



Draft proposal for the eHealth Network

D5.3 – Guidance for the implementation of common principles for practical governance of big data

**with a special focus on data to be used (and the implementation of data access and use) in
public health, research and quality assurance in healthcare on a European scale**

WP 5 – Innovative Use of Health data

22-10-2020

Version 1.0

18th eHealth Network meeting, November 2020

Draft for Discussion

Grant Agreement n° 801558



Co-funded by the European
Union's Health Programme
(2014-2020)

To ease the uptake of innovative usage of data across the healthcare sector for the benefits of society, individuals and performance of Member State health systems, this guidance proposes recommendations for the support of creation of good governance principles, practices and methods in handling use of health data, including big data.

CONTROL PAGE OF DOCUMENT	
Document name	D 5.3 – Draft proposal for the eHealth Network on the guidance for the implementation of common principles for practical governance of big data with a special focus on data to be used (and the implementation of data access and use) in public health, research and quality assurance in healthcare on a European scale
Work Package	WP 5 – Innovative Use of Health data
Dissemination level	
Status	Draft
Author(s)	
Beneficiary(ies)	

Dissemination level:

PU = Public, for wide dissemination (public deliverables shall be of a professional standard in a form suitable for print or electronic publication) or CO = Confidential, limited to project participants and European Commission.

REVISION HISTORY				
Version	Date	Author	Organisation	Description
0.1	01-04-2020	István Csizmadia, Robert Láng, Márton Kis, Anna Feller, Krisztina Sipos	NHSC, SU	First version of the document.
0.2	13-07-2020	István Csizmadia	NHSC	Chapter on definitions, introduction added
0.3	27-07-2020	Kathrin Trunner	GOeG	Results from FGM and first version of principles included
0.4	04-09-2020	István Csizmadia, Robert Láng, Márton Kis, Anna Feller, Antal Bódi, Kornél Tóth, László Bencze, Claudia Habl, Kathrin Trunner, Alexander Degelsegger- Márquez, Cátia Pinto	NHSC, SU, GOeG, SPMS	Restructured version of v0.3 with details of principles, updated for QM
0.4a	13-09-2020	Hugo Agius Muscat	MFH	QM review of version 0.4
0.4b	16-09-2020	Anna Feller, István Csizmadia	NHSC, SU	Updated version after QM review

REVISION HISTORY				
Version	Date	Author	Organisation	Description
0.4c	20-09-2020	Hugo Agius Muscat	MFH	QM feedback on version 0.4b
0.4d	20-09-2020	István Csizmadia	NHSC	Updated version after QM feedback
0.4e	12-10-2020	László Bence, Kornél Tóth, Kathrin Trunner, Diogo marins, Cátia Pinto, Arto Vuori, István Csizmadia	SU, GOeG, SPMS, THL, NHSC	Updated version addressing EC comments
0.4f	20-10-2020	Hugo Agius Muscat	MFH	QM review of v0.4e
1.0	22-10-2020	István Csizmadia	NHSC	v0.4f approved by SC

Disclaimer

The content of this deliverable represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use of its contents.

Table of Contents

TABLE OF CONTENTS	4
ACRONYMS	5
LIST OF FIGURES AND TABLES	6
EXECUTIVE SUMMARY	7
INTRODUCTION	9
BACKGROUND	10
WORKING DEFINITIONS	12
VISION, PURPOSE AND OBJECTIVES	12
AUDIENCE AND SCOPE	15
METHODOLOGICAL APPROACH	16
Problems	16
Domains and use cases	16
Results from the Focus Group Meetings	18
Conclusions from the FGMs on the design of the principles	19
PRINCIPLES	19
PRINCIPLES FOR TURNING PROBLEMS INTO OPPORTUNITY, ACTION AND VALUE	19
CLUSTERING OF PRINCIPLES AND PRINCIPLE MAPPING	20
RELATIONS, INTERDEPENDENCES AND PRIORITISATION AMONG PRINCIPLES	21
CLUSTER 1: INCREASE TRUST IN PRIVACY PROTECTION AND CYBER SECURITY	24
Principle 1 – Work with anonymised or pseudonymised data	24
Principle 2 – Develop digital health literacy	25
Principle 3 – Disseminate data reuse purpose well	27
Principle 4 – Make a great effort to secure and protect data	28
CLUSTER 2: INCREASE EFFICIENCY OF USING FUNDS AND FINANCIAL RESOURCES	30
Principle 5 – Use existing technical standards	30
Principle 6 – Involve stakeholders in co-creation	31
Principle 7 – Consider current and future needs	33
Principle 8 – Foster data reuse	34
CLUSTER 3: FOSTER COMMON INTERPRETATION OF LEGAL AND ETHICAL RULES	36
Principle 9 – Adapt model to the settings	36
Principle 10 – Create code of conduct	38
Principle 11 – Speak the same language	40
Principle 12 – Aim for win-win of all stakeholders	41
GUIDANCE FOR THE IMPLEMENTATION OF COMMON PRINCIPLES	43
APPROACH AND CONTENT	43
NEXT STEPS	44
APPENDICES	45
1. MAIN FINDINGS	45
2. WORKING DEFINITIONS	47
3. TABLES	52

Acronyms

Acronym	Description
AP	Associated Partner
CEF	Connecting Europe Facility
CoC	Code of Conduct
D5.1	Deliverable of Task 5.1 in Work Package 5 of eHealth Action
D5.2	Deliverable of Task 5.2 in Work Package 5 of eHealth Action
D5.3	Deliverable of Task 5.3 in Work Package 5 of eHealth Action
DCF	Data Conversion Framework or 'the stakeholder value chain analysis framework for data conversion'
DHL	Digital Health Literacy
eHAction	eHealth Action – 3 rd Joint Action supporting the eHealth Network
eHDSI	eHealth Digital Service Infrastructure
EHDS	European Health Data Space
eHN	eHealth Network
eIDAS	electronic Identification, Authentication and trust Services
EU	European Union
FAIR data	Findable, Accessible, Interoperable and Reusable data
FGM	Focus Group Meeting
GDPR	General Data Protection Regulation
HP	Health Professional
MWP	Multiannual Work Programme
PA	Priority Area
PLA	Policy Level Action
T	Task
TL	Task Leader
WP	Work Package
WPCo-L	Work Package Co-Leader
WPL	Work Package Leader

List of figures and tables

Figure 1 - The way how principles affect practical governance of big data domains in health...	7
Figure 2 – Innovative use of health data.....	13
Figure 3 - Root causes of major problems.....	17
Figure 4 - Use cases and domains.....	18
Figure 5 – Principle ratings before FGMs.....	19
Figure 6 – Principle ratings before FGMs.....	19
Figure 7 - Principle mapping.....	21
Figure 8 - Accelerating and enabling relations between principles.....	22
Figure 9 - Action timing.....	23
Figure 10 – Approach to prepare guidance.....	43
Table 1 - Overview Work Package 5.....	52

Executive summary

One of eHAction's main objectives is to improve the knowledge base for health and healthcare policy, envisioning development of new ways in the usage of data across the healthcare sector. Tasks related to this objective are included in Work Package 5 (WP5): 'Innovative use of health data'.

This document includes Deliverable 5.3 (D5.3): Draft proposal for the eHealth Network – Guidance for the implementation of common principles for practical governance of big data with a special focus on data to be used (and the implementation of data access and use) in public health, research and quality assurance in healthcare on a European scale.

The draft proposal provides information about the background, objectives, planned structure, working definitions and methods of D5.3, and contains recommendations on common principles for practical governance of big data.

According to the main findings of WP5 deliverables D5.1 and D5.2¹, a series of challenges and obstacles were identified, like confidentiality issues, legal uncertainties, lack of trust, gaps in the legal environment, the level of digital literacy, lack of business models, interoperability issues, safety of health data, cybersecurity, intellectual property issues, unsolved issues of sharing benefits among stakeholders, and scarce resources including the number of data scientists. These hurdles will need to be solved or dealt with before the aim to create a European Health Data Space can be achieved.

WP5 has been looking for principles to mitigate or eliminate the root causes (and through them major problems) hampering Member States/countries from implementing measures to foster innovative use and reuse of health data in key use cases and domains. Figure 1 below illustrates this approach.



Figure 1 - The way how principles affect practical governance of big data domains in health

Principles were drafted, based on literature, with three main pillars: The Study on Big Data in Public Health, Telemedicine and Healthcare (EU Big Data Study)², the OECD Ministerial Statement from 17 January 2017³ and previous results from Work Package 5, namely deliverables D5.1 and D5.2. In the second phase of the work, the draft principles were discussed

¹ For further details see Appendix 1

² https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf

³ <https://www.oecd.org/health/ministerial/ministerial-statement-2017.pdf>

with actors and stakeholders in a two-step process: a survey, followed by a workshop held online on 23-25 June 2020 containing six focus group meetings with different stakeholder groups.

Following the idea to recommend principles that eliminate or mitigate root causes of challenges, WP5 formed the following three clusters to group the 12 principles, each related to one root cause:

- Cluster 1: Increase trust in privacy protection and cyber security

- Principle 1 – Work with anonymised or pseudonymised data
- Principle 2 – Develop digital health literacy
- Principle 3 – Disseminate data re-use purpose well
- Principle 4 – Make a great effort to secure and protect data

- Cluster 2: Increase efficiency of using funds and financial resources

- Principle 5 – Use existing technical standards
- Principle 6 – Involve stakeholders in co-creation
- Principle 7 – Consider current and future needs
- Principle 8 – Foster data re-use

- Cluster 3: Foster common interpretation of legal and ethical rules

- Principle 9 – Adapt model to the settings
- Principle 10 – Create code of conduct
- Principle 11 – Speak the same language
- Principle 12 – Aim for win-win of all stakeholders

Finally, D5.3 proposes guidance for the implementation of the above mentioned common principles.

The final milestone of work package 5 – “Discussion Paper for the eHN on the implementation of common principles for practical governance of big data with a special focus on data to be used in public health, research and quality assurance in healthcare on a European scale with guidance on implementation of data access and use (Deliverable D5.3)” – shall be reached by April 2021, by submitting the final report containing the proposal for the guidance for implementation of the principles as well.

Introduction

eHAction is a Joint Action that is supporting the eHealth Network (eHN) on some specific objectives related to the priorities of the current Multiannual Work Programme of eHN (MWP 2018-2021). The main eHAction work is based on establishing interoperability for enhancing safe cross-border exchange of health data between health professionals and healthcare providers. This kind of initiative can generate access for patients to their health data, wherever they are located. It is of utmost importance to be able to sustain a mechanism after the present Joint Action, as well as after the end of the Connecting Europe Facility (CEF) instrument. Further co-funding mechanisms could be designed and proposed, including those regarding modernisation of digital infrastructures (at European, national or regional level) to enable and maintain the cross-border exchange of health data, as well as the access of patients and professionals to this data for primary and secondary use.

One of the eHAction's main objectives is to improve the knowledge base for health and healthcare policy, visualising development of new ways in the usage of data across the healthcare sector. Tasks related to this objective are included in Work Package 5 (WP5): 'Innovative use of health data'. WP5 focuses on priority areas B.1, B.2 and B.3 of MWP 2018-2021. Secondary use of data and big data can provide value for research, teaching, managing and planning healthcare systems. It can also be a great opportunity for the development of personalised medicine, the improvement of the effectiveness of medicines, efficiency of health systems, and continuity of care. It is also a much-needed opportunity for improvement of traditional public health surveillance systems, prevention and control strategies. The COVID-19 pandemic stressed the need to use data to improve early warning and response systems, adding precision and timeliness to disease surveillance. WP5 faces the problem of lack of awareness of these potential benefits and the need to share expertise.

This draft proposal provides information about the background, objectives, planned structure, working definitions and methods of Deliverable 5.3, and contains recommendations on common principles for practical governance of big data. The final report, containing proposal for the guidance for implementation of the principles, shall be delivered by May 2021.

The document builds upon:

- Deliverable D5.2 – 'Report on identified cross-border use cases, including assessment of pros & cons of stakeholders, and practical solutions with potential for European scale benefits', which was delivered and adopted during the 16th eHealth Network meeting in November 2019;
- Deliverable D5.1 – 'Report on policy-level actions on innovative use of big data in health', which was adopted at the 17th eHealth Network meeting in June 2020.

Background

WP5 had its kick-off meeting in Brussels in October 2018, met stakeholders at DIBSS Conference in Dubrovnik in May 2019 and held a workshop in Prague in September 2019 with the participation of Member State representatives, where experts tested and validated a canvas tool intending to capture the rationale behind the lack of use of already collected health data for better health outcomes. Using this canvas tool, named 'the stakeholder value chain analysis framework for data conversion' (in short: Data Conversion Framework or DCF tool) WP5 mapped privacy aspects, as well as identified obstacles that prevent Member State/country policies from being replicable either in other Member States/countries or on an EU level. The workshop contributed to finalising the DCF tool to identify key stakeholders, their needs, goals, offers/evaluation, behaviour and attitude, as well as interdependencies, obstacles and possible interventions and expected results.

In order to make a smooth transition from tasks T5.1 and T5.2 to task T5.3, WP5 joined the workshop on health data governance for secondary use, named 'Towards the European Health Data Space - National Strategies for Secondary Use of Data in the Context of National and EU Digital Health Networks', organised by the co-leader of WP8 in Lisbon in January 2020. From 23rd until 25th of June six focus group meetings (FGMs) on common governance principles for the reuse of health data were held by the T5.3 team as virtual workshops.

Summarising the results of our analysis, the following conclusions will be considered:

- Confidentiality issues and legal uncertainties are highly interdependent; without adequate regulation patients are more reluctant to give consent to share their data. Thus, more broadly promoting a FAIR data approach (i.e. Findable, Accessible, Interoperable, Reusable) and patient-curated data principles and strengthening digital health literacy among citizens would be beneficial.
- Lack of trust can be traced back also to a lack of normative regulation, rules, terms or codes of conduct (beyond GDPR) on the one hand, and lack of willingness (to establish regulations, rules, terms or codes of conduct) or empowerment (to follow them) and fear of abuse on the other hand.
- GDPR in many cases did not accelerate, but rather impeded, the process of data sharing and processing, and caused confusion in governance of data access (unclear implementation and lack of knowledge about technology at policy making and regulatory levels, as well as misunderstanding needs of stakeholders). More work is necessary to identify and overcome the actual barriers.
- A solid legal environment is a prerequisite to setting up suitable data infrastructure and education/training plans.
- Educational efforts are still necessary, also with so-called "digital natives", to have a digitally literate population; but also, measures targeted at health professionals are required: here one building block could be to revise existing curricula and add at least one semester block on digital health.
- Safety of healthcare data and transparency of analytical tools are of major concern in establishing trust over privacy; cybersecurity threat is imminent in several sections of the healthcare data value chain.

- Much data exists still only in paper format or as “scanned” but not searchable documents, and databases are not interoperable even on national level; governments also must recognise this and provide funds to enhance digitisation, but there is a limited number of valid use cases.
- Policy makers and market regulators together with stakeholders in the data value chain should set up and understand the business models of healthcare data management to provide enough financial resources in subsidies or in other formats (public-private partnership, etc.); without sustainable business models, results and the inclination to provide broader funding could be constrained.
- Standardisation of data and fulfilment of the criteria of FAIR data principles is in the interest of all stakeholder groups but it needs time, money and proper design thinking with the right staff; all these are scarce resources.
- There is considerable reluctance to learn at care giver and payer levels, due to no clear vision about the short-term benefits of using healthcare data and lack of enough financial resources.
- In most Member States/countries, the level of digital health literacy of patients and care givers is low, which damages both awareness and trust. Efforts have to be made to improve digital health literacy on the one hand and to build up trust in the safe and justified use of individual, pseudonymised, data in the field of health and social care. To achieve this, a joint EU-wide public awareness campaigns seems to be a good instrument for the task team.
- There is lack of data for scientists participating in big horizontal healthcare projects contributing to successful scientific results and valid business cases which could function as proof to enhance state and EU funding.
- There is misunderstanding of the difference between ‘analysis’ and ‘reporting’, followed by uncertainties about implications and unclear impacts on intellectual property issues, liability, transparency or costs of mitigating negative effects;
- There are unsolved issues regarding the sharing of benefits among stakeholders in accordance with consent and ownership and invested time and resources in data capture, storage, development, processing, analysis and use (e.g. intellectual property, wages, fees, knowledge, better working or care conditions).

Considering that T5.1 has been delivering information on the obstacles (and reasons) preventing Member States/countries from implementing (more) effective actions in (more) efficient ways to foster successful transformation of data into innovation, we started to match this information with the plans or results and experiences of national strategies which could contribute to making the value we want to deliver in T5.3 as a next step.

Besides all the research and analysis above, widening our scope is inevitable, as health digitalisation is not a separate effort; it should be aligned with other changes induced by the Digital Single Market efforts, and also Member State level changes in digitalisation of public administration, as well as developments and interdependencies with other industries (IT, cybersecurity, energy, environment, transport, logistics, etc.). New possibilities and challenges are arising from the continuous upgrading of underlying infrastructure (eIDAS, eHDSI, etc.) and their legal framework. This should be also taken into consideration when working at T5.3.

Furthermore, the opportunity to remove limits to scale, scope and learning opens the door to a fairly unlimited set of use cases for innovative use of health data and big data too. This trend has been highlighting the importance of innovative use of health data, since the 2018-2021 Multiannual Work Programme (MWP) of the eHealth Network was adopted. In parallel, the European Commission has identified the need and importance of EU level data spaces and repositories, as potential enablers and catalysts of innovative use of health data.

Working definitions

Working definitions were elaborated by WP5 members at the kick-off meeting in order to lay down statements of the meaning of certain phrases which have been explained in various ways in the literature. These working definitions were introduced to the eHealth Network in November 2018, and are considered as continuously evolving definitions based on the evolutionary and constantly changing nature of the field. Definitions reflect the value-based approach followed in WP5, where value refers to satisfaction of a specific need and replicability at an economical cost. When elaborating deliverable D5.3, WP5 contributors added definitions for some new terms to the list. The full set of WP5 working definitions are available in Appendix 2.

Vision, purpose and objectives

Work package 5 has a vision to develop innovative usage of data across the healthcare sector knowledge base for healthcare policy and other purposes.

WP5 looks at data as key driver of disruptive innovation in the economic, societal and social systems. Therefore, WP5 is committed to make an impact on the use of health data, as well as assisting data-driven innovation leading to patient-centred health systems, evidence-based health policies and decision-making.

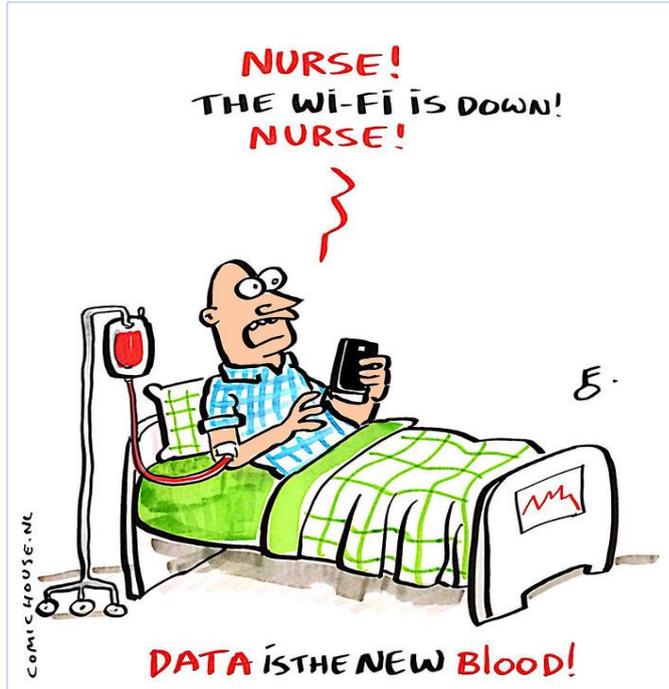
WP5 is committed to assist decision makers understanding and exploiting main advantages of the innovative use of health data.

The purpose of this document is to assist bringing closer altering approaches, adherence and empowerment of decision makers. There is much uncertainty on how to go forward on benefiting from reuse varying datasets and apply artificial intelligence on the practical level.

On the policy level it is important to settle common approach and strengthen the awareness, adherence or empowerment to identify the strengths, opportunities, weaknesses and threats of the reuse of data, in order to be able to realise possibilities to deliver societal and economic benefits of innovative use of data and big data.

It is also important to raise awareness not only of the value big data offers to improve health, but also of the several operational barriers to reuse data to realise its potential. Data collections, extractions, quality, semantics, interoperability – to name a few – are not a given in the big data era. There is a need to raise health policy makers' awareness of the data related services needed, as digital transformation changes current data assets.

What are the main advantages of the innovative use of health data?



Source: Floris Oudshoorn - ComicHouse / EHFG

Figure 2 – Innovative use of health data

Innovative use of health data, regardless of whether the use is primary or secondary, fosters innovation in the field of public health interventions, prevention strategies and health system management, as well as in the organisation and provision of health services and medical care, including health promotion and disease prevention interventions.

It has the potential to improve public health outcomes, enhance the quality of care to patients and respond to unmet needs, and also to foster the competitiveness of stakeholders and to improve the cost-efficiency and sustainability of health services and medical care. Innovative use of health data will also uncover new ways for surveillance and control of global health threats and emerging infectious diseases, upscaling health systems surveillance and response capacity.

The overall objective of WP5 is to support the application of good practices in Member States and provide guidance at European Union level on handling big data in health within the existing EU regulatory framework, on secondary use of personal health data, and consequently to ease the uptake of innovative usage of data across the healthcare sector for the benefits of society, individuals and performance of Member State health systems.

Task 5.3 **'Towards an attempt to define common principles for practical governance'** foresees to deliver documents on:

- Making available guidance on practical governance for eHN and Member States.
Providing a framework for the implementation of common principles for practical governance of big data, including privacy protection and security, aiming at improving health data transferability across borders, with a special focus on data to be used in public health, research and quality assurance in healthcare on a European scale.
- The guidance will include recommendations on the implementation of data access and focus on helping Member States to utilise the potential of harnessing new opportunities arising from big data and improved data analytics capabilities, as well as from personalised medicine, use of clinical decision support systems by health professionals,

and use of mobile health tools (applications and technology) for individuals to manage their own health and chronic conditions, in order to:

- facilitate preparation of actions to improve the comparability, accuracy and reliability of health data and to encourage the use of health data to enable more transparent and patient-centred health systems focusing on health outcomes and evidence-based health policy and decision-making, as well as to promote data-driven innovation;
 - to enable the use of health data for research and innovation, in full compliance with data protection requirements and FAIR data principles;
 - apply network optimisation to cross-border IT infrastructure and data flow management;
 - foster patient-centred interoperability;
 - improve service effectiveness for the individual patient in which benefits are experienced locally;
 - enhance interdisciplinary and openness that removes barriers (silos) between data sources and infrastructure to provide 'fit for purpose' data platforms.
- The Proposal (Paper) on common principles for big data governance will contain practical guidance on governance of big data and knowledge.
 - Common principles to facilitate the development of innovative use of data projects at European level. Principles will assist in improving recognition of the practical conditions for rational governance of big data in eHN and Member States/countries.
 - The use of big data with best practices of good governance on handling big data complying to current regulations and at the same time bringing patients at the centre of the digital health should be promoted. Common approaches are needed to create enabling conditions for data integration and analysis in a secure environment, as well as to achieve necessary scale for advanced data analytics, for example in the field of genomics and personalised medicine. Another field for common work should be improving data infrastructure, analytics and decision support to predict, prevent and control serious cross-border health threats. It is necessary to share best practices on how issues concerning privacy, protection and security, transparency around criteria and automatic processes/algorithms in the use of big data in healthcare – technical and governance – are being addressed nationally or at European level . The focus will be about facilitating the innovative use by designing and testing appropriate methods and governance.
 - T5.3 aims to build practical guidance on practical governance of big data and knowledge to Member States/countries based on their big data efforts and practices – improving recognition of the practical conditions for rational governance of big data in eHN and Member States/countries (in order to ensure patient-centred health systems, evidence-based health policy and decision-making, as well as data-driven innovation).
 - There is a special focus in our guidance on laying the basis for bridges over the gaps between different practices and legal interpretations among data owners and users around Europe (and the world). These differences have necessitated a common set of

rules and terms. Therefore, D5.3 provides guidance for Member States/countries how to prepare and foster encouraging and supporting production of a code of conduct for health data sharing, accessing and processing for primary and secondary use. By this deliverable T5.3 contributes to prepare the work of the next relevant joint action⁴ as well.

Audience and Scope

Data governance entails defining, implementing and monitoring strategies, policies and shared decision-making over the management and use of data assets.⁵

Therefore, governance of big data in health is the process of managing the availability, usability, integrity and security of the consolidated data from existing fragmented data sources for the purpose of understanding, forecasting and improving health and health system status, needs and performance. It is based on standards and policies that control secondary use of data, including respect for privacy, quality and integrity of data, and access to data.

These principles and guidance for their implementation are addressed to national and Community legislators and policy makers. They have a wide focus on data to be used (and the implementation of data access and use) in public health, research and quality assurance in healthcare on a European scale.

Governance of data reuse shall focus on:

- assisting academic and sponsored research
- providing tools and evidence in public health
- strengthening quality assurance in healthcare:
 - supporting institution management and system governance
 - improving patient care by assisted decision making, diagnostic analytics, robots (software and equipment), smart devices, telehealth, mHealth and remote care
 - reducing (hospital) re-admission rate
 - forecasting patient demand, attitude, behaviour and need
 - predicting outbreaks and spread of the epidemics
 - optimising administration, finance, procurement, inventory, infrastructural capacities, workforce management and investment
 - etc.

⁴ JA-04-2020, AWP Ref.: 1.2.2.8, Topic: Ironing out differences in national General Data Protection Regulation (GDPR) implementation in the health sector — development of a code of conduct for data processing (Article 40 GDPR).

⁵ https://ec.europa.eu/info/sites/info/files/summary-data-governance-data-policies_en.pdf

Methodological Approach

Principles were drafted, based on literature, with three main pillars: The Study on Big Data in Public Health, Telemedicine and Healthcare (EU Big Data Study)⁶, the OECD Ministerial Statement from 17 January 2017⁷ and the results from Work Package 5, namely the deliverable D5.1 and D5.2.⁸ Then followed by a large-scale discussion with actors and stakeholders in the field (e.g. public health, legal experts, digital health, technical persons, EU-wide networks and EU services) in a two-step process: 1) A survey with a first prioritisation was conducted, and 2) a series of six workshops took place, with altogether attendees clustered as follows: 50% from public bodies, 22% from European Commission and Agencies, 20% from international organisations, and 8% from other sectors (academia, R&D).

WP5 has been looking for principles to mitigate or eliminate the root causes (and through them major problems) hampering Member States/countries to implement measures to foster innovative use and reuse of health data in key use cases and domains. Figure 1 illustrates this approach (see page 7).

Problems

Major issues holding the EU back from realising its potential in the data economy were introduced in the latest EU Data Strategy⁹. These problems have their impact on healthcare and health economy, too. According to the assessment made by WP5, the root causes beyond the obstacles faced in the governance of secondary use of health data and big data, having been identified by the DCF tool, have been considerably affecting the problems listed in the EU Data Strategy as well (see Figure 3 on page 17).

Domains and use cases

Successful implementation of the principles to develop and sustain a friendly environment for secondary use of health data and big data depends on the recognition of the specificities of the domains and use cases where they will be used. Therefore, development of the guidance for implementation is based on having these specificities into consideration. The term 'domain' refers to one or several data sets that covers particular topics.¹⁰ Principles recommended by WP5 affect practical governance of big data in the following topics (domains):

1. Policy, Governance and Management (PGM)
2. Education and Training (ET)
3. Legal, Ethical and Regulatory Issues (LER)

⁶ https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf

⁷ <https://www.oecd.org/health/ministerial/ministerial-statement-2017.pdf>

⁸ LINK to D5.1 and D5.2 (ehaction.eu)

⁹ A European strategy for data, Brussels, 19.2.2020 COM(2020) 66 final (<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>)

¹⁰ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1596807664480&uri=CELEX:32019R1700>

4. Interoperability and Technical Standards (ITS)
5. Dissemination and Stakeholder Involvement (DS)

<u>Problems</u>	Root causes		
	Lack of trust	Lack of financial and funding resources	Legal uncertainties
Not enough data available for reuse	XX	X	X
No real user empowerment	XX		
Lack of data processing and storage solutions	X	XX	
Absence of comprehensive data governance approaches (legal & technical barriers)		X	XX
Skills shortage and low data literacy	X	XX	X
Fragmentation of the single market		X	XX
Healthcare systems are separated, what is not the case in other systems	X		XX

Figure 3 - Root causes of major problems

WP5 identified 14 use cases of secondary use of health data and big data. Each of these use cases can be related to two or more domains. WP5 identified following relations (see Figure 4 below):

<u>USE CASES</u>	PGM	ET	LER	ITS	DS
1. Forecasting patient demand, attitude, behaviour and need	X				X
2. Hospital re-admission reduction	X	X	X		
3. Public health, surveillance and prediction of health threats and diseases		X	X		X
4. Clinical decision support, assisted diagnostic analytics and robots		X	X	X	
5. Personalised medicine	X		X		X

<u>USE CASES</u>	PGM	ET	LER	ITS	DS
6. Robots, smart devices, telehealth, mHealth and remote care	X	X	X	X	X
7. Administration, finance, inventory and investment optimisation	X		X		
8. Prediction of procurement demand and workforce needs	X	X	X		
9. Big data assisting legislation, governance and regulatory activities (e.g. disease epidemiology or product related)	X		X	X	X
10. Hidden citizen behaviour patterns analysis		X	X		X
11. Insight and evidence for policy-making and investments	X		X		
12. Prediction of health and health related needs, expenditure and costs		X	X	X	
13. Academic, clinical and translational research and sponsored research			X		X
14. Provision of tools and evidence			X	X	

Figure 4 - Use cases and domains

Results from the Focus Group Meetings

From 23rd to 25th June 2020, six focus group meetings (FGMs) on common governance principles for the reuse of health data were held by the T5.3 team as virtual workshops. The three days were organised for three different target groups/audiences:

DAY 1. Technology and Interoperability, Tuesday, 23 June 2020

DAY 2. Public Health and Research, Wednesday, 24 June 2020

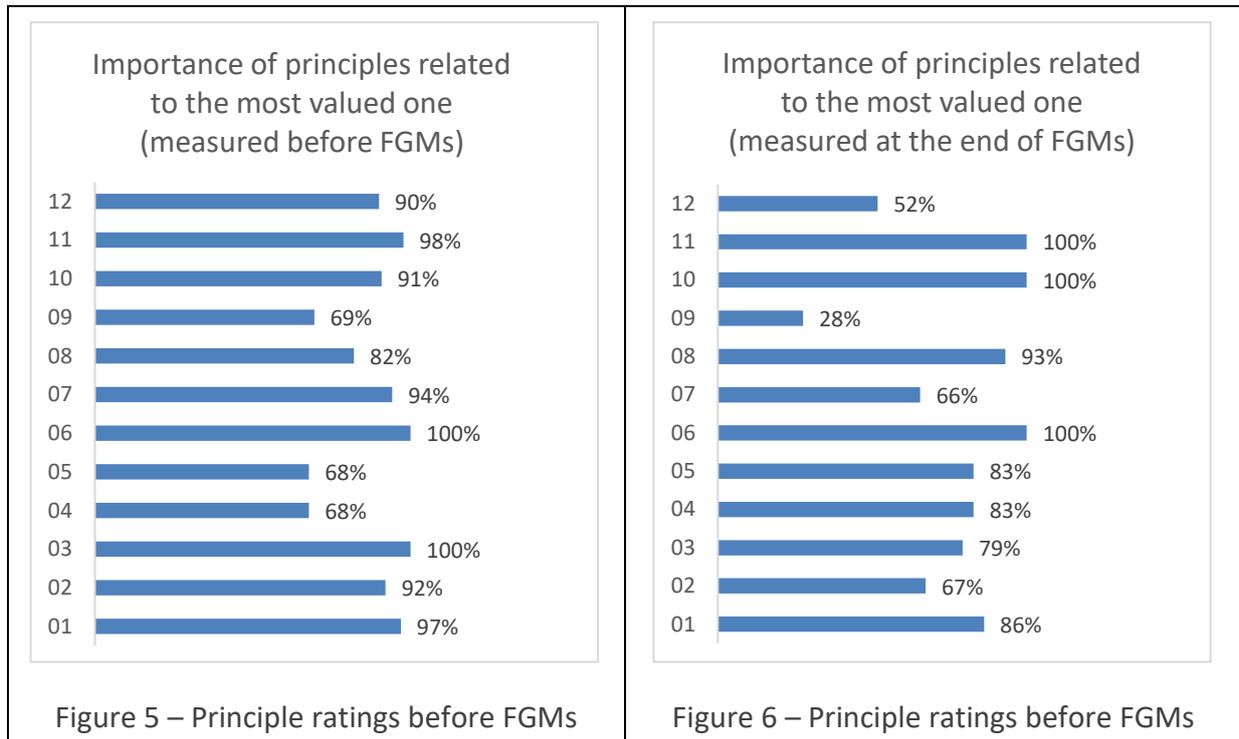
DAY 3. Policy and Regulation, Thursday, 25 June 2020

A total of 101 participants from 26 European countries attended the 6 events.

FGMs helped WP5 to find replies to questions regarding policy level unmet needs for common principals and practical guidance why and how to foster innovative use of health data, big data and AI, as well as how to implement available relevant recommendations. In addition, 12 draft principles, developed by WP5 prior to the FGMs, were discussed allowing the WP team to further develop and reshape them in order to prepare their submission to the eHealth Network. Principles were also rated and prioritised by participants putting them in order of importance.

Participants could rate the preliminary principles during the registration period, and the result of the rating was introduced at the PGMs (see Figure 5).

Participants were asked to prioritise the further developed and reshaped principles at the wrap-up of the FGMs. The aggregated result of the voting is introduced in Figure 6.



Conclusions from the FGMs on the design of the principles

- ❖ Principles shall be designed for policy makers – Policy/macro level + Best practices
- ❖ Principles shall support policy makers to offer win-win data reuse governance model for their key stakeholders – Interoperable and aligned with stakeholders’ principles and programmes (e.g. Data Saves Life)
- ❖ Principles shall be easy to start and implement - Low hanging fruit
- ❖ Principles shall be effective – Eliminate or mitigate root causes of challenges
- ❖ Principles shall be efficient – Effective + no side effects + minimised costs

Principles

Principles for turning problems into opportunity, action and value

WP5 has intended to recommend common principles for governance of big data which turn problems into opportunity (furthermore, into action and value for stakeholders) by eliminating or mitigating root causes of the issues facing privacy protection and security or transferability and secondary use of health data in public health, research and quality assurance across borders. The opportunities and resources to improve health data transferability and management across borders ensure that citizens have control over their own personal data.

WP5, by identifying the principles, has been looking for responses to the following questions in order to turn problems into opportunity, action and value:

- **How can we make use of, maintain and develop the opportunities and resources?**
- **Is it possible and useful to agree on international terms of sharing risks of liability and costs of cross-border data sharing and access?** – Can we use the Incoterms® rules¹¹ as best practice, which are the world's essential terms of trade for the sale of goods?

The below issues also need further consideration to turn problems into opportunity, action and value:

- Content opportunities and core values of smart information
Registries
Delivery of high-quality information
- Infrastructural opportunities - Efficiency
Physical infrastructure and building blocks - integration and interoperability
- Data-driven innovation in healthcare – Unmet needs and expanding demand
Available and affordable technology
- Knowledge, skills, capabilities, capacities
- Legal opportunities and elimination of legal uncertainties
- Creation, maintenance and development of trust
- Increase and concentration of funding and financial resources

Clustering of principles and principle mapping

Following the idea to recommend principles that eliminate or mitigate root causes of challenges, WP5 formed 3 clusters to group the principles, each related to one root cause.

Figure 7 shows the allocation of principles in the map of the following 3 clusters:

- Cluster 1: Increase trust in privacy protection and cyber security
- Cluster 2: Increase efficiency of using funds and financial resources
- Cluster 3: Foster common interpretation of legal and ethical rules

Some principles can be related not only to one root cause, but more.

When grouping principles into clusters, each principle was linked to the root cause that could be most influenced by that principle. E.g. Principle 2 (Develop digital health literacy) is most linked to the goal to increase trust in privacy protection and cybersecurity, but it also contributes to increase efficiency of using funds and financial resources on the one hand, and it helps to foster common interpretation of legal and ethical rules on the other hand. Principle

¹¹ <https://iccwbo.org/resources-for-business/incoterms-rules/incoterms-2020/>

5 (Use existing technical standards) recommends to invent or develop new standards if the use case requires, otherwise use available ones to save time or costs and make use of global or European interoperability. In this way, if you deploy existing standards, your system would be more efficient. In addition, you can communicate that you chose standards that have proven to strengthen privacy, data security and safety. In this way implementation of Principle 5 can contribute to increase trust in privacy protection and cybersecurity as well.

Another aspect of the grouping was whether the elimination of a root cause by the principle also contributes to mitigating the effect of another one. E.g. Principle 3 (Disseminate data reuse purpose well) could contribute to increase efficiency of using funds and financial resources if the level of trust in privacy protection and cybersecurity had been raised.

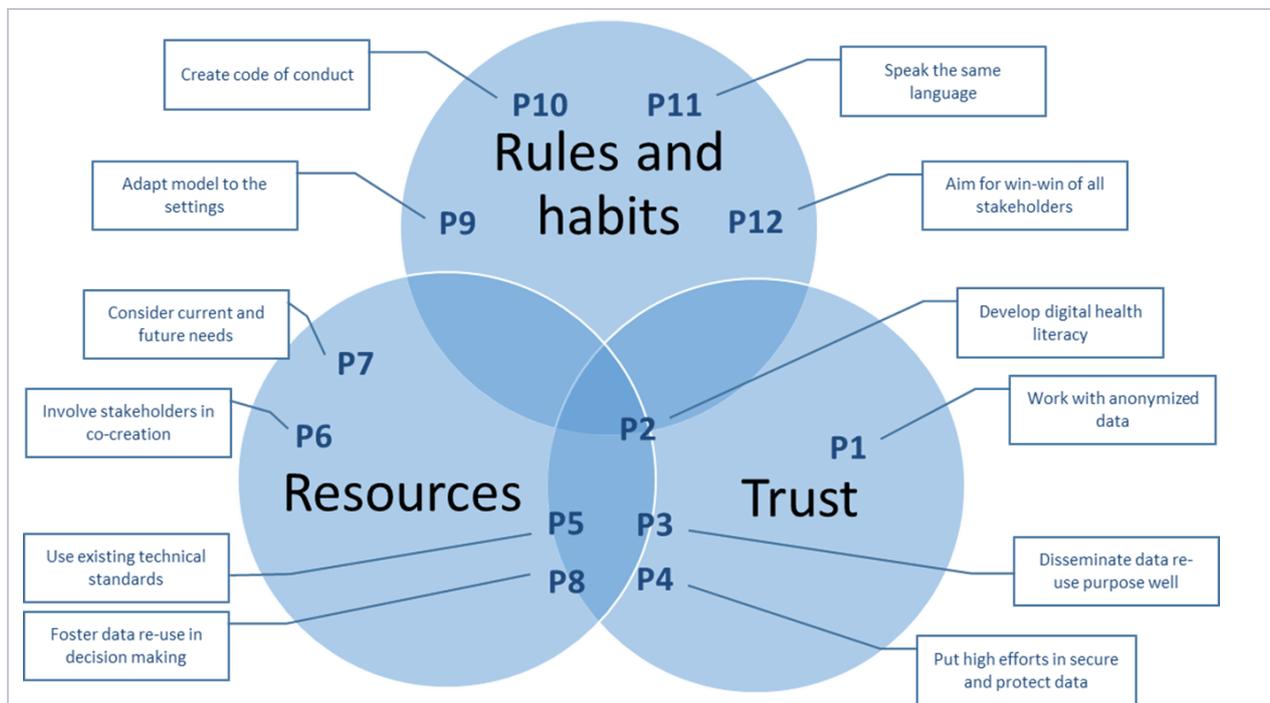


Figure 7 - Principle mapping

Relations, interdependences and prioritisation among principles

Any single principle cannot solve all issues related to the major problems but should be part of a complex toolbox. They should also have links to some other principles in the same cluster or another.

Visualising how our efforts with realising one principle affect the value/success factor of another one, helps to decide which principle should be implemented first. In Figure 6, “accelerating relations” between two principles, where principles listed in the rows have quick positive impact on the other ones listed in the columns by encouraging or promoting the implementation of the affected ones, are indicated by green colour. E.g. creating a Code of Conduct (CoC) quickly could have a positive effect on carrying out a lot of other principles. Red

colour reflects “enabling relations”, where principles listed in the rows eliminate obstacles to implement other ones listed in the columns. For example, trying to aim for win-win for all stakeholders might result in slowing down the process of a lot of other principles to take effect. Therefore, not all the “accelerating relations” are low hanging fruits. For instance, it might need a long time to develop a CoC because stakeholders do not aim for win-win.

It is quite interesting to observe that, while all principles are quite important, still the relation between them is quite dynamic. This relationship matrix in Figure 8 below offers the opportunity to try to establish an order of the principles realisation. WP5 has identified a possible scenario illustrated in Figure 9 below. Of course, alternative scenario(s) are recommended to be designed to assess possible implementation versions at national and cross-border levels.

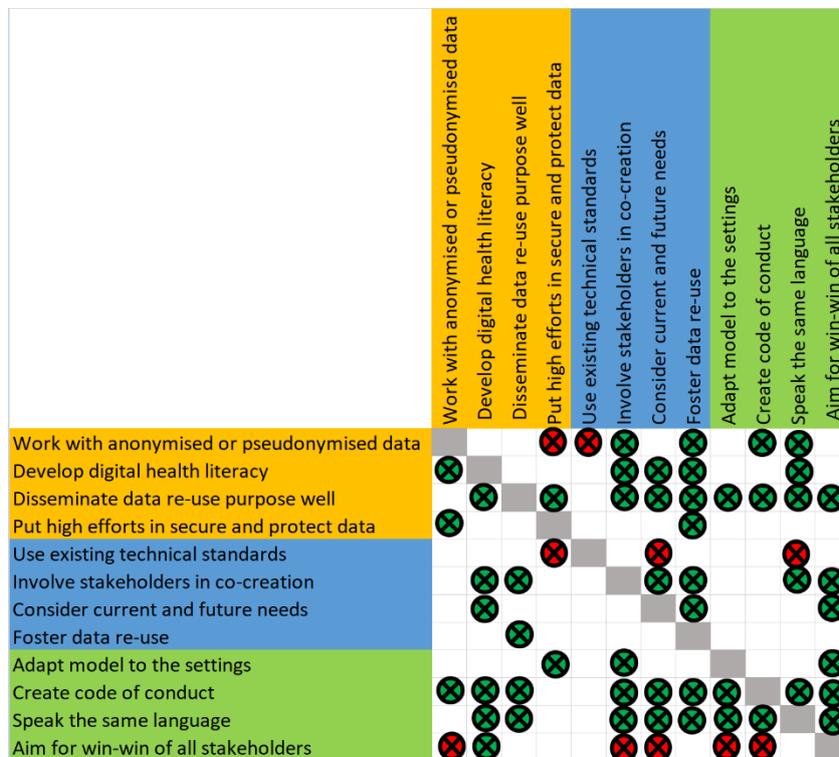


Figure 8 - Accelerating and enabling relations between principles

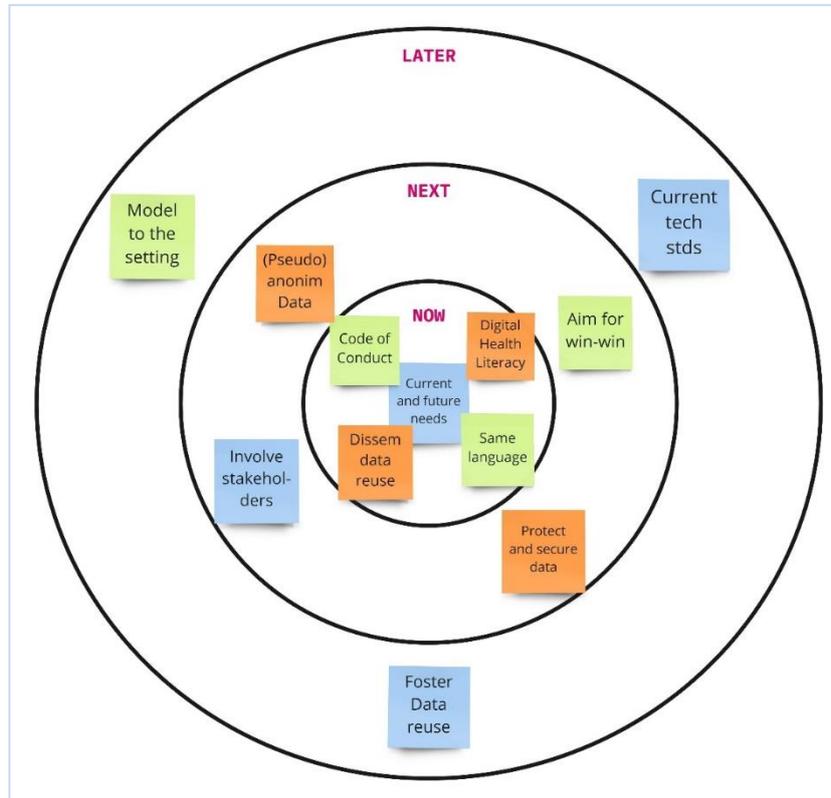


Figure 9 - Action timing

Cluster 1: Increase trust in privacy protection and cyber security

The principles presented in Cluster 1 target the 'lack of trust' that was introduced in D5.1. These principles intend to increase trust in privacy protection and cyber security.

Principle 1 – Work with anonymised or pseudonymised data



Work preferably with anonymised or pseudonymised data for the purpose of secondary use of data.

Cross-sectoral and cross-country data is needed for research purposes (a typical example is cooperation within European Reference Networks). Better protection has to be ensured for individuals in terms of data processed about them and less risks for individuals in relation to their right to data protection and privacy when re-use of their personal data occurs, in compliance with Art. 9 and Art. 89 of GDPR. It is recommended to preferably work with anonymised or pseudonymised data for the purpose of secondary use of data¹². The decision to use either anonymised or pseudonymised data needs to be analysed for each specific use case separately and depends on the particular national situation with the legal framework.

In the European Parliament resolution¹³ of 14th March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement the, the Parliament under point 10: Underlines that the intrinsic purpose of big data should be to achieve comparable correlations with as few personal data as possible; [and] stresses, in this regard, that science, business and public communities should focus on research and innovation in the area of anonymisation“. Under point 11 in the same resolution, the European Parliament “Recognises that the application of pseudonymisation, anonymisation or encryption to personal data can reduce the risks to the data subjects concerned when personal data are used in big data applications; further highlights the advantages of pseudonymisation provided for by the GDPR as an appropriate safeguard; recalls that anonymisation is an irreversible process by which personal data can no longer be used alone to identify or single out a natural person; takes the view that contractual obligations should ensure that anonymised data will not be re-identified using additional correlations by combining different data sources; calls on the private and the public sector and other actors involved in the analysis of big data to regularly review such risks in the light of new technologies and to document the appropriateness of the measures adopted; calls on the Commission, the European Data Protection Board and other independent supervisory authorities to prepare guidelines on how to properly anonymise data in order to avoid future abuses of these measures and to monitor practices”.

¹² See our working definitions section below for a definition of pseudonymisation and the difference to anonymisation

¹³ European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)) <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1596809628647&uri=CELEX:52017IP0076>

While data anonymisation provides the highest degree of data protection for the data subjects, anonymous data will not be useful in many big data applications. Hence the relevance of pseudonymisation as a strategy towards data minimisation and data protection in the regulatory context of GDPR. In its European Strategy for Data, the European Commission commits to developing standards, tools and best practices on pseudonymisation.¹⁴

With regard to Principle 1, T5.3 also recommends to check with the GDPR survey follow up on current discussions on how the reuse of data is currently organised in the different Member States. The study conducted in 2020 by the EUHealthSupport Consortium has the objective to examine EU Member States' rules governing the processing of health data in light of the GDPR, thus highlighting possible differences in legal interpretation and identifying elements that might also affect the cross-border exchange of health data in the EU.¹⁵

Principle 2 – Develop digital health literacy

	<p><i>Make efforts to improve digital health literacy in the population and healthcare providers.</i></p>
---	---

In the digital era, it is expected that being helped by technology, patients will eventually become their physicians' partners in medical decision-making.¹⁶ Considering that decision-making in public health, research and improving quality of healthcare provision and systems is expected to be more assisted by artificial intelligence, and the role of digital technology is continuously and rapidly growing, it is obvious that stakeholders in generating, capturing, sharing, accessing processing and using data have to be partners in developing and using digital health technology and services. Therefore, it is inevitable that digital health literacy (DHL) of all stakeholders, especially health professionals, shall be improved.

DHL helps people to assimilate the information they obtain regarding new technology and protocols, and it enables them to put up questions, reply them and understand the answers. In this way, DHL is a key tool to increase trust.

To improve DHL, including knowledge on data, primary and secondary use of data, data sharing and accessing, privacy and security issues, Big Data and AI, in both the general population and among healthcare providers, WP5 proposes the following actions:

1. Curricula of healthcare providers (including health professionals and healthcare managers) and other stakeholders should be adapted to include eSkills/eAbility and big data in all EU countries to develop digital health literacy of stakeholders to strengthen empowerment and adherence to data sharing and reuse.

¹⁴ COM(2020) 66 final

¹⁵ Consultation Document for Workshop on 19 May 2020. Experts' Workshops Assessment of the Member States' rules on health data in the light of GDPR.

¹⁶ Mesko, B., Drobni, Z., Benyei, E., Gergely, B., Gyorffy, Z.: Digital health is a cultural transformation of traditional healthcare. *Mhealth* 3, 38 (2017). <https://doi.org/10.21037/mhealth.2017.08.07>

2. DHL must be monitored and measured by evidence-based tools adopted to EU Member States and European countries to update or further develop curricula and design interventions appropriately, measure their impact, and take care that no one is left behind.¹⁷
3. Curricula of teachers of healthcare providers, as well as the capacities to educate them, should meet the increased needs for developing digital health literacy of stakeholders.
4. All data protection statements need to be written in an easy-to-understand way! - To achieve this goal population and healthcare providers should receive a clear view about the definition and features of 'data'. Differences between national rules for interpretations of GDPR can be treated more easily if stakeholders understand that "the idea of owning data is challenging because data is not like other goods that we can own. It is non-rivalrous – I can both give it to you and still have it myself without it costing me any of the original good".¹⁸ Exploring these differences helps to better understand and learn the common EU rules laid down in GDPR itself. Therefore, key legal and ethical features of data, preparing data protection statements, asking for consent or data access (as well as providing it) shall be taught.¹⁹
5. Start teaching at 14 years! Data generated by citizens is becoming more and more important. Therefore, improving basic DHL of young people is essential to pave the way for higher level and/or specialised education and training.
6. Use tools decreasing resistance and demotivating factors to learn (especially gamification, loyalty programs, social interaction, public competitions etc.)! Such tools can increase efficiency of education and training through effective motivation.

Two interesting good practices:

1. Embedding the teaching of digital communication skills and encouraging the practice of co-production and self-management with patients, are core learning outcomes for pharmacy students at Liverpool John Moores University (LJMU). The Patients Know Best platform, which became the first personal health record to be fully integrated into the NHS App, has been incorporated into the curriculum to facilitate simulated interactions between patients and pharmacists.²⁰
2. The Australian Library and Information Association (ALIA), the ALIA Australian Public Library Alliance (APLA), and ALIA Health Libraries Australia (HLA) have partnered with the Australian Digital Health Agency to deliver a train the trainer program from January to October 2020 to support library staff in responding to queries about government digital health programs. This nationwide initiative will provide training to library staff on

¹⁷ Zrubka, Z., Hajdu, O., Rencz, F. *et al.* Psychometric properties of the Hungarian version of the eHealth Literacy Scale. *Eur J Health Econ* 20, 57–69 (2019). <https://doi.org/10.1007/s10198-019-01062-1>

¹⁸ <https://royalsociety.org/~media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf>

¹⁹ Zrubka, Z., Hajdu, O., Rencz, F. *et al.* Psychometric properties of the Hungarian version of the eHealth Literacy Scale. *Eur J Health Econ* 20, 57–69 (2019). <https://doi.org/10.1007/s10198-019-01062-1>

²⁰ <https://www.digitalhealth.net/2020/08/patients-know-best-education-programme/>

how users can navigate and understand the new Federal Government digital health initiatives, such as My Health Record.²¹

Principle 3 – Disseminate data reuse purpose well



Communicate in a socially acceptable and timely manner the purpose of reusing data.

Trust in use and reuse of data can be easily destroyed or strengthened by the way of communicating purposes and results of data controllers. However, it may be difficult to articulate, identify or predict the role and importance of a specific piece of information or even a data subject (a concrete person), thus the purpose of asking consent for controlling one's data can face difficulties. Therefore, to make it obvious that the data subject has consented to the particular processing, data controllers have to develop or use proper methods to disseminate and communicate their general and specific scientific, social or business goals. The more difficult is to characterise and specify a purpose in advance, the more important is to find and disseminate socially acceptable purposes and make it clear that these purposes are serving 'public good'. WP5 accepts the economic definition of this term: "Public good, in economics, a product or service that is non-excludable and nondepletable (or "non-rivalrous")."²² In this context, data as 'public good' can be: "Using data for public benefits by applying tools and techniques that are generally used in business applications".²³ In addition we also have to take into consideration that a 'public good' has two characteristics: 'non-rivalry' and 'non-excludability'. This means that it must be checked if the amount available for others could be reduced when data is consumed, and/or it can be possible to provide data without it being possible for others to enjoy.²⁴

In order to be able to develop successful dissemination activities to communicate socially acceptable purposes providing public benefits, data controllers, as well as policy makers who intend to improve innovative use of health data and big data, are recommended to identify:

- who the key stakeholders, having influence over the decision on using results of data processing, are. Either because they're paying for it, or because they can persuade whoever is paying;
- who can end up experiencing the product or service developed or improved by using or re-using data;
- whose life would improve because of the product or service developed or improved by using or re-using data.

²¹ <https://www.alia.org.au/which-sector-are-you-interested/digital-health-literacy>

²² <https://www.britannica.com/topic/public-good-economics>

²³ <https://towardsdatascience.com/data-for-public-good-1414cbc99335>

²⁴ <https://www.economicshelp.org/micro-economic-essays/marketfailure/public-goods/>

These three groups of stakeholders are the customers, end-users and beneficiaries for non-for-profit organisations²⁵. Creation of public good, e.g. using data for public benefits, may not be paid directly but financed (e.g. from taxes or donations), therefore, the customer is not a payer but a financier who also decides or influences decisions on the use of money.

Communication of the public good to the three groups requires development and use of different strategies, ways and content.

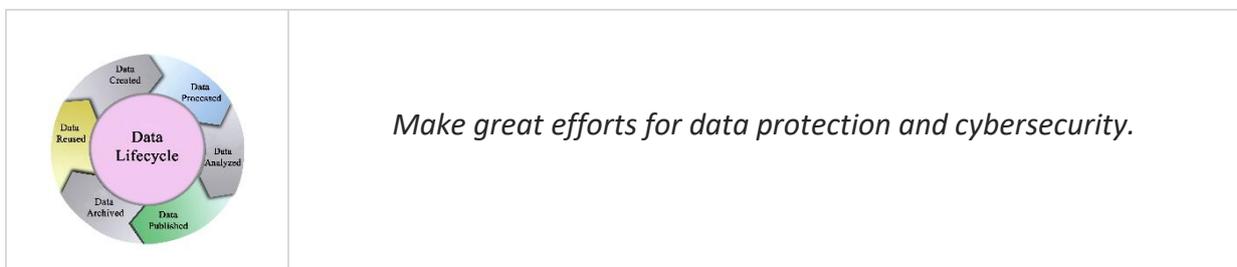
Strategies can partly build on data altruism in order to facilitate decisions on which data generated and held by data subjects can be used, how and by whom can these data be used, and what the public good purposes are.

Strategies can also build on the identification and introduction of the value that the public good offers to the data subject personally, therefore personalisation of the public good is also important.

In order to increase the level of trust, it also has to be ensured that the way how data will be processed is transparently communicated. This applies for safety, security and privacy solutions, as well as the use of interoperability standards and the efficiency issues. Strategies shall be implemented in a manner compliant with the GDPR and FAIR data principles.

Communication strategies and implementation plans shall be elaborated in line with developing, measuring and updating digital health literacy of these groups, as well as the number and the level of acceptance of codes of conduct in different domains and use cases.

Principle 4 – Make a great effort to secure and protect data



Do everything possible to reduce the risk and avoid attacks and misuse or abuse.

This principle builds on the so called ‘comprehensive data governance model’ where system/organisation-wide data governance aggregates historic and the newest data together. It requires full engagement and adherence at leadership level, and clear management of data control and access rights.

Data security and data protection is a multidimensional model. It is a functional space where it is necessary to ensure the physical protection, administrative protection, technical protection of the data, as well as compliance with legal regulations.

There are general rules to:

²⁵ <https://isaacjefries.com/blog/2016/3/1/customers-end-users-and-beneficiaries>

- ensure the confidentiality, integrity, and availability of all data they create, receive, maintain or transmit;
- identify and protect against reasonably-anticipated threats to the security or integrity of the information;
- protect against reasonably-anticipated impermissible uses or disclosures;
- ensure compliance by their workforce.

Responsible data governance approaches:

- Integrity – All entities involved in data governance must have integrity in their communication with each other.
- Transparency – Data governance processes need to be transparent in order to make it clear when and how data decisions and controls are implemented in the processes.
- Auditability – Data decisions and processes must be auditable, with documentation of compliance-based and operational control requirements
- Accountability – One must define who is accountable for data-related decisions and controls, and for data management activities.
- Breaks and counterweights – One must have a data-governance-based data lifecycle, and one needs to define data source and data cryptography protection, and who makes data, collects data, uses data and manages data.
- Standardisation – Data governance supports quantum resistance standardisation of data.
- Change management – Data governance supports proactive and reactive change management activities.
- Risk management – Data governance requires security rules, what covers administrative, technical and physical dimensions, as well as the data owner and data controller's activities. Artificial Intelligence should be mentioned as a risk factor, because we do not know its limits and effects. It can be managed through understanding the 'transparency paradox'²⁶ (transparency makes methods vulnerable while it can help mitigate issues of fairness, trust and discrimination).
- Resilience solution – One needs data governance to be able to manage disaster and force majeure situations.
- Confidentiality – Data governance must ensure data security, to serve privacy, protect individuals' medical records and other personal health information.

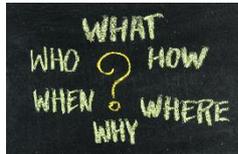
Recommendations to implement the principle:

- "Responsible data governance" approach
- Data ownership models and tailored solutions
- Different types of data should be handled differently, with regard to their lifecycle and purpose of use (personalised, anonymised or pseudonymised form).

²⁶ Andrew Burt: The AI Transparency Paradox, published on hbr.org, December 13, 2019

Cluster 2: Increase efficiency of using funds and financial resources

Principle 5 – Use existing technical standards



Choose appropriate existing technical standards instead of inventing new ones from scratch.

The intention behind this principle is to foster the use and harmonisation of existing technical standards and to avoid duplication of efforts by inventing new standards when it is not justified. Interoperability is key when it comes to big data and secondary use of health data. In particular, relying on the definitions of the 'Refined eHealth European Interoperability Framework', we refer to the interoperability level of *information*, i.e. the data models, their functional description and links to terminologies.

The reuse of existing technical standards and exploitation of means already available like terminologies and data models allows an efficient use of resources and contributes to further increase interoperability and the efficient use of resources. Examples include:

- Consider whether existing standard data models are useful for your big data application, e.g. if you are aiming to combine administrative data with data from health records (e.g. OMOP Common Data Model²⁷)
- Consider whether your data can be coded according to widely used terminologies that allow linking to other healthcare data sources (e.g. SNOMED CT)
- Consider standards in describing your metadata to comply with FAIR principles
- Consider making use of/contributing to existing data repositories (e.g. in the context of the European Health Data Space and the European Open Science Cloud)
- Support efforts to provide health-related data in a machine-readable format

Before entering into the implementation phase, an intensive research and design phase is recommended to investigate existing technical standards available for the specific big data use case that is determined for the secondary use of health data. Countries shall also aim to have health-related data in a machine-readable format.

It is also important to be aware of and communicate the standardised semantic meaning of the clinical data when exchanging electronic health data. In order to achieve semantic interoperability of clinical data, it is important not only to specify clinical entries and documents and the structure of data in electronic health records but also to use clinical terminology to describe clinical data. This is essential not only in the use cases of direct patient care, but for research, reimbursement, quality assurance, epidemiology, public health or policy and governance development as well. There are three types of clinical terminology: 1) interface

²⁷ <https://www.ohdsi.org/data-standardization/the-common-data-model/>

terminology to support a user-friendly structured data entry; 2) reference terminology to store, retrieve, and analyse clinical data; and 3) classification to aggregate clinical data for secondary use. In order to use electronic health records data in an efficient way, healthcare providers first need to record clinical content using a systematic and controlled interface terminology, then clinical content needs to be stored with reference terminology in a clinical data repository or data warehouse, and finally, the clinical content can be converted into a classification for reimbursement and statistical reporting. It is necessary to map reference terminology with interface terminology and classification. It is necessary to adopt clinical terminology in electronic health records systems to ensure a high level of semantic interoperability.

Principle 6 – Involve stakeholders in co-creation



Identify and involve stakeholders in the secondary use of health data and big data and creating value.

Efficiency of using funds and financial resources for the innovative use of health data and big data in health is closely related to who is taking part in this process and how.

On the one hand, acceptance of an innovation depends on how key stakeholders evaluate it. Therefore, focusing on the needs and expectations of only one stakeholder is likely to result in failure and waste of time and money.

On the other hand, innovative use of health data and big data in health means various data sources that must be reliable, safe, findable, accessible, etc. This requires co-operation with varying data subjects and controllers.

Thus, identification of key stakeholders, as well as their needs and expectations, is essential. WP5, using the Data Conversion Framework tool, identified the following key groups of stakeholders for innovators and researchers: 1) patients and citizens; 2) care providers, including professionals, organisations and authorities; 3) payers. In this context, innovators and researchers are those data controllers who use or reuse health data in order to create something new or to renew or develop existing things, services, solutions, protocols or systems.

Asking them about their needs is the first step, however, success needs continuous co-operation along the whole process of translating an idea or invention into goods or services that create value, or for which customers will pay. Innovators and researchers shall assess how stakeholders feel about key features of existing and innovated or invented healthcare services and protocols or solutions, devices or infrastructure, etc., and shall ask which of the following three categories each feature falls into ²⁸:

- **Non-negotiables:** Performance features that make an offer minimally acceptable.

²⁸ How To Get Ecosystem Buy-In - A tool kit for assessing the way an innovation will affect each stakeholder, by Martin Ihrig and Ian C. MacMillan, Harvard Business Review, March-April 2017

- **Differentiators:** Features that positively distinguish an offering from the competition.
- **Dissatisfiers:** Stakeholders are not happy about these attributes but may be willing to endure them for a time if compensating differentiators exist.

Innovators and researchers are recommended to involve their stakeholders in looking for a differentiator to add at each link in the consumption chain (from suppliers or resource providers to the end-users) and try to find at least one differentiator for each of the stakeholders. Removing a dissatisfier, which effectively creates differentiation from competitors, is also an option. Once they've come up with new differentiators to improve how to satisfy a given need, they have to assess how they might affect other stakeholders — particularly whether they might conflict with a dominant non-negotiable. Naturally this is work in progress which can be updated based on the changes or findings in all stages of the innovation process.

Involvement of all the key stakeholders, in most cases, faces the challenge that people who need to be involved in the co-creation process or data governance are really busy. To be successful, a data governance initiative requires time and input from subject matter experts, with a deep understanding of the data or the purpose and opportunities of data reuse, and key managers who can make operational changes happen. They will be the innovation scouts²⁹ or data stewards.³⁰

Involvement of stakeholders in co-creation processes can be carried out anywhere, however, innovation hubs (places and tools for professional-user led co-creation) and living labs (centres for individual-user led innovation) offers more friendly environment for value creation.

Implementation of Principles 6 and 12 can be carried out together. While Principle 12 assists to eliminate misunderstandings around data sharing and access in a more effective way, Principle 6 can be a tool to use data to innovate something in a more efficient way.

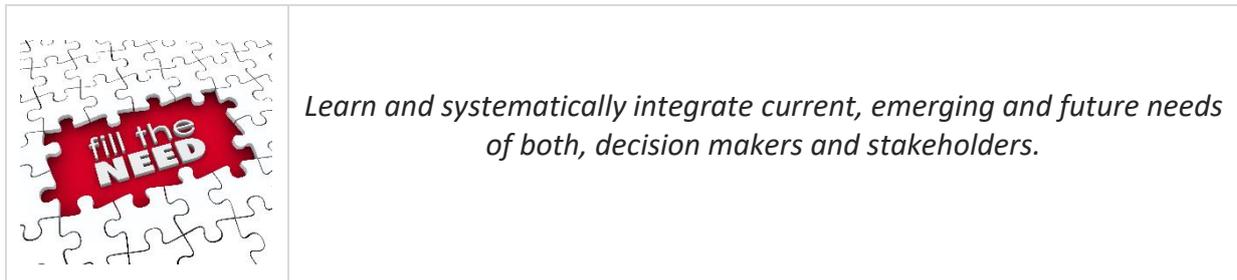
WP5 recommends the policy makers and implementing bodies of funding programmes in the Member States/countries and the European Union to assess which actions or measures could be used:

1. to encourage beneficiaries to use models for identification and involvement of stakeholders in the secondary use of health data and big data;
2. to develop innovation hubs and living labs to improve the digital health ecosystem. Supporting these forms can be implemented separately, but WP5 believes that using both of them in a complex program can deliver even more increased efficiency at policy level;
3. to develop and run innovation scout and data stewardship programmes at healthcare organisations.

²⁹ <https://www.innovationagencyncw.nhs.uk/media/Images/Innovation%20Scout%20Programme%20Overview.pdf>

³⁰ https://www.ehdc.org/sites/default/files/resources/files/Data-Gov-in-Healthcare_092016.pdf

Principle 7 – Consider current and future needs



A long-term vision of big data use in health sector requires that a continuous process is established to develop big data services as a new asset for health and health services (policy making, management and care provision) based on scalable and sustainable governance. Particularly, in the health sector, data governance mechanisms should be identified to use these data for the benefit of the society and to support different organisations in improving their decisions, while minimising societal risks, and be continuously reviewed to ensure adequacy.

There is a growing acknowledgement that governance is a key aspect for managing and getting value out of big data, to more effectively implement big data tools and to maximise the rewards for the organisational and institutional point of view. The framework for implementing data governance can have different components: policy and organisational, technological and standards related. In fact, due to the number of different health systems models and organisations, with their individual culture, numerous possibilities to implement data governance exists. However, any data governance framework to deal with big data in healthcare organisations, and in the health system as a whole, must ensure stakeholders have access to timely, consistent and accurate data resulting in quality data-driven decisions.

One of the main drivers for implementing data governance must be a proactive, rather than a reactive, approach to deal with ever-growing needs of data and the dynamic nature of healthcare, public health and research needs. As such, having an organisational structure within the healthcare system that aligns all stakeholders on value (including matters experts, managers, health professionals, IT and data specialists and citizens), through an efficient leadership and communication strategy, is a requirement for an organisation-wide approach and dedicated organisational initiatives to optimise data assets to deliver both sophisticated and practical information (NOESIS³¹, Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services).

Engaging stakeholders in a data-driven approach to health requires a trusted environment in the healthcare system to perform its functions effectively, based on data insights; maintaining a trusted reputation amongst its many data providers, data recipients and stakeholders engagement is also a critical requirement to achieve a continuous governance process that is sensitive to the ever-growing needs of data for decision making in different settings and for different purposes. Data needs of citizens, health professionals, healthcare institutions and

³¹ <https://cordis.europa.eu/project/id/769980>

central agencies must be accounted for since the beginning through a customer-centred approach to secondary use of data.

Communication, in a clear and proactive manner, and information to health professionals about the data and limitations, can enable realistic expectation setting about the way the requirements are implemented. Communication channels should also be ensured with the public and/or public oversight groups, sharing best practices and discussing ethical concerns. As it cannot be expected that everyone is able to fully understand the workings and effects of (complex) data-based systems, consideration should be given to organisations that can attest to the broader public that a system is transparent, accountable and fair. Involvement of stakeholders throughout the data life cycle also needs training and education so that all stakeholders are aware and have the appropriate skills to integrate data-driven technologies into daily operations³².

As such, a common principle for practical implementation of (big) data governance is to consider the dynamic nature of the healthcare system, which can be only be achieved by ensuring stakeholders play key leadership roles in the development of governance frameworks and are empowered to support innovative data-driven technologies when they arise.

Recommendations:

Establish a framework for data governance that ensures continuous engagement of key stakeholders in needs assessment, monitoring and review process of data assets.

Principle 8 – Foster data reuse



Find means to make existing data sources visible and preferably use open source data models to allow for reuse.

It is also important to include data to support the decision-making process.

One must address the right decision makers, users and beneficiaries of the business model for open data and data sharing to foster data reuse.

The main purposes of reuse of health and healthcare data are to speed up medical innovation and improve citizens' health. Reusing data can provide important insights into results of innovation in technology, development of new tools and also new skills.

There are several aspects of reuse of health and healthcare data:

- focusing on strictly health and health related outcomes;
- economic aspects;

³² European Commission High-Level Expert Group on Artificial Intelligence. Ethics guidelines for Trustworthy AI [Internet]. 2019. Available from: <https://ec.europa.eu/futurium/en/ai-alliance-consultation>

- social impacts; and
- the mix of all above aspects.

These considerations top the notion of sole access to information by citizens.

To improve the reuse of health and healthcare data, WP5 proposes the following actions (over and above the actions of primary use):

Reuse of health data should be incentivised by public entities on the regional and EU levels (most relevant fields: continuity of care delivery, personalised medicine, clinical research, pharmacovigilance, quality improvement, infectious disease, bio-surveillance, financial analysis). Outcome-oriented incentives are preferable which should contribute to results (above direct healthcare-related benefits) in the field of science and informatics. Those can improve patient-clinician partnership and cross-sectoral collaboration (healthcare, information technology, manufacturing etc.). It should be a dynamic process with a continuous learning curve.

When setting up a framework for reusing health data, key – and relevant – stakeholders should be included and brought together in the process (i.e. public authorities, healthcare policymakers, professionals, pharmaceutical companies, research organisations, patient organisations, enterprises, etc.).

Reusing health data should facilitate actions in setting up research, quality and public health strategies. Benefits of reuse of data should be estimated if possible before taking action.

Business models can be constructed taking into consideration health data governance frameworks. Charges for the reuse of such data can be considered depending on the purpose of reuse (i.e. marginal costs for usage and dissemination). Externalities should be internalised in models if reasonable and possible.

Application of the Open Data Directive³³ should be considered in the field of health data. The Open Data Directive requires the adoption by the Commission of a list of high-value datasets to be provided free of charge. These datasets, have a high commercial potential and can speed up the emergence of value-added EU-wide information products. They will also serve as key data sources for the development of Artificial Intelligence. Reuse of these datasets is associated with important benefits for the society and economy. They are subject to a separate set of rules ensuring their availability free of charge, in machine readable formats, provided via Application Programming Interfaces (APIs) and, where relevant, as bulk download.

Data quality should be continuously improved to get more usable health data for reuse. Therefore, WP5 recommends to assess the opportunity of taking over the data quality model

³³ Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

provided by the ISO/IEC 25012 standard^{34,35} which links data quality to the degree to which a set of data characteristics meets the requirements. It helps not only to disseminate data reuse purposes (recommended in Principle 3), but can be a proper tool to increase efficiency of using funds and financial resources for innovative use of health data and big data as well.

Data and big data quality is an essential topic for public health and research, healthcare management and governance and ‘businesses, providing accurate information in order to make correct decisions accordingly.’³⁶

It is also elementary and an important prerequisite for reuse to build a trusted digital health data sharing environment. Trustworthy and transparent use of health data is essential.

Security, legal and ethical aspects could be different from those applied at primary use due to:

- different stakeholders,
- different purposes and fields of application,
- different processes, and
- different space–time structures^{37,38,39}.

It should be considered to launch a ‘time-and-space-changing data space’ for European healthcare, which would be able to satisfy artificial intelligence (AI) requirements to show changes and deviations in the space–time structures and would be able to predict future trends and assist decision-making related to varying purposes of different stakeholders. It is, however, especially important to protect and manage sensitive personal health data in accordance with the GDPR.

Cluster 3: Foster common interpretation of legal and ethical rules

Principle 9 – Adapt model to the settings



Treat health, health-related and non-health data with a special use case / business model for data management & data governance.

Shall health challenges in different environments be treated with special strategies and models for data reuse management or data governance? If you understand where people live, you can develop and implement data reuse governance and financing models and communication

³⁴ <https://www.iso.org/standard/35736.html>

³⁵ <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>

³⁶ <https://www.sciencedirect.com/science/article/pii/S1877050919305915>

³⁷ <https://link.springer.com/article/10.1186/s40537-018-0116-9>

³⁸ Space-time Data Science for a Speedy World, Harvey J. Miller, Published 2015

³⁹ <https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis>

strategies which are adopted to the settings of people, and in this way you can develop schemes to foster common interpretation of legal and ethical rules that help you raise the level of trust and improve resource management as well.

Urban environments, for example, play an essential role in shaping human health and well-being, as 55% of the population worldwide live in urban areas. 97% of cities in low-and middle-income countries with more than 100,000 inhabitants do not meet air quality guidelines (49% in high-income countries).⁴⁰

Rural areas often have the highest rates of premature death due to tobacco, nutrition and alcohol issues. Almost half the world's people live in rural and remote areas. The problem is that most health workers live and work in cities⁴¹.

Despite the significant advances that have been made at the EU level in terms of health inequality measurement, there is still a lack of comparable health-related data across countries and regions. Existing measures and indicators of health status and quality of life are considered inadequate to capture the holistic understanding of population health, with multiple determinants involved.⁴²

Therefore, following the logic and findings of the PULSE project⁴³, shifting public health from a surveillance-based system to an inclusive and collaborative system via citizen engagement through a citizen science approach using digital technology is strongly recommended. This requires creation of environment specific health and wellbeing datasets.

Data generated and held (1) in health systems, (2) in research activities and (3) by citizens shall be federated for secondary use purposes and, together with (4) cross-sector data about determinants of health and geographic features, shall be shared and could be accessed.

New investments shall be made in combined information sources deriving greater context. 'This is needed to more effectively and securely leverage patient data, recognise patterns faster, and manage the health of populations more effectively. (...) Population health management is fundamental to the transformation of healthcare delivery. For providers, the term translates to knowing what's going on with your patients and taking evidence-based, standardized and, where appropriate, automated actions to proactively achieve the best outcomes.'⁴⁴

This requires use-case-specific decisions on following issues in policy and implementation levels:

- which data generated and held by the four sources above can be used;

⁴⁰ María Fernanda Cabrera Umpiérrez, PULSE Coordinator, UPM, <http://www.project-pulse.eu/events-2/pulse-final-workshop/>

⁴¹ WHO Global policy recommendations on Increasing access to health workers in remote and rural areas through improved retention (2010)
https://apps.who.int/iris/bitstream/handle/10665/44369/9789241564014_eng.pdf;jsessionid=928CD4C30FC14287A105522A5B93DED7?sequence=1

⁴² https://www.researchgate.net/publication/321361767_Atlas_of_Population_Health_in_European_Union_Regions

⁴³ See footnote 20

⁴⁴ Femi Ladega and Gurdip Singh, Digital Revolution Enables Population Health Management
https://www.himssasiapac.org/sites/default/files/HIMSSAP_Whitepapers_CSCDigitalRevolutionEnablesPopulationHealthManagement.pdf

- who can use the data and for what purposes;
- what the common interpretations of different legal and ethical rules can be;
- how to communicate the value of the use of artificial intelligence and machine learning solutions that can easily remove limits to scale, scope, and learning⁴⁵;
- how to transform public health from a reactive to a predictive system focused on both risk and resilience by the digital population health model.

The most concerned use cases are '1. Forecasting patient demand, attitude, behaviour and need' and '3. Public health, surveillance and prediction of health threats and diseases', as well as '10. Hidden citizen behaviour patterns analysis', '11. Insight and evidence for policy-making and investments' or '12. Prediction of health and health related needs, expenditure and costs'.

Principle 9, being part of Cluster 3, does not concentrate on technical interoperability but on how to develop proper models and use terms of co-operation and codes of conduct to predict, mitigate and manage public health problems, and promote community health in different environments where people live.

Principle 10 – Create code of conduct



Create international terms of data ownership, sharing, exchange, access and control to mitigate risks regarding sensitivity and privacy.

The European Parliament, in its resolution adopted in 2017, urged 'the private and public sectors and other data controllers to make use of instruments provided for by the GDPR, such as codes of conduct and certification schemes, in order to seek greater certainty over their specific obligations under Union law and to bring their practices and activities into compliance with the appropriate EU legal standards and safeguards.'⁴⁶

The legal basis for the adoption of codes of conduct is provided by GDPR. According to Article 40, associations and other bodies representing categories of controllers or processors may prepare codes of conduct for the purpose of specifying the application of GDPR with regard to various issues including collection and pseudonymisation of personal data, fair and transparent processing, the exercise of the rights of data subjects, as well the information provided to them

⁴⁵ Competing in the Age of AI - How machine intelligence changes the rules of business by Marco Iansiti and Karim R. Lakhani: <https://hbr.org/2020/01/competing-in-the-age-of-ai>

⁴⁶ European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI))

and the public. Drawing up codes of conduct shall be encouraged by the EU and national authorities, under Article 40.

Guideline No. 1/2019 of the European Data Protection Board (EDPB) on Codes of Conduct adds that codes of conduct can help to bridge harmonisation gaps between Member States in their application of data protection law. Codes of conduct also provide an opportunity for specific sectors to reflect upon common data processing activities and to agree to bespoke and practical data protection rules, which will meet the needs of the sector as well as the requirements of the GDPR.

In the health sector, a code of conduct or a network of codes of conduct can foster common interpretation of the legislation applicable to personal data, including secondary use of data for the purposes of public health, health research, and quality assurance in healthcare.

From a technical perspective, the approach can either be the adoption of a code of conduct for the health sector, or rather a set of corresponding codes of conduct.

The scope of a possible code of conduct or codes of conduct could cover:

- key terms and definitions including data controller, consent, anonymisation and pseudonymisation;
- the legal basis for data collection and processing;
- ownership of data (e.g. differences between legal provisions regarding ownership of things in general and the rights to control somebody's data or intellectual property);
- specific domains or use-cases;
- interoperability issues and technical standards.

A code of conduct cannot solve all issues related to legislation but should be part of a complex toolbox, and address different challenges. It should also have a link to the vocabulary described under Principle 11.

Associations and other bodies representing categories of controllers or processors can be fostered and assisted by EU and national funds through calls for projects to elaborate, disseminate and educate or train in the use of their code of conduct. Financial sustainability as well as updates of the content of recommended specifications and maintenance of educational and training infrastructures shall be central issue and requirement of granting. Experiences of development and usage of the International Commercial Terms (Incoterms⁴⁷) can be assessed.

⁴⁷ Incoterms is a series of pre-defined commercial terms published by the International Chamber of Commerce (ICC) relating to international commercial law. For more information see: <https://iccwbo.org/resources-for-business/incoterms-rules/incoterms-2020/>

Principle 11 – Speak the same language



*Share common vocabulary and meta-data to avoid misunderstandings
and lead to exploit the value of data.*

People who need to collaborate on data governance speak different languages. Years of traditional information silos in healthcare have resulted in language barriers. Finance speaks one language, clinicians another, supply chain another, and IT yet another. Terms and definitions are provided for the by the legislation but there are gaps in their interpretation and application in practice. Breaking down these language barriers takes time and culture change. Because of these silos, people don't always understand the other sides. Each department faces unique challenges and constraints to make the cultural and operational changes needed to develop a more data-driven culture and to support data governance.

The following actions should be taken:

To create a strong foundation for practical implementation of data governance across the healthcare sector, there is a need for people from different organisations and departments to collaborate. These people could then act as 'translators' to facilitate conversations and break silos, helping to tackle data problems together. A goal of the data governance program should be to develop more of these 'multilingual people'⁴⁸.

Data quality initiatives (in which data-related vocabulary and standards are included) must emerge from this broad integration and collaboration strategy. Having data stewards in every organisation and/or department of the healthcare system to spend time on data governance work: proactive data quality management, definition/adaptation, review and enforcement of data standards, and updating of role-based access policies, as requirements (legal, social and technological) inevitably change.

A shared data-related vocabulary (information, technological and data standards) is a key achievement to render operational any data governance model. In fact, one of the main challenges healthcare systems are facing is data quality due to the complexity of the clinical-systems' data structure, massive growth in clinical data volume and the lack of standardisation between the clinical systems in terms of naming and modelling⁴⁹.

⁴⁸ Paula J Edwards; Why is Data Governance in Healthcare so Difficult? Common challenges and ideas for overcoming them; Himinformatics. https://www.ehdc.org/sites/default/files/resources/files/Data-Gov-in-Healthcare_092016.pdf

⁴⁹ Alofaysan S., Alhaqbani B., Alseghayyir R. and Omar M.. The Significance of Data Governance in Healthcare - A Case Study in a Tertiary Care Hospital. DOI: 10.5220/0004738101780187 In Proceedings of the International Conference on Health Informatics (HEALTHINF-2014), pages 178-187

A common data-related vocabulary is therefore required to ensure that data management activities are in line with the overall objectives of the healthcare system, are understood by stakeholders, and reflect and preserve the value to society from the sharing and analysing datasets as a collective resource. It is an operational requirement for practical implementation of data governance, as data quality refines the basis of how data are interpreted (metadata) as well as accessed (data access) by users.

The vocabulary is also related to the code of conduct described under Principle 10, mainly its part clarifying key terms and definitions.

Principle 12 – Aim for win-win of all stakeholders



Measure, understand and communicate why stakeholders of data sharing and access are important to each other, as well as their value for them.

With progress comes new challenges, such as maintaining patient privacy in a data-driven healthcare world and using data to predict rather than just diagnose and treat patients' diseases.⁵⁰ Legal uncertainties related to privacy issues can be mitigated by codes of conduct and other tools. However, developing and using them can be made more effective if stakeholders could find win-win solutions to their problems.

Don't pigeonhole your data scientists.⁵¹ Data scientists can be in the centre of the co-creation processes with their colleagues and stakeholders (such as patients, care providers and payers), where they can scale up a proof of concept.

This co-operation allows all the participants of the co-creation team to learn about the needs and expectations of the others by identifying key features and attributes (e.g. non-negotiables, differentiators and dissatisfiers) of data sharing, access and control.

The team can define indicators to measure how important (valuable) are:

- certain purposes (or benefits and impacts) for the participants (stakeholders) to share, assess and control data;
- the identified features and attributes;
- the level of compliance with the importance of these purposes (or benefits and impacts) or features and attributes.

⁵⁰ <https://hbr.org/sponsored/2017/11/three-keys-to-unlocking-data-driven-health-care>

⁵¹ <https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture>

Stakeholders should quantify not only the level of the state or intensity of changes, but uncertainty as well⁵². For example: Is the data reliable? How reliable it is? Is privacy or data safety and security guaranteed? How much are they guaranteed?

Implementation of Principle 12 can be carried out together with Principles 1, 4 and 6. While Principles 1 and 4 help to increase trust (mainly in data controlling and processing) and Principle 6 is a tool for using data to innovate something in a more efficient way, Principle 12 assists to eliminate regulatory, contractual or ethical misunderstandings around data sharing and access in a more effective way.

⁵² See footnote 32

Guidance for the implementation of common principles

As mentioned in the 'Introduction', the current draft proposal provides information about the background, objectives, planned structure, working definitions and methods of Deliverable 5.3, and contains recommendations on common principles for practical governance of big data.

The final milestone of work package 5, "Discussion Paper for the eHN on the implementation of common principles for practical governance of big data with a special focus on data to be used in public health, research and quality assurance in healthcare on a European scale with guidance on implementation of data access and use (Deliverable D5.3)", shall be reached by April 2021 by submitting the final report, containing a proposal for the guidance for implementation of the principles as well.

Approach and content

The guidance will use practical examples and/or similar solutions to illustrate how policy makers could implement and make use of the implementation of the 12 principles to eliminate root causes of the obstacles to utilise the potential of harnessing new opportunities arising from big data and improved data analytics capabilities (see Figure 10).



Figure 10 – Approach to prepare guidance

WP5 plans to organise its final stakeholder workshop, face-to-face or online, in the fourth quarter of 2020 to:

- select practical examples wholly or partly covering the principles,
- discuss stakeholders' expectations and questions regarding implementation, and identify concrete needs of Member States/countries and stakeholders about what they are looking for in the guidance,
- discover opportunities and required resources affecting the finalisation of the implementation scenario drafted in the current document, and

- prioritise common potential cross-border use cases of innovative use of health data and big data, including the issues of creating a European Health Data Space.

Next steps

WP5 intends to recommend 'next steps of national and common implementation' as well. The guidance will contain a suggestion for an endorsement of eHN on the way that the members will commit themselves to:

1. go into a national dialogue on how to use the final report (common principles + implementation guidance) with appropriate actors in the field,
2. use as input to future work in the field (e.g. joint actions or coordination and support actions),
3. convey principles to decision makers of local, regional, national and international funding programmes, and
4. consider the principles in their decision making in planning and implementing development programmes or integrating results in healthcare system or health economy.

Appendices:

- Main findings
- Working definitions

1. Main findings

Findings from D5.1 and D5.2

1. What are the main challenges regarding big data and artificial intelligence in health?

There are many policy and operational level recommendations in this topic. Most of them are still relevant in most Member States and organisations. The result of our mapping showed that three general obstacles appeared as reasons slowing down or hampering translation of policy-level recommendations into actions: lack of trust, legal uncertainties, and lack of funding and financial resources.

All these obstacles can be traced back to a general lack of data governance related priorities in health policy at Member State/country level. Many recognise that implementing effective data governance is critical to meet increasing demand for information to support value-based care and population health, however most countries are only at the beginning of a complex journey to encourage the development and safe use of health data. Health Ministry leadership is necessary to ensure that delivering the data to manage this important sector is at the forefront of government policy and action. Optimal decision making about potential statistical and research uses of personal health data can only be achieved if there is an overarching data governance framework in the country that has been aligned to minimise societal risks and to maximise societal benefits from data uses. Mechanisms of collaboration must be designed to support countries in developing data governance frameworks and engaging in regulatory and legislative reforms, including those necessary as the result of the EU Data Protection Regulation).⁵³

2. Lack of trust is one of the identified barriers in the D5.1 Draft report on policy-level actions on innovative use of big data in health. What can be done to overcome this?

Common efforts are needed to eliminate obstacles caused by lack of willingness, unavailability of comprehensive data or fear of abuse. The same applies to resistance to learn, lack of motivation for further education and lack of expertise or resources (not only financial, but trained personnel as well).

In addition, the unmet need about transparency in capturing, cleaning, storing, sharing or using data shall be satisfied, and a significant amount of data still kept somewhere in paper format, shall be traced and digitised. At the same time, key stakeholders have to understand the difference between analysis and reporting.

⁵³ OECD (2015), Health Data Governance: Privacy, Monitoring and Research, OECD Publishing, Paris

Last, but not least, stakeholders have to confront the ‘transparency paradox’. To do so, they need assistance in finding or updating solutions on how transparency could make AI methods less vulnerable, while they can help mitigate issues of fairness, trust and discrimination.

3. Within the scope of D5.2 - Report on identified cross-border use cases, WP5 conducted interviews to collect information from key stakeholders. How do you assess the achieved feedback?

The in-depth interviews enlarged the knowledge pool on innovative cross-border use of health data. Even though the projects selected for in-depth analysis had their foci on the intervention and implementation of innovative health technologies from very different disease areas (schizophrenia, rare diseases, MCC, etc.) most of them faced rather similar challenges when implementing their project.

In particular, issues of privacy protection, ethics, data security, health assessments, data quality, interoperability of health data systems, and demonstrating added value to the key actors (such as citizens, patients and professionals) were brought up.

The analysis of the conducted interview results shows that there already exist a plethora of experiences and project results that are accessible and available from different projects. The projects could be consulted to develop a knowledge base and a framework for continuous exchange of best practices at the EU level.

4. Considering the related challenges and opportunities, what should be the next steps to advance the innovative use of health data in Europe?

It is quite important to define priority areas and compose the optimal set and order of the most challenging and most promising use cases. Finding, implementing or developing use cases that provide optimal set of gains at different levels shall be fostered by launching dedicated regulatory and specific funding programmes to:

- reuse and further develop existing and technical capacities for exchange data
- foster empowerment, adherence and grit to increase level of knowledge, skills and competences related to sharing and accessing data for analysis and innovation purposes
- process good/best practices for optimised use cases
- develop and implement new curriculum and training programmes
- prepare and run communication, and dissemination and CSR programmes.
- accelerate product, procedure and service development and bring new innovations and advances to patients
- mitigate uncertainties about implications and unclear impacts on intellectual property issues, fairness, safety, security, trust, liability, algorithmic transparency, social inclusion
- offer clarity about the roles of varying agencies in the Member States/countries and the EU.

For example, the creation of a European Health Data Space (EHDS) may help to launch these dedicated regulatory and specific funding programmes, and in this way, it can foster innovative use of health data. It can be built not only on the network of national, regional, local or

organisational data spaces, but on the cooperation of key stakeholders throughout Europe as well. It is likely that EHDS could be an important, fundamental part of the digitalised healthcare ecosystem, therefore it requires EU level co-ordination.

2. Working definitions

Working definitions were elaborated by WP5 members at the kick-off meeting in order to lay down statements of the meaning of certain phrases which have been explained in various ways in the literature. The kick-off meeting of Work Package 5 was held on 16 October 2018 in Brussels. The meeting included a workshop to review available definitions and to propose for adoption those that are important to empower patients, policy makers and professionals about the innovative use of health data. Experts agreed on defining health data, big data in health, big data analytics in health and innovative use of health data. These working definitions were introduced to the eHealth Network in November 2018, and are considered as continuously evolving definitions based on the evolutionary and constantly changing nature of the field. Definitions reflect the value-based approach followed in WP5, where value refers to satisfaction of a specific need and replicability at an economical cost. The following definitions were used in deliverable D5.1, endorsed by the eHN in June 2020.

1. **'Data'** means information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.
2. **'Pseudonymisation'** means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person⁵⁴. In a GDPR context, pseudonymised data is considered personal data. GDPR does not apply to anonymised data, which is not considered personal data. (GDPR "Whereas: (26) The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner

⁵⁴ 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), Article 4 (5) Definition of Pseudonymisation. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.”)

3. **‘Anonymisation’** means the process of changing data into anonymous data which do not relate to an identified or identifiable natural person, or the process of rendering personal data anonymous in such a manner that the data subject is not or no longer identifiable.⁵⁵
4. **‘Machine-readable format’** means a file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure.
5. **‘Health data’**: Patient data in health records (records kept by health professionals and care providers, as well as self-reported health data), data from apps and wearables, any background data that will give insights on the social determinants of health.
6. **Primary and secondary use of health data**: In many cases the term ‘innovative use of health data’ is defined as secondary use of health data and big health data. However, it is worth considering whether there was a tangible border between primary and secondary use, or there are other data usable for primary and/or secondary use.

In our interpretation, primary use of health data is related to the care or treatment of a person (a data subject⁵⁶), while secondary use covers every other case related to any goals regarding policy making, system governance or planning, regulation, authorisation, control, monitoring, governance, management, research, innovation, development, etc. In this context, secondary use of health data is not only use (reuse) of health data again or more than once, but the use of the data of one data subject for the benefit for more or other data subjects other than the initial purpose.

The pellucid border between primary and secondary use is well reflected in the introductory description of the UK NHS Innovative Uses of Data Team: ‘Our Innovative Uses of Data (IUoD) team aims to improve our information analysis and reporting, by using novel data science techniques. This will enable new insights from data that work to improve health and social care. Products will be focused on the needs of patients, clinicians and organisations within the health and social care sector, to increase the likelihood of delivering real benefits that will improve patient outcomes.’⁵⁷

7. **‘Open Data’** as a concept is generally understood to denote data in an open format that can be freely used, reused and shared by anyone for any purpose. Open data policies which encourage the wide availability and reuse of public sector information for private or commercial purposes, with minimal or no legal, technical or financial constraints, and which promote the circulation of information not only for economic operators but

⁵⁵ Based on the definition of Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, Article 2 (7).

⁵⁶ One shall be very careful to use the notion of ‘owner’ or ‘ownership’ in relation to any personal data without deeper legal elaboration. There is an ongoing professional debate about the concept of ownership in this regard. In this context ownership refers to the origin and the beneficiary of data. Therefore, the working definition uses the term ‘data subject’ who can be an identified or identifiable natural person accordingly the GDPR.

primarily for the public, can play an important role in promoting social engagement, and kick-start and promote the development of new services based on novel ways to combine and make use of such information.⁵⁸

8. **'Health literacy'** is defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions".⁵⁹
9. **'eHealth literacy'**, according to the Lily model of Norman and Skinner, covers a broader concept, encompassing traditional literacy (basic ability to read and comprehend written text), information literacy (the ability to find and use information), media literacy (the ability to think critically about media content and context), computer literacy (the ability to use computers for problem solving) and scientific literacy (understanding how knowledge is created with its aims, methods, limitations, and politics), in addition to traditional health literacy. eHealth literacy has been defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem"⁶⁰.⁶¹
10. **'Digital Health Ecosystem'** integrates mHealth, Telehealth and telemedicine, eHealth governance, Digital health literacy, Electronic health records, Social networks, Standards and interoperability and big data.⁶²
11. **Digital health literacy:** "is the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to preventing, addressing or solving a health problem"⁶³. Better digital health literacy can lead to enhanced prevention models, better observance of healthier behaviours and improved wellbeing. Digital health literacy is one of the characteristic elements of 'Digital Health Ecosystem'⁶⁴ eHAction work package 4 dealing with digital health literacy of people has recently emphasised that "People who are digitally health literate can find, appraise and understand health information from electronic sources, before applying it in their own lives to achieve better health outcomes. Improving the level of digital health literacy empowers people to take a more active role in the management of their health and wellbeing." WP5 adds that "as the proliferation of wearables, connected medical devices, personal health records and mobile apps continues, consumers are becoming increasingly interested in capturing, analysing and sharing their own health data. While providing both challenges and opportunities for physicians, this information, also known as patient-generated health data (PGHD), can help people become more engaged in their own

⁵⁸ <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX:32019L1024>

⁵⁹ <https://www.ncbi.nlm.nih.gov/books/NBK216035/>

⁶⁰ <https://www.jmir.org/2006/2/e9/>

⁶¹ <https://link.springer.com/article/10.1007/s10198-019-01062-1>

⁶² https://www.who.int/global-coordination-mechanism/working-groups/digital_hl.pdf

⁶³ <https://www.who.int/global-coordination-mechanism/activities/working-groups/17-s5-rowlands.pdf>

⁶⁴ https://eurohealthnet.eu/sites/eurohealthnet.eu/files/publications/PP_Digital%20Health%20Literacy_LR.pdf

care.”⁶⁵ Therefore, WP5 stresses that digital health literacy also covers ability to share health data and understanding the importance and possibilities of innovative use of health and health related data.

12. **Big data in health:** Consolidated data from existing fragmented data sources for the purpose of understanding, forecasting and improving health and health system status, needs and performance. (This working definition was developed at the kick-of meeting of WP5, having into consideration the following definition of the Study on Big Data in Public Health, Telemedicine and Healthcare (hereinafter referred as ‘EU Study’)⁶⁶: “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.”)
13. **Big data analytics in health:** Statistical learning methods and algorithms applied to big data in health, which include descriptive analytics, mining/predictive analytics to support evidence-based decision making, analytical techniques that are ideal for analysing a large proportion of text-based health documents and other unstructured clinical data (e.g. physicians’ written notes and prescriptions and medical imaging).
14. **Artificial intelligence:** Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems), or can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications). Many AI technologies require data to improve their performance. Once they perform well, they can help improve and automate decision making in the same domain.⁶⁷
15. **‘Innovation’** is the process of translating an idea or invention into goods or services that create value, or for which customers will pay. To be called an innovation, an idea must be replicable at an economical cost and must satisfy a specific need. Innovation involves deliberate application of information, imagination and initiative in deriving greater or different values from resources, and includes all processes by which new ideas are generated and converted into useful products.⁶⁸
16. **‘Innovative use of health data’:** The use of health data is considered ‘innovative’ if this use results in better patient outcomes and/or higher quality of healthcare delivery and/or higher productivity and performance. Our approach in defining innovative use of health data is also based on the definition of innovation of the World Health Organization (WHO): ‘Health innovation identifies new or improved health policies, systems, products and technologies, and services and delivery methods that improve people’s health and

⁶⁵ https://letstalkrespiratory.eu/the-opportunities-and-challenges-of-patient-generated-health-data/?gclid=Cj0KCQjwhb36BRCfARIsAKcXh6HHdL7nQfWTcVLPrcXsjs5Mm3NdTdownzrUMoZutsJJOZiihVlzlQMcaAqX5EALw_wcB

⁶⁶ https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf

wellbeing. Health innovation responds to unmet public health needs by creating new ways of thinking and working with a focus on the needs of vulnerable populations. It aims to add value in the form of improved efficiency, effectiveness, quality, sustainability, safety and/or affordability. Health innovation can be preventive, promotive, curative and rehabilitative and/or assistive care. The WHO engages in health innovation to achieve universal health coverage within the context of the Sustainable Development Goals.’⁶⁹ Innovative use of health data is determined by the ways of converting unstructured, separated datasets into new or renewed things, services, solutions, organisations or systems.

3. Tables

WP	Task	Deliverables	MWP 2018-2021
WP-5. Innovative use of health data	T5.1 Mapping, awareness raising and policy relevant actions on innovative use of big data in health (Lead: NHSC)	D5.1 Report for the information of the eHN on policy level actions (M24)	B.1 Awareness raising of using Big Data in healthcare Expected outcome: Increase awareness on the possible impacts, challenges, risks and directions of Big Data in healthcare.
	T5.2 Sharing and learning best practices on European level (Lead: THL)	D5.2 Report on identified cross-border use cases, including assessment of pros & cons of stakeholders, and practical solutions with potential for European scale benefits (M18)	B.2 Develop common vision of innovative use of data in healthcare Expected outcome: Common vision and priorities for innovative use of data in healthcare. Clear criteria are needed to define which use cases / good practices of Big Data use in healthcare have the best potential to improve the sustainability of healthcare system and provide examples for benchmarking.
	T5.3 Towards an attempt to define common principles for practical governance (Lead: NHSC)	D5.3 Proposal for the eHN on the guidance for the implementation of common principles for practical governance of big data with a special focus on data to be used (and the implementation of data access and use) in public health, research and quality assurance in healthcare on a European scale (M36)	B.3 Governance and methodologies for innovative use of health data, including big data Expected outcome: Common principles to facilitate the development of innovative use of data projects at European Level. (+ Guidance for the implementation of common principles)

Table 1 - Overview Work Package 5