IDEA-FAST

Identifying Digital Endpoints to Assess FAtigue, Sleep and acTivities in daily living in Neurodegenerative disorders and Immune-mediated inflammatory diseases.

Grant Agreement No. 853981

**WP 8 – Data Protection,
Ethics and Legal challenges**

D8.4: Report on the challenges of sharing of digital device data from legal and IPR perspectives in the context of a learning healthcare system

Lead contributor	P15 MLCF (until 1-2-2023) P49 LYG (as of 1-11-2022)
Other contributors	

Due date	31 JAN 2022
Delivery date	24 AUG 2023
Deliverable type	R
Dissemination level	PU

Document History

Version	Date	Description
V0.1	22 AUG 2023	Frist draft based on D8.3 after removing confidential information
V1.0	24 AUG 2023	Final version

Table of Contents

1 List of abbreviations

AI	Artificial Intelligence
AI Act	Artificial Intelligence Act COM/2021/206 final
BMI	Body Mass Index
CA	IDEA-FAST Consortium Agreement
COS	Clinical Observational Study
DA	Data Act, COM (2022) 68 final
DGA	Data Governance Act, (EU) 2022/868
EHDS	European Health Data Space
EHDSR	EHDS Regulation, COM (2022) 197 final
EMA	European Medicines Agency
EU	European Union
FS	Feasibility Study
GDPR	General Data Protection Regulation (EU) 2016/679
ICD	Implantable Cardioverter-Defibrillators
IMID	Immune-Mediated Inflammatory Diseases
IMP	Investigational Medical Product
IPR	Intellectual Property Rights
MDCG	Medical Device Coordination Group
MDR	Medical Device Regulation (EU) 2017/745
NDD	Neurodegenerative Diseases
ODD	Open Data Directive (EU) 2019/1024
SPE	Secure Processing Environment
TS	Trade Secrets
UDI	Unique Device Identifier

2 Abstract

In this Deliverable raw device data are seen as the electronic signals which a device generates before they are via software translated into human intelligible data. This Deliverable discusses the legal possibilities to open up such raw data for research outside the context of a research plan and consortium agreement where the manufacturers had agreed in advance that the raw data would be shared. Such projects are rare while with the digital transformation in health care more and more devices will come onto the market. There will always be a certain loss of information in the translation of raw data into human intelligible data. The Deliverable assumes that in certain research projects one would need the full data for a better clinical understanding or simply to control the software which translates raw data to the human understandable data or clinical interventions as happens which devices such as ICD's.

The discussion in the Deliverable seems novel. We did not find directly applicable literature. Hence, this Deliverable is explorative in nature. We found that the upcoming European Data Act and the EHDS seem to provide possibilities to open-up raw data for research. The MDR or the GDPR on the other hand do not. That is also the conclusion of the trade organisation of manufacturers, MedTech Europe. MedTech Europe warns against such an extensive application of the coming legislation.

The case for opening up raw data would be strengthened if 'real world research scenarios' could be described where the access to raw data would be beneficial to a learning health system. Such scenarios exist already for opening-up raw data from clinical trials (meaning the original data in the clinical records instead of the case record forms as submitted to trial database). The next step in this line of research should be to describe such scenarios together with researchers involved. Also, more legal research will be needed, especially to align the novel concepts of data holder and data user in the coming EU Data Act with GDPR concepts of personal data, controller, processor and data subject in the context of raw data where the device will be offered by a health care provider.

3 Scope and outline

3.1 Introduction

IDEA-FAST aims to identify digital endpoints that provide reliable, objective and sensitive evaluation of fatigue, sleep problems and activities of daily living (ADL) for a number of immune-mediated inflammatory diseases (IMID) and neurodegenerative diseases (NDD). These will be identified through remote assessment of fatigue and sleep disturbances using monitoring sensors and mobile or residential technology. This will be done in the context of a 'clinical observational study' (COS) which was preceded by a feasibility study (FS). For both the FS and the COS, the raw data of the devices are analysed together with other measurements. As far as we know, manufacturers do not share the raw data of devices outside such clinical studies performed in the context of a consortium agreement (CA) or grant agreement. In the case of the IDEA-FAST CA, the device makers agreed in advance that the raw data would be made available to the partners in the consortium and are usually also funded for that.

This report is about whether this raw data can *also* be made available outside the context of specific studies where the device makers agreed beforehand that the raw data would be available for research.

3.2 Raw data

In the context of this report, we define 'raw device data' as all the electronic machine-readable signals which the device generates. These data are usually converted into human intelligible data. As a simple example: an electronic weighing scale converts the raw data (electronic signals) into numbers

expressed in grams and kilograms (on a western European scale). The algorithm behind this is fairly straightforward. A sophisticated scale can also provide a readout for BMI or fat rate. The algorithms will become more complex, however. The devices used in IDEA-FAST are far more complicated and the conversion from raw data to human intelligible data is much more complex as well. Hence, the interest in the raw data of these devices in the context of the mentioned studies.

However, outside a specific study where the devices makers agreed beforehand that the raw data would be shared, the raw data and the conversion to human intelligible data remain a black box. That is the background of the question of this deliverable.

In the context of health research, raw data can also have a different meaning. The European Medicines Agency (EMA) refers to 'raw data' as the data in the case record forms (CRFs). Those are the data on which the trial report is based as submitted to the EMA (EMA, 2022). There can be very good reasons to investigate the CRF data as well, also when there has been sufficient monitoring that the CRFs reflect the actual patient file (Doshi et al., 2013). In genetics, 'raw data' are results of whole genome sequencing (GWAS) analysis without the clinical data or their analysis of the (possible) influence on the health of the patient or participant, if the GWAS data had been collected in the context of research (Shabani et al., 2018).

From a philosophy of science point of view, 'raw data' is an oxymoron (Gitelman, 2013) if raw data are seen as 'something' given by nature, unchangeable, existing before our discovery and telling the truth about the object measured. We discover and measure in the context of a paradigm and others, starting from a different paradigm, might discover other data relating to the same object. So that raw data as a relative concept starts already before the following cascade which is often mentioned: raw data, information, knowledge, wisdom. The latter stages are obviously even more prone to our view of how things can be understood, but in this philosophical approach also the 'raw data' are subject to our context of discovery. For this report, there would not be 'raw device data' if we did not invent those devices for a certain purpose. That recognition is in a way sobering but not relevant for our report.

In conclusion: where we speak of raw data in this report, that is a shorthand for "raw device data seen as the output of the device in electronic machine readable signals before their conversion into 'human intelligible data' ". One could also speak of the 'unadulterated device data' which would be more exact.

3.3 Raw device data and intellectual propterty rights (IPR)

The relationship between data and IPR is complex. That is even more the case for the 'raw device data'. In general, IPR balance the economic or intangible interests of the inventor of a creation of the mind and the general economic or intangible interests of society that these inventions become available for others to use or to enjoy. Hence, a prerequisite for the application of IPR is that the invention is made public.

As the 'raw device data are not made public, that might be the end of the discussion in this section. However, the situation is somewhat more complex. It starts with the observation that 'data' as such are, as yet, not subject to IPR even if they were published. Europe has a 'sui generis' IPR for databases, the database right.[1] The Database Rights Directive defines a protective database as a 'collection of independent works, data or other materials arranged in a systematic or methodological way and individually accessible by electronic or other means (art. 1.2). The Directive has given rise to much debate and many court cases. In short, in order to profit from the Database Rights Directive, there must have been a substantial investment in the creation of the database, and it must also have a creative

[1] DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 1996 on the legal protection of databases.

element (hence more than just numbering the data in a certain order), and the protection is limited to data which are already there, in other words data which have been observed and then made available in the database.[2]

The latter raises the question whether device generated data are 'created' or 'observed'. Hugenholz critically discusses the first proposals of the European Commission of 2016 for new IPR in 'machine generated data' (Hugenholtz, 2018). However, that approach has been abandoned. It has been substituted by what has been called: "a rather unequivocal twist towards the abandonment of the traditional (intellectual) property arena as a regulatory device for data exchanges, in favor of a more granular data governance regime" (Margoni et al., 2023).

The following is based on that new approach. In that context the legal analysis will not cover IPR but the (many) Regulations which are directly or indirectly relevant to the governance of raw device data.

3.4 The main question

In this report we focus on the legal arguments via which 'raw device data' can potentially be opened up for research. As discussed in the preceding section, IPR will not be discussed in the context of the legal analysis.

We place that question in the context of a learning health care system being in essence that we learn from real world patient data to improve health care and health protection (more about a learning health care system in section 2). Patient data in this context mean in the first place the human intelligible data as the patient or treating doctor will see them. However, if those data only tell part of the story as some data were not taken up in the conversion, we might not get the full picture of what happened to this patient or why it happened. Hereinafter, we refer to this absence of the full picture and lack access to raw data in general as 'the gap'.

3.5 Focus on specific devices

Modern health care is unthinkable without devices. In the next section we will expand on the turn in healthcare towards more device generated data.

In the context of this study, we are focussing on a sub-class of devices which are in general comparable with those used in the IDEA-FAST FS or COS study. That is the background of this paper. For the FS and COS, researchers get the raw data. However, the intention is that at a later stage such devices are used on a much wider scale, using the results from IDEA-FAST and similar studies[3] to monitor patients either in the context of clinical trials and to assess the effects of the investigational medical product (IMP) on a more fine grained level than would be possible without the device[4] or to improve disease management of the patient in daily care using only the human intelligible data of the devices and not the original raw data before the conversion. And then there will be 'the gap' referred to earlier.

This means that we are interested in devices which measure patients on a more or less constant basis during their daily living or sleep.[5] That still leaves a broad array of devices open for our study. The second limitation is that the focus is on *medical* devices. The EU medical device regulation (MDR, see section) defines what those are.[6] That definition is quite complex with inclusion and exclusion criteria. Suffice to say here that the many devices such as smart watches which, even though they claim to improve the health of the wearer if he or she uses all functions, do not claim to be medical devices

[2] This brief summary is based on (Hugenholtz, 2018)
[3] Such as Radar AD and Mobilise D
[4] In which case the IMP will be the focus of research and not the device
[5] In IDEA-Fast some devices 'measure' patients during their sleep.
[6] In vitro medical devices have not been investigated in this report.

in the sense of the MDR are not object of this study. Which does not mean that what follows could be correspondingly applicable to 'health trackers' as well.

The third limitation is that the raw data must be stored somewhere. With a device which directly translates electronic signals into human readable output after which the electronic signals are deleted, access to raw data at a later stage is impossible. Where the data is processed by software, either or not a Medical Device as well, and are stored, at a platform of the device maker, in the cloud, in the device or at the healthcare provider, is irrelevant for this report. The same applies to the duration of the data storage. If the raw data are needed at a later stage, one could make arrangements that they will be stored longer.

3.6 Why this exploratory report?

The scope is situations when the raw data of medical devices are there already and afterwards the question of access to these raw data arises. Hence, the scope of this exploratory report is explicitly not limited to devices used in IDEA-FAST or to studies such as IDEA-FAST. Those are situations where a consortium agreement already requires data sharing by the partners. But as mentioned, those situations are rare.

IDEA-FAST and all similar projects should look beyond the horizon of those projects and even beyond the 'sustainability' of further use of the data assembled in the project. This was the main reason for including this deliverable in the IDEA-FAST work plan. Unlike most of the efforts of WP8 which are aimed to give ethical and legal guidance to IDEA-FAST as it unfolds, we were also given the opportunity to look beyond the horizon and to investigate other aspects of the digital transformation.

With the digital transformation in healthcare, more and more monitoring devices will be used in patient care, changing also the patient-physician interaction (Leclercq et al., 2022). A strong case for this transformation is made by Rosenberg (Rosenberg, Lawrence, 2023). Healthcare systems aiming equitable access[7] as all European healthcare systems do, will become unsustainable, financially but also regarding the necessary staff involved, if they do not change. According to Rosenberg a major necessary change will be decentralisation of healthcare via much more home monitoring. Patients will become much more in control of their health and needing hospital visits and physicians less. Rosenberg also warns against 'big tech' taking over without in my opinion offering concrete solutions against this possible negative side effect of the digital transformation. One of those could be access to raw data.

Some devices do not only monitor but can also intervene automatically when certain measurements get below or above a certain threshold. Defibrillators and implantable cardioverter-defibrillators (ICDs) are a case in point. Insulin pumps which react without human interference are another example. These exist already for a long time and it may assumed that more such devices will come on the market. The safety of these devices will be subject to strict testing, see the paragraph on the Medical Device Regulation. The literature on cyber-security threats of connected devices (Cilliers, 2020; Coventry & Branley, 2018; Jiang & Shi, 2021; Kapoor et al., 2019) shows that many devices are connected to a platform which will capture the raw data, This underscores that it is worthwhile to investigate how these raw data can be opened up for other purposes than the business interests of the device maker.

3.7 More devices at home is not only about 'the gap'

This paper only discusses 'the gap'. We are quite aware that the potential proliferation of devices which monitor patients on a more or less constant basis, which raises other questions as well, such as about intrusiveness of that nearly constant monitoring which should then be weighed against the advantages of better management of the patient's disease. Of course, there should be informed consent

[7] In his words 'universal healthcare'

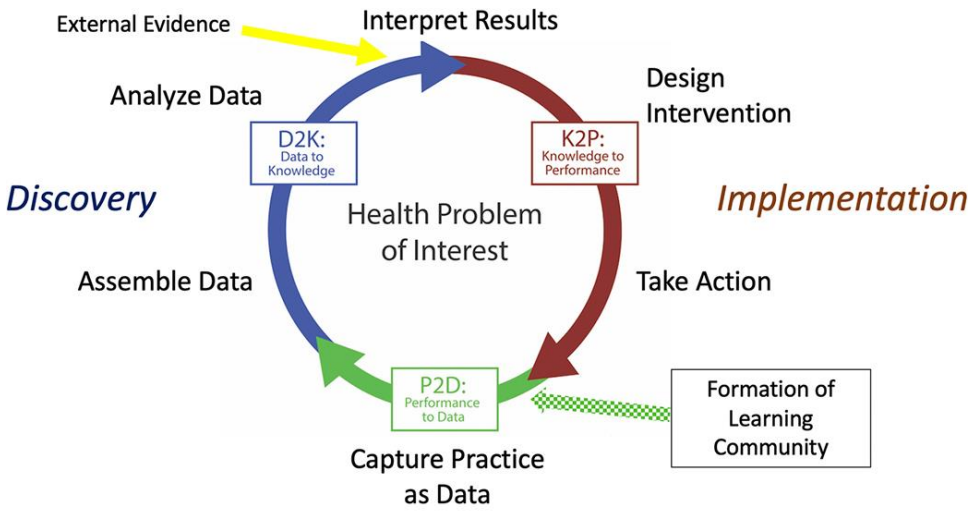
of the patient but how much choice does he or she have in the situation of a severe disease? There are obviously also relevant questions of data protection in the flow of data from device to human intelligible data on a reader. We may refer to a host of ethical literature about these issues (for example: Mittelstadt, 2017) However, those questions are not the scope of this report.

4 Brief description of a learning healthcare system

4.1 A learning health system and data

A learning health system has been defined as a "health system in which progress in science, informatics, and care culture align to generate new knowledge as a natural by-product of the care experience, and refine and deliver best practices for continuous improvement in health and health care" (Grossmann et al., 2011). Reuse of health data is at the core of such systems.

Establishing learning health systems, in which health data is constantly captured, generated, reused and learned from is essential to maintain acceptable levels of quality, accessibility and sustainability of care, can improve health protection, can underpin decisions about appropriate use of limited resources, and will further what had been called 'appropriate care' in general. Foley describes the continuous learning circle of a learning health system as follows (Foley & Vale, n.d.):



European health care systems must meet certain standards such as equitable access, high quality, and long-term sustainability while at the same time they are faced with an ageing population, increasing burden of disease and staff shortages (for the Netherlands see: 'Kiezen voor houdbare zorg. Mensen, middelen en maatschappelijk draagvlak WRR-Rapport 104'). Hence European healthcare systems face many challenges. A learning health system may help us to face these challenges. The actions resulting from a learning health system can relate to better treatment options, better knowledge of which treatments are more cost-effective than other treatments with a comparable effect of health or better prevention.

The efforts for a learning health system are often associated with 'real world data' or 'real world evidence'. A host of initiatives are taking place under that name, such as by the European Medicines Agency[8], [9], [10]or the BigData@Heart project[11] and its outputs (Kotecha et al., 2022).

The data captured, P2D in the figure, usually refer to data in electronic health records (EHRs), other data which are captured in healthcare or social care or to administrative data. To use those for a learning health system is quite challenging already with practical hindrances and privacy concerns standing in the way (Grossmann et al., 2011). Yet, 'raw device data' can be seen as part of the practice as well.

4.2 Are 'raw' data from devices also relevant for a learning health system?

In February 2022, we interviewed one device maker. The spokesperson of this manufacturer did not see the added value of access to the raw data outside very specific protocols. In general, human intelligible data should be sufficient.

This is not what we heard from researchers. We also briefly inquired about access to raw data among several researchers. It should be said that those researchers were all involved in studies with devices which measure patients' activities. Hence, it was little surprise that access to raw data was welcomed outside such very specific protocols where the device maker had agreed beforehand that the raw data would be shared.

We did not ask researchers who are reusing patient data in the context of a learning health system. However, we may submit the following. A learning health system can also be seen as uncovering and challenging existing practices in healthcare and health protection in order to improve each (which will depend on the study). If the devices would have a 'bias' built in the algorithm which can be uncovered by access to the raw data, that access is needed. We do not know that in advance. The same would apply if the raw data would show more then what is translated into human readable data but which might help to improve the health system. The devices are there, the signals are there, and we should try to make best use of them.

On July 14, 2023, we discussed the question of opening raw device data with the 'Ethical Legal Advisory Board' (ELAB) of IDEA-FAST. In general, the ELAB members:

- Questioned whether access to raw device data would help patient's autonomy and trust. We share that scepticism. Hence, we focus on access for research (the context of learning health system). When discussing the EU AI Act we will briefly come back to the issue of trust in device generated output.
- The ELAB also questioned the concept of 'the more data, the better' (for a learning health system) paradigm. Acquiring and storing data also carries costs and privacy risks. There should be a good reason for further use of data and not that the data might be relevant in the future.

We will come back to these remarks at various points in this paper. Suffice to mention here that when we discuss access to raw data for research, we assume a research protocol which clarifies why the raw data are relevant for the research question or questions.

[8] https://www.ema.europa.eu/en/news/vision-use-real-world-evidence-eu-medicines-regulation
[9] https://www.ema.europa.eu/en/news/high-quality-data-empower-data-driven-medicines-regulation-european-union
[10] https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu
[11] https://www.bigdata-heart.eu/

5 Legal analysis

In this and the following sections the various regulations will be discussed which might influence access to raw data for research.

5.1 Raw data in the context of the MDR

Regulation 2017/745[12] (hereinafter: the MDR) regulates in short the safety of medical devices entering the European market. Medical devices which are only used in research are exempted from most of the provisions of the MDR. The MDR replaces earlier applicable directives and became fully applicable on March 26, 2021. The long transition period is mainly due to the difficulties encountered in the limited number of "approved" Notified Bodies with the required expertise. Another reason was setting up the Eudamed database on medical devices that integrates different electronic systems to collate and process information regarding devices on the market and the relevant economic operators, certain aspects of conformity assessment, notified bodies, certificates, clinical investigations, vigilance and market surveillance.

Insofar as relevant here, the safety of a medical device is via the MDR assured by:

- A quality management system at the manufacturer.
- Assessment of the safety and function of the device by a notified body, unless for very low risk devices.
- A unique device identifier (UDI).
- For medium to high risk devices, the assessment must be based on a clinical investigation.
- Once on the market the device maker is required to organise post marketing surveillance.

The MDR contains detailed provisions on the clinical investigations and the report to be submitted to the notified body. Those provisions do not give details about the type of data which must be submitted, in particular whether they must contain the raw data. The Medical Device Coordination Group (MDCG) issues further guidance about the clinical evaluation.

The MDR mentions raw data in one article, being article 76.6. In addition to the full clinical investigation report, a summary report must be drawn up. That report is public. Article 76. 6 reads as follows (our emphasis).

> The Commission shall issue guidelines regarding the content and structure of the summary of the clinical investigation report.
>
> In addition, the Commission may issue guidelines for the formatting and *sharing of raw data, for cases where the sponsor decides to share raw data on a voluntary basis*. Those guidelines may take as a basis and adapt, where possible, existing guidelines for sharing of raw data in the field of clinical investigations.

The Commission Guidance on the content and structure of the summary of the clinical investigation report has been published May 8, 2023.[13] These are not (yet) about the 'raw data' as mentioned in article 76.6 of the MDR.

Even if the Guidance would mention 'raw data', those would then be 'raw data' as the EMA uses that term. Not the raw device data but the raw clinical data. This also follows the definition of clinical data in article 2.48 of the MDR which reads:

- 'clinical data' means information concerning safety or performance that is generated from the use of a device and is sourced from the following:

[12] In full: Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017
[13] Official Journal of the European Union 8.5.2023, C.163/7

— clinical investigation(s) of the device concerned,

— clinical investigation(s) or other studies reported in scientific literature, of a device for which equivalence to the device in question can be demonstrated,

— reports published in peer reviewed scientific literature on other clinical experience of either the device in question or a device for which equivalence to the device in question can be demonstrated,

— clinically relevant

— information coming from post-market surveillance, in particular the post-market clinical follow-up.

Hence, not data generated by the device. The 'raw data' will in general not be submitted to the notified body. The notified body will base its assessment on the clinical data and, as the devices of this study always use software, on whether the software has been developed in accordance with the applicable ISO norms. The guidance of the MDCG refer to those norms as well. Those end in software verification and validation.

Annex II of the MDR states about the software the following:

> describing the software design and development process and evidence of the validation of the software, as used in the finished device. This information shall typically include the summary results of all verification, validation and testing performed both in-house and in a simulated or actual user environment prior to final release. It shall also address all of the different hardware configurations and, where applicable, operating systems identified in the information supplied by the manufacturer.

Annex II describes the information which should be provided with the device to the users. The notified body will see more than the summary of the results. But -as far as we could see-, given the norms about software development, verification and validation, that does not include the raw data themselves.

The fact that the notified bodies do not have the 'raw data' as defined in our study (they do have the raw data as the term is used by EMA) also means that they cannot give access to them. That is relevant for the access opportunities offered by the new EU 'data acts' as discussed in paragraph 6.

5.2 Raw data and the right to data portability under the General Data Protection Regulation (GDPR)

Under the GDPR, personal data is any information that in short relates to an identified or identifiable natural person (article 4.1 GDPR). Recital 26 of the GDPR expands on this definition. There is considerable debate as to when data which under certain extreme circumstances might be indirectly identifiable can still be considered personal data (Groos & Veen, 2020). Recently the EU Court in first instance decided that also pseudonymised data are not necessarily personal data for the receiver of those personal data (Case T-557/20, 26 April 2023). According to the Court, that depends. If the receiver cannot re-identify the data subject without illegal or excessive mean so that the risk of re-identification is in fact negligible, then the pseudonymised data are not considered to be personal data.

Groos and van Veen attempt to operationalise these rather vague terms in the 6 safes model. These safes are (slightly adopted to a more general, non-research scenario):

1. Safe transfer: can the data not be intercepted during transfer?
2. Safe projects: is this use of the data appropriate?
3. Safe people: can the authorised users be trusted to use it in an appropriate manner?
4. Safe data: is there a direct disclosure risk in the data itself?

5. Safe settings: does the access facility limit unauthorised use?
6. Safe outputs: are the statistical results non-disclosive?

The model also takes into consideration that other parties using illegal means may want to access the data. A database which can be easily hacked, will more easily hold personal data than a database which meets all the safety requirements. The hacker might afterwards use means to re-identify which are considered excessive to the recipient who has no economic interest whatsoever to re-identify. On the contrary, that would seriously undermine its reputation.

This is important as, in healthcare situations, the manufacturer will hardly ever process directly identifiable data of the patient. The usual scenario is that the healthcare provider supplies the device to the patient. The device will have the UDI. The UDI is linked to the patient record at the healthcare provider. The manufacturer knows to which healthcare provider the device has been delivered and will know when it comes into action, namely when it starts to send out signals. But not to which specific patient. It will process the raw data under the UDI, hence will only process pseudonymised data.

Given the decision of the court, it may very well be doubted whether those pseudonymised data are, merely because they are pseudonymised[14], personal data. Yet, it also unclear whether the data under the UDI meet the six safes. The technical safety procedures will be met but how 'appropriate' will the use of the data be?

However, let us for the sake of the argument assume that the data can be considered anonymous data. Then there is another hurdle for the patient to request the raw data from the manufacturer, based on the GDPR.

The GDPR distinguishes between controller and processor. The data controller is the entity which determines the purposes and means of the processing of personal data (article 4.7 GDPR). The processor is an entity which acts on behalf of the controller, according to their instructions. The healthcare provider is certainly the controller of the data which it receives from the device. This distinction is relevant as almost all GDPR rights can only be exercised against the controller.

It is a very complex discussion whether the manufacturer should be considered a processor or controller. The usual opinion is that the manufacturer is a mere processor. This is different when the product is data the user submitted. Then the user would be controller or, depending on the circumstances, one could speak of joint controllers (Finck, 2021). Here we have product which does something for you. It cannot function without processing data but neither without all the other technical aspects of the device. The data and their processing are an intermediary step for the functioning of the device. They bring the measurements of the device into output. When you purchase a device, the user (we will come back to that term later) decides about the measurement and output. Hence the user also decides about purposes and means of the data processing. Even though the product cannot function without those data as decided by the manufacturer.

Dahi and Copagnucci challenge that paradox for cases where the patient directly purchases the device (Dahi & Compagnucci, 2022). They plead for an extensive interpretation of the concept of controller in that context. In their view the manufacturer should be considered a controller. The extensive interpretation is based on certain assumptions, such as that medical devices in a solidarity based health care system are bought be the patient and that when the manufacturer is a controller, the patient will have a better handle for compliance of the manufacturer than as a processor.

But again, for the sake of argument, let us assume that the manufacturer is the sole controller.

[14] It should be mentioned that data can be personal under the pseudonym because of the granularity and identifiability of those data and when he six safes are not met.

If the raw data are to be considered personal data and the manufacturer is a controller, the patient can in principle request a copy of those data (article 15.3 GDPR). The data should be provided in a 'commonly used electronic form'. Such a commonly used form does not exist in this context. The legislator had more standard cases in mind with this provision. As far as we can see, the large and complex datasets which this copy of the data would entail, would be novel.

In theory also another article of the GDPR could be applicable, being article 20, the right to data portability, meaning that the data subject can request the controller to transfer the data to another controller. The background of this article is to prevent vendor-lock in for data subjects, especially for data on social media platforms. The application of article 20 has inherent limitations (Vanberg & Ünver, 2017) and published case law is still absent.

In order to be applicable, in essence four conditions must be fulfilled:

1. The data processing must be based on consent or a contract with the controller;
2. The data must be processed by automatic means;
3. The data must be submitted by the data subjects themselves;
4. The transfer must be technically feasible.

It should be mentioned that the proposed EU Data Act (DA) aims to lift the barriers 1, 3 and 4 to a large extent. We will come back to the DA in section 6. As it might take time before the DA becomes applicable, let us first look into article 20 GDPR.

Of the above conditions, only 2 is applicable without further discussion.

For condition 3 it can be questioned whether the patient submitted those data. Certainly not in the traditional sense, meaning that patient him- or herself entered the data. The device generated the data based on certain bodily functions of the patient. Condition 1, however, is the main obstacle for the applicability of article 20.

Consent in the sense of the GDPR means that it is freely given. Withholding consent should only mean that the personal data may not be processed. If a certain activity cannot take place without data processing annexed to it, then the legal basis must be found in the legitimacy of that activity and not in consent for data processing. (art. 4.11 and 7 GDPR, (European Data Protection Board & EDPB, 2020), such as a contract or a legal obligation.)

As argued, in this case the main purpose of the device is not data processing but measuring and sometimes interfering with the body of the patient in the context of the treatment of the patient. Data processing is annex to that main function, a necessary intermediary step. The patient cannot say 'I want or need the device but not the data processing'. Hence, the data processing of the device is not based on consent. The annexed data processing should be notified to the patient (art. 13) and preferably explained to the patient as part of the informed consent process to use the device. But it is function of device which is ultimately decisive. We submit that the legal basis of the data processing will be 'necessary for the performance of a contract with the healthcare provider 'or the legal obligation of the healthcare provider to provide appropriate care.

The second legal basis, a contract might be more apt. If the patient would purchase the device, there will be a contract with the manufacturer. However, is that contract about data processing or about the functioning of the device and will it even mention the data processing inherent to its functioning? Another question is whether that is really necessary. There is no case law on this point yet.

The conclusion is that using article 20 GDPR to open-up raw device data for research, as the patient would then request that the data are transferred to the research database, is subject to many hindrances. There might be more direct route as will be discussed below.

6 The new European 'data acts'

6.1 Introduction

The European Commission aims to foster a vibrant digital economy (European Commission, 2020). Following that policy, the European Commission has issued an avalanche of legislation in order to increase responsible reuse of data. Parts of these proposals have already made into law, others are still being debated. What data and under what circumstances parties who had not assembled the data originally, can have access to data from parties who had collected them originally, differs between these various instruments.

The following legislation seems at first sight relevant for this exploration about access to raw data. Each will be briefly discussed to assess whether the legislation is on closer inspection indeed relevant for researchers or patients wanting to access to raw data of a manufacturer.

1. The Open Data Directive 2019/1024, applicable as of 17 July 2021.
2. The Data Governance Act (DGA), applicable as of 24 September 2023
3. The Data Act (DA), under discussion
4. The European Health Data Space Regulation (EHDS), under discussion

A study commissioned by the European Commission showed that the first three pieces of legislation leave much ambiguity for research organisations and research funders alike (Directorate-General for Research and Innovation (European Commission) & Eechoud, 2022). We will take that study into account as well in the following. However, the focus is on opening up data from manufacturers. These are research organisations only if special conditions apply as will be explained below.

6.2 Open Data Directive (ODD)

6.2.1 In general

This Directive amends earlier 'open data' Directives of 1998 and 2003. It applies to public sector and publicly funded information. There are nuances in the ODD for certain publicly funded organisations such as public libraries.

Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose (Recital 16 of the ODD). This concept shows that:

- 'open data' cannot be personal data, as the essence of the GDPR is that personal data cannot be freely used but are subject to the restrictions and warranties for data protection;
- The so-called contextual approach to whether data are personal data or not, does not apply to releasing open data. The conceptual approach states that it depends on the circumstances of the holder of the data whether the data are anonymous or still personal data (Groos & Veen, 2020). These circumstances must then be very well established in advance. With open data there is no context and established circumstances. The data might end anywhere and also illegal means to re-identify the data subject could in theory be used.

This is relevant as the ODD mentions the usual formula in the new data governance Directives and Regulations, being that it is applicable without prejudice to the GDPR (Recital 52). This means that the threshold for releasing open data as really anonymous data should be very high. The ODD, however, is not limited to such open data. The ODD supports that certain documents can be made available under a license and hence are not open data as defined before. As stated in Recital 44:

> However, in some cases justified by a public interest objective, a licence may be issued imposing conditions on the re-use by the licensee dealing with issues such as liability, the protection of personal data, the proper use of documents, guaranteeing non-alteration and the acknowledgement of source.

There are several instances in the ODD where licenses play a role. In the next section possible license for research data will be discussed.

6.2.2 Application to raw device data

In general, the manufacturer as a private company, is exempt from the ODD. This would be different when the development of the device would be publicly funded. It is not completely clear what 'publicly funded' means in this respect. There might be mix of public funding and 'in kind contributions' of the manufacturers. We did not find clear indications about the threshold. If the publicly funded research would aim to reach the device, starting from technology readiness level 4, to technology readiness level 6, obviously the data generated in those stage 5 and 6 are publicly funded. But the preceding stages were not, and the research agreements will have clauses about this background knowledge.

However, in a less complicated and perhaps hardly existing case where the whole device development would be fully publicly funded, the research data, which may encompass 'raw device data', would fall under the remit of the ODD. It should be noted that this means only the data processed during the research phase. Not the raw data once the device is on the market which is the general scope of this paper.

Article 10 of the ODD applies to research data. The ODD defines research data as follows (article: 2.9):

- documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results.

If the raw device data are used in the research and as evidence for the results, those data are research data.

Article 10 which relates to opening research, is difficult to interpret. It states as follows:

1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.
2. … research data shall be re-usable for commercial or non-commercial purposes in accordance … insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account.

The second part of this article is in a way relatively easy to interpret. Data which have been made publicly available already, shall be re-usable subject to conditions. Those conditions can also entail legitimate commercial interests of the manufacturer. It is, however, subject to question how such licenses can be imposed when the data were already *publicly* available.

The first part of this article seems most of all a strong recommendation for FAIR data reuse. FAIR does not mean 'open data' as described before. Amongst other things, data protection stands in the way (see also (Boeckhout et al., 2018). Section 1 of article 10 also mentions other limitations to reuse,

such as intellectual property rights and legitimate commercial interests. Hence, it will very much depend on the balance which is being struck in the governance of the controller of the original research data which raw device data can become available to third parties and under what conditions. We propose that it would be contrary to FAIR if the original data could not be reused in a pre-competitive research setting and after the results have been published, accommodating the legitimate interest of the parties who collected the data.

It should be mentioned that the ODD is a directive in the sense of EU law. That means that member states must implement it into national law. Within certain limits member states can impose higher standards for FAIR than as laid down in the ODD. However, those cannot be contrary to established EU law such as the GDPR.

6.3 The Data Governance Act (DGA)

6.3.1 In general

The DGA regulates three issues of potential relevance for this deliverable:

1. Further requirements for 'public sector bodies' to make protected data (for an explanation see later in the text) available for reuse.
2. Notification and supervisory framework for the provision of data intermediation services.
3. A framework for voluntary registration of entities which collect and process data made available for altruistic purposes (data altruism).

Additionally, the DGA establishes a "European Data Innovation Board".

Ad 1: public bodies and protected data

This part of the DGA can be seen as an expansion of the ODD. It applies to public sector bodies. Public sector bodies are broadly defined in EU law. In addition to regulated by public law, strong government involvement and/or funding makes an entity a public sector body. Universities in EU member states should in general be considered public sector bodies. However, this part of the DGA makes an exception for – insofar as relevant here, there are more exceptions – data held by cultural establishments and educational establishments. The reason is that the works and other documents they hold are predominantly covered by third party intellectual property rights (Recital 12). If a research performing organisation is instituted as a public sector body, it falls under the DGA. However, Recital 12 also states (our emphasis):

> The exchange of data, purely in pursuit of their public tasks, among public sector bodies in the Union or between public sector bodies in the Union and public sector bodies in third countries or international organisations, *as well as the exchange of data between researchers for non-commercial scientific research purposes*, should not be subject to the provisions of this Regulation concerning the re-use of certain categories of protected data held by public sector bodies.

What is probably meant here that when there is another public obligation to share already amongst public sector bodies, the DGA does not apply. This is linked to a second element of what seems to be the main purpose of the DGA: to make protected public data available to businesses.

The DGA does not institute new obligations to make data available for reuse but elaborates on the conditions how hat should be done for the so called 'protected data'. Data at public bodies can be protected data because of:

- commercial confidentiality, including business, professional and company secrets;
- statistical confidentiality;
- the protection of intellectual property rights of third parties; or

- the protection of personal data, insofar as such data fall outside the scope of Directive (EU) 2019/1024.

The relevant part of DGA then suggests tools to make these data can still be available such as via licenses (which may grant an exclusive right to re-use the data for a certain limited period) or by providing a 'secure processing environment'. It poses obligations how these tools should be used, in short in a fair, transparent and non-discriminatory way.

A national 'competent body' should be established which can assist public sector bodies in re-use of data or can perform re-use operations itself.

Ad 2: data intermediation services

Data intermediation services providers, which may include public sector bodies, will offer services that connect the different actors. As such they have the potential to contribute to the efficient pooling of data as well as to the facilitation of bilateral data sharing. They should be a separate legal entity, independent of the data holders and data users.

As we are either discussing research to research or manufacturers to research situations in this paper, the data intermediation services are not relevant. A research database instituted by a consortium which has a governance stating how data can be reused by third parties, is not a data intermediation service also when the third parties can be commercial parties. That would become different if the governance would be outsourced to a separate legal entity. Hence it would be interesting to assess certain repositories for research data such as dbGAP[15] against the provisions of this chapter of the DGA. However, that is beyond the scope of this deliverable.

And there is still time. The DGA will apply as from 24 September 2023. However, the obligation of the 'data intermediation services' will only be applicable as from 24 September 2025.

For the data intermediation services one or more a competent authorities should be established. Their remit is in short to monitor compliance of the data intermediation services with the provisions of the DGA. If data intermediation service does not comply, the authority can after having followed an administrative procedure, impose penalties, order the postponement of the activities of the data intermediation service or its cessation.

Ad 3: data altruism

The DGA defines data altruism as follows:

- the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, etc, etc. public policy making or scientific research purposes in the general interest.

It should be mentioned that in many EU countries, opening up patient data for health research is not based on the patient's individual consent (Hansen et al., 2021). Data sharing for health research under the EHDS (see section 7 infra) will in its present version not be based on consent either. Hence, the applicability for data altruism in health care might be limited.

The DGA states that member states should foster data altruism. Data altruism can be organised by 'data altruism organisations'. These organisations must meet a number or requirements such as being

[15] See: https://www-ncbi-nlm-nih-gov.eur.idm.oclc.org/gap/ This link seems only to open if one access to a research library.

a not-for profit legal person. In order to act as such the data altruism organisation must be recognised by a competent authority for such organisations.

6.3.2 Consequences of raw device data

The DGA has consequences for notified bodies in the sense of the MDR. Those are also public bodies in the sense of EU law. However, as discussed earlier, the notified bodies generally do not receive the raw device data in the application of the manufacturer. If there are no raw device data at the start, also the other parts of the DGA cannot be applied in practice.

6.4 The (proposed) Data Act (DA)

6.4.1 Introduction

The DA was proposed by the European Commission on 23 February 2023. According to the Commission it aims at: 'ensuring fairness in the digital environment, stimulate a competitive data market, open opportunities for data-driven innovation and make data more accessible for all'.[16] In a way it can be seen as the counterpart of the ODD and the DGA. While the latter mainly apply to obligations of public bodies to make data available for re-use, the DA applies to enterprises and how those should make data available for re-use subject to the conditions of the DA.

The DA is meant to be 'horizontal regulation'. That means that is applicable to all kinds of sectors where data re-use may play a role. But each sector can have different rules which, though starting from the principles of the DA, might contain different legal obligations and procedures for re-use. For healthcare, that would be the proposed Regulation on the European Health Data Space (EHDSR). The EHDSR will be discussed in a separate section. When there are no sector specific rules (yet), the DA applies.

More concretely the DA sets rules for, in so far as relevant here:

- Facilitate access to and the use of data by consumers and businesses, while preserving incentives to invest in ways of generating value through data. This includes increasing legal certainty around the sharing of data obtained from or generated by the use of products or related services, as well as operationalising rules to ensure fairness in data sharing contracts.
- The proposal clarifies the application of relevant rights under Directive 96/9/EC on the legal protection of databases (the Database Directive) to its provisions.
- Provide for the use by public sector bodies and Union institutions, agencies or bodies of data held by enterprises in certain situations where there is an exceptional data need. This primarily concerns public emergencies, but also other exceptional situations where compulsory business-to-government data sharing is justified, in order to support evidence-based, effective, efficient, and performance driven public policies and services.

Additionally, the DA:

- Facilitates switching between cloud and edge services.
- Puts in place safeguards against unlawful data transfer without notification by cloud service providers.
- Provides for the development of interoperability standards for data to be reused between sectors.

[16] See: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

- Supports the setting of standards for 'smart contracts' to facilitate access to and the use of data by consumers and businesses.

Those policy goals are of less interest to the scope of this paper and will not be discussed.

6.4.2 Applicability of the DA applicable to raw device data

In our opinion the DA is applicable to raw device data. Data are very broadly described.

- 'data' means "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording" (art. 2.1 DA)

This should be read in conjunction with which generates the data, being the product in terms of the DA:

- 'product' means "a tangible, movable item, including where incorporated in an immovable item, that obtains, generates or collects, data concerning its use or environment, and that is able to communicate data via a publicly available electronic communications service and whose primary function is not the storing and processing of data" (art. 2.2 DA).

The 'publicly available electronic communications service' might be seen as a problem here. The manufacturer will (and should not) not store the raw device using a service to which everybody has access. But in the end the human intelligible data should reach its user. That can only be achieved via a 'publicly available electronic communications service'.

The broad scope of the DA is also underlined in Recital 31 where it explained how the DA compliments article 20 GDPR.

- The right under this Regulation complements the right to receive and port personal data under Article 20 of Regulation (EU) 2016/679 in several ways. It grants users the right to access and make available to a third party to any data generated by the use of a product or related service, irrespective of its nature as personal data, of the distinction between actively provided or passively observed data, and irrespective of the legal basis of processing. Unlike the technical obligations provided for in Article 20 of Regulation (EU) 2016/679, this Regulation mandates and ensures the technical feasibility of third party access for all types of data coming within its scope…

Hence, the limitations of article 20 GDPR, as discussed in section 5.2, do not apply when requesting data under the DA.

6.4.3 The details, making available to whom and by whom

Regretfully, we are not there yet. Though the Recitals of the DA often refer to the manufacturer of the product, that term does not come back in the definitions. There it is the data holder:

- a legal or natural person who has the right or obligation, in accordance with this Regulation, applicable Union law or national legislation implementing Union law, or in the case of non-personal data and through control of the technical design of the product and related services, the ability, to make available certain data (2.6 DA).

The data should be made available to the user, being:

- a natural or legal person that owns, rents or leases a product or receives a services.

Or another data recipient:

- a legal or natural person, acting for purposes which are related to that person's trade, business, craft or profession, other than the user of a product or related service, to whom the data holder makes data

available, including a third party following a request by the user to the data holder or in accordance with a legal obligation under Union law or national legislation implementing Union law

These terms should be mapped against the flow of the device data as described earlier and the description of that flow in terms of the GDPR. The explanatory memorandum and the Recitals are somewhat ambiguous about these issues. The suggestion is that data holder = a manufacturer and controller in the sense of the GDPR. Earlier we concluded that the manufacturer should usually be considered a processor.

The user of the product is the "legal or natural person, such as a business or consumer which has purchased, rented or leased the product or receives services" (article 2.5 DA).

It will very much depend on the organisation of the health care system and the specific device under what conditions the patient receives the device. It might be ownership but lease or receiving a service from the healthcare provider who owns the device are more likely scenarios. In that case there would be two users: the healthcare provider and the patient.

These are scenarios which must be assessed in more detail. However, in general we disagree with the conclusions in Nature (Fleming, 2023) that the DA 'leaves researchers in the cold' and that 'the DA is primarily meant for business to business opening up of data'. The ODD and the DGA are to a large extent, but the DA is about opening up data *from businesses to users*. The user as a patient or healthcare provider could liaise with a research organisation and perform research together and may also request that the data are sent to a data recipient such as research organisation as is explicitly acknowledged in Recital 29 of the DA.

Hence research can be done on the raw device data before the other scenario of the DA for access of public bodies, which may involve research organisations, becomes applicable, being in case of public emergencies. The first scenario, a data user requesting raw data to be transferred to a research organisation, will involve a lot of paperwork but seems in principle feasible.

7 The European Health Data Space Regulation (EHDSR)

7.1 Introduction

The EHDSR proposal was launched by the European Commission in May 2022. It is meant as major game changer for health data use and reuse in Europe. The EHDS has two aspects:

1. rules for health data when delivering healthcare;
2. rules for research with health data.

The first type of rules contains in essence:

- standards for electronic health records (EHRs) to increase interoperability of EHR systems and data portability of patient data;
- access of patients to data in their EHR;
- easier exchange of EHR for the delivery of care.

For this report the second type of rules are more interesting as laid down in Chapter 4 of the EHDSR. Those also got the most attention in the many comments on the EHDS proposal.

Yet it is interesting to note the MedTech Europe's reaction to this part of the proposal. MedTech Europe states that EHR data should be clearly defined and that:

- "Including raw data and technical parameters, might not be in the interest of users (citizens, patients or HCPs, - meaning healthcare professionals, authors note) as it will be irrelevant to

the needs of patient care or (personal) health management and might overload them" (MedTech Europe, 2023).

7.2 EHDS and research

7.2.1 Chapter 4 of the EHDS proposal summarised:

- a new separate legal basis for research with health data;
- the institution of national health data access bodies;
- unless data from only one health care provider are being used, all research proposals must be made to the national data access body;
- the data access body can give access, a 'data permit', to the health data for research in an approved secure processing environment. The original data cannot be downloaded from this environment, only the statistical scientific output;
- It is forbidden to use the data to take decisions detrimental to individuals, to increase insurance premiums, to market health products towards health professionals or patients or to design harmful products or services;
- Health data access bodies must ensure transparency: information will be published about data access applications. In addition, data users must make public the results of their electronic health data uses and inform the health data access bodies of any significant findings relevant for the health of individuals.

7.2.2 Health data covered by chapter 4 of the EHDS

Article 33 section 1 of the EHDSR contains a long list of types of data falling within the scope of Chapter IV. Point k mentions:

- electronic health data from medical devices and from registries for medicinal products and medical devices.

This seems to encompass 'raw device data'. The EHDSR does not exempt data entailing protected intellectual property and trade secrets from private enterprises from its scope. These shall also be made available for secondary use. However, where such data is made available for secondary use, all measures necessary to preserve the confidentiality of IP rights and trade secrets shall be taken (33.4 EHDSR). The EHDSR is less clear what those measures entail. The proposal states:

- Public sector bodies or Union institutions, agencies and bodies that obtain access to electronic health data entailing IP rights and trade secrets in the exercise of the tasks conferred to them by Union law or national law, shall take all specific measures necessary to preserve the confidentiality of such data (34.4).

Probably the same would apply to non-public bodies.

7.3 Obligations of the data holders

The EHDSR is not without obligations of the holders of the data mentioned in 33.1. These obligations differ slightly between datasets collected and processed with the support of Union or national public funding, and other datasets.

The first type of datasets must have a 'data quality and utility label' encompassing:

a. for data documentation: meta-data, support documentation, data model, data dictionary, standards used, provenance;
b. technical quality, showing the completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;
c. for data quality management processes: level of maturity of the data quality management processes, including review and audit processes, biases examination;
d. coverage: representation of multi-disciplinary electronic health data, representativity of population sampled, average timeframe in which a natural person appears in a dataset;
e. information on access and provision: time between the collection of the electronic health data and their addition to the dataset, time to provide electronic health data following electronic health data access application approval;
f. information on data enrichments: merging and adding data to an existing dataset, including links with other datasets.

The Commission can detail and amend this list via delegated Acts (article 56 EHDSR).

For *all* data holders the following applies:

- contribute to a metadata catalogue. Each dataset shall include information concerning the source, the scope, the main characteristics, nature of electronic health data and conditions for making electronic health data available (41.2 plus 55.1). The metadata catalogue will be published by the national data access body. The Commission shall, by means of implementing acts, set out the minimum information elements data holders are to provide for datasets and their characteristics (55.2 EHDSR)
- the data holders shall put the electronic health data at the disposal of the health data access body within 2 months from receiving the request from the health data access body. In exceptional cases, that period may be extended by the health data access body for an additional period of 2 months (41.4 EHDSR) *if*:
- a data permit has been issued to the data user (such a research organisation) comprising the data in question (46.4 and 45.2.b EHDSR combined).
- the data holder shall cooperate in good faith with the health data access bodies, where relevant (41.1).
- data access bodies can fine data holders when they withhold the electronic health data from health data access bodies with the manifest intention of obstructing the use of electronic health data, or do not respect the deadlines set out by the body (43.5 EHDSR).

7.4 Follow-up of the EHDSR in the (grey) literature

A host of comments have followed the proposal. For a general overview also taking the Acts and proposals mentioned earlier and focusing on the consent issue see, Shabani and Yilmaz 2022. In general, the EHDSR is welcomed as a game changer but then the comments follow. The European Patient Federation stressed that that reuse of health data should be based on opt-out (instead of no consent modality at all as seems to be implied in the EHDSR) and that the public interest criterion for which health data may be re-used should be strengthened (European Patients Forum, 2022). A very broad assembly of stakeholders issued a more general comment, stressing the need for stakeholder

involvement in the further development of the EHDS, better alignment with the other Acts, as described earlier, a clearer definition of the legal basis for secondary use, applicable throughout Europe, and a more tailored definition of health data which could be re-used (EPIC, n.d.).

MedTech Europe, while being part of the broad assembly of stakeholders, issued a more specific statement stressing the need to align the EHDS with all other Acts and proposals and not to create new sometimes conflicting and overlapping administrative obligations. More specifically related the scope of this paper, it was mentioned (MedTech Europe, 2023):

- we recommend to only include validated and actionable output data of the products rather than device-generated data and raw data sets which often are too large and granular and therefore not fit for purpose.
- At the same time, it was acknowledged that prescription data as well as post-market and surveillance data should be part of the EHDS secondary use system.

8 AI Act

8.1 Introduction

The European Commission released the proposed Regulation on Artificial Intelligence (the EU AI Act) on 21 April 2021.[17] In June the European Parliament (EP) released its amended version of the AI Act (European Parliament, 2023). Now, the so called 'trialogue' will start where the European Commission, the EP and the Council of Ministers must agree upon a compromise text.

The AI Act aims to address the ethical and legal concerns associated with AI technologies and their impact on individuals and society. It intends to provide a harmonized framework for AI regulation across the EU member states. It sets out rules and requirements for the development, deployment, and use of AI systems, with a focus on ensuring transparency, accountability, and respect for fundamental rights (see the various Recitals in the AI Act and (European Commission, 2021)). The AI Act revolves around a classification system designed to evaluate the potential risks posed by an AI technology to the health, safety, and fundamental rights of individuals. This framework categorizes AI systems into four tiers of risk: unacceptable, high, limited, and minimal.[18]

Medical devices that are AI-systems or consist of AI-systems as defined in article 3, under 1, AI Act fall under the scope of the AI Act. This means that additional rules apply to the medical devices containing artificial intelligence from the draft AI Act on top of the requirements in the MDR and the IVDR. Medical devices / IVDs in scope of the MDR and IVDR will in basically all cases constitute high-risk AI systems in the meaning of the AI Act. This means that they will be subject to the requirements of, among other things:

- Risk management system (similar to MDR and IVDR) (article 9)
- Data governance and data management practices (similar to MDR/IVDR and GDPR) (article 10)
- Technical documentation (similar to MDR/IVDR article 11)
- Logging capabilities (similar to GDPR) (article 12)
- Transparency and information to users (similar to GDPR) (article 13)
- Human oversight requirements (similar to MDR/IVDR) (article 14)
- Accuracy, robustness and cybersecurity (similar to MDR/IVDR and GDPR) (article 15)

[17] European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM/2021/206 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[18] The EU's Artificial Intelligence Act, explained | World Economic Forum (weforum.org)

- Obligations very much like article 10 MDR/IVDR (device manufacturer obligations plus QMS) (articles 16 and 17)
- Economic operator requirements (similar to MDR/IVDR) (articles 25 to 28)
- MDR and IVDR PMS systems must integrate AI Act PMS elements (Article 61 (4))

AI systems necessitate a significant volume of precise data for their construction and training (Bak et al., 2022). For this purpose, access to raw medical device data will likely be used in conjunction with clinical outcomes. In domains like healthcare services and medical research, this data can often contain sensitive medical information. Consequently, ensuring data protection becomes a crucial legal concern. Regarding the protection of personal data, the AI Act intersects with the GDPR and reinforces its principles. The AI Act emphasizes the importance of protecting individuals' personal data when AI systems are involved. It requires that AI systems processing personal data comply with the GDPR and other relevant data protection laws.

8.2 Relation of the AI Act with the GDPR

The relationship between the GDPR and the proposed AI Act is multifaceted. They have the potential to complement each other by sharing common definitions and provisions concerning data protection, including aspects like biometrics and special categories of data. The AI Act explicitly states that it does not serve as a legal basis for processing personal data, including special categories of personal data. For that the GDPR applies.

The GDPR imposes restrictions on the processing of health data and prohibits the sole reliance on automated decision-making (ADM) unless specific circumstances apply (article 22 GDPR). Therefore, the utilization of health data in AI systems for ADM is said to encounter notable legal limitations (Meszaros et al., 2022). We should add, only if the device would act -as were- on its own such as the ICD's mentioned earlier. And in that case, it could be argued that this 'ADM' is based on consent which is of the legitimate bases for ADM (article 22.2 under c GDPR).

In most cases, even if AI is running on the background, the device will inform the physician or patient. That is one aspect of responsible AI, in the end humans decide (Muller et al., 2021) as is also reflected in the AI Act.

The AI Act proposal of the EP reinforces data protection in the AI context, stating in article 10.5 of the AI Act:

- To the extent that it is strictly necessary for the purposes of ensuring negative bias detection and correction in relation to the high-risk AI systems, the providers of such systems may exceptionally process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving. In particular, all the following conditions shall apply in order for this processing to occur: (a) the bias detection and correction cannot be effectively fulfilled by processing synthetic or anonymised data; (b) the data are pseudonymised; (c) the provider takes appropriate technical and organisational measures to ensure that the data processed for the purpose of this paragraph are secured, protected, subject to suitable safeguards and only authorised persons have access to those data with appropriate confidentiality obligations; (d) the data processed for the purpose of this paragraph are not to be transmitted, transferred or otherwise accessed by other parties; (e) the data processed for the purpose of this paragraph are protected by means of appropriate technical and organisational measures and deleted once the bias has been corrected or the personal data has reached the end of its retention period; (f) effective and appropriate measures are in place to ensure availability, security and resilience of processing systems and services against technical or physical incidents; (g) effective and appropriate measures are in place to ensure physical security of locations where the data are stored and processed, internal IT and IT security governance and management, certification of processes and products; Providers having recourse to this provision

shall draw up documentation explaining why the processing of special categories of personal data was necessary to detect and correct biases.

That is quite a lot and stricter than the GDPR. Only one goal seems to be legitimate basis, namely to detect and correct biases, to process sensitive data while in the GDPR there are many (see article 9.2 GDPR).

At the same time the EP version adds an important research exemption to the AI proposal, being:

- This Regulation should help in supporting research and innovation and should not undermine research and development activity and respect freedom of scientific research. It is therefore necessary to exclude from its scope AI systems specifically developed for the sole purpose of scientific research and development and to ensure that the Regulation does not otherwise affect scientific research and development activity on AI systems. Under all circumstances, any research and development activity should be carried out in accordance with the Charter, Union law as well as the national law.

The Union law is the GDPR and national law. In conjunction with the national regulations using the latitude offered by the GDPR for research exemptions (Hansen et al., 2021; van Veen, 2018) research with (raw) device data using AI is a real possibility, whether in the development stage of the device or later when the device is functioning already. In the latter case the research could use the raw device data to gain insights in what the algorithm does not translate into actions or human intelligible data.

8.3 Access to raw device data under the AI Act

The AI Act does not give any concrete tools for such access. However, via the angle of research such access would be possible. Obviously, the research organisation would need to cooperate with the 'provider' of the AI system in the sense of article 3.2 AI Act which in this case would also be the manufacturer of the device.

9 Conclusions

Opening up raw device data for research is an underexplored subject. We did not find any literature which directly relates to this issue. Yet, given the digital revolution in healthcare and the necessity to learn from experience as a learning health system aims for, the issue will become more and more important. Opening up raw device data is also important to control and reassess the algorithms which translate the raw data into human intelligible data as designed by the manufacturer.

The DA might be most the most logical legislative instrument to open-up raw device data for research. However, our analysis was based on the proposal by the European Commission. The European Parliament and the European Council must together the Commission reach a compromise on a common text, the so called 'trialogue'. The published texts of the Parliament and that of the Council seem to be more restrictive in opening up data which the data holders consider proprietary information (Margoni et al., 2023). We will only know what the DA specifically entails after the final agreed text has been published. Also, the EHDS proposal could influence the opening up of raw device data. However, it might take quite a while before the final text of the EHDS will be published.

In addition, to reassess the very preliminary conclusions of this paper in the light of the final legislative texts, further research is needed into:

- Use cases where researchers need access to raw device data based on a more or less concrete research protocol which clarifies the public interest if such access would be given.
- The 'mapping' of the terms used in the various legislation (data controller, data processor, holder, data user, third party) against how devices are actually made available in a solidarity-

based healthcare system when the patient needs the device for the disease management as supervised by the physician, and the actual data flows in this respect.

- Further exploration of the issue in the light of the fundamental rights in the European Charter and what will ultimately serve the public good and patients in the EU healthcare systems most (for a first exploration, not taking into account that of patients as that was not the subject of the paper by Margoni and the references there (Margoni et al., 2023).

In addition to this further research which would explore issues beyond IDEA-FAST, the analyses here also have direct bearing to the final governance document for the data generated in IDEA-FAST.

10 References

Bak, M., Madai, V. I., Fritzsche, M.-C., Mayrhofer, M. Th., & McLennan, S. (2022). You Can't Have AI Both Ways: Balancing Health Data Privacy and Access Fairly. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/articles/10.3389/fgene.2022.929453

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, *26*(7), 931–936. https://doi.org/10.1038/s41431-018-0160-0

Cilliers, L. (2020). Wearable devices in healthcare: Privacy and information security issues. *Health Information Management Journal*, *49*(2–3), 150–156. https://doi.org/10.1177/1833358319851684

Coventry, L., & Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas*, *113*, 48–52. https://doi.org/10.1016/j.maturitas.2018.04.008

Dahi, A., & Compagnucci, M. C. (2022). Device manufacturers as controllers – Expanding the concept of 'controllership' in the GDPR. *Computer Law & Security Review*, *47*, 105762. https://doi.org/10.1016/j.clsr.2022.105762

Directorate-General for Research and Innovation (European Commission), & Eechoud, M. van. (2022). *Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/71619

Doshi, P., Goodman, S. N., & Ioannidis, J. P. A. (2013). Raw data from clinical trials: Within reach? *Trends in Pharmacological Sciences*, *34*(12), 645–647. https://doi.org/10.1016/j.tips.2013.10.006

EMA. (2022, July 11). *EMA launches pilot project on analysis of raw data from clinical trials* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/news/ema-launches-pilot-project-analysis-raw-data-clinical-trials

EPIC. (n.d.). Joint Statement: Stakeholders' consensus response to the proposed European Health Data Space. *DIGITALEUROPE*. Retrieved 24 November 2022, from https://www.digitaleurope.org/news/joint-statement-stakeholders-consensus-response-to-the-proposed-european-health-data-space/

European Commission. (2020). *Shaping Europe's Digital Future*. Publications Office of the European Union. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/shaping-europes-digital-future_en

European Commission. (2021). *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*.

European Data Protection Board & EDPB. (2020). *Guidelines 05/2020 on consent under Regulation 2016/679 | European Data Protection Board*. EDPB. https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en
European Parliament. (2023, June). *Texts adopted—Artificial Intelligence Act—Wednesday, 14 June 2023*. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
European Patients Forum. (2022). *EPF's recommendations on the European Health data Space (EHDS)*.
Finck, M. (2021). Cobwebs of control: The two imaginations of the data controller in EU law. *International Data Privacy Law*, *ipab017*. https://doi.org/10.1093/idpl/ipab017
Fleming, N. (2023, May 9). *Proposed EU data laws leave researchers out in the cold* [Nature Index]. Nature. https://www.nature.com/articles/d41586-023-01572-2
Foley, T., & Vale, L. (n.d.). A framework for understanding, designing, developing and evaluating learning health systems. *Learning Health Systems*, *n/a*(n/a), e10315. https://doi.org/10.1002/lrh2.10315
Gitelman, L. (Ed.). (2013). *'Raw data' is an oxymoron*. The MIT Press.
Groos, D., & Veen, E.-B. van. (2020). Anonymised Data and the Rule of Law. *European Data Protection Law Review*, *6*(4), 498–508. https://doi.org/10.21552/edpl/2020/4/6
Grossmann, C., McGinnis, J. M., & Powers, B. (2011). *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. National Academies Press.
Hansen, W. J., Wilson, P., Verhoeven, E., Kroneman, M., Verheij, R., & van Veen, E.-B. (2021). *Assessment of the EU Member States' rules on health data in the light of GDPR* (p. 262). European Commission. https://ec.europa.eu/newsroom/sante/items/702120/en
Hugenholtz, P. B. (2018). Against 'data property'. In H. Ullrich, P. Drahos, & G. Ghidini (Eds.), *Kritika: Essays on Intellectual Property*. Edward Elgar Publishing. https://doi.org/10.4337/9781788971164.00010
Jiang, D., & Shi, G. (2021). Research on Data Security and Privacy Protection of Wearable Equipment in Healthcare. *Journal of Healthcare Engineering*, *2021*, 6656204. https://doi.org/10.1155/2021/6656204
Kapoor, A., Vora, A., & Yadav, R. (2019). Cardiac devices and cyber attacks: How far are they real? How to overcome? *Indian Heart Journal*, *71*(6), 427–430. https://doi.org/10.1016/j.ihj.2020.02.001
Kiezen voor houdbare zorg. Mensen, middelen en maatschappelijk draagvlak WRR-Rapport 104. (n.d.). 414.
Kotecha, D., Asselbergs, F. W., Achenbach, S., Anker, S. D., Atar, D., Baigent, C., Banerjee, A., Beger, B., Brobert, G., Casadei, B., Ceccarelli, C., Cowie, M. R., Crea, F., Cronin, M., Denaxas, S., Derix, A., Fitzsimons, D., Fredriksson, M., Gale, C. P., … Grobbee, D. E. (2022). CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *BMJ*, e069048. https://doi.org/10.1136/bmj-2021-069048
Leclercq, C., Witt, H., Hindricks, G., Katra, R. P., Albert, D., Belliger, A., Cowie, M. R., Deneke, T., Friedman, P., Haschemi, M., Lobban, T., Lordereau, I., McConnell, M. V., Rapallini, L., Samset, E., Turakhia, M. P., Singh, J. P., Svennberg, E., Wadhwa, M., & Weidinger, F. (2022). Wearables, telemedicine, and artificial intelligence in arrhythmias and heart failure: Proceedings of the European Society of Cardiology Cardiovascular Round Table. *EP Europace*, *24*(9), 1372–1383. https://doi.org/10.1093/europace/euac052
Margoni, T., Ducuing, C., & Schirru, L. (2023). Data Property, Data Governance and Common European Data Spaces. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4428364
MedTech Europe. (2023). *MedTech Europe's position on the proposed European Health Data Space Regulation*. MedTech Europe.

Meszaros, J., Minari, J., & Huys, I. (2022). The future regulation of artificial intelligence systems in healthcare services and medical research in the European Union. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/articles/10.3389/fgene.2022.927721

Mittelstadt, B. (2017). Ethics of the health-related internet of things: A narrative review. *Ethics and Information Technology*, *19*(3), 157–175. https://doi.org/10.1007/s10676-017-9426-4

Muller, H., Mayrhofer, M. T., Veen, E.-B. V., & Holzinger, A. (2021). The Ten Commandments of Ethical Medical AI. *Computer*, *54*(07), 119–123. https://doi.org/10.1109/MC.2021.3074263

Rosenberg, Lawrence. (2023). *Patients Matter Most. How healthcare is becoming personal again*. Forbes Books.

Shabani, M., Vears, D., & Borry, P. (2018). Raw Genomic Data: Storage, Access, and Sharing. *Trends in Genetics*, *34*(1), 8–10. https://doi.org/10.1016/j.tig.2017.10.004

Shabani, M., & Yilmaz, S. (2022). Lawfulness in secondary use of health data: Interplay between three regulatory frameworks of GDPR, DGA & EHDS. *Technology and Regulation*, *2022*, 128–134. https://doi.org/10.26116/techreg.2022.013

van Veen, E.-B. (2018). Observational health research in Europe: Understanding the General Data Protection Regulation and underlying debate. *European Journal of Cancer*, *104*, 70–80. https://doi.org/10.1016/j.ejca.2018.09.032

Vanberg, A. D., & Ünver, M. B. (2017). The right to data portability in the GDPR and EU competition law: Odd couple or dynamic duo? *European Journal of Law and Technology*, *8*(1). https://ejlt.org/index.php/ejlt/article/view/546