic Reviews 2005 to January 20, 2016,

EBM Reviews - ACP Journal Club 1991 to January 2016,

EBM Reviews - Database of Abstracts of Reviews of Effects 2nd Quarter 2015,

EBM Reviews - Cochrane Central Register of Controlled Trials December 2015,

EBM Reviews - Cochrane Methodology Register 3rd Quarter 2012,

EBM Reviews - Health Technology Assessment 4th Quarter 2015,

EBM Reviews - NHS Economic Evaluation Database 2nd Quarter 2015

1	(large adj3 public adj3 health adj3 data$).ab,ti.	15
2	(big adj3 data$).ab,ti.	1339
3	(large adj3 health adj3 care adj3 data$).ab,ti.	117
4	1 or 2 or 3	1467
5	exp Public Health/	6434788
6	4 and 5	451
7	limit 6 to yr="2000 -Current" [Limit not valid in DARE; records were retained]	438
8	remove duplicates from 7	**431**

Search strategy Scopus

Search date: 27th January 2016

Database: Scopus

20 ((TITLE-ABS (large PRE/3 health PRE/3 care PRE/3 data)) OR (TITLE-ABS (big PRE/3 data)) OR (TITLE-ABS (large PRE/3 public PRE/3 health PRE/3 data)) OR (INDEXTERMS ("big data"))) AND (INDEXTERMS (public health)) AND NOT INDEX (medline) AND PUBYEAR > 1999
148 document results

19 ((TITLE-ABS (large PRE/3 health PRE/3 care PRE/3 data)) OR (TITLE-ABS (big PRE/3 data)) OR (TITLE-ABS (large PRE/3 public PRE/3 health PRE/3 data)) OR (INDEXTERMS ("big data"))) AND (INDEXTERMS (public health)) AND NOT INDEX (medline) 148 document results

18 ((TITLE-ABS (large PRE/3 health PRE/3 care PRE/3 data)) OR (TITLE-ABS (big PRE/3 data)) OR (TITLE-ABS (large PRE/3 public PRE/3 health PRE/3 data)) OR (INDEXTERMS ("big data"))) AND (INDEXTERMS (public health))
148 document results

17	INDEXTERMS (public health)	306,825 document results
13	(TITLE-ABS (large PRE/3 health PRE/3 care PRE/3 data)) OR (TITLE-ABS (big PRE/3 data)) OR (TITLE-ABS (large PRE/3 public PRE/3 health PRE/3 data)) OR (INDEXTERMS ("big data"))	13,735 document results
12	INDEXTERMS("big data")	7,025 document results
10	TITLE-ABS(large PRE/3 public PRE/3 health PRE/3 data)	9 document results
9	TITLE-ABS(big PRE/3 data)	11,403 document results
8	TITLE-ABS(large PRE/3 health PRE/3 care PRE/3 data)	28 document results

Search strategy Embase

Search date: 27th January 2016

Database: Embase

#7	#6 AND [embase]/lim AND [2000-2016]/py	**51**
#6	#4 AND #5	59
#5	'public health'/exp	138,538
#4	#1 OR #2 OR #4	1,595
#3	(large NEAR/3 public NEAR/3 health NEAR/3 data$):ab,ti	14
#2	(large NEAR/3 health NEAR/3 care NEAR/3 data*):ab,ti	163
#1	(big NEAR/3 data*): ab,ti	1,420

6.2 Annex 2: Extraction and quality evaluation

As most of the information on implementation examples was derived from grey literature and websites not from studies, classic quality assessment of the sources was not suitable. Due to the heterogeneity of sources it was decided not to assess each source separately, but to assess the sources identified per implementation example. Thus, the results of the quality assessment refer to the quality information as a whole which was identified for an implementation example. Due to the lack of classic study types (i.e. systematic reviews, interventional studies, observational studies, etc.) included in the analysis of the implementation example's added value, classic quality assessment tools (i.e. Cochrane quality assessment tool) could not be applied. Therefore, the criteria used for the quality assessment have been adapted for the purpose of the study.

Table 16: Comet K-Project Dexhelpp

Author(s)	
Sources	Dexhelpp (http://www.dexhelpp.at/?q=de/node/93) (http://www.dexhelpp.at/?q=de/home) IMEHPS.research (http://www.imehps.at/index.html)
Aim of the implementation example	Development of new methods, models and technologies for Big Data
Commissioning party/sponsor of the implementation example	▪ City of Vienna ▪ Austrian Ministry for Transport, Innovation and Technology (*BMVIT)* ▪ Ministry of Science, Research and Economy *(BMWFW*)
Partners involved in the initiative	10 Austrian partners (dwh GmbH; Austrian Public Health Institut; Main assosiation of social insurance; IMEHPS.research, SBA Research; Synthesis research; Technical University of Vienna; Umit; VRVIS)
Country	Austria
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)

	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			X (various projects and therefore most likely various patient groups)
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High quality** • Clearly stated aim • Comprehensive description of parties involved as well as funding received		

Source: GÖ-FP own presentation

Table 17: The Shared Care Platform

Author(s)	
Sources	Bruun-Rasmussen, M. et al. (report) Good Practice Region of Southern Denmark (short report) News Medical (http://www.news-medical.net/news/20130515/Southern-Denmark-to-use-IBMs-Shared-Care-Platform-to-improve-care-for-patients-with-chronic-illness.aspx)
Aim of the implementation example	▪ Supporting cross sector collaboration within healthcare ▪ Facilitating the coordination between the general practitioner, the municipality and the hospital
Commissioning party/sponsor of the implementation example	National fund for chronic diseases
Partners involved in the initiative	▪ The Region of Southern Denmark ▪ IBM
Country	Denmark
Limitations	Not mentioned.

Criteria to assess the quality of the sources identified (studies and grey literature)

	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	X (chronic diseases)		
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High quality** ▪ Clearly stated aim ▪ Comprehensive description of parties involved as well as funding received		

Source: GÖ-FP own presentation

Table 18: E-Estonia – National Identity Scheme

Author(s)	
Sources	eHealth Foundation (http://www.e-tervis.ee/index.php/en/) E-Estonia (https://e-estonia.com/) Estonia.eu (http://estonia.eu/about-estonia/economy-a-it/e-estonia.html) Rannala, 2007 (Case study) Widén & Hasteltine, 2015 (Case Study)
Aim of the implementation example	▪ decrease bureaucracy in physician's work process ▪ more efficient distribution of work time ▪ comprehensive access to medical information for physicians ▪ more patient friendly health care services
Commissioning party/sponsor of the implementation example	Estonian Ministry of Social Affairs
Partners involved in the initiative	eHealth Foundation (i.e. Ministry of Social Affairs of Estonia, North Estonia Medical Centre, Tartu University Hospital Foundation, East Tallinn Central Hospital, Estonian Hospitals Association, The Estonian Society of Family Doctors, Union of Estonian Emergency Medical Services)
Country	Estonia
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?			x
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?	x (partly)		
Results of quality assessment	**High quality** ▪ Clearly stated aim ▪ Comprehensive description of parties involved as well as funding received ▪ A number of independent sources		

Source: GÖ-FP own presentation

Table 19: AEGLE (An analytics framework for integrated and personalized healthcare services in Europe)

Author(s)	
Sources	AEGLE (http://www.aegle-uhealth.eu/en/) Soudris, D. et.al. (report)
Aim of the implementation example	Advance transnational medicine as well as integrated and personalized healthcare services
Commissioning party/sponsor of the implementation example	Partially funded by EU
Partners involved in the initiative	13 partners from different EU Countries (i.e. Exodus, S.A., ingston University Higher Education Corporation, University Vita-Salute San Raffaele, Centre for Research and Technology Hellas, MAXELER Technologies, UPPSALA UNIVERSITY, TIME.LEX, Erasmus Universiteit Rotterdam, Croydon Health Services NHS Trust, GLOBAZ S.A)
Country	EU (UK, IT, GR, S, B, NL, P, F)
Limitations	Not mentioned.

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	X (hospital patients)		
Are issues related to data security fully described?			X (not applicable, as focus on data management/provision of data)
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High quality** ▪ Clearly stated aim ▪ Information about involved parties and funding is available		

Source: GÖ-FP own presentation

Table 20: The Greek E-Prescription System

Author(s)			
Sources	Pangalos & Asimakopoulos, n.d. Pangalos, et al., 2013 Pangalos, et al., 2014		
Aim of the implementation example	▪ Cost reductions ▪ Improving transparency ▪ Compatibility to other e-health systems		
Commissioning party/sponsor of the implementation example	Greek National Organization for Provision of Healthcare Services (EOPYY)		
Partners involved in the initiative	EOPYY, IDIKA (developer, state company)		
Country	Greece		
Limitations	Not mentioned		
Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?			x
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?			x (some authors affiliated to the development firm)
Results of quality assessment	**Moderate quality** ▪ Information on aim, target population and data security issues well in place ▪ No information provided on the funding ▪ Involved parties not explicitly mentioned ▪ Not fully clear, if information from independent source, as some of the authors have an affiliation with IDIKA, the development firm		

Source: GÖ-FP own presentation

Table 21: PASSI

Author(s)			
Sources	Baldissera, S. et al. Quarchioni, E. et. al EpiCentro (http://www.epicentro.iss.it/passi/en/organization.asp)		
Aim of the implementation example	Estimate the frequency and evolution of behavioural risk factors		
Commissioning party/sponsor of the implementation example	Italian Ministry of Health		
Partners involved in the initiative	National Institute of Public Health		
Country	Italy		
Limitations			
Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?			X (partners mentioned but not in detail)
Is the funding of the implementation example clearly stated?	X (only initial funding)		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Moderate Quality** ▪ Clearly stated aim and target population ▪ Information about involved parties and funding is available but not very exhaustive and not independent sources		

Source: GÖ-FP own presentation

Table 22: ARNO Observatory

Author(s)	
Sources	CINEA (http://www.cineca.it/en/content/arno-observatory) ARNO Observatory (https://osservatorio-arno.cineca.org/arnoeng.htm)
Aim of the implementation example	providing advanced IT resources to the local and regional programs dedicated to the monitoring of medical prescriptions
Commissioning party/sponsor of the implementation example	Not mentioned
Partners involved in the initiative	CINEA, Local Health Units of Italian regions involved
Country	Italy (7 regions)
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			x
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?			x
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Low quality** ▪ very little information identified which does not covers essential quality criteria and does not stem from independent source		

Source: GÖ-FP own presentation

Table 23: The Swedish Big Data Analytic Network

Author(s)	
Sources	Görnerup, O. et al.
Aim of the implementation example	Construction of a new strategic national research and innovation agenda in Big Data Analytics
Commissioning party/sponsor of the implementation example	Not mentioned
Partners involved in the initiative	26 partners
Country	Sweden
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?		x	
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Low Quality** ▪ no information about funding ▪ no independent source of information		

Source: GÖ-FP own presentation

Table 24: Clinical Practice Research Datalink

Author(s)	
Sources	Herrett, E. et al. CPRD (https://www.cprd.com/intro.asp)
Aim of the implementation example	▪ Provide an ongoing primary care database ▪ Improve the way clinical trials can be undertaken
Commissioning party/sponsor of the implementation example	▪ NHS National Institute for Health Research (NIHR) ▪ Medicines and Healthcare products Regulatory Agency (MHRA)
Partners involved in the initiative	Not mentioned.
Country	United Kingdom
Limitations	

Criteria to assess the quality of the sources identified (studies and grey literature)

	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			x
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High Quality** ▪ Aim is clearly stated and partners as well as funding is mentioned ▪ Data security is described ▪ No information from independent sources		

Source: GÖ-FP own presentation

Table 25: Sentinel Stroke National Audit Programme

Author(s)	
Sources	Royal College of Physicians (https://www.rcplondon.ac.uk/projects/outputs/sentinel-stroke-national-audit-programme-ssnap)
Aim of the implementation example	Improve stroke care by auditing stroke services against evidence based standards as well as national and local benchmarks
Commissioning party/sponsor of the implementation example	Not mentioned
Partners involved in the initiative	Royal College of Physicians
Country	UK (England, Wales, Northern Ireland)
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?			x
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Low quality** ▪ Information identified stems from Royal College of Physicians, the founder of the Programme ▪ Aim and target population is described ▪ Funding and data security issues are not mentioned		

Source: GÖ-FP own presentation

Table 26: Hospital Episode Statistics

Author(s)			
Sources	Health and Social Care Information Centre (http://www.hscic.gov.uk/hes)		
Aim of the implementation example	Not explicitly mentioned		
Commissioning party/sponsor of the implementation example	Not mentioned		
Partners involved in the initiative	Health and Social Care Information Centre National Programme for IT		
Country	UK (England)		
Limitations	Not mentioned		
Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?			x
Was a target population stated?	x		
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Moderate quality** ▪ Information identified stems not from an independent source. ▪ No information provided regarding funding ▪ Also, the aim of the implementation example is not explicitly stated		

Source: GÖ-FP own presentation

Table 27: Yale University Open Data Access (YODA) Project

Author(s)	
Sources	Krumholz, 2013 The YODA Project (http://yoda.yale.edu/)
Aim of the implementation example	facilitating access to participant-level clinical research data and/or clinical research reports to promote scientific research
Commissioning party/sponsor of the implementation example	Yale University
Partners involved in the initiative	The YODA Project Medtronic Inc. and Johanson & Johnson (as data providers)
Country	US
Limitations	Dependency on collaboration with clinical research data providers

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			x (not applicable, as focus on research data in general, not on population groups)
Are issues related to data security fully described?			X (not applicable, as data security issue has to be fulfilled already during the trial phase by the data providers)
Are the involved parties adequately mentioned?			x (not clear if more data providers than those mentioned above)
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?	x (partly)		
Results of quality assessment	**Moderate/unclear quality** ▪ Information stems partly of independent source and partly of an scientific background ▪ As the YODA project is only the distributer of already collected trial data and focused on research, the questions regarding target population and data security are not applicable ▪ Information regarding funding was not provides		

Source: GÖ-FP own presentation

Table 28: FDA Adverse Event Network Analyser

Author(s)			
Sources	Bostsis T. et al.		
Aim of the implementation example	Make AENA data usable for pattern recognition in the area of medical product safety and clinical data		
Commissioning party/sponsor of the implementation example	Not mentioned.		
Partners involved in the initiative	▪ Office of Biostatistics and Epidemiology, CBER, FDA, Rockville, MD, USA. ▪ Booz Allen Hamilton, McLean, VA, USA.		
Country	USA		
Limitations			
Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			x
Are issues related to data security fully described?			x
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Low Quality** • Funding is not mentioned and only one source which is not independent • Aim and involved parties are clearly stated		

Source: GÖ-FP own presentation

Table 29: Google Flu Trends

Author(s)	
Sources	Ginsberg, et al., 2009 Lacer, et al., 2014 Google (https://www.google.org/flutrends/about/; http://googleresearch.blogspot.co.at/2015/08/the-next-chapter-for-flu-trends.html)
Aim of the implementation example	Faster detection of influenza outbreaks in order to reduce the impact of seasonal and pandemic influenza
Commissioning party/sponsor of the implementation example	Google Inc.
Partners involved in the initiative	Google Inc.
Country	US (international)
Limitations	Limited accuracy of predictions

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?			x
Are issues related to data security fully described?			x
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?	x (partly)		
Results of quality assessment	**Moderate quality** ▪ Information stems partly from independent sources ▪ No information is provided with regards to funding ▪ No statement can be made regarding information on the target population and on data security issues		

Source: GÖ-FP own presentation

Table 30: CEPHOS-LINK (Comparative Effectiveness Research on Psychiatric Hospitalisation by Record Linkage of Large Administrative Data Sets)

Author(s)	
Sources	CEPHOS-LINK
Aim of the implementation example	▪ to compare treatment of psychiatric patients in psychiatric vs. non psychiatric hospitals or departments. ▪ to examine and compare continuity of care after discharge from a psychiatric department/hospital vs. no continuity of care.
Commissioning party/sponsor of the implementation example	European Commission
Partners involved in the initiative	National Institute for Health and Welfare (project lead; FI) IMEHPS Research (AT) Dwh GmbH (AT) National School of Public Health, Management and Professional Development (RO) SINTEF (NO) The University of Verona (IT) Research Centre of the Slovenian Academy of Sciences and Arts (SI)
Country	EU (FI, AT, IT, NO RO, SI)
Limitations	Not mentioned

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?			
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High quality** ▪ Information relevant for the assessment of the source's quality is available ▪ Non-independence of information provision is not regarded as major quality limiting factor, as the project is ongoing and its focus is research		

Source: GÖ-FP own presentation

Table 31: Twitter for adverse drug reactions and Public Health

	adverse drug reactions	**public health**
Author(s)		
Sources	Ginn, et al., 2014 Freifeld, et al., 2014 O'Connor, et al., 2014	Paul & Dredze, 2011 Broniatowski, et al, 2013 Paul et al., 2015
Aim of the implementation example	Analyse Twitter's value as source for signals for adverse drug reactions	To investigate what kind of public health information can be learned by using Twitter as source
Commissioning party/sponsor of the implementation example	Arizona State University Regis University Denver	John Hopkins University
Partners involved in the initiative	Not mentioned	Not mentioned
Country	US	US (international)
Limitations	Availability of enough data sample (i.e. Tweets)	Availability of enough data sample (i.e. Tweets) Analysis only possible on higher aggregation level (i.e. population rather than individual level; state rather than county level)

Criteria to assess the quality of the sources identified (studies and grey literature)						
	Yes	**No**	**Unclear**	**Yes**	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x			x		
Was a target population stated?	x					x
Are issues related to data security fully described?			x			x
Are the involved parties adequately mentioned?			x			x
Is the funding of the implementation example clearly stated?	x (partly)					x
Is the information on the implementation example based on an independent source?	x			x		
Results of quality assessment	**Moderate quality** ▪ Except for data security issues and involvement of partners for which no statement can be given based on the information identified, all information relevant for assessing the quality of the implementation project is mentioned and/or described			**Moderate/Unclear quality** ▪ No statement can be made regarding target population, data security issues, parties involved and the funding of the implementation project ▪ The aim of the example was stated and information was independently provided		

Source: GÖ-FP own presentation

Table 32: Flatiron

Author(s)	
Sources	▪ Flatiron (http://www.flatiron.com) ▪ Fortune ▪ (http://fortune.com/2014/07/24/can-big-data-cure-cancer/) (http://fortune.com/2014/06/12/flatiron-healths-bold-proposition-to-fight-cancer-with-big-data/) ▪ Foundation Medicine, Inc. ▪ (http://investors.foundationmedicine.com/releasedetail.cfm?releaseid=885539) ▪ Techcrunch (http://techcrunch.com/2015/08/19/guardant-health-and-flatiron-health-team-up-to-cure-cancer-with-big-data/)
Aim of the implementation example	▪ Make EMR data usable for research and analytics ▪ Building a larger cancer database
Commissioning party/sponsor of the implementation example	Google (main sponsor)
Partners involved in the initiative	Guardant Health, Flatiron Health
Country	USA
Limitations	

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	X (cancer patients)		
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?			x
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High Quality** ▪ Clearly stated aim and information regarded target population available ▪ Clear information on funding		

Source: GÖ-FP own presentation

Table 33: UK Biobank

Author(s)	
Sources	Biobank (http://www.ukbiobank.ac.uk/) Thermo Scientific Inc.
Aim of the implementation example	▪ Improving prevention, diagnosis and treatment of different diseases
Commissioning party/sponsor of the implementation example	▪ Wellcome Trust medical charity ▪ Medical research council ▪ Department of Health ▪ Scottish Government ▪ Northwest Regional Development Agency
Partners involved in the initiative	▪ University of Manchester ▪ Supported by the National Health Service
Country	UK
Limitations	

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x(general population)		
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High quality** ▪ Exhaustive information on the aim, the funding and the involved parties		

Source: GÖ-FP own presentation

Table 34: SEMCARE

Author(s)	
Sources	▪ SEMCARE Consortium ▪ Averbis (https://averbis.com/presse/neues-eu-projekt-bringt-patienten-und-klinische-forschung-naeher-zusammen/) ▪ SEMCARE (http://semcare.eu/the-project-2/the-project/)
Aim of the implementation example	▪ Making electronic medical records usable for research
Commissioning party/sponsor of the implementation example	EU funding from European Union's Seventh Programme for research, technological development and demonstration
Partners involved in the initiative	▪ Averbis GmbH ▪ Erasmus Unisversitair Medisch Centrum Rotterdam ▪ Medical University of Graz ▪ Saint George´s University of London ▪ Synapse Research Management Partners S.L.
Country	International (DE, NL, AT, UK, ES)
Limitations	

Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?		x	
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High Quality** ▪ Aim is clearly stated and funding as well as involved parties are adequately addressed		

Source: GÖ-FP own presentation

Table 35: Integrated BioBank of Luxembourg (IBBL)

Author(s)			
Sources	IBBL (http://www.ibbl.lu/about-ibbl/)		
Aim of the implementation example	▪ Provision of high quality biospecimens ▪ Fostering scientific excellence ▪ Catalysing partnerships ▪ Supporting research		
Commissioning party/sponsor of the implementation example	Not mentioned		
Partners involved in the initiative	Not mentioned		
Country	Luxembourg		
Limitations			
Criteria to assess the quality of the sources identified (studies and grey literature)			
	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?		x	
Are issues related to data security fully described?			x
Are the involved parties adequately mentioned?	x		
Is the funding of the implementation example clearly stated?		x	
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**Moderate Quality** ▪ information on funding not clearly stated ▪ aim and involved parties are stated sufficiently		

Source: GÖ-FP own presentation

Table 36: Spanish rare diseases registries research network (Spain RDR)

Author(s)	
Sources	Alonso, V. et al., Institute of Rare Diseases Research International rare disease research consortium (http://www.irdirc.org/spainrdr-aims-to-build-the-national-rare-diseases-registry-in-spain-qa-with-dr-manuel-posada/)
Aim of the implementation example	▪ provision of a central platform with access to information and data for health policy making and clinical research ▪ facilitating research in the area of rare diseases
Commissioning party/sponsor of the implementation example	Institute of Health Carlos III
Partners involved in the initiative	▪ Health Departments of the Autonomous Communities (regions) of Spain ▪ Spanish Ministry of Health ▪ Spanish Centre of Reference of People and Families affected by RD (CREER) ▪ six Spanish Medical Societies ▪ four research networks ▪ pharmaceutical and biotechnological organizations (ASEBIO and FARMAINDUSTRIA) ▪ the Spanish Federation of RD (FEDER) and its foundation (FEDER TELETHON FOUNDATION)
Country	Spain
Limitations	

Criteria to assess the quality of the sources identified (studies and grey literature)

	Yes	**No**	**Unclear**
Was the aim of the initiative clearly defined and stated?	x		
Was a target population stated?	x		
Are issues related to data security fully described?	x		
Are the involved parties adequately mentioned?			x
Is the funding of the implementation example clearly stated?	x		
Is the information on the implementation example based on an independent source?		x	
Results of quality assessment	**High Quality** ▪ aim, funding and target population is stated ▪ funding is adequately adressed		

Source: GÖ-FP own presentation

6.3 Annex 3: List of participating experts

Name	Institution	Country
7 experts in health issues and in health policy working for government of 5 different MS		
Pedro Batista	Ministry of Health	Portugal
Ilmo Keskimäki	Outcomes and Equity Research Group, National Research and Development Centre for Welfare and Health	Finland
Piret Hirv	Ministry of Social Affairs, E-services Development and Innovation Policy	Estonia
Robert Scharinger	Ministry of Health	Austria
Miklós Szócska	Health Services Management Training Centre, Semmelweis University Budapest	Hungary
Zdenek Gütter	Palacky University Olomouc - National eHealth Center	Czech Republic
4 stakeholders a) patient/citizens b) health professionals and c) health care payers		
Ian Banks	European Cancer Organisation (ECCO)	United Kingdom
Birgit Beger	European Cancer Organisation (ECCO)	Belgium
Evert Jan van Lente	European Social Insurance Platform (ESIP)	Germany
Bernard Maillet	Standing Committee of European Doctors (CPME)	Belgium
Carole Rouaud	Standing Committee of European Doctors (CPME)	Belgium
5 experts of Health Information issues, Big Data and telemedicine from industry or multi-stakeholder associations		
Isabelle Hilali	Orange Healthcare	France
Elizabeth Kuiper	European Federation of Pharmaceutical Industries and Associations	Belgium
David Manset	Gnubila / MAAT France	France
Niki Popper	Comet-K project Dexhelpp – Decision Support for Health Policy and Planning, dwh GmbH	Austria
Dietrich Rebholz Schuhmann	Insight - a Joint initiative between Dublin City University, NUI Galway, University College Cork & University College Dublin	Ireland
Mário Romão	Intel Corporation	Belgium

6.4 Annex 4: Added-value of shortlisted examples of Big Data use in Health

AEGLE - UK, IT, GR, SE, BE, NL, PT, FR

AEGLE's field of application falls within *health system and service research* and deals with routinely collected data on adverse events.

The analytics framework for integrated and personalized healthcare services in Europe called AEGLE aims to advance transnational medicine as well as integrated and personalized healthcare services by developing innovative ways to handle Big Bata. Ultimately, it is supposed to cover the complete healthcare data value chain ranging from storage of large volume data to big data analytics, cloud computing and the provision of integrated care services, and to derive value from it [71]. The Big Data analytics framework consists of two levels in order to reflect the requirements of different stakeholders such as health care providers and researchers.

On the local level data originates from patient monitoring services such as data obtained by Intensive Care Units (ICU). Big Data analytic services for real-time processing will be applied to this fast generated and multiple-formatted raw data. In this way AEGLE can offer real–time analytic services, for example to detect unusual or deteriorating states of patients connected to an ICU. The second level of the Big Data analytics framework is the cloud level, which will offer a Big Data research platform consisting of semantically-annotated and anonymized healthcare data of different origin. Visualization tools will be implemented helping data scientists or doctors to analyze this data. By using the advanced mechanisms for analysis doctors can be supported in terms of pattern recognition of the severity and the countermeasures for patient's condition [72]. The project unites 13 partners from 10 different European countries and is partially funded by the EU. It is coordinated by EXUS, a software house with experience in delivering complex systems and solutions [73].

Table 37: The added value of AEGLE

Patient	Provider	Policy	Research
▪ improved monitoring during hospital stay	▪ improved diagnosis by use of real-time analytic services and visualization tools	not applicable for this example	▪ a wide range of quantitative, patient focused research

Overall added value: Improving quality and effectiveness of treatment

Source: GÖ-FP based on AEGLE [71-73]

The Greek E-Prescription System - GR

Greek's e-Prescription System falls within *health system and service research, surveillance (pharmaceutical)* as well as e*pidemiology* and deals with routinely collected prescription data.

The Greek e-Prescription System was first introduced in 2010 aiming to effectively control and rationalize expenses and to improve transparency in the social insurance system. Thus, it was designed as a tool for combating the challenges related to rising medication costs such as system abuse, over-prescribing and fraud. Further, the system aims to provide an e-prescription environment which is compatible and interoperable with other national e-health applications and third-party information systems [74]. As an example, data of the e-Prescription System is combined with data of the Business

Intelligence System of the Greek National Organization for Provision of Healthcare Services (EOPYY) for national analyses in the field of healthcare [75, 76]. The system is web-based and provides access for authorized users (i.e. physicians, pharmacists). Necessary security infrastructure and a security policy is in place. An e-prescription contains the at least the following data: patient's and doctor's social identification number, the ICD-10 code, details of the pharmaceuticals prescribed (e.g. quantity, dosage), the patient's share of payment for each medication. By collecting this information, the e-Prescription System provides valuable information on prescription processes (i.e. who prescribed what to whom for which illness) which can be subsequently used for administrative and public health advancements. In 2014, the e-Prescription System covered more than 95 % of all prescriptions prescribed monthly in Greece, as 95 % of Greek pharmacies and 90 % of prescribing doctors used the system for electronic prescriptions. The system was developed by IDIKA, a state company responsible for e-health applications in Greece [74].

Table 38: The added value of the Greek E-Prescription System

Patient	Provider	Policy	Research
▪ reduction of medical errors	not applicable for this example	▪ reduction of unnecessary health care spending for pharmaceuticals ▪ better monitoring and planning of public health ▪ better administrative control and transparency	▪ data pool which is compatible to others
Overall added value: Improving sustainability of health systems			

Source: GÖ-FP based on the Greek e-Prescription System [74, 77, 78]

Semantic Data Platform for Healthcare (SEMCARE) – DE, NL, AT, UK, ES

The initiative's field of application falls within *health system and service research, epidemiology and clinical research* and deals with electronic health records (EHRs).

The Semantic Data Platform for Healthcare (SEMCARE) is a pan-European project that operated for two years (2014-2015). Its primary aim was to build a semantic data platform, which can identify patient cohorts based on clinical criteria dispersed in heterogeneous clinical resources [79]. The definition of patient cohorts is highly relevant for different purposes like feasibility studies, patient enrolment for clinical studies, quality assurance in hospitals and the identification of undiagnosed patients. In case of clinical studies almost 80 % of clinical trials fail to enrol enough patients on time which brings delays to drug development [80].

SEMCARE offers the possibility to search electronic healthcare documents using different criteria such as diagnosis, symptoms, regulations and special finding features. This can be done through a single interface where heterogeneous patient data, which is available in structured and unstructured form, is harmonized, analysed and made searchable. [79, 81]. In order to do so, text mining technologies and multilingual semantic resources such as terminologies, nomenclatures, classifications, and domain vocabularies were integrated to capture the specific idiosyncrasies of medical language including ambiguous terms, acronyms, derivations and spelling variants.

The semantic data platform was tested in three different hospitals which served as pilot sites for the local implementation of the system and the use of the system mainly in the area of rare diseases. The future goal is to build a pan-European supported platform, which can be used by hospitals all over Europe to identify patients for clinical studies and to support diagnostics [79].

SEMCARE was carried out by an interdisciplinary team of researchers including the Averbis GmbH (AVERBIS) in Germany, the Erasmus Universitair Medisch Centrum Rotterdam (EMC) in the Netherlands, the Medical University of Graz (MUG) in Austria, the Saint George's University of London (SGUL) in the United Kingdom and the Synapse Research Management Partners S.L. (SYNAPSE) in Spain[81]. The project was mainly funded by the European Community's Seventh Framework Programme.

Table 39: The added value of Semcare

Patient	Provider	Policy	Research
▪ faster diagnosis especially for rare diseases	▪ support for diagnosis of rare diseases ▪ earlier completion of clinical trials through facilitated patient enrolment	▪ lower costs due to faster diagnosis	▪ simplified identification of suitable candidates for clinical studies
Added value overall: improving the quality and effectiveness of treatment			

Source: GÖ-FP based on SEMCARE[79-81]

Integrated BioBank of Luxembourg (IBBL) - LU

The initiative's field of application falls within *clinical research* and deals with biospecimen and associated data.

The Integrated BioBank of Luxembourg (IBBL) is an autonomous institute, which is working not-for-profit and is organised within the Luxembourg Institute of Health (LIH). Its starting point was the launch of the "Health Sciences and Technologies Action Plan" by the government of Luxembourg in 2008. The creation of IBBL as well as the LCSB (Luxembourg Centre for Systems Biomedicine) and the establishment of a large lung cancer programme followed this launch [82, 83].

The IBBL opened in 2010 but the first human biospecimens were already collected in 2009. The collection grew to 67.000 samples in 2012 and today the IBBL has collected over 245,000 samples, which were derived from urine, stool, blood, tissue, salvia, and cerebrospinal fluid.

IBBL aims at providing high quality biospecimens, fostering scientific excellence, catalysing partnerships and supporting research, which strives for the implementation of scientific discoveries into new healthcare solutions. It offers biobanking services such as the collection, processing, analysis and storage of biological samples and associated data. Moreover, research is conducted with the aim to optimize biospecimen processing and to biospecimen quality. The field of research is ranging from programmes on cancer, diabetes and Parkinson´s disease to microbiome [83].

The IBBL has several international partnerships and collaborations and is a member of the Personalized Medicine Consortium (PMC) of Luxembourg which is an initiative aiming to adopt personalized medicine to the national healthcare system. The PMC recently extended the scope of its research to other diseases and areas such as for example Big Data. [84]

Table 40: The added value of IBBL

Patient	Provider	Policy	Research
not applicable for this example	not applicable for this example	not applicable for this example	▪ provision of high quality biospecimens and other biological data ▪ support of partner-ships and networks
Overall added value: improving the quality and effectiveness of treatment			

Source: GÖ-FP based on IBBL [82-84]

Twitter - international

Twitter's field of application falls within *surveillance (pharmaceutical*), when Tweets about adverse drug reactions (ADR) are analysed*,* and within *surveillance (public health)*, when the focus lies on public health topics in general.

Twitter[17] represents an enormous data pool for analysing population trends. However, Twitter's challenge in this context is its brevity, informal and unstructured information which requires automatic processing and extraction additionally to simple natural language processing. Recently, efforts have been made to detect adverse drug reactions via Twitter. In one study, Tweets on 74 medicinal products were collected to analyse their value as source for signals for ADRs. The used list of medicinal products covered brand names, generic names as well as potential misspellings. From 187,450 Tweets in total, 10,822 Tweets have been included in a manually annotated corpus. This corpus is freely available from http://diego.asu.edu/downloads and can be used to train automated tools to mine Twitter for ADRs [85, 86].

Another study using Twitter investigated the level of matching between Tweets mentioning ADR and spontaneous reports received by a regulatory agency and concluded that there is some causal association between ADRs identified via Twitter and spontaneous reports received by the Federal Drug Association (FDA) when analysed by a broad classification system. However, for a clear statement regarding causal association, bigger sample sizes are necessary [87].

Besides pharmacovigilance topics like the mining of adverse drug reactions, Twitter can be utilized for public health research as well. In one of the identified studies, an Ailment Topic Aspect Model was applied to over 1.5 million health related Tweets. By means of this, Tweets mentioning over a dozen ailments in total (e.g. allergies, obesity, and insomnia) could be identified. Furthermore, the Ailment Topic Aspect Model was extended to investigate its applicability for different surveillance approaches using Twitter mining: tracking illnesses over time (syndromic surveillance), measuring behavioural risk factors (sentinel surveillance), and localizing illnesses by geographic region (geographic syndromic surveillance), and analysing symptoms and medicinal product's usage. In the field of syndromic surveillance, a popular research area is the tracking of influenza infections using Twitter data as source [88, 89]. The popularity of this approach can be explained by the episodic character and the wide spread of the influenza infection which suits syndromic surveillance research best. The results of the study suggest that Twitter's applicability for public health research has considerable limitations such as the

[17] Twitter was launched in 2006, and has since grown to be one of the largest social media websites with about 320 million active users per months (as of 30. September 2015), producing about 9,100 Tweets per second [85].

requirement of large data sets and analyses on highly aggregated level (i.e. population rather than individual level; state rather than regional level) [90].

Table 41: The added value of Twitter (Adverse drug reactions)

Patient	Provider	Policy	Research
▪ not applicable for this example	▪ not applicable for this example	▪ supplement to traditional sources ▪ potential for substituting current more expensive and time consuming methods	▪ extensive pool of informal data ▪ easy access to AER data
Overall added value: Improving quality and effectiveness of treatment and supporting of healthy life styles and sustainability of health systems. Limitation of added value by size of data set (i.e. number of Tweets)			

Source: GÖ-FP based on [85-87, 90]

Sentinel Stroke National Audit Programme (SSNAP) - UK

The SSNAP's field of application falls within *clinical practice* as well as *health system and service research* and deals with routinely collected clinical stroke data.

SSNAP is an audit programme developed by the Royal College of Physicians in order to improve stroke care by auditing stroke services against evidence based standards as well as national and local benchmarks. The programme runs in England, Wales and Northern Ireland and is designed as prospective and longitudinal audit measuring the quality of care of stroke patients. It covers the whole care pathway up to six months post admission. SSNAP is the single source of stroke data in England, Wales and Northern Ireland and supplies among others the statutory data collection of the NICE[18] Quality Standard and the NHS Outcomes Framework. SSNAP comprises three components: 1.) clinical audit, 2.) acute organisational audit and 3.) post-acute organisational audit. In the prospective clinical audit, a minimum dataset for every stroke patient is collected. It includes acute care, rehabilitation, and six-month follow-up and outcome measures. The biennial acute organisational audit is a snapshot audit measuring the structures of stroke services in acute hospitals [91]. In the post-acute organisational audit, data about the structures and types of post-acute stroke services is collected. Since January 2013, 185,000 patient records representing 95 % of expected stroke cases have been submitted to SSNAP [92].

Table 42: The added value of the Sentinel Stroke National Audit Programme

Patient	Provider	Policy	Research
▪ patient empowerment in order to ask search questions	▪ benchmarking with other providers – nationally and regionally ▪ identification of need for improvement	▪ regular monitoring of stroke services to improve quality ▪ efficient data collection, no duplication of data	▪ quality measurement in the field of stroke care ▪ availability of comprehensive and robust data ▪ data on entire patient pathway
Overall added value: Improving quality and effectiveness of stroke treatment			

Source: GÖ-FP based on Royal College of Physicians [91, 92]

[18] National Institute of Clinical Excellence in the UK

The Swedish Big Data Analytic Network - SW

The network's field of application falls within *health system and service research* and comprises different cooperation projects in the field of Big Data.

The Swedish Big Data analytic network consists of a consortium of over twenty partners in the field of Big Data Analytics and its related areas and includes large established companies, small and medium sized private companies, universities, research institutes and other stakeholders in the public sector. It aims to construct a new strategic national research and innovation agenda in Big Data Analytics (including healthcare related topics) in order to create a fertile ground for future businesses services and societal applications based on Big Data. Moreover, it aims to create value out of data streams that are already existing and to connect the different players in the field.

The agenda displays unique national competences and strengths, and the needs of Swedish industry and society within Big Data analytics. In the future suggestions for actions to improve national competitiveness will be given. The network captures academic excellence within the central research areas, but also includes industrial perspectives from both established and growth industries. It therefore forms a basis for future cooperation in Big Data analytics across sectors by linking stakeholders in industry, academia and the public, including the health care sector [93].

Table 43: The added value of the Swedish Big Data analytic network

Patient	Provider	Policy	Research
not applicable for this example	not applicable for this example	not applicable for this example	▪ facilitates connection and exchange in the field of Big Data analytics
Overall added value: Potentially contributing to sustainability of health systems			

Source: GÖ-FP based on The Swedish Big Data Analytic Network[93]

Clinical Practice Research Datalink (CPRD) – UK

The field of application falls within c*linical research* as well as *epidemiology* and deals with primary health care records.

The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service in the United Kingdom (UK), which is in place since 1987. It aims to provide an ongoing primary care database of anonymized medical records from general practitioners with the objective "to maximize the health gain that can be achieved through the use of anonymized linked NHS data in research studies and help improve the way clinical trials of innovative medicines can be undertaken"[94]. For over half of the included patients, linkage with datasets from secondary care, disease-specific cohorts and mortality records enhance the range of data availability [95].

The NHS[19] National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA), a part of the Department of Health, are jointly funding the project. CPRD covers over 11.3 million patients from 674 practices in the UK. With 4.4 million active (alive, currently registered) patients meeting quality criteria, approximately 6.9 % of the UK population are included [95]. The CPRD

[19] National Health Service in the UK

is widely used for research on a national and international level. Data of CPRD has been used to produce over 1,500 research studies, published in peer-reviewed journals across a broad range of health outcomes. Areas of research cover epidemiology, pharmacovigilance, pharmacoepidemiology as well as economic evaluation studies [94, 95].

Table 44: The added value of CPRD

Patient	Provider	Policy	Research
not applicable for this example	▪ benchmarking with other providers	▪ target-oriented health care planning ▪ appropriate resource allocation	▪ wide range of quantitative, patient focused research ▪ linkage with data sources other than patient records
Overall added-value: Contributing to quality and effectiveness of treatment and to the sustainability of the health care system			

Source: GÖ-FP based on CPRD [94, 95]

FDA Adverse Event Network Analyser (AENA) – USA

The initiative's field of application falls within *surveillance (pharmaceutical)* and deals with routinely collected data on adverse events.

The Food and Drug Administration's (FDA) Adverse Event Network Analyzer (AENA) presents a new approach for analyzing the US Vaccine Adverse Event Reporting System (VAERS). VAERS is the national vaccine safety surveillance program where adverse events, reported by health care providers, vaccine recipients or other interested parties after immunization, are pooled. Medical experts at the FDA usually review the data following two approaches. The first includes statistical data mining algorithms, which aim at assessing whether adverse events (AEs) occur disproportionally more often in connection with one product compared to others. The second approach includes manual reviews of reports in order to identify unusual patterns in the data that might indicate safety issues. However, with these approaches interactions among co-administered vaccines and multiple AEs are not evaluated. Moreover, manual reviews are time intensive and dependent on human pattern recognition. VAERS is an example for a spontaneous reporting system and, as traditional statistical methods are not suitable for this kind of data, Big Data data mining approaches are necessary. Therefore, AENA was developed for improved data analysis and visualization (including three network approaches). The approaches include a weighting scheme based on co-occurring triplets in reports (three words/adverse events close to each other), a visualization layout and a network growth methodology for the detection of outliers. The approaches were verified with data on intussusception after the administration of RotaShield vaccine (RV) in 1999. The aim of this project was to make AENA data usable for pattern recognition in the areas of medical product safety and other clinical data [96].

Table 45: The added value of the FDA Adverse Event Network Analyser

Patient	Provider	Policy	Research
▪ increased vaccination safety	▪ not applicable for this example	▪ fast and clear presentation of AE ▪ improved monitoring of AE	▪ avoidance of human mistakes during data analysis
Overall added value: Improving the quality and effectiveness of treatment			

Source: GÖ-FP based on FDA Adverse Event Network Analyser (AENA)[96]

6.5 Annex 5: Feasibility study of policy actions

Table 46: Results of the SWOT analysis – Legal aspects

Internal factors	Strengths	▪ ***Enables the use of Big Data in Health within a defined and clarified legal framework.*** Directives and / or regulations for a common legal framework need to be set at EU level and implemented in Member States increasing the likelihood of utilizing the full potential of Big Data in health care. ▪ ***Enables patients, health care providers and IT to securely generate, access and share data.*** In order to utilize this, especially patients and health care providers need to be certain about their data being protected in order to feel their rights are protected and secure. Communication concepts that inform about new / adapted legislation in an easily understandable way need to be in place.	Weaknesses	▪ ***Alignment of existing fragmented (national) policies might be difficult and thus might jeopardise an EU-wide uniform approach of Big Data in healthcare.*** In order to tackle this, it is suggested to: (1) Review national legislation on collection, storage, analysis, use and dissemination of Big Data / routinely stored data. (2) Find common ground but also contradictions and discrepancies in the national legislations. (3) Revise and expand existing EU legislation (e.g. on data protection) if necessary. Draft new legislations that set the minimum legal requirements that have to be implemented by each Member State, which should give clear directions on at least the above mentioned aspects (privacy and data-ownership, secondary use of health data, cloud services and institutions hosting and managing EHR). (4) The EU legislation have to be revised on a regular basis in order to incorporate the changing technological environment. ▪ ***Legislation might leave room for interpretations.*** In order to tackle this, existing legislation should be as clear as possible. In the drafting of new EU legislation emphasis should be given to clearness and preciseness in order to ensure collective understanding and action. Involving different stakeholders (i.e. national and EU decision makers, providers, patients' representatives, academia, industry) in order to secure their support.
External factors	Opportunities	▪ ***Existing legal frameworks (e.g. directives) can be the foundation of new / extended legislation on Big Data use in healthcare.*** To utilize this, existing legal frameworks at EU as well as on national level need to be (1) identified, (2) assessed regarding legal loopholes and / or grey areas which need to be closed and (3) adapted in order to facilitate secure generation, sharing and access to health data. ▪ ***Support of patient organisations and other parties representing patient's interests promoting the establishment of a comprehensive legal framework.*** In order to secure the support of patients' representatives (at EU and national level) regarding the establishment of new / adapting existing legal frameworks, early stage involvement is crucial. Further, transparent communication highlighting the benefits of Big Data use in healthcare could be supportive.	Threats	▪ ***No political commitment / different political priorities and interests hindering the agreement on existing and new policies and amendments of directives.*** To tackle this, the benefits of Big Data analysis in health (i.e. social, economic and budgetary) compared to conventional analysis, and the necessity of adopting new policies (e.g. on privacy and data protection) or the agreement on expanding existing ones facilitating Big Data analysis in health have to be communicated to decision makers. One strong argument paving the way for Big Data application in health is its potential to increase health system efficiency. If Member States refuse to implement EU legislation, adequate procedures need to be in place. ▪ ***Adoption of EU policy / framework in Member States is jeopardised by certain stakeholders.*** In order to tackle this, it is suggested to transparently communicate the content and scope of new / adapted legislation to stakeholders, i.e. patient organisations and other parties representing patient's interest and the public, in order to secure their understanding and support.

Table 47: Results of the SWOT analysis – Stakeholders

Internal factors	Strengths	▪ ***Increased acceptance and risk sharing by involving different public and private stakeholders. As different stakeholders might have different agendas/interests, their involvement increases the probability of successful Big Data analysis.*** The Swedish Big Data Analytics Network, for instance, involves partners in the field of Big Data Analytics and its related areas including the industry, academia, the public sector. This kind of collaboration can facilitate the set-up of a strategic national research and innovation agenda in Big Data Analytics which is accepted by all involved parties.	Weaknesses	▪ ***Higher need for coordination in case of shared responsibilities and involvement of different stakeholders/experts.*** This can be tackled by the establishment of a national coordinating body for Big Data Analysis, which can act as the interface of all activities related to Big Data Analysis. In order to avoid duplication of work and to utilize synergies, the coordination body should cover sub-divisions for the various application fields of Big Data (i.e. health, security, finance, etc.). This coordination body can also encompass governing functions including the definition and execution of standards and procedures for Big Data analysis in healthcare in accordance to standards set at EU level. Questions addressed by this body should refer to: How to support the availability of health data? How to support the usability of data? How to ensure integrity and security of Big Data analysis? ▪ ***Complexity of relationship and interest management between actors and stakeholders.*** Subtle resentments (not necessarily related to Big Data analysis) between different stakeholders might hinder the work in the field of Big Data analysis. There is no "one fits all"-solution for this issue, rather it has to be tackled individually in the specific context of each Member State. Transparent communication and a neutral mediating body (e.g. national Data Officers or a Big Data in Health watchdog) might be supportive.
External factors	Opportunities	▪ ***Big Data might already be on the agenda of some stakeholders who are therefore interested in fostering European cooperation.*** Identify public and private stakeholders already interested in Big Data analysis and use them as levering bodies for those parties still in doubt and thus hindering the application of Big Data analysis in healthcare. ▪ ***Existing Big Data expert networks in academia can be utilized to establish expert teams and/or decision bodies at EU level.*** Set up working groups on Big Data in Health at EU and national level consisting of different public and private stakeholders. These should be in regular dialogue in order to bring Big Data analysis into the Member States and to define and follow a common Big Data strategy across all EU Member States.	Threats	▪ ***Resistance of healthcare professions and/or stakeholders to open up to Big Data as they might fear an undermining of their positions.*** For future generations of health professionals, resistance against Big Data analysis can be reduced by integrating Data/Big Data into their training (i.e. Big Data as part of curricula).

Table 48: Results of the SWOT analysis – Privacy and data protection

Internal factors	Strengths	▪ ***Shift of collective mind-set towards a sharing of data with protection enables a more transparent and efficient data value chain. Supported by a well-designed and aligned privacy policy a comprehensive application of Big Data analysis can be facilitated.*** The Greek e-Prescription System contributes to increased transparency in the social insurance system and increased efficiency of the prescription process. Supported by its security infrastructure and security policy it interoperates with other e-health applications and third party information systems. ▪ ***Reduction of data access barriers for researchers and decision makers in order to establish (new) policies on reliable and comprehensive research information in the fields of healthcare, telemedicine and public health.*** An example for this would be the ARNO Observatory in Italy which is designed to combine and aggregate masses of administrative patient data (e.g. on pharmaceutical prescriptions, hospital discharges, laboratory analyses) of seven Italian regions. By means of this data regional health policy can be tailored according to the analysis results.	Weaknesses	▪ ***Reaching consensus with respect to privacy harms between different stakeholders involved might be difficult. Also, balancing the priorities of maintaining and promoting public health and R & D in health against the privacy of personal data might be a challenge.*** At EU level an independent advisory panel should be established consisting of Member State Data Officers (i.e. representatives who are responsible for Big Data issues in one Member State). Within this advisory panel privacy issues should be discussed and consensus found (e.g. on privacy harms, on trade-offs between personal data protection and public health research). National Data Officers who participate in the panel should disseminate issues discussed in their countries in order to secure national acceptance. ▪ ***Aligning privacy protection with existing privacy legislation might not be feasible for certain kinds of personal data. Also, the technical implementation of privacy regulations might be difficult for certain kinds of data sources (e.g. from Telemedicine devices).*** In order to tackle this, it is possible to install a digital identity which is interoperable not only at national but even at EU level. This digital ID should comprise all information that uniquely describes a person, an entity or a device. Thereby, it would include similar properties as ID cards with the purpose of identity verification and data authentication.

External factors	Opportunities	▪ ***Cooperation with all stakeholders interested in privacy policy (also outside the field of health care) in order to increase the pressure for establishing comprehensive and aligned privacy policy across Member States.*** Stakeholders interested in privacy policy at EU as well as at national level should be identified and connected in order to build a strong alliance, in order to enforce the alignment of national privacy policies. Emphasis should be given that stakeholders of various resorts – not only health – are represented in this alliance to increase political pressure.	Threats	▪ ***Different privacy policies in Member States which might be difficult to align.*** In order to tackle this, it is suggested to: (1) Review national legislation on privacy and data protection. (2) Find common ground but also contradictions and discrepancies in the national legislations. (3) Revise and expand existing EU legislation (e.g. on data protection) if necessary. Draft new EU legislation that set the minimum legal requirements that have to be implemented by each Member State, which should give clear directions on at least the above mentioned aspects (standards for data ownership and control (i.e. opt out options), purpose specification and use limitation, confidentiality, data access (i.e. for patients and health professionals), cloud services, storage and processing, including archiving durations of data, and reuse and cross-border flow of data). (4) The EU legislation have to be revised on a regular basis in order to incorporate the changing technological environment. ▪ ***Opposition of patient organisations and other parties representing patient's interests jeopardising wide use of Big Data due to privacy concerns.*** To overcome this threat it is important to have an open communication strategy which helps to inform patients and their representatives in simple language about the importance and benefits of Big Data in Health and also to still patients' fears. ▪ ***No political commitment/different political priorities and interests.*** To tackle this, the benefits of Big Data analysis in health (i.e. social, economic and budgetary) compared to conventional analysis and the necessity of privacy policies facilitating Big Data analysis in health have to be communicated to decision makers. The EU can enforce this by adopting EU legislation to be implemented at Member State level.

Table 49: Results of the SWOT analysis – Open data and data sharing

Internal factors	Strengths	▪ ***Access to complementary data sources enables improved analytical insights and facilitates data analysis as more data will be available. As a consequence of improved research in healthcare, the effectiveness of health services might be improved.*** An example which already makes use of complementary data sources is the ARNO Observatory in Italy. It links administrative patient data with other data flows of different data bases, such as population registry, GP's registry, National Vital Statistics, etc. By this epidemiological and economic indicators can be built for the participating seven Italian regions for analysing health services and outcomes.	Weaknesses	▪ ***Privacy of information might be compromised.*** Privacy and data protection policies should facilitate secure data sharing and access. Therefore, open data and data sharing needs to be adequately addressed in national privacy policies in Member States, taking into account the balance between cooperation and competition between different data holders.

External factors	Opportunities	▪ ***The application of open data and data sharing might be facilitated by clear rules and standards for accessing data (e.g. governmental data) as well as legal frameworks and policies ensuring personal data protection.*** Data Officers should be appointed in each Member State who oversee all Big Data activities in their country. They should be the official body for setting standards and rules. Regular exchange at EU-level (i.e. in form of an independent advisory panel consisting of Data Officers of all Member States) can facilitate learning. ▪ ***Communication with the public can increase awareness of open data's and data sharing's importance.*** In this context, it is important to have an open communication strategy which helps to inform the public in simple language about the importance and benefits of Big Data in Health and also to still individual's fears regarding open data and data sharing.	Threats	▪ ***Resistance of stakeholders to share their data and / or reluctance of stakeholders / Member States to agree on rules and standards for data sharing.*** For avoiding opposition on data sharing on behalf of stakeholders and / or Member States, their interests should be factored in at the early stage of formulating standards. Therefore, stakeholders should be consulted by national Data Officers when setting-up standards. An open communication strategy stating the benefits of open data and sharing of data should go along with this. ▪ ***Restrictive legal frameworks and data protection stipulations might hinder data sharing.*** In order to tackle this, it is suggested to: (1) Review national legislation on privacy and data protection in the context of open data and data sharing. (2) Find common ground but also hindering factors for accessing public and private data. (3) Revise and expand existing legal frameworks (e.g. on data protection and access) emphasising at least the above mentioned aspects (common protocols for users and resources, data architecture, governance models for managing and sharing data and mutual recognition procedures). (4) The legal frameworks have to be evaluated on a regular basis in order to incorporate the changing technological environment. ▪ ***Media reports about misuse of private data might increase reluctance towards open data*** To tackle this an open communication and media strategy needs to be in place which provides reliable and easy to understand information about the use of private data. If misuse of data is evident, sanctions should follow. At national level the Data Officers, at EU-level the independent advisory panel should have a monitoring function. ▪ ***Different technological levels and advancements of involved national bodies and stakeholders.*** In order to tackle this, it is suggested to: (1) Review levels of technological advancement in each Member State. (2) Specify similarities in technological advancement levels on which one can build on but also differences. (3) Define a minimum level of technological advancement which need to be met by each Member State in order to participate in open data and data sharing (4) Initiate technological developments in lagging Member States following role models in this area such as E-Estonia with the X-Road application which allows the harmonious linkage and operating of various e-health services.

Table 50: Results of the SWOT analysis – Standards and protocols

Internal factors	Strengths	▪ ***Existing standards, policies and best practices can be utilized.*** To utilize this, existing standards and protocols at EU as well as on national level need to be (1) identified, (2) assessed regarding their compliance (3) adapted in order to ensure interoperability of different data (sets). Focus should be given not only to the health sector but also to other fields, in which the application of Big Data arose earlier. Thus, lessons could be learned. ▪ ***By setting common standards across the Big Data value chain pooling and exchanging data will be more efficient and conducting cross-country comparisons will be facilitated.*** Examples benefitting from common standards are AEGLE, which covers the complete healthcare data value chain ranging from storage of large volume data to Big Data analytics, cloud computing and the provision of integrated care services with the aim to advance transnational medicine in ten EU Member States. Also, CEPHOS-LINK benefits in linking EHR derived from administrative electronic healthcare databases of six Member States.	Weaknesses	▪ ***Standards and protocols have to be continuously updated in order to keep up with the rapidly changing technological environment.*** At EU level an independent advisory panel should be established consisting of Member State Data Officers (i.e. representatives who are responsible for Big Data issues in one Member State). Within this advisory panel Big Data standards and protocols are consensually set and regularly revised. ▪ ***Health care data is not always available in a digitized form / the right format; transformation into usable format might be costly.*** The example of SEMCARE provides an approach of how to use heterogeneous data. It can search electronic healthcare documents of various formats through a single interface by using text mining technologies and multilingual semantic resources. ▪ ***Current discussions focus on IT-standards in order to guarantee interoperability of different data, and do not consider quality and outcomes-based protocols.*** When setting new standards / revise existing standards the independent advisory panel should also consider clinical standards related to Big Data in Health care. Therefore, quality and outcome-based protocols can serve as basis.
External factors	Opportunities	▪ ***Involvement of all relevant stakeholders in the formulation of standards and protocols might secure comprehensive adoption.*** In order to secure stakeholders' support, they should be involved for consultation in the process of setting-up and revising standards and protocols. At national level involvement is possible by Data Chief Officers, at EU level they could be consulted by the independent advisory panel.	Threats	▪ ***Variety of existing national standards and protocols difficult to align. Different national legislation (e.g. on privacy) hindering the adoption of cross-country standards.*** In order to tackle this, it is suggested to: (1) Review national legislation and standards in each Member State. (2) Specify promoting and hindering legislation for adopting cross-country standards. Specify differences in national standards. (3) Revise cross-country standards in order to fit national legislations. Revise existing legislation (e.g. on data protection) if necessary. (4) The standards have to be revised on a regular basis in order to incorporate the changing technological environment. ▪ ***Reluctance of Member States' to adopt and implement international standards.*** Each Member State should nominate a Data Officer, who participates in the independent advisory panel, which sets and revises standards at EU level. By participation of Member State's representatives in the process of setting standards, the likelihood of adopting those might be improved.

Table 51: Results of the SWOT analysis – Technological development

Internal factors	Strengths	▪ ***New technologies and software developments can improve the use and security of Big Data and the possibility of data analysis in real-life settings. Also, developments in technology can increase the usability of Big Data for all stakeholders.*** Examples for this are background infrastructures ensuring not only the interoperability but also the security of various e-health applications (e.g. E-Estonia, Greek e-Prescription System). User-friendly interfaces and visualization tools like those AEGLE provides support data scientists and doctors to analyse data in terms of pattern recognition of the severity and the countermeasures for patient's condition by the implemented visualization tools. ▪ ***Support of a varied technological development of Big Data applications which will work against a monopolization of software/technologies in this area.*** Software and technology monopolies can be countered by financial incentives for innovative start-ups as well as universities investigating and developing technological Big Data applications. Therefore, Member States as well as the EU should tender grants.	Weaknesses	▪ ***Decisions about funding priorities have to be made because there are several areas where technological input is needed***. This might be a critical process, as investments in technologies other than Big Data in Health might be of higher priority and thus more urgent. Investment options in Big Data technologies need to be analysed according to their social and economic benefits and priorities have to be set. ▪ ***Different technological solutions might complicate data sharing if they are not interoperable.*** In order to tackle this, adequate background infrastructure needs to be established which is interoperable with different/existing systems. The X-Road infrastructure of E-Estonia is not a centralized national database, but retrieves data from various providers using different systems, and presents it in a standardized format.
External factors	Opportunities	▪ ***Higher number of graduates in computational social sciences and data sciences might lead to more research on technologies for Big Data.*** In order to utilize this, national governments need to adapt existing curricula of computational studies in terms of integrating Big Data science. With respect to Big Data in Health, Big Data should also be integrated into the curricula of studies in health sciences in order to ensure a common understanding of future generations of health professionals and data scientists. ▪ ***Shift in mind-set of the general population might increase the acceptance of new technologies and their application.*** This can be supported by an open communication strategy which helps to inform the population in simple language about the importance and benefits of Big Data in Health and also to still individual fears.	Threats	▪ ***Resource intense funding of technological development.*** In order to secure funding and sharing the risk of high financial investments, the financial burden of Big Data and its application should be split by several partners in form of public (-private) investment partnerships. The UK Biobank was funded by a combination of Government, NGO and Research (i.e. the Wellcome Trust, the Medical Research Council, the Department of Health, the Scottish Government and the North-West Regional Development Agency). ▪ ***Existing technological solutions provided by commercial companies might be expensive.*** To tackle this threat, Member States as well as the EU should promote competition between different providers. By providing financial support for small start-ups in the field of Big Data technology, a more diverse market can be created. Another option for Member States for reducing costs and using synergies would be the establishment of public private partnerships consisting of governmental institutions, companies providing the technology and research institutions analysing the data generated. ▪ ***Technology developers might fear potential loose of intellectual property.*** To tackle this threat, EU wide legislation on intellectual property rights need to be adapted to the needs of new technological developments.

Table 52: Results of the SWOT analysis – Data sources

Internal factors	Strengths	▪ ***Expanding existing sources for the purpose of Big Data analysis might save resources because already existing infrastructure and know-how is used instead of building up databases from scratch.*** In principle, this should be feasible in all Member States as health data is very often routinely collected (e.g. mortality statistics, patient records, data for reimbursement of providers). Projects that are using existing data bases and linking them in order to get new insights are amongst others the Austrian Comet K-Project, the Danish Shared Care Platform and the E-Estonia Identity Scheme. ▪ ***By exploiting new and complementary data sources improved analytical insights can be gained and data analysis facilitated.*** Newly collected data can help to contain costs as was shown by the Greek E-Prescription System which collects data on the prescription process. It was designed as a tool for combating system abuse, over-prescribing and fraud.	Weaknesses	▪ ***Diverse quality of different data sources can be a problem especially if different databases are linked.*** To overcome this weakness, it is of high importance to routinely monitor the data of all sources and make their quality and reliability transparent for the data users. For example data supplied by individuals might be prone to selection bias and attrition, and is of course always subjective (e.g. data coming from a survey such as the Italian PASSI). Other data, such as data that was collected for reimbursement purposes, might be unreliable because of coding errors or "optimized" coding by health care providers.
External factors	Opportunities	▪ ***The creation of new bodies of data can be facilitated by partnering of different stakeholders that hold different databases.*** An example for the cooperation of different data holders is the Spanish Rare Diseases Registries Research Network which is a central platform for data on rare diseases that involves partners from all Health Departments of the Autonomous Communities, the Spanish Ministry of Health, the Spanish Centre of Reference of People and Families affected by rare diseases, six Spanish medical societies, four research networks, pharmaceutical and biotechnological organizations and many more. A similar network system that does not focus on rare diseases exists in Italy, the ARNO observatory, in which seven Italian regions (30 local health units) are involved. ▪ ***Using new technological development can help to establish new data sources.*** Especially, new telemedicine applications used by the patient offer new opportunities for data collection. But also diagnostic devices, such as the Dual Energy X-ray Absorptiometry (DEXA) bone-scanning enhancement, can be utilized to collect more relevant health data (as it is done for the UK Biobank). Furthermore, social media applications such as Twitter are extensively researched for their potential in public health research. Especially detecting adverse drug reactions. However, it has been shown that using Twitter for pharmacovigilance or public health research has severe limitations.	Threats	▪ ***Restrictive privacy law might hinder the gathering of personal health data.*** To overcome the threat of privacy violations, highest standards of data security and patient confidentiality have to be attained. The UK Hospital Episode Statistics, which collects individual health records since 1987, applies strict statistical disclosure control and stored the data in a secure data warehouse. ▪ ***Expanding existing datasets, establishing new data sources, securing compatibility of different data sources, integrating data, and monitoring its quality needs (additional) resources.*** Funding needs to be clarified, whether at national or at European level, to create sustainable datasets that contain data for several years. An example of sustainable funding is the Clinical Practice Research Datalink which was established in the UK in 1987 and is jointly financed by the NHS National Institute for Health Research and the Medicines and Healthcare products Regulatory Agency, which is part of the Department of Health.

Table 53: Results of the SWOT analysis – Data analysis

Internal factors	Strengths	▪ ***By using up-to-date/innovative analytical methods for data analysis Big Data's full potential can be realized.*** Various examples of Big Data in Health are using innovative tools of analysis, such as the FDA's Adverse Event Network Analyser that was developed to improve data analysis and visualization. Another US based initiative, the Flatiron, developed the cancer-focused data analytics platform "OncologyCloud" with the aim to build up the world's largest cancer database. ▪ ***Combining data of different and complementary sources provides improved analytical insights.*** New analytical methods can facilitate the integration of different data. The pan-European Semantic Data Platform for Healthcare for example made it possible to search different electronic healthcare documents through a single interface and analyse them using text mining technologies and multilingual semantic resources. ▪ ***Utilizing the predictive power of Big Data can increase effectiveness and efficiency in health care.*** This can be facilitated by carrying out health care research on a large scale with high quality data. An example is the case of the UK where data from the Clinical Practice Research Datalink (UK) has been used in over 1,500 research studies published in peer-reviewed journals.	Weaknesses	▪ ***New analytical methods might not be fully developed yet and distorting analysis's results.*** This weakness, which has been observed in the past (Google influenza trends), can be tackled by publishing the methods and results of data analysis in peer-reviewed journals. Furthermore the networking of the international research community should be supported to facilitate the development of new methods. ▪ ***Comparability with previous research might be jeopardized when using new and more sophisticated methods.*** This weakness should not keep researchers from advancing new methods, but rather encourage them to develop tools and techniques to utilize and analyse "old" databases to make them comparable to the more recent ones.
External factors	Opportunities	▪ ***"Innovators" and "early adopters" can foster the use of innovative Big Data analytics.*** This also includes lessons learned in other fields which might benefit the methodological application of Big Data in Health. In order to utilize this opportunity research networks that focus on data analytics in general, without a special focus on health care, should be established or supported if they already exist. ▪ ***High awareness and understanding of the demand for and the added-value of Big Data in Health can promote the development of Big Data analysis.*** This high awareness should not only be found within the IT community but also within the health care community, because interdisciplinary research teams are of special relevance in the health care sector. This interdisciplinary cooperation was utilized by the Flatiron project in the US, where a team of leading oncologists and software engineers jointly designed a software.	Threats	▪ ***Integrating data with new tools might be hindered by a lack of rules and standards for data sharing and / or data privacy issues.*** This is why a legal framework that is transparent and consistent across the EU's Member States is of high importance if methods and tools ought to be shared among researchers. ▪ ***Bad reputation of Big Data (analysis) by producing invalid results caused by new but non-sophisticated analytical methods could threaten the development of new data analysis tools.*** For the success of such methods it is necessary that they are developed by interdisciplinary teams, including clinicians. To tackle this threat information about the potentials of Big Data analysis for improving health care should be distributed among health care professionals.

Table 54: Results of the SWOT analysis – Applications

Internal factors	Strengths	▪ ***Results from studies using Big Data become more reliable if they deal with this specific kind of data in the correct way.*** It is for example important for the researcher to know how data was collected in order to estimate the risk of bias. This leads to an improved knowledge about the value of Big Data in terms of its impact on the society and the economy. ▪ ***Application of Big Data in new fields of health care research will yield new insights.*** An example for this is the Spanish Rare Disease Registries Research Network which accumulate information about patients with rare diseases across all of Spain. This larger data pool facilitates the recruitment of suitable patients for studies on disease aetiology, pathogenesis, diagnosis and therapy.	Weaknesses	▪ ***Methods and technologies for Big Data Analysis continue to evolve and therefore its application need constant revision.*** The methods used in the different fields can be quickly outdated, and new data sets might become available that are better suited for a certain field of application or open up an entirely new field of application. The field of Biomedicines for example is expanding rapidly as more powerful computers become available. ▪ ***Lack of regulations / applications may lead to myriads of available solutions.*** It is therefore important that researchers in all fields of Big Data connect and exchange methodological insights. Furthermore, a research framework (such as Horizon 2020) could be established on EU level in order to coordinate research on Big Data in Health. This would require coordination between the health care research community and the IT community. For this purpose a pan-European scientific research panel could be established.
External factors	Opportunities	▪ ***Adoption of methods and standards for Big Data applications that are already used by data scientists in other fields than health.*** This opportunity arises because in some fields the use of Big Data started earlier and is not as much impeded by data protection issues (e.g. logistics and transport). ▪ ***Due to pervasive budget constraints the applications of Big Data in Health economics is important for political decision makers.*** This opportunity is utilized by the Greek E-Prescription system, where system abuse, over-prescribing and fraud is being detected. But also other projects such as the Comet K-Project Dexhelpp (AT) or the Clinical Practice Research Datalink (UK) are using health care data for economic evaluation studies as a decision support.	Threats	▪ ***Different views on the correct way to deal with Big Data between the different disciplines might complicate an agreement on specific ways of application.*** This threat is specific to Big Data in Health as an interdisciplinary approach is necessary in almost all applications. To tackle this communication between the disciplines is of high importance. Facilitating this communication at EU level with networking grants should be considered. ▪ ***Some stakeholders / parties and even parts of the research community might not be interested in the use of Big Data and might therefore not accept the research outcomes.*** This threat can be overcome by an emphasis on sound and accepted research methods that are non-disputable. Furthermore, commonly agreed standards and regulations should be in place and followed rigorously to avoid accusations of ill research practice.

Table 55: Results of the SWOT analysis – Communication

Internal factors	Strengths	▪ ***Increased awareness of the added-value of Big Data through active public communication.*** An appropriate communication strategy should focus on target groups that are not naturally technophiliac, but who hold reservations against Big Data. However, there should be room for an open public debate where all concerns can be raised and discussed. ▪ ***Improved communication regarding privacy issues.*** This ensures the rightful usage of Big Data and can reduce prejudices against the use of Big Data. In countries were health data has been collected, stored and analysed for several years	Weaknesses	▪ ***Communication might be distorted by interests of different stakeholders.*** To tackle this it is important to communicate the potential of Big Data for all stakeholders. This needs continuous efforts from a coordinating body such as an advisory panel made up of national Data Officers, or a sub-unit of this panel that is solely installed for the public communication of Big Data in Health matters. ▪ ***Unforeseen consequences of inadequate communication.*** This can be a problem if information is wrongly interpreted, and is especially problematic regarding privacy issues. This weakness can be tackled by developing a sound communication strategy that takes specific needs and conceptions of different target populations into account.
External factors	Opportunities	▪ ***Opinion leaders in different stakeholder groups can facilitate communication.*** In order to utilize this opportunity the opinion leaders have to be identified and encouraged to participate in the development and implementation of a communication strategy. ▪ ***People's increased interest in technology might be helpful for mutual understanding between the general population and Big Data experts*.** To seize this opportunity it is very important to target population groups that are less up-to-date with new technologies, such as elderly citizens as it is very often their data that is most relevant in health care.	Threats	▪ ***Users of Big Data in Health might not be interested in an open discussion about legal and data privacy issues because they fear that this leads to a closing rather than an opening up of data sources.*** To tackle this threat, it is important to include users of Big Data in Health (health professionals, data analysts etc.) as a target group in the communication strategy and make them understand that an open debate is beneficial in the long run as this is the only way to sustainably utilize the benefits of Big Data in Health. ▪ ***Difficulties to reach target population.*** This can either be due to a lack of interest in the general population because the topic might seem too abstract and high-tech, or because of reservations against the use of Big Data in Health in general. It is important to find out the true reason behind the difficulties in reaching target populations and adapt the communication strategy accordingly.

Table 56: Results of the SWOT analysis – Human capital

Internal factors	Strengths	▪ ***Resources can be saved by integrating existing training and education programs as a basis for new curricula.*** It is not necessary to develop entirely new fields of study, but existing programmes in the field of public health or health care should integrate data analysis into their curricula and offer specialisations for students that are particularly interested. ▪ ***Minimum level of cross-linked knowledge of involved parties facilitates the whole Big Data in Health value chain.*** A basic understanding of all involved parties facilitates communication and make the whole process more efficient. For example, if the health care professional is aware of what the data that he/she collects and inserts into the system can be used for, he/she might be more careful during the process. The same is true for the data analyst, who should know under what circumstances the data was collected (e.g. high pressure due to an overcrowded waiting room)	Weaknesses	▪ ***Planning and steering human capabilities is a complex undertaking and requires coordinated action and continuous efforts.*** Even if it is achieved, the effects are only perceptible in the long run as most education and training programmes last several years. One way to tackle this could be to offer additional training courses for professionals that are already working in the field. This would speed up the process, however, a strategy would still be needed on how to recruit people for these courses. It might be wise to include professional associations (e.g. the medical association) in the planning and development of such courses. ▪ ***Resource consuming development of new training and educational programmes.*** This can be tackled by using (human) resources that are already available, e.g. by collaborations of technical and medical universities.
External factors	Opportunities	▪ ***People's technology affinity might increase interest in data sciences even if their primary profession lies within the field of health care.*** The spread of technology in everyday life might decrease barriers for using Big Data applications amongst health professionals. This opportunity is best utilized in younger students or professionals as they are generally more open to technological developments. ▪ ***Existing occupational fields and their working definitions can be altered regarding new requirements without substituting them altogether.*** This opportunity could be realized by offering extra trainings and payments for the new tasks (e.g. coding of ICD-10 codes in a hospital) to health professionals.	Threats	▪ ***Due to the rapid technological changes educational programmes might be outdated by the time of their implementation.*** This threat needs to be tackled and in order to do so, the curricula should remain flexible and additional courses and trainings for health professionals should be continuously offered. It might be possibility to appoint a pan-European group of national representatives that are experts in the field of Big Data in Health to review the most important changes in this field. This group should agree regularly on a set of key competences for health professionals and data, computational, social and public health scientists for the use of Big Data in Health. ▪ ***Unwillingness of traditional professions and / or stakeholders to open up to Big Data and refusing a change in the curricula for health professionals.*** This can be tackled by an appropriate open communication strategy that includes opinion leaders across Europe as peers. For this purpose it might be relevant to publish articles about the use of Big Data in Health in journals of traditionally sceptical disciplines, but also to place the topic on the agenda of health related conferences.

Table 57: Results of the SWOT analysis – Funding

Internal factors	Strengths	▪ ***Investments should eventually yield a return to society, if cost-effectiveness principles are applied.*** For this strength to be realized it is important that the application of Big Data in Health supports the decision making process by evaluating the economic effects of different investment options and their follow-up cost. This is already done in the UK where cost-effectiveness studies are using data from the Clinical Research Datalink and the Hospital Episode Statistics. ▪ ***Increase stability by including different public and private stakeholders and sources of financing.*** Spreading the financial risk across more than one stakeholder also increases the likelihood that the involved partners are willing to share their datasets, if they hold any. Using different sources of funding is rather common for projects or initiatives of Big Data in Health. The funding sources span from EU structural funds (E-Estonia from 2005 to 2008), over national Ministries (e.g. Comet K-Project Dexhelpp, AT) and institutions (e.g. Integrated Biobank of Luxembourg) to private investors such as Google (Flatiron, US).	Weaknesses	▪ ***Decisions about funding priorities have to be made at national level.*** This might be a lengthy process where other investments are a higher priority or seem more urgent for some of the financing parties. This weakness should be tackled by analysing the social and economic potential of Big Data in Health compared to other investments. ▪ ***Multi-source financing might involve higher need for coordination.*** This threat occurs because stakeholders might feel that if they pay, they want their interests prioritized. To tackle this weakness, it is important to lie down the rights and duties that come with a financial engagement. Clear contracts need to be in place. Templates of such contracts could be provided by the EU.
External factors	Opportunities	▪ ***The added-value of existing pilot programmes is not lost but can be accumulated in a coherent strategy.*** To realize this opportunity it is important to gain knowledge about already existing pilot programmes on Big Data in Health in the Member States. This could be done by sending out questionnaires to ask the Member States about their activities. This way, existing infrastructure can be utilized and is not lost. ▪ ***The economy as a whole might benefit from the investments in Big Data systems and infrastructure.*** Increased employment and reduced inefficiencies eventually benefit the whole society. However, it needs to be carefully considered in which projects / activities to invest. To fully utilize the potential of Big Data in Health policy makers in the Member States would probably benefit from guidelines regarding an efficient and welfare-enhancing use of Big Data in Health.	Threats	▪ ***Political opposition against investments in Big Data in Health and/or lack of available funds.*** To tackle this threat the social, economic, and budgetary impacts of an investment in Big Data in Health, compared to other health investments, but also compared to general investments, have to be analysed and communicated to the decision makers. This can be done at national level, but also at EU-level when the transferability of the results to the Member States is proven. If funds in general are unavailable, the short- or mid-term budgetary impacts of an investment in Big Data in Health have to be analysed as an aid for decision makers. A strong argument for investments in Big Data in Health, would be savings due to higher efficiency in health care provision generated by the use of Big Data. ▪ ***Investments might not yield returns due to unforeseen technological or political changes.*** This threat can be tackled by a rather rapid decision making process, including impact analysis as decision support. These impact analysis should include scenarios of political and technological change as sensitivity analysis.

6.6 Annex 6: Minutes of the Expert Workshop

3 May 2016

Expert Meeting on Big Data in Public Health, Telemedicine and Healthcare

MINUTES

DG HEALTH (4 Rue Breydel, 1040 Brussels)

Conference room: B232 08/120

Facilitator: Claudia Habl, GÖ FP

The following minutes summarize the discussions during the workshop, whose main objective was to discuss the policy actions and SWOT-analysis on Big Data in Health that was shared with the delegates prior to the workshop.

Based on this policy actions, points that should be presented as policy recommendations were collected and put on flip-charts. These recommendations will be shared by the delegated once the project team has compiled it and will be sent out for feedback and commenting (electronic Delphi method).

1.) Welcome by the European Commission/DG SANTE,

The Head of Unit, Mr. Piha Tapani stressed the focus on Big Data in Public Health (as other Directorates and units were in charge of research topics) and asked "What is the real purpose of big data in every day health care and its daily routines?" resp. "How to best use big data applications". He pointed out that the aim of today's workshop is to come up with practical recommendations for the e-Health network – a body for cross-border health care – as an endorsement by them will have more visibility in the Member States. It would be especially important to point out the added value of Big Data in Public Health, Telemedicine and Healthcare and also show that innovation is not necessarily costly.

He welcomed the representatives of other Directorate Generals and thanked the experts for their coming, especially as the originally planned workshop had needed to be postponed to today because of the tragic events in April 2016.

2.) Tour de Table

The introduction was followed by a brief presentation of all participants, who consisted of, besides the project contractor and EC Services representatives of

a. experts in health issues and in health policy working for governments of five different Member States,
b. stakeholders representing health professionals and health care payers (patient representatives were invited but were not able to participate because of the change of date)
c. experts of Health Information issues, Big Data and telemedicine from industry or multi-stakeholder associations

Names and affiliations may be found in the attached list of signature.

It was also announced that Mr. Roger Lim of DG SANTE would take over the file from Ms. Meta Geibel soon after the workshop. Ms. Geibel mentioned that experts are welcome to voice their personal opinions which not necessarily need to be identical with the opinion of their professional affiliation.

3.) Introduction to the project "Study on Big Data in Public Health, Telemedicine and Healthcare"

Ms. Anna Renner of the contractor team introduced the objectives and the state-of-play of the project. She pointed out that the first challenge was to develop, together with the commissioning party the working definition for big data in health with regard to the project tender specifications.

- Large routinely or automatically collected datasets
- Electronically captured and stored
- Reusable and multipurpose data
- Fusion and connection of existing databases
- Purpose of improving health and health system performance
- Not: data collected for a specific study

Mr. van Lente considered the definition as a bit too broad, even this was defined to be broad in the tender specifications, whereas Mr. Schuhmann endorsed the definition be reminding all to consider "how much open data has to be opened, and when we draw the line. Mr. Manset and Mr. Romao draw the attention to the new General Data Protection Regulation (EU\2016\679) that went into force on 27/4/16 and Mr. Manset asked also to consider the work of UN + US on Digital Health (matrix developed by WHO/ITU); he offered to provide contact details. Some delegates called for some form of expert committee involving stakeholders and industry as such type of working group is currently not in place on EU level. Mr. Szócska reminded on the ethical aspect of data use, also regarding secondary use of data; Mr. Romao endorsed this and mentioned the influence of national "data culture" which is quite diverse in Europe. These issues should be better stressed in the SWOT analysis. Ms. Kuipers added that we shall be able to see from the recommendations in which direction our vision points.

It was also asked why the method of choice was literature analysis; the reason is that it was demanded in the Terms of Reference but there are also qualitative elements foreseen such as the current expert workshop, the planned Delphi phase or the discussion of the recommendation on congresses and conferences. Mr. Schuhmann, Mr. Banks, Mr. Popper, Mr. Manset and Mr. Batista explicitly considered the policy action drafts as good work.

Further literature and information that were mentioned by experts are

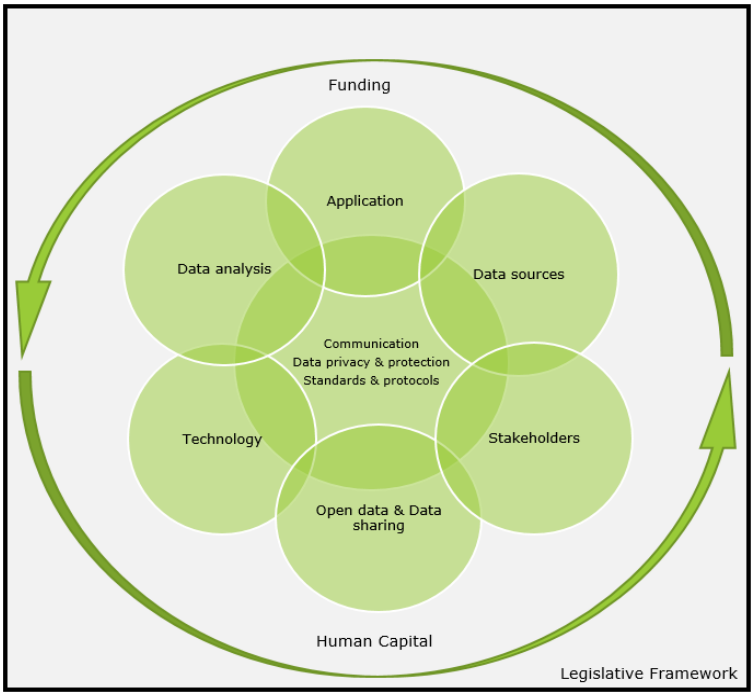
- Report of the eHealth Taskforce (https://ec.europa.eu/digital-single-market/en/news/eu-task-force-ehealth-redesigning-health-europe-2020)
- Strategy tool called "PESTLE analysis".
- WHO: Be he@lthy be mobile (mHealth: https://www.itu.int/en/ITU-D/ICT-Applications/eHEALTH/Be_healthy/Documents/Be_Healthy_Be_Mobile_Annual_Report%202013-2014_Final.pdf and http://www.itu.int/en/ITU-D/ICT-Applications/eHEALTH/Be_healthy/Pages/Be_Healthy.aspx)
- Documents from meeting of the European Alliance for Personalized Medicines http://euapm.eu/events,47.html
- governance models in other sectors (banking)

Time table and next steps: Ms. Renner explained that delegates will receive the compiled set of policy recommendations after the workshop and will be asked to again comment them in electronic format. Based on the comments the contractor will up-date the recommendations.

No further meeting is planned as of now, but the recommendations will be discussed more broadly with authorities, experts, industry and stakeholders on at least three different public congresses and meetings (e.g., E-Health Conference, Amsterdam and International Health Conference, King's College London, both in June). By end of October the final draft shall be available.

4.) Drafting of concrete Policy Recommendations based on Policy Actions

For the sake of time recommendations were not completely formulated or phrased during the session, but topics were collected on the base of the documents (policy actions and SWOT-analysis, see Diagramm outlining the framework) sent out prior to the workshop.



Besides developing topics and bullet points for the recommendations (that will be shared in a seperate document) experts agreed on the **following basic principles** that should to be adhered.

The recommendations will be shared in a separate document therefore the bullet notes are just briefly summarised at the end of the minutes.

- Recommendations shall be **pragmatic, operational** and **practical** and shall e.g. keep the global perspective in mind (e.g. what is going on in the USA and RoW).
- Recommendations shall be rather **generic** than too specific (ie. What would apply to all Member States)
- On the other hand recommendations depend on the purpose (what should be achieved) with big data, otherwise recommendations will be too broad and not implementable → we need to focus on some aspects in the beginning to at least achieve something – **step-by-step approach**
- Recommendations shall have clear statements on the <**Fair Use of Data>; Type of data** (personal health, individual vs. cohort) and **national behaviour/culture towards use of data** shall be considered
- Main **target groups** of the recommendations are: administration and clinical sector on national and EU level in the first instance, followed by providers (mainly doctors), → scope could be increased or at least it should be made clear who is the target group (patients are an indirect target group only).
- Recommendations shall consider **the <risk-benefit> profile** of big data m-health applications and clearly mention the **value of big data .**
- Speed of development in the field has to be accounted for in recommendations
- strategic plan is needed (pathway to the future) to build up a structure for big data in the future; forward looking is important! → look for good practices, and perhaps collect and assess the potential benefits and limitations in the inclusion of genomics data into Electronic Health Records.
- Define / structure the **processes** rather than the specific results or methodologies
- Decision makers need to be aware that <c**reeping> ownership of data by google, facebook** etc. happens already
- Big data should not be used as a proxy for data, it has got specific problems and should therefore have specific recommendations
- Regroup policy actions maybe by:

    - Horizontal topics such as the legal framework and communication including the privacy issues.
    - Vertical topics such as standards, training of data analysts, application and sources.
- Leverage the good things that are going (reference networks, research cloud)

Regarding **priorities**, it was made clear that all elements regarding the **communication of the benefits and challenges** of big data in health are very important, and likewise all **legal aspects** – some of the overarching level, others on very detailed level are of uttermost importance. One point to consider is the implementation of the mentioned new general data protection regulation that also concerns health data.

Finally, Ms. Renner and Ms. Geibel thanked all delegates for their active contributions and closed the session.