EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in silico* models for personalized medicine – EU-STANDS4PM

Harmonization and integration of big data of relevance for personalized medicine into in silico modelling? – Recommendations for technically feasible, and ethico-legal sustainable avenues

Deliverable 3.3

Imprint

Authors

Alphabetical: Katharina Ó Cathaoir[1], Eugenijus Gefenas[2], Mette Hartlev[1], Miranda Mourby[3], Vilma Lukaseviciene[2]

[1] University of Copenhagen, Denmark

[2] Vilnius University, Lithuania

[3] University of Oxford, United Kingdom

Contact information

Mette Hartlev

University of Copenhagen, Faculty of Law, Denmark

Mette.Hartlev@jur.ku.dk

Publisher

EU-STANDS4PM administrative office on behalf of the EU-STANDS4PM consortium, March 2020

Forschungszentrum Jülich GmbH, Project Management Jülich, Germany

Contact: Marc Kirschner (m.kirschner@fz-juelich.de)

Using this content

Table of content

1 Introduction 4
2. Data Protection & Technical Avenues for Data Integration 6
2.1 Terminology 6
2.2 Introduction 6
2.3 Data Governance Act & the Health Data Space 7
2.3.1 Consent & the DGA 8
2.3.2 Collective Governance & the DGA 9
..... 11
2.4 GDPR & Technical Means of Integration 11
2.5. Technical Avenues: A Possible Hierarchy 15
2.5.2 Technical Avenues & Data Control 17
2.6 Recommendations 19
3. Interaction between research ethics guidelines and data protection regulation 20
3.1 Consent misconception and other consequences of a 'dual' meaning of consent 20
3.2 Diverging interpretations of consent: GDPR vs research ethics regulations 22
3.3 Waivers of consent and research condition 25
3.4 Implications for researchers, research institutions, RECs and DPOs 26
3.5 Public trust and the role of consent 27
3.6 Recommendations 28
4. Rights of Patients & Research Subjects - Aspects of harmonisation & integration 29
4.1 Harmonisation and individual rights 29
4.2.1 Research subjects' right to health information 30
4.2.2 Patients' Right to Information 31
4.2.3 Application to harmonisation/integration 31
4.2.3.1 Reflections on Anonymisation of different data sources 34
4.3 Transparency 35
4.4 Patients' Rights to Freedom from Discrimination & Equitable Access to treatment 37
4.5 Recommendations 39
5. Summary of recommendations 40
6. Acknowledgements 43

1 Introduction

This report is the third deliverable of WP3 in the EU-STANDS4PM project. WP3 is responsible for providing guidance on the legal, ethical and policy considerations arising from data integration for *in silico* modelling for personalized medicine. In our first deliverable, we provided a legal and ethical analysis based on the data sources relevant for development of in silico models. The report included:

1. a survey of the international and European data protection and clinical trials regulation as well as legal and ethical regulation relevant for protection of equal access to health, and right to information and self-determination relevant for harmonization and integration of data in in silico modelling, and
2. an assessment of the challenges and opportunities this regulation provides for harmonization and integration of data for in silico modelling.

The aim of our second report and deliverable was to explore *solutions* to the challenges identified in the first report and outline technically feasible as well as legally and ethically sustainable avenues for harmonization/integration of big data of relevance for personalized medicine into *in silico* modelling.

The major challenges identified in our second deliverable and discussed in our first deliverable relates first to the *lack of harmonization of national laws* both within and beyond EU borders. Despite the 'harmonization' of EU data protection law through the GDPR, important differences remain permitted in national law, leading to legal fragmentation. This fragmentation is an obstacle for data integration across borders. Another complication relates to uncertainty regarding the *role of consent for sharing of personal data*; especially the position of consent as both an approval to bodily intervention in clinical care or research (interventional consent) and a permission to processing of personal data (informational consent). The role of consent as a legal basis for processing of personal data has been questioned due to the consent requirement stipulated in the GDPR. If consent in itself cannot justify processing of personal data for in silico modelling, another legal basis is required. Consequently, clarity regarding a proper *legal basis for data sharing* is needed. Finally, we also draw the attention to the *impact patients' rights* may have on data sharing for the development of in silico models. Principles of confidentiality and rights to information, including information regarding secondary findings, need to be taken into consideration when developing in silico models. Issues related to discrimination and protection of vulnerable groups are also concerns.

In our second report we explored and outlined potential *solution*s to these challenges. Some of the solutions can be established within the existing national and European legal and ethical framework. Others will need adjustments or amendments of national or EU-law. In section 2 of the report we looked into potential legal and technical avenues for data integration and considered both, practical models for data integration and outlined possible legal mechanisms, by which such use of data can be legitimized. We also discussed the potential for data integration in the new Data Governance Act, which aims at providing a common framework for data sharing within the EU. Section 3 of the report explored and discussed the role of consent, and possible models for combination of interventional and informational consent solutions. It also addressed the role of consent in sustaining public trust in research and the need of stronger integration of research ethics and data protection mechanisms. Finally, section 4 focused on the importance of and potential models for integrating patients' and data subjects' *privacy and confidentiality rights* as well as *rights to information* in the development of in silico models. It was recommended that rights of patients and research subjects should travel with the data, and that *transparency* for both clinicians and patients should be ensured when developing in silico models. Risks of bias and discrimination should also be addressed.

This third deliverable is an revised and updated version of our second report.The recommendations in the second report were discussed during a virtual stakeholder workshop 19. April 2022, where we received prepared comments from six experts and questions and comments from other participants. We have also updated the report with recent legislative developments. We are grateful for the very valuable comments

we received during the workshop. They have provided important input for our revision and further development of the final recommendations in this report. A special thanks to the commentators senior consultant Catherine Bjerre Collin, Dr. Iur Fruzsina Molnar-Gabor, Professor Barbara Prainsack, Professor Dominique Sprumont, Dr. Edward Dove, and Senior Lecturer Santa Slokenberga. We would also like to thank participants at the EU-STANDS4PM annual meeting 22 May 2022 for valuable input.

2. Data Protection & Technical Avenues for Data Integration

2.1 Terminology

This section explores concepts defined within the General Data Protection Regulation ('GDPR'). It is therefore useful to summarise the terms that will be used throughout:

Data Controller: *an organisation (or, occasionally, an individual) which determines the purposes and manner for which data are processed.*

Joint Data Controller: *where two organisations make a common decision about how & why data will be used, or where they make 'converging' decisions that are dependent on each other (e.g. where one decides what they would like to do with the data, and the other approves this use in granting access).*

Legal Basis: *under Article 6 GDPR, a basis for processing personal data must be identified. For scientific research, this basis is often public interest or consent. An additional basis must be identified if personal data are to be transferred outside the EEA.*

Condition for processing: *in the case of health or genetic data, or any other 'special category' of personal data, as well as a legal basis an additional 'condition' for processing must also be demonstrated. This could be explicit consent, or the scientific research condition which requires safeguards such as pseudonymisation.*

2.2 Introduction

This section considers the technical means by which data from across the EU can be integrated – i.e. either physically combined or rendered sufficiently interoperable, so that a user can access them across multiple sites in response to a particular query.

As other sections touch upon clinical trials data, this section also acknowledges the increasing importance of real-world data for novel therapies and personalised medicine. 'Real World Data' are, in practice, often collected through routine healthcare records at a local or national level. The reliability of 'RWD' as an evidence base can be variable,[1] which can make the use of such data for innovative regulatory practices highly controversial.[2] Nevertheless, RWD can fill in evidential gaps which conventional clinical trials do not address: providing information about populations typically excluded from trials (such as pregnant women[3]), supplementing clinical trial data for rare disease populations, validating surrogate endpoints used in trials,[4] and enabling post-marketing surveillance (particularly for novel therapies[5]).

There are thus key ethical drivers to overcome the limitations in RWD (at least in its raw state) to serve sub-populations who have been 'orphaned' by conventional therapies, or by conventional trial design. This

[1] K Facey et al, 'Real-world evidence to support Payer/HTA decisions about highly innovative technologies in the EU—actions for stakeholders' (2020) 36 International Journal of Technology Assessment in Health Care 4

[2] K Syrett, 'Regulation, innovation and disruption: the European Medicines Agency and adaptive licensing of pharmaceuticals' (2020) 12 Law, Innovation and Technology 2

[3] B George et al, 'Application of physiologically based pharmacokinetic modeling for sertraline dosing recommendations in pregnancy' (2020) 6 npj Systems Biology and Applications 36

[4] J Shafrin et al, 'The value of surrogate endpoints for predicting real-world survival across five cancer types' (2016) 32 Curr. Res Med Opin. 4

[5] European Medicines Agency, 'Draft Guideline on safety and efficacy follow-up and risk management of Advanced Therapy Medicinal Products' 25 January 2018, available from: <https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-safety-efficacy-follow-risk-management-advanced-therapy-medicinal-products-revision_en.pdf>

should, however, be achieved without undue sacrifice of data standards to prove medicinal safety and efficacy.

To this end, the joint HMA-EMA Big Data Taskforce has made as its top recommendation the creation of the 'Data Analysis and Real World Interrogation Network.' The initial ambition is for the platform to include healthcare data of known quality, which could in time be expanded to include registry and claims data.[6] The 'DARWIN' platform is currently building its business case, and we do not yet know its operational details, which are likely to pose regulatory as well as technical challenges.[7]

In anticipation of the launch of this platform in 2023, in order to provide greater regulatory accessibility and acceptability to RWD, this section considers the technical and legal avenues for the integration of health-related data within the EU.

2.3 Data Governance Act & the Health Data Space

While this section provides an overview of the current technical infrastructure & legal mechanisms to integrate data in the EU, it should be noted that these legal mechanisms are due to be significantly updated.

In our first-year report, we highlighted the differences between national implementations of the GDPR, and the strain this places on attempts to jointly govern health-related data sourced from different EU countries (of which, more below in section 2.4). This has since been acknowledged in a study published by the European Commission in February 2021:

It is clear from the views shared in the workshops and by country correspondents to the legal and technical survey that while the GDPR is a much appreciated piece of legislation, variation in interpretation of the law and national level legislation linked to its implementation have led to a fragmented approach which makes cross-border cooperation for care provision, healthcare system administration or research difficult.[8]

The national implementations of the GDPR culminate in what has been described as a fragmented landscape, including fault-lines along the question of informed consent as a legal basis for processing data, which inhibits data sharing for health research.[9] This creates confusion and uncertainty for data protection officers, and members of research ethics committees, who are tasked with ensuring the lawful & ethical use of any data shared for research. At the same time, individual citizens in member state countries report a willingness to share their data for health-related research, but do not feel there are adequate tools for them to donate their data.[10]

The EU has now passed a new Data Governance Act ('DGA')[11] in May 2022. The Regulation aims to create a framework for sharing personal and non-personal data from a number of sectors—within a number of 'European data spaces'—in a way which is compatible with data protection and intellectual property rights.

---

[6] Heads of Medicines Agencies—European Medicines Agency, 'HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data Driven Regulation' (March 2020) available from: < https://www.ema.europa.eu/en/documents/annual-report/2019-annual-report-european-medicines-agency_en.pdf>

[7] B Scrace, 'EU: On the Origin of Big Data – is D.A.R.W.I.N the future?' Allen & Overy Digital Hub, 10 February 2020, available from: <https://aodigitalhub.com/2020/02/10/eu-on-the-origin-of-big-data-is-d-a-r-w-i-n-the-future/>

[8] European Commission, DG Health and Food Safety, 'Assessment of the EU Member States' rules on health data in the light of the GDPR' 12 February 2021, available from: 'https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf' accessed 12 March 2021

[9] Olga Tzortzatou, Santa Slokenberga, Jane Reichel et al, 'Biobanking Across Europe Post-GDPR: A Deliberately Fragmented Landscape' in: Santa Slokenberga, Olga Tzortzatou & Jane Reichel (eds) *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe* (Springer, 2021) Law, Governance and Technology Series, vol 43. https://doi.org/10.1007/978-3-030-49388-2_9

[10] European Commission, 'Summary report of the public consultation on the European strategy for data', 24 July 2020, available from <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data>

[11] European Commission, 'Regulation EU 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and amending Regulation (EU) 2018/1724 (Data Governance Act)' L151/2, hereafter cited as 'DGA.'

As the framework will apply directly across a number of EU states and third countries, less room is allowed for national derogation. The explanatory memorandum states that a horizontal regime that cuts across Member States should be established, but acknowledges that some sector-specific modifications may be required. Deorgation at sector rather than national level is very welcome within research. The international community benefits from modifications for the *type* of processing involved (i.e. secondary uses of health information for research) more than discrepancies stemming from the politics of the country or countries in which the processing takes place.

The mechanisms for data altruism aim at empowering those individuals who wish to share their data for (e.g.) health-related research but do not currently have the means to do so. This sits alongside provisions aimed at companies and public authorities, allowing them to share their data as well.

2.3.1 Consent & the DGA

The role of the individual within the DGA is therefore inconsistent: they are actively involved where they provide information directly, but less so in other contexts. Even though the DGA supplements the GDPR, and claims it will only serve to help individuals exercise their GDPR rights,[12] it is notable that individual consent will play an even more reduced role in some of its provisions than under its contested status within GDPR. Under Article 5 DGA, which governs the use of public sector data, consent will be used where there is no other legal basis available:

Where the re-use of data cannot be granted in accordance with the obligations laid down in paragraphs 3 to 5 and there is no other legal basis for transmitting the data under Regulation (EU) 2016/679, the public sector body shall support re-users in seeking consent of the data subjects.

This makes it clear that consent will be used as a last resort, and suggests a continuing trend away from consent as a legitimating mechanism for the secondary use of data. However, in the context of data altruism, consent is clearly envisaged as very important. Article 25 requires the Commission to develop a uniform European data altruism consent form, and the requirement in 22(3) that this consent should be revocable to a GDPR standard implies consent may be deemed the relevant GDPR basis for processing in this context. As such, the DGA does not eliminate reliance on consent altogether, but circumscribes its use to a context where the individual clearly participates on a voluntary basis.

The development of a uniform consent form by the Commission is a useful opportunity to identify a 'gold standard' for GDPR-compliant consent for the secondary use of data for research (at least in its documented form). Wide consultation across different EU member states—who currently have different attitudes to GDPR consent—is advisable. It is also a welcome opportunity to provide a standard form of consent to transfer of data outside the European Economic Area under GDPR Article 49(1)(a). This would be particularly helpful for researchers in lower to middle-income countries, who are often excluded from personal data access at present.[13] The prospect of providing high-level clarity on a form of consent sufficient for Articles 6 and 9 of the GDPR under Article 25 DGA is a hopeful prospect for an aspect of GDPR compliance which has proven complex and controversial to date. Transfer of health-related personal data outside the EEA for research purposes has also been the subject of much uncertainty within the research community (particularly to jurisdictions which have not been granted adequacy status by the European Commission). While the GDPR positions consent as an option of last resort for international transfers, the question of a standard consent form for third-country transfers should at least be considered.

[12] DGA, Explanatory Memorandum, para 1

[13] S Slokenberga & J Reichel, 'EU data transfer rules and African legal realities: is data exchange for biobank research realistic?' (2019) 9 International Data Privacy Law 1

Recommendation 1: Maximise the impact of the template consent form

1.1 During the grace period for the application of the Data Governance Act, the European Commission should consult widely among stakeholders across Member States to achieve clarity & consensus as to a standard form of GDPR compliant consent, with consideration given as to whether this consent form could also legitimate transfers to third countries under Article 49 GDPR.

1.2 Data intermediaries and data altruism organisations established under the DGA should explore the use of dynamic consent to ensure that data subjects are kept up to date about research uses of their data, and the identity of any other data controllers who use their information for research.

1.3 Those involved in the designing of the DGA consent form, European Health Data Space and BBMRI-ERIC Code of Conduct for Health Research should be brought into contact to ensure their requirements align and do not perpetuate the uncertainty and confusion within the research community.

In addition to the recommendations made above, we also note that harmonisation efforts within the research sector will themselves need to be brought into aligment. For example, those working on the BBMR-ERIC Code of Conduct should have some overlap with the groups engaged in designing the European Health Data Space. These groups need to have common positions on—for example—the role of consent in processing health data for research to prevent inconsistency and confusion among researchers attempting to comply with the GDPR.

2.3.2 Collective Governance & the DGA

Another interesting aspect of the DGA is its regulation of non-personal data—i.e. data which does not identify natural, living people through any means reasonably likely to be used per the GDPR's definition of personal data. Therefore, data are defined as:

'*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 audiovisual recording*.'[14]

The GDPR is broader in the sense that it also applies to non-digital information which identifies individuals;[15] reflecting the DGA's emphasis on the digital single market. Otherwise, the DGA has a wider scope in its regulation of digital information, including non-personal data in a bid to reflect the trade secrets and intellectual property rights of 'legal persons.' Theoretically, the application of the DGA to non-personal data could help divorce data-interests from the rights of individuals, and towards a more group-based, or even societal, view of the implications of data processing.[16]

It also takes a small step towards a more collective approach by allowing individuals & Small and Medium Enterprises to join Cooperatives in respect of their (non)personal data.[17] However, these cooperatives will only represent the interests of their members, and not of society as a whole in their scrutiny of data processing. This may mean the interests of more marginalised people, who are less likely to come forward for these governance models, can be left out of the equation. More work would need to be done (outside

[14] DGA, Article 2(1)
[15] As long as it is processed within a filing system, GDPR Article 2(1)
[16] See, for example, L Floridi, 'Open Data, Data Protection, and Group Privacy' (2014) 27 Philosophy & Technology 1
[17] DGA, Article 9

the text of the legislation itself) to ensure that any data cooperatives under the DGA (like Data Trusts) do not end up serving only the least vulnerable members of society.[18]

The DGA should help support the European Health Data Space ('EHDS'): a framework for the pooling & secondary use of health-related data that encompasses both governance rules and legal infrastructure.[19] The remainder of this section therefore reviews the governance rules and technical infrastructure currently available for the integration of health-related data, even if these will soon be supplemented by the DGA and EHDS.

When the recommendations in this report were discussed at a Legal & Ethical expert workshop on 19 April 2022, participants were particularly concerned about the idea of 'data donation.' While some considered the use of the phrase appropriate to describe the actions of data subjecs contributing to data altruism organisations, others stressed the distinction between altruism and donation. While altruism implies the gifting of one's information for purposes other than one's own personal gain, the idea of a 'donation' implies that the gift has been surrendered and the giver has no ongoing control. We therefore recommend that organisations who facilitate the altruistic use of information (including for research) make it clear that a data subject is not 'donating' their data in the sense of giving up all say in its use, and that they can withdraw their consent wholesale (or at a granular level) at any time without suffering detriment. This revocability should be made clear in the consent form, and be made easy for individuals to put into practice through the tools offered by the data-sharing platform.

Clarity about the nature of the relationships these new structures will create will be necessary beyond data altruism organisations. For example, the DGA acknowledges in Recital 31 that data cooperatives can help individuals in making choices about their information, but that rights under the GDPR are personal and cannot be waived (or, it follows, pooled). Similarly, therefore, it needs to be clear to a data subject joining a cooperative that they are not simply 'handing over' their information, but can (and, potentially, should) continue to shape how it is used by giving and withdrawing their consent as they see fit over time.

[18] S Delacroix and N D Lawrence, 'Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance' 9 International Data Privacy Law 4

[19] European Commission, 'European Health Data Space', available from: https://ec.europa.eu/health/ehealth/dataspace_en accessed 12 March 2021

Recommendation 2: **Make data collectives as transparent and inclusive as possible**

2.1 Investment should be made at an EU and Member State level to ensure that any Data Cooperatives (or Data Altruism organisations) that contribute to the European Health Data Space do not only serve less vulnerable members of society, but are accessible for a wide range of groups and demographics.

2.2 Data altruism organisations should make it clear that participatnts are not 'data donors', in the conventional sense of donation, and that they retain the right to control how their information is used.

2.3 Data Cooperatives should also be clear with their members about the limitations of their remit, and that members still need to exercise their data subject rights for themselves.

2.4 GDPR & Technical Means of Integration

This sub-section considers the practical models for data integration, and the GDPR mechanisms by which such use of the data can be legitimised. In reality, there are many legal obligations that need to be satisfied in order for data to be shared lawfully, including:

- Confidentiality (medical or otherwise);
- Human rights, including the right to privacy;
- Intellectual property rights, including trade secrets, copyright and database rights.

The multitude legal obligations at play is the very reason theData Governance Act attempts to provide an overarching framework to navigate all of these rights. As this section focuses on data protection, however, the following table only provides an overview of how different infrastructures could be used in combination with different GDPR provisions.

Where data are integrated beyond the parameters of the EU & the EEA for more globally accessible international research, this is a complex prospect in and of itself (even before other laws come into play). A relevant GDPR basis & condition for processing must be selected, as well as an appropriate basis for the transfer to the third country (NB even if data are only made available to someone outside the EEA, and not physically moved or copied, this still counts as a transfer). Table 1 sets out a combination of:

- Technical or infrastructural model of integration;
- The GDPR bases & conditions for processing personal data
- The GDPR bases for transferring data outside of the EEA (where applicable).

Some combination of the three will be necessary to integrate personal data for in silico modelling in personalised medicine. While effort has been made to propose GDPR bases/conditions suitable for each integration model, it should be stressed that the table is for illustration purposes only. Multiple bases for processing/ transfer could be used for each model, depending on the particular circumstances in which the model is used.[20]

[20] As noted later, much will depend on the nature of the data controller-subject relationship, and the country in which the controller operates.

Table 1: The table below is illustrative of the combination of technical and legal avenues for integrating data, but not exhaustive. For example, broad consent for ethical and confidentiality purposes may be used within a number of data sharing models (e.g. remote access).

Technical Model	GDPR Basis/ Condition	Third Country Transfer Mechanism
Remote Access A platform through which data are analysed on the host's server (this may be a private cloud,[21] meaning there is overlap with Cloud Based Access). *Example:* The IMI-funded DIRECT project stores project data on a single server (a 'private cloud'), with data access determined by a committee representing different data contributors.[22] *Limitations:* *How the data can be processed for analysis may be limited, depending on the platform functionality. This could also be strength, however, where the data controller wishes to maintain control. *Joint data control between different contributors difficult due to disparities in GDPR implementation across EU member states.[23]	**The Public Interest Basis** A legal basis for processing under Article 6 GDPR must be satisfied before a condition for special category data can be considered (see below). Article 6(1)(e)—processing which is necessary for a task carried out in the public interest—is likely to be one such basis available for scientific research. A task in the public interest also requires a corresponding basis in EU or member state law—the same legal power (e.g. a university's charter to conduct research). *Limitations:* *Ideas of the public interest are vague & often ill-specified; *Can continue to apply if data subject consent is withdrawn (this may be an advantage in some ways, but can raise ethical concerns); *National laws setting out 'public interest' bases may differ, and not intersect perfectly.	**Adequacy Decision** An adequacy decision adopted by the European Commission under Article 45 GDPR in respect of a third country is the preferred basis for transfer, where available. *Limitations:* *Relatively small number of decisions compared to the number of non-EU countries; *Slow process for adoption; *Only available for a minority of countries; *Are kept under review, and could be withdrawn if the country's laws change significantly.
Cloud-Based Access A means by which large volumes of data can be analysed internationally without the need for local download.[24] *Example:*	**The Research Condition** Once an Article 6 basis is satisfied (see above) an Article 9 condition must also be fulfilled if personal data are genetic or health-related. Such 'special category' data can be processed on the scientific research	**Code of Conduct** If mutually agreed and sanctioned by the European Commission, an international Code of Conduct could provide an infrastructure for sharing

[21] H Teare et al, 'The governance structure for data access in the DIRECT consortium: an innovative medicines initiative (IMI) project' (2018) 14 Life Sciences, Society and Policy 20

[22] H Teare et al, 'The governance structure for data access in the DIRECT consortium: an innovative medicines initiative (IMI) project' (2018) 14 Life Sciences, Society and Policy 20

[23] F Molnar-Gabor & J Korbel, 'Genomic data sharing in Europe is stumbling—Could a code of conduct prevent its fall?' (2020) 12 EMBO Molecular Medicine, available from <https://doi.org/10.15252/emmm.201911421>

[24] P Campbell et al, 'Pan-cancer analysis of whole genomes' (2020) 578 Nature 82

The Personal Genomes Project offers cloud-based access.[25] *Limitations:* *Has become cheaper and have transformed genetics, but still requires funding. *Can rely on data pseudonymisation for access—ambiguous as to whether this results in anonymity or if data remain personal.	condition (Article 9(2)(j) GDPR) as long as: 1) There is an accompanying basis in EU/member state law; 2) Safeguards are in place. Pseudonymisation is named as a relevant safeguard, under which third parties are prevented from identifying individuals, but the data controller can still relate information to them and enforce their rights. *Limitations:* *As with 'public interest,' the parameters of scientific research can be ambiguous (e.g. is a strict hypothesis needed?). There may also be differences in the national laws founding the public interest.	data across national jurisdictions.[26] *Limitations:* *A Code could not smooth over the differences in national interpretations of the GDPR.[27] It remains to be seen whether the DGA can provide a more suitable framework.
Controlled access Data may be physically transferred, subject to conditions (often contractual).[28] This can help parties identify their obligations as data controllers. *Example:* The harmonised data access agreement developed by Work Package 4 of EU-STANDS4PM,[29] although this can be deployed in conjunction with other models. *Limitations:* *Standard conditions can be inflexible; *Bespoke conditions can be time-intensive to negotiate;	**Manifest Publicity Condition** Article 9(2)(e) also provides for special category data to be processed where they have 'manifestly been made public' by the data subject. This could theoretically be via a research data platform. This might operate as an alternative basis to consent, where the GDPR standard cannot be satisfied.[30] *Limitations:* *This condition for processing has been under-explored, but its application in practice may be limited. *Scant guidance is available as to when it is sufficiently clear that the data have been made 'manifestly' public by the data subject, although their request for an intermediary (e.g. research data platform) to make their data available to	**Standard contractual clauses** A basis on which data can be transferred outside the territorial application of the GDPR. Would be part of a DAA where data are transferred outside the EU. These clauses are still available following the European Court of Justice's decision in *Schrems II.* *Limitations:* Cannot be adapted to the needs of the individual parties and must be used wholesale. They also require bilateral time-limited contractual relationships, so it is a

[25] O Chervova et al, 'The Personal Genome Project-UK, an open access resource of human multi-omics data' (2019) 6 Scientific Data 257
[26] M Phillips et al, 'Genomics: data sharing needs an international code of conduct' (2020) 578 Nature 31
[27] F Molnar-Gábor & J Kobel, 'Genomic data sharing in Europe is stumbling-Could a code of conduct prevent its fall?' (2020) 12 EMBO Molecular Medicine 3
[28] I Lappalainen, 'The European Genome-phenome Archive of human data consented for biomedical research.' (2015) 47 Nature Genetics 692
[29] Available from https://www.eu-stands4pm.eu/publications
[30] E Dove and J Chen, 'What does it mean for a data subject to make their personal data 'manifestly public'? An analysis of GDPR Article 9(2)(e)' (2021) International Data Privacy Law (forthcoming) available from: https://doi.org/10.1093/idpl/ipab005

*Trust is required for the parties to take on joint liability	others for certain purposes *might* qualify.	'fragmented' form of data integration.
Federated access Data are analysed locally and only computational results made available.[31] It is possible that the results of the local analysis are sufficiently aggregated that they will not be personal data. However, a GDPR basis/ condition will still be required for the analysis itself, where this uses personal data, even if personal information is not provided to a third party. *Example:* ELIXIR has a 'federated ecosystem of interoperable services.'[32] *Limitations:* *Lower risk, although still some data security issues.[33] *Need to ensure the technical infrastructure is available at each local site.[34] *This may present a bar for less well-resourced centres from acting as 'nodes,' raising wider questions of access to medical innovation across the EU.	**Consent (Condition & Basis)** Consent is both a legal basis for processing under Article 6 and a condition for processing special category data under Article 9 GDPR**.** Consent must be sufficiently granular and revocable to satisfy the requirements of consent as a basis & condition for processing under the GDPR, as long as there are no concerns about power dynamics, or that the consent can be freely given/ withheld without detriment. Dynamic Consent, a digital consent model under which data subjects can modify their consent preferences at any time, and receive a record of how their information has been accessed and used,[35] might be more suitable when data remain within their original local setting. There is some evidence which indicates individuals would want a high-level of control over their health-related data,[36] and so an advantage of this model is to help provide this individual control. *Limitations:* *Not appropriate where consent is too broad, or not capable of revocation without detriment; *If consent is too granular it can be labour-intensive to implement and bring about 'consent fatigue.'	**Consent** Explicit consent to transfer outside the EEA, pursuant to Article 49(1)(a) could also form part of the Dynamic Consent within federated data platform(s)—this would only be necessary where results of local analyses are sufficiently granular to identify individuals, however. *Limitations:* *Consent is only available as an option if an adequacy decision is not in place, and standard contractual clauses or approved code of are not available. It is therefore low on the GDPR's hierarchy of third country transfer options.

[31] G.A Kaissis et al, 'Secure, privacy-preserving and federated machine learning in medical imaging' (2020) 2 Nature Machine Intelligence 305

[32] ELIXIR 'ELIXIR Federated Human Data (2019-21)' available from: <https://elixir-europe.org/about-us/commissioned-services/federated-human-data> accessed 16 March 2021

[33] Georgios A. Kaissis et al, 'Secure, privacy-preserving and federated machine learning in medical imaging' (2020) 2 Nature Machine Intelligence 305

[34] N Rieke et al, 'The future of digital health with federated learning' (2020) 3 npj Digital Medicine 119

[35] H Teare et al, 'The RUDY study: using digital technologies to enable a research partnership' (2017) 25 European Journal of Human Genetics 816

[36] S Courbier, R Dimond & Virginie Bros-Facer, 'Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations' (2019) 14 Orphanet Journal of Rare Diseases 175

Synthetic data	**No GDPR Basis/ Condition**	**No Basis Needed**
Information which does not relate to any real people recreates the characteristics of real-world personal data.[37] *Example:* The UK's Medicines and Healthcare products Regulatory Agency has a synthetic dataset to assist COVID-19 and cardiovascular research.[38] The European Data Incubator creates synthetic datasets from random noise.[39] *Limitations:* Can be labour-intensive to create & update. There is a risk that it may inadvertently identify or misrepresent individuals from the source data.	If data do not identify individuals, no basis or condition under the GDPR is needed Broad consent may be appropriate in these lower-risk scenarios. Where personal data are used to generate synthetic data—but there is no reasonable risk that an individual will be identifiable from the synthetic information—broad consent to the generation of synthetic data, which would then be used for a wide range of purposes may suffice. The broad consent is useful for medical confidentiality purposes, as well as demonstrating GDPR transparency. *Limitations:* Broad consent may not be granular or revocable enough to function as a GDPR basis/ condition (e.g. once the synthetic data has been generated, the consent cannot be revoked). Another basis/ condition may be needed for personal data processing.	Once the synthetic data are created, they should not be classified as personal data and no basis for transfer outside the EEA should be necessary.

As indicated above, the GDPR provides its own hierarchy of options for transferring data to a third country (adequacy decision -> standard contractual clauses/ code of conduct -> consent). It is less clear from the text of the GDPR which *technical* means of integration should be prioritised under its provisions. The next subsection therefore considers these options in more detail, and which could be said to be optimal for GDPR purposes (particularly with reference to data minimisation & purpose limitation).

2.5. Technical Avenues: A Possible Hierarchy

As a starting point, it is worth observing that the above options involve differing levels of data access, meaning differences in the level of control over the means by and purposes for which data are used. This is key for two key principles under the GDPR:

- Data minimisation: whereby data should be limited to the purposes for which they are processed (Article 5(1)(c));

[37] See, for example, the Horizon 2020 project YDATA, focused on making synthetic data safe for sharing with third parties: < https://edincubator.eu/edi-call2-s028/> accessed 15 December 2020

[38] MHRA, 'New synthetic datasets to assist COVID-19 and cardiovascular research' (29 July 2020) available from <https://www.gov.uk/government/news/new-synthetic-datasets-to-assist-covid-19-and-cardiovascular-research> accessed 16 March 2021

[39] https://edincubator.eu/edi-call2-s028/ accessed 16 March 2021

- Purpose limitation: whereby data should only be collected for limited purposes, and only re-used for purposes compatible with the original purpose (Article 5(1)(b)).

In other words, the data accessed should be the minimum necessary for the envisaged purpose, and the uses to which the data are put should be carefully managed to ensure they are still compatible with the purpose of collection. The above avenues of integration all allow a level of control over the amount of information disclosed about individuals, and how it is used. However, it is important to recognise that these different means of making information available involve different levels of interference with data subjects' privacy, and different levels of security as to how their information will be used. As a rough guide, a hierarchy could be suggested as follows:

1. Synthetic data, once created, should not pose any risk of identifying individuals (at least theoretically) and therefore it is not personal data, or information which impacts upon individual privacy. In this sense, it is the least intrusive (and most data-minimised) option it no longer requires personal data for its development.
2. Federated queries, in which data never leave their local context, offer the next level of protection, as only aggregate answers to responses are made available to researchers, even if processing of personal data is involved.
3. Remote/cloud-based access may be the next level of interference with privacy and data protection. Researchers have direct access to data, but only via the controller's platform, and so the options as to how they can use the data are technologically managed and recorded.
4. Contractually controlled access—in which data are physically transferred and downloaded—offer users the greatest degree of freedom in how they use the data. As personal data may be transferred across national boundaries in the process, compliance with national laws will need to be closely monitored. The parties will be bound by contractual terms as to how they use the information, but this requires a level of trust that they will comply with the contract, rather than technologically managing the situation. This option is most suitable where the research question requires a high degree of access to the data, which makes the above options less suitable.

As a very general proposition, the above list could represent an order in which infrastructure options are prioritised for compliance with data minimisation & the purpose limitation principle. The extent to which the research use for the health-related data justifies a greater level of access should be a guiding principle in selecting access options. For example, scepticism was raised at our Legal & Ethical Stakeholder workshop on 19 April 2022 about the adequacy of synthetic data for developing medical-grade models. It was suggested that such data may be better suited for sandbox-stage hypothesis generation, in which case the use of real patient data (albeit in federated/pseudonymised form) would be justified in latter stages of model training and validation.

As shown in the table 1 above, all of these technical avenues for integration will need to be combined with:

- A GDPR basis for processing any personal data (e.g. consent or public interest)
- An additional condition for processing any health-related or genetic personal data (e.g. scientific research, explicit consent or manifest publicity)
- A basis for transferring personal data outside the EEA, where applicable.

This very general preferential ranking of technical avenues for integration sits alongside the GDPR's hierarchy of options for transfer out of the EEA. We do not propose a hierarchy of GDPR bases/conditions for processing as this will be a very fact-specific determination, depending on (for example) the nature of the data controller's relationship with the subjects, the feasibility of revocable & granular consent, and the national data protection laws & guidance to which the data controller is subject. It may be necessary for parties to the same data sharing relationship to select different GDPR bases/ conditions for processing, according to their location and circumstances.

Recommendation 3: **Prioritise the least intrusive infrastructure**

If multiple infrastructure options are available—as may be the case under the European Health Data Space, for example— care should be taken to select the least intrusive option necessary for each research access. Synthetic data could be seen as the least privacy invasive method of querying information, but this will need to be assessed on a case-by-case basis. In many cases, real-world data is necessary for in silico modelling, in which case remote/technologically controlled access offers more privacy protection than a download with contractual controls.

2.5.2 Technical Avenues & Data Control

Another relevant factor to consider when choosing between the above options is whether EU-integrated data are controlled by a single committee, per the IMI DIRECT project,[40] or whether the different national implementations of the GDPR point towards a 'country of origin' approach, in which the data remain within the country of origin and subject to its local version of the GDPR.

In our first year report;[41] we provided a detailed review of the ways in which different member states have used national derogations to implement the GDPR in their own way. These include:

- Different exercises of the scientific research exceptions in Article 89, with some member states making fuller use of the exemptions than others.[42] [43]
- Guidance and national legislation placing different emphasis on the role of consent as a GDPR basis for processing in scientific research.
- Differing standards for anonymization across Europe.

These discrepancies make a joint approach towards control of health-related data difficult if data controllers are based in different EU states, and subject to different GDPR interpretations. For this reason, we suggested in our first-year report that federated data access models might be a preferable option wherever possible. If cloud-based/remote access models are used, we would still recommend that data remain under the control of the original data provider wherever possible.

A complication occurs where personal data are accessed by researchers or regulators in a different EU member state from that where the data originated. Under the General Data Protection Regulation, an individual or an entity processing personal data will do so as a controller or a processor. Although the GDPR also refers to 'third parties,' it is clear that data recipients will also be data controllers for any personal data they process where they determine the purposes of the processing.[44] The researchers are clearly not acting

[40] Teare et al, note 21 **Fehler! Textmarke nicht definiert.**
[41] Available from <https://www.eu-stands4pm.eu/publications>
[42] R Boardman R & F Molnár-Gábor F, 'GDPR brief: how is article 89 implemented across the EU/EEA?' (2019) available from: https://www.ga4gh.org/news/how-is-article-89-implemented-across-the-eu-eea/
[43] M Mourby et al, 'Governance of academic research data under the GDPR—lessons from the UK' (2019) 9 International Data Privacy Law 3
[44] European Data Protection Board, 'Guidelines 07/2020 on the concepts of controller and

on the instructions of the data provider; the only question which remains is whether they are joint controllers or single controllers in their own right?

Guidelines adopted by the European Data Protection Board[45] offer the following example:

Several research institutes decide to participate in a specific joint research project and to use to that end the existing platform of one of the institutes involved in the project. Each institute feeds personal data it holds into the platform for the purpose of the joint research and uses the data provided by others through the platform for carrying out the research. In this case, all institutes qualify as joint controllers for the personal data processing that is done by storing and disclosing information from this platform since they have decided together the purpose of the processing and the means to be used (the existing platform). Each of the institutes however is ***a separate controller*** *for any other processing that may be carried out* ***outside the platform*** *for their respective purposes.*

For the reasons explained above, such joint control may be difficult where the institutes are based in different EU countries and subject to different rules. The example of a cloud-based/ remote access platform, through which researchers access data, the question of data control is not immediately clear. There are two interpretations:

1) The researcher has devised their own purpose for using the data, so they are a single controller for processing carried out via the platform.
2) The data provider has authorised the researcher's proposed use of the data, and sets the terms on the 'manner' of processing within the platform. They are thus joint controllers of the researcher's use of the data via the platform.

The same could be said of data transferred under a data access agreement, in which it is perhaps more likely that joint control will be established as the parties have jointly determined the manner and purposes of the processing via the contractual terms.

Data control is a question of fact—of who, in practical terms, is making these decisions about personal data— but it is clearly feasible that researchers accessing data across the EU will become joint data controllers with an institution subject to a different GDPR implementation. The legal differences this would raise with different combinations of jurisdiction may be difficult to overcome on an ad-hoc basis.

At page 41 of their draft guidelines, the EDPB state:

Although the GDPR does not preclude joint controllers to use different legal basis for different processing operations they carry out, it is recommended to use, whenever possible, the same legal basis for a particular purpose.

This suggests it is not a strict necessity for two data controllers based in different countries to rely on different legal bases for the same purpose. However, the Board clearly stress that a common approach to data governance should be used '*wherever possible.*' Further clarification from the Board as to whether differing national legislation and guidance could serve as an adequate justification for different legal bases might help to illustrate this point. But, as it currently stands, the divergences created by national deviation seem to create an undesirable rift in the common approach that would normally be expected for joint data controllers.

It is also worth considering that, in some cases, an alternative legal basis may not be available for the proposed research use. For example, where data are to be re-used on the basis of public interest & scientific research, and no specific consent has been obtained for that further use, an EU jurisdiction which

processor in the GDPR' version 2.1, 20 September 2022, p 22. Available from: <https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf >

[45] Ibid

is less liberal in its research exemptions may not permit the research use at all, in which case the data cannot be integrated.

Consequently, we maintain our recommendation from our first-year report, that federated access to data—in which personal data are only processed in their country of origin—may be a more promising model of integration.

2.6 Recommendations

To summarise, this section has made three main recommendations:

Recommendation 1: Maximise the impact of the template consent form

Recommendation 2: Make data collectives as transparent and inclusive as possible

Recommendation 3: Priotise the least instrusive data infrastructure

These recommendations aim to ensure that the opportunities arising from the Data Governance Act are harnessed to bring greater harmony to GDPR compliance in different territories. The GDPR's emphasis on data minimisation should be built into the design of the European Health Data Space, and other research data infrastructures. Investment of time and resources is required to ensure that data cooperatives and altruism organisations are open to everyone in practice as well as theory, to support the collection of fair and representative data.

3. Interaction between research ethics guidelines and data protection regulation

3.1 Consent misconception and other consequences of a 'dual' meaning of consent

Informed consent remains a fundamental normative core of such well known global research ethics guidelines as the WMA Declaration of Helsinki (2013) and the Council for International Organizations of Medical Sciences ('CIOMS') guidelines (2016). These guidelines followed the Nuremberg code (1949) in emphasizing a fundamental role informed consent must play in research involving physical or other form of intervention ('interventional consent'). A similar centrality of consent seems to be also retained by these guidelines in case of research that only involves processing of health data ('informational consent') because any modifications of consent in these guidelines are only allowed in the exceptional circumstances.

On the other hand, within the GDPR consent has no predefined priority for health data processing. It is only one among other legal mechanisms, such as a distinct research condition for processing special category data under Article 9(2)(j) (abbreviated further as 'research condition'), which provides an alternative to re-consent for the use of previously collected personal data and biological materials.

Another distinctive feature of the GDPR is that it sets up very high requirements for informed consent to process personal data: it needs to be clear, concise, specific and granular, freely given and revocable[46]. However, such a high standard of consent is also the source of divergence from the conceptual framework of consent in the research ethics regulations[47]. A strict compliance with such GDPR consent standards, as specificity, granularity or revocability might not always be met in case of collecting and sharing big sets of data as well as secondary use of personal data for research purposes. As a result, stricter GDPR requirements of consent have encouraged research institutions and researchers to assess whether the mentioned activities comply with a valid consent according to the GDPR and therefore was an impetus to search whether the exemptions from the explicit consent as well as another legal basis (such as public interest) would be more appropriate for the mentioned purposes (see Table 2).

[46] European Data Protection Board. Guidelines 05/2020 on consent under Regulation 2016/679. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf

[47] In this report **research ethics regulations** refer to global RE guidelines (CIOMS guidelines, Declaration of Helsinki, CoE recommendations), as well as CTR and other EU documents as far as they correspond to the role and meaning of consent in human research

Table 2: Comparison of legal regimes governing informed consent

	Research ethics regulations		**General Data Protection Regulation**
Concept of consent and its modifications	Consent for clinical intervention ("Interventional" consent) – Specific only	Consent for health data processing ("Informational" consent) – Specific – Dynamic – Broad	Consent as a legal condition for health data processing (Art 9 (2) a) – Specific (granular, revocable) – Dynamic – Broad?
Data collection without consent	Waiving of consent is recommended in exceptional cases only, when informed consent is not feasible		Other legal conditions for health data processing (e.g., research condition under Art 9 (2) j) are equal to consent, thus become a prevailing scenario for data processing in practice in many EU countries
Responsible actors	RECs, researchers, research institutions		DPOs, DPAs, researchers, research institutions

Changing consent requirements have been the reason why some ambiguities and misconceptions have emerged. For example, a researcher conducting a particular clinical trial still must follow research ethics regulations to obtain explicit consent for clinical intervention. However, s/he can process the data obtained in the course of this intervention following the mentioned 'research condition', which allows data processing without explicit consent when this is necessary for scientific purposes. According to some authors, failure to make a clear distinction between consent as it is defined in the research ethics regulations and consent in the context of the GDPR can lead to the so-called "consent misconception where research participants can think that consent to participate in a research project also extents to the consent to process their personal data [48]and therefore mistakenly believe that s/he is still able to access the data, restrict its further processing or erase it (in case these rights are restricted either by the national or EU law). Any such misunderstanding could threaten the clarity needed for public trust in future research.

Consequences of restricting the conceptual framework of informational consent are also important in other contexts of personalized medicine.

One important area of concern is related to the debate on the use of the concept of *'broad consent'*, which has been particularly popular in the field of biobanking and other areas of prospective research use of big data. No less problematic is another interplay between data protection and research regulations related to the *secondary use of data* and *waiving of consent*. The GDPR 'research condition' scenario combined with the tighter definition of consent has a potential to make research without explicit consent a much more usual rather than exceptional scenario for data processing as compared to what is still recommended in the current research ethics guidelines.

[48] Edward S Dove, Jiahong Chen. Should consent for data processing be privileged in health research? A comparative legal analysis, *International Data Privacy Law*, Volume 10, Issue 2, May 2020, Pages 117–131, https://doi.org/10.1093/idpl/ipz023

To summarize, after the GDPR entered into force, the areas of human research involving interventional aspects and those dealing exclusively with processing of personal data became legally more separated due to the introduction of a distinct 'research condition' for processing under Article 9(2)(j) as well as stricter requirements for consent under Article 7 GDPR. As a result, in contrast to the consent-centered research ethics regulations, the GDPR made consent a much more challenging option for data processing. Therefore, research participants, researchers and research ethics committees should be well informed about this change in the role and meaning of consent, which transformed a rather uniform pre-GDPR consent framework. A clear distinction between the interventional and informational aspects of research and corresponding types of consent is an important precondition to safeguard the rights of research participants and to prevent consent misconception. It is also equally important not to skip consent as a legal basis when it is methodologically feasible because consent 'puts the data subject on a more symmetrical informational and communicative plane with the data controller'[49].

Recommendation 4: **Clearly distinguish between interventional and informational aspects of research as well as different legal grounds for processing personal data for research purposes**

We recommend paying close attention to the changes of the pre-GDPR consent framework. In particular, clear distinctions between the interventional and informational aspects of research as well as different legal grounds of data processing should be made. These circumstances should be duly considered by all the different parties, which are supposed to navigate between two parallel regimes of data processing regulations and requirements envisaged in research ethics guidelines.

It should be noted however that the mentioned contorversies between research ethics and data protection regulations could be reduced if we look at personalized medicine form the broader relational, social perspective rather than the narrow individualistic approach. This has been a very relevant observation raised by the particiants of the Legal & Ethical expert workshop on 19 April 2022. It has been emphasized that such a relational and solidarity based aprroach is much more relevant when we deal with the use of health data that relates not only to the person from whom the data originated, but also to the "secondary data subjects", such as family members. In addition, solidarity based approach encourage people to share their health data, which is a fundamental precondition for the development of personalized medicine[50].

3.2 Diverging interpretations of consent: GDPR vs research ethics regulations

Due to the mentioned difficulties to apply the GDPR concept of consent for health data research, European Data Protection Board (EDPB) [51] and national bodies like Medical Research Council[52] and and Health Research Authority [53] in the UK) recommend choosing other legal basis for data processing than consent.

[49] Edward S Dove, Jiahong Chen. Should consent for data processing be privileged in health research? A comparative legal analysis, *International Data Privacy Law*, Volume 10, Issue 2, May 2020, Pages 117–131, https://doi.org/10.1093/idpl/ipz023
[50] Prainsack B. The "We" in the "Me": Solidarity and Health Care in the Era of Personalized Medicine October 2017, Science, Technology & Human Values 43(1):016224391773613 DOI:10.1177/0162243917736139
[51] The EDPB considers that as an alternative to data subject's consent, the lawful grounds of processing provided under Article 6(1)(e) or 6(1)(f) are more appropriate. (European Data Protection Board. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR)(art. 70.1.b)). (https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf)
[52] https://mrc.ukri.org/documents/pdf/gdpr-lawful-basis-research-consent-and-confidentiality/
[53] https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/consent-research/

On the other hand, some authors[54;55;56]argue that consent, in particular broad consent, is still an option for collecting data in research and such consent is in line with the requirements for consent set in the GDPR. Subsequently, some international research organizations[57] still see broad consent as a legal basis for data processing.

In addition to the contradicting legal viewpoints, the GDPR and research ethics regulations also provide rather different interpretations of what are the legitimate modifications of consent in the context of human research based exclusively on data. In particular, the need to clarify the role and meaning of broad consent at the EU emerged because two recent European Regulations (i.e. the GDPR and Clinical Trial Regulation (CTR)) provided different and, in some cases, contradicting interpretations of this principle.

For example, the GDPR Recital 33 reads: '*It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to* ***certain areas of scientific research.***' However, the Article 29 Working Party provides a restricting interpretation of the Recital 33 noting that '*Considering the strict conditions stated by Article 9 GDPR regarding the processing of special categories of data, WP29 notes that when special categories of data are processed on the basis of explicit consent, applying the flexible approach of Recital 33 will be subject to a stricter interpretation and requires a high degree of scrutiny*.'[58]

On the other hand, a much more relaxed interpretation is provided by Recital 29 and Article 28 (2) of CTR,[59] enabling universities and other research institutions to collect data from clinical trials to be used for future scientific research, for example for medical, natural or social sciences research purposes, if the subject gives consent to use his or her data outside the protocol of the clinical trial and has the right to withdraw that consent at any time. Similarly, the CIOMS (2016) Guidelines 11, 12 introduced **broad informed consent '***extending to a number of* ***wholly*** *or partially undefined studies*'[60]. Some authors seem to rely on these more relaxed interpretations of broad consent and claim that the GDPR '*limiting the use of broad consent for research has significantly narrowed the extent to which consent could ever be a realistic basis for processing in the context of banked personal data and associated biospecimens.'* [61]

Of note, the relationship between professional ethical standards and the GDPR is still under discussion within the European Data Protection Board[62]. It is feasible that the outcome of these discussions may impact upon the role ethical review can play in legitimizing broad consent as a basis for data processing.

---

[54] Hallinan, D. Broad consent under the GDPR: an optimistic perspective on a bright future. *Life Sci Soc Policy* **16,** 1 (2020). https://doi.org/10.1186/s40504-019-0096-3

[55] Reichel J. (2021) Allocation of Regulatory Responsibilities: Who Will Balance Individual Rights, the Public Interest and Biobank Research Under the GDPR?. In: Slokenberga S., Tzortzatou O., Reichel J. (eds) GDPR and Biobanking. Law, Governance and Technology Series, vol 43. Springer, Cham. https://doi.org/10.1007/978-3-030-49388-2_23

[56] Hansson M.G. (2021) Striking a Balance Between Personalised Genetics and Privacy Protection from the Perspective of GDPR. In: Slokenberga S., Tzortzatou O., Reichel J. (eds) GDPR and Biobanking. Law, Governance and Technology Series, vol 43. Springer, Cham. https://doi.org/10.1007/978-3-030-49388-2_3.

[57] https://www.ga4gh.org/news/gdpr-brief-is-consent-for-genomic-and-health-related-research-specific-enough-to-constitute-a-valid-consent-under-the-gdpr/

[58] 'Considering the strict conditions stated by Article 9 GDPR regarding the processing of special categories of data, WP29 notes that when special categories of data are processed on the basis of explicit consent, applying the flexible approach of Recital 33 will be subject to a stricter interpretation and requires a high degree of scrutiny.' (Article 29 Working Party, 'Guidelines on Consent under Regulation 2016/ 679' WP 259 rev.1, as last revised 10 April 2018, p 28.)

[59] Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials in medicinal products for human use and repealing Directive 2001/20/EC. Off. J. 2014;L 158(27.05.2014):1

[60] Commentary to CIOMS Guide 12

[61] Peloquin, D., DiMaio, M., Bierer, B. *et al.* Disruptive and avoidable: GDPR challenges to secondary research uses of data. *Eur J Hum Genet* **28,** 697–705 (2020). https://doi.org/10.1038/s41431-020-0596-x

[62] European Data Protection Supervisor. A Preliminary Opinion on data protection and scientific research https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf

The debate that has sparked around the broad consent also raises more theoretical questions on the meaning of this concept and its continued use in Europe and world-wide. For example, broad consent has been rather recently introduced into the US research legislation[63]. Its modifications to fit bio-banking purposes and to better comply with the requirements of informed consent have also been widely discussed in the academic literature. On the one hand, there are those proposing to 'improve' broad consent through '*strong ethical review and continuous communication'* by the means of regular newsletters with the biobank participants[64]. On the other hand, there are those who argue that more structured and specific consent approaches, such as dynamic consent[65] as well as meta-consent[66], would be better options to protect personal autonomy of people**.** All these attempts to further develop the concept of broad consent show that there is also a deeper discrepancy between the legal and more philosophically oriented discourses on the issue.

Recent discussion on the Data Governance Act (DGA) also shows the complexity of this discussion. As it was already mentioned the DGA introduces the concept of 'data altruism' and defines it as 'the consent by data subjects to process personal data pertaining to them, <…or > without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services' (DGA Article 2.10)[67]. By introducing the concept of 'processing for purposes of general interest', the DGA may seem to also allowing in certain circumstances to process personal data for not strictly defined research purposes, which serve the general interest (e.g. scientific research) and therefore challenges a restricting position of the A29WP on broad consent referred to earlier.

To sum up, there are currently different approaches on the ethical and legal acceptability of the consent modalities to be employed in the context of prospective research on health data.

Recommendation 5: **Consider different consent modalities such as broad broad or dynamic consent to fulfill research ethics standards**

There are currently few divergent approaches dealing with ethical and legal acceptability of broad consent:

- it might be possible in some situations to accept a broader consent for data processing for research purposes under the GDPR, provided the consent is truly voluntarily (i.e. there is no significant disparity in the power relation between data subjects and controller, or risk of detriment from refusal of processing) and can effectively be revoked).
- another option is to stick to a stricter GDPR based position and to use public interest rather than consent as a legal ground for prospective collection and research on health data and biomaterials. This does not preclude from the need to obtain (broad) consent as required by the research ethics regulations, but in this case consent would be regarded only as an additional safeguard rather than a legal ground for data processing.
- these two opposing options could be reconciled by such consent modality as dynamic consent, which seems to meet both the GDPR and the research ethics requirements.

[63] https://www.brany.com/the-revised-common-rule-and-informed-consent-broad-consent/

[64] Mikkelsen, Rasmus Bjerregaard, M. Gjerris, G. Waldemar and P. Sandøe. "Broad consent for biobanks is best – provided it is also deep." *BMC Medical Ethics* 20 (2019): n. pag. (https://www.semanticscholar.org/paper/Broad-consent-for-biobanks-is-best-%E2%80%93-provided-it-is-Mikkelsen-Gjerris/ff5e92aeee48daefb00ecfbb3bd7607e70f2e5b9)

[65] Teare, H.J.A., Prictor, M. & Kaye, J. Reflections on dynamic consent in biomedical research: the story so far. *Eur J Hum Genet* (2020). https://doi.org/10.1038/s41431-020-00771-z

[66] Ploug T, Holm S. The 'Expiry Problem' of broad consent for biobank research - And why a meta consent model solves it. *Journal of Medical Ethics* 2020;**46:**629-631. (https://jme.bmj.com/content/46/9/629).

[67] Ibid note 11

3.3 Waivers of consent and research condition

Still another area of interplay between data protection regulations and research ethics regulations might have emerged regarding the option of waiving consent for the use of identifiable human material or data for research purposes. To use the GDPR specific terminology, in these cases we refer to the already mentioned 'research condition' scenario, which allows secondary data processing without explicit consent when this is necessary for the scientific research purposes.

Let's start from the research ethics guidelines. For example, the latest version of the Declaration of Helsinki (2013), which is one of the most influential research ethics guidelines, notes that in case of research on identifiable human material or data, an option to waive consent is only allowed in '**exceptional situations** where consent would be impossible or impracticable to obtain for such research' and with an approval of a research ethics committee. Similarly, the Recommendation on research on biological materials of human origin (2016)[68] as well as CIOMS Guideline 12: Collection, storage and use of data in health-related research [69] also seem to be presenting the option of waiving of consent as an exemption allowed by RECs only in case other necessary conditions are satisfied[70].

On the other hand, some interpretations of the GDPR research related provisions can lead to much more relaxed positions. For example, according to the EDPB[71], scientific research may not require a legal basis envisaged in Article 6 of the GDPR and therefore seems to allow the re-use of data for scientific research without an identified lawful basis, such as consent or public interest. Some European countries also encourage researchers not to rely on consent as a basis for processing personal health data. However, most of the countries take a more cautious approach and require taking public interest as a legal basis for data processing in health research under the 'research condition' scenario. In this case they also provide for some specific conditions to be satisfied. For example, the UK Health Research Authority requires to assure that '*substantial damage or distress is not caused to the subjects*'. In Denmark scientific and statistical studies require '*significant importance to society*' to process special category data under Article 9(1)(j), and in Germany research condition is allowed only if 'the interests of the controller in processing substantially outweigh those of the data subject in not processing the data'[72].

[68] Council of Europe: Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin, 2016, 2016. Available from: https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=090000168064e8ff

[69] Council for International Organisations of Medical Sciences. International Ethical Guidelines for Health-related Research Involving Humans. Geneva: CIOMS;2016. Available from: http://cioms.ch/ethical-guidelines-2016/WEB-CIOMS-EthicalGuidelines.pdf

[70] When researchers seek to use stored data collected for past research, clinical or other purposes without having obtained informed consent for their future use for research, the research ethics committee may consider to waive the requirement of individual informed consent if: 1) the research would not be feasible or practicable to carry out without the waiver; and 2) the research has important social value; and 3) the research poses no more than minimal risks to participants or to the group to which the participant belongs

[71] https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf

[72] EU-STANDS4PM report: Legal and ethical review of in silico modelling (March 2020) p.39)

The divergent interpretations of research use of personal data and biological materials also point to some broader questions. When should consent be prioritized as a legal basis for data processing? What are the criteria to decide when research condition scenario is justified and public interest overrides the need to get consent? What role should be played by the RECs and DPOs in these circumstances?[73].

Recommendation 6: **Take public interest as a legal basis for data processing in health research under the 'research condition' scenario**

Although there are arguments supporting the re-use of data for scientific research without an identified lawful basis, it is suggested to keep a cautious approach and consider that Articles 6 and 9 GDPR as well as national requirements still need to be satisfied in all cases of further processing for research purposes. In the case of secondary use and waiving of consent, a legal basis for processing data for research use most likely would be public interest.

3.4 Implications for researchers, research institutions, RECs and DPOs

It seems that the questions posed at the end of the previous section show the need for closer cooperation between RECs and data protection bodies when searching for ethical and legal justification of research studies using personal data and biological materials. The involvement of RECs in this cooperation seems to be very relevant because the importance and social value of the research, benefit and risk ratio as well as distress to the subjects are the issues typically examined during the review procedure by RECs and elaborated in detail by different research ethics guidelines[74]. A more extended involvement of RECs might serve as a useful tool to determine criteria on how to define 'public interest' and to assess proportionality when balancing the aims of the research and the protection of personal data[75]. There are however some important organizational challenges to such a cooperation between RECs and data protection bodies arising in the context of overlapping research and data protection regulatory regimes.

On the one hand, the GDPR does not give a legal mandate for RECs to look at data processing issues as all these issues are a prerogative the disposal of data controllers, data protection authorities and data protection officers (DPOs). On the other hand, CTR explicitly states that the II part of the assessment of the clinical trial application (which is usually assigned to research ethics committees) should include a review whether the application is in 'compliance with Directive 95/46/EC' (currently the GDPR). This shows that the obligation to review the relevant aspects of personal data protection is not automatically removed

Recommendation 7: **Foster further integration between research ethics and data protection frameworks**

We suggest a better integrated framework for processing of data, where research ethics regulation operates along with the GDPR. In such a framework, RECs should have a meaningful role in interpreting GDPR relevant issues such as "public interest", risks and benefits to the research subjects whose data are used for research purposes, especially having in mind that DPOs are not always involved in the development of research protocols and do not directly interact with the researchers. Research ethics regulations could provide a relevant set of criteria to decide whether the justification in terms of public interest could be employed and the waiver of consent justified.

[73] Mezinska, Signe & Buka, Arnis & Bankava, Agnese & Barzdins, Juris. (2020). Legal and Ethical Issues in Secondary Use of Administrative Health Data: The Case of Latvian Healthcare Monitoring Datalink. Studies in health technology and informatics. 270. 1138-1142. 10.3233/SHTI200340.

[74] E.g., CIOMS, Guideline 1: Scientific and social value and respect for rights https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/

[75] Mezinska, Signe & Buka, Arnis & Bankava, Agnese & Barzdins, Juris. (2020). Legal and Ethical Issues in Secondary Use of Administrative Health Data: The Case of Latvian Healthcare Monitoring Datalink. Studies in health technology and informatics. 270. 1138-1142. 10.3233/SHTI200340.

from RECs and therefore it may fall within the scope of the assessment done by the RECs. In addition, according to some authors RECs can take various steps to ensure that researchers share individual-level data from clinical trials responsibly by assessing data sharing plans to make sure that privacy and security safeguards are appropriate, and that researchers commit themselves to making data available ensuring that an appropriate consent language is used that enables sharing. RECs could even take steps to hold researchers accountable for not sharing data[76].

As already discussed above, there are also other areas of concern where an interaction between research institutions (legal departments, research support units, DPOs) and RECs is needed. This interaction seems to be particularly important where justification of a broad consent or waiving of consent options for secondary data use (usually assessed by RECs) as well as where the assessment of the compatibility of secondary data use (usually done by researchers in consultation with the DPO and other personnel of research organisation) are considered, and whether RECs could object the legal basis for data processing chosen by the data controllers.

3.5 Public trust and the role of consent

An important step for data sharing is ensuring that public trust is obtained for the use of the data.

Studies show that there is still a prevailing interest of individuals to have control over personal data in terms on what data are collected, who has access to this data, how and with whom data are shared and for what purposes the data are used. In this context informed consent is still regarded as an important mechanism for facilitating individual control over personal data[77]. The data from public surveys demonstrates that some of participants preferred a thorough initial consent process with agreed terms as opposed to "inferred consent", or inferred opt-in where researchers neglected to ask for explicit consent to share[78]. While there are still ongoing discussions on what would be the best option for data sharing: the use of dynamic consent, broad consent or data donation without consent, we argue that public preferences for a certain level of control over their data, regardless of the lawful basis relied upon under the GDPR, support a continued role for consent to maintain trust in research, represent respect of study participants, as well as its default necessity under research ethics guidelines.

Establishing clear consent mechanisms could provide greater clarity for all parties involved in the data sharing debate. Ensuring that appropriate consent for future research, including secondary data analysis and sharing and linking of datasets, is gained at the point of data collection, would continue to promote research transparency and provide healthcare professionals and researchers with knowledge that an individual is aware that their data may be used for other research purposes. [79].

[76] A. Thorogood, B.M. Knoppers, Can research ethics committees enable clinical trial data sharing?, Ethics, Medicine and Public Health, Volume 3, Issue 1, 2017, Pages 56-63, ISSN 2352-5525, https://doi.org/10.1016/j.jemep.2017.02.010

[77] Aitken, M., de St. Jorre, J., Pagliari, C. *et al.* Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* **17,** 73 (2016). https://doi.org/10.1186/s12910-016-0153-x

[78] Howe, Nicola & Giles, Emma & Newbury-Birch, Dorothy & Mccoll, Elaine. (2018). Systematic review of participants' attitudes towards data sharing: A thematic synthesis. Journal of Health Services Research & Policy. 23. 123-133. 10.1177/1355819617751555

[79] Hutchings, E., Loomes, M., Butow, P. *et al.* A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. *Syst Rev* **9,** 240 (2020). https://doi.org/10.1186/s13643-020-01485-5

On the other hand, public trust derives from a number of sources, including the existence of regulatory systems to enforce rights[80]. As some authors argue control may be facilitated through transparency and public engagement rather than direct or specific opt-in consent[81]. Thus, it may be argued that well established policy, which supports and promotes the secondary use of data and data sharing together with the use of patient education and awareness campaigns might decrease concerns regarding privacy and consent.

Recommendation 8: **In case of research initiatives for data sharing and secondary data analysis based on other legal grounds than consent, ensure that patients are aware of their rights to privacy**

Consent of those sharing data for future research helps to maintain trust in research, represents respect of study participants and it also reflects the fundamental principle of research ethics regulations. Therefore, any research initiatives for data sharing and secondary data analysis based on other legal grounds than consent should ensure that patients are aware of their rights to privacy.

3.6 Recommendations

To summarise, this section has made six core recommendations:

Recommendation 4*:* Clearly distinguish between interventional and informational aspects of research as well as between different legal grounds for processing personal data for research purposes.

Recommendation 5: Consider different consent modalities such as broad broad or dynamic consent to fulfill research ethics standards.

Recommendation 6. Take public interest as a legal basis for data processing in health research under the 'research condition' scenario.

Recommendation 7. Foster further integration between research ethics and data protection frameworks.

Recommendation 8. In case of research initiatives for data sharing and secondary data analysis based on other legal grounds than consent s ensure that patients are aware of their rights to privacy.

[80] James Scheibner, Marcello Ienca, Sotiria Kechagia, Juan Ramon Troncoso-Pastoriza, Jean Louis Raisaro, Jean-Pierre Hubaux, Jacques Fellay, Effy Vayena, Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies, *Journal of Law and the Biosciences*, Volume 7, Issue 1, January-June 2020, lsaa010, https://doi.org/10.1093/jlb/lsaa010

[81] Aitken, M., de St. Jorre, J., Pagliari, C. *et al.* Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* **17,** 73 (2016). https://doi.org/10.1186/s12910-016-0153-x

4. Rights of Patients & Research Subjects - Aspects of harmonisation & integration

In this section, we summarise aspects of the **rights of patients and research subjects** (analysed in detail in the first report), and make recommendations for how these rights can and should be included when selecting avenues for data harmonisation and integration.

4.1 Harmonisation and individual rights

Fundamentally, we acknowledge the importance and centrality of data sharing, including in fulfilling the right to benefit from scientific progress.[82] When selecting avenues for data harmonisation and integration, researchers should weigh the rights of the data subject (such as, privacy and data protection) with the interests of society (new scientific discoveries and treatments, the right to scientific progress). This is often a challenging exercise. We note that the balancing of individual rights and societal interests is subject to the principle found in the Biomedicine Convention and Additional Protocols, that *the interests and welfare of the human being shall prevail over the sole interest of society*. [83] This means that in the event of conflict between the human being and the interests of society or science, the individual must take precedence in both treatment and research contexts.[84] However, as raised at our Legal & Ethical Stakeholder workshop on 19 April 2022 this balancing is not always straight forward, as patients may have an interest in scientific progress as well, and protecting patient's privacy is also of societal interest to ensure trust and confidence in the health care services and medical research.

In this report, we recommend that the data subject's **rights to information of significance to their health** should also be part of this calculus. In other words, data processors should reflect on the implications of the avenues of harmonisation/ integration on the broader rights of data subjects, not only their right to privacy.

As detailed in the 12-month report, data subjects hold different rights depending on whether their data was gathered in connection with treatment (e.g. electronic health records) or research (i.e. enrolled in a study as a research subject). Different treaties and recommendations are at play depending on whether data was collected in the clinic, as a research project, from registries, as a clinical trial, from social media or mobile apps. However, as highlighted in the first report and the EU-STANDS4PM survey, these categories can overlap. For example, electronic health records may contain results of genetic sequencing conducted as part of a research project; is data gathered through mobile apps patient data or research data or neither? In the case of registry data, national law in some instances requires that health professionals submit patient data to registries without patient consent, and this category thereby becomes subject to how registry data is classified.

As described in the first report, (health) data may be subject to a **duty of confidentiality** and the protection of medical data forms part of the right to private life under the European Convention on Human Rights.[85] The question of whether information is subject to confidentiality often depends on the domestic law

[82] The right to scientific progress is recognised under article 15 of the International Covenant on Economic, Social and Cultural Rights. States that have ratified the Convention are bound to respect, protect and fulfil this right. Non-state actors should respect the right. (The Covenant has been ratified by 171 states and signed by 4, see here: https://indicators.ohchr.org/). See further: Knoppers BM, Joly Y. Introduction: the why and whither of genomic data sharing. Human Genetics. 2018 Aug;137(8):569-574. DOI: 10.1007/s00439-018-1923-y.

[83] The Biomedicine Convention is a binding Council of Europe Treaty, ratified by 29 states and signed by 6. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (Biomedicine Convention), ETS no. 164, Oviedo, 4.IV.1997.

[84] Explanatory Report – ETS 164 – Human Rights and Biomedicine (Convention), para 21.

[85] *Z v. Finland* App. No. 22009/93, 25 Eur. H.R. Rep. 371 (1997).

framework. If confidential patient data is to be reused for research, a legal basis is required to share the data, such as, patient consent. Legislation can provide for circumstances that allow for limiting confidentiality without patient consent. In exceptional circumstances, if a person cannot give consent, patient confidentiality can be set aside.[86] For example, in Denmark, regions can provide a researcher with data from patient records for use in a specific research project of 'important societal interest' without patient consent.[87] Personal data collected during biomedical research is also confidential and should be protected from inappropriate disclosure.[88]

Recommendation 9: **Ensure transparency regarding exemptions from confidentiality**

Clinicians and researchers should be transparent with patients/ research subjects regarding the different possible legal bases that can set aside confidentiality (e.g. when consent will be required or which basis are available beyond consent, as well as opportunities to opt out).

4.2.1 Research subjects' right to health information

Rights to health and information were mapped in the first report and summarised in the executive summary. In the case of research subjects, states that have ratified the Biomedicine Convention's Additional Protocol have obligations to ensure that conclusions of research and findings of relevance to a research subject's health/ quality of life are available through, for example, adopting legislation to guarantee same.[89]

A further question is how far the "**duty of care**" of researchers stretches, does it only relate to the findings of the original research project or to further use of research data (post integration/ harmonisation)? Henney and Kerr note that data sharing and access can lead to a 'moral distance' between new contexts for data use and sharing, and compromising the ability of the original researchers to fulfil commitments to research subjects.[90] The duty of care and entitlements of research subjects should be clarified in the interests of transparency and societal trust in research. This is all the more important, as GDPR provides for use of personal data for research purposes without the data subject's consent, which may collide with promises to research participants, when consenting to research participation.

Recommendation 10: **The rights of research subjects should travel with the data**.

This requires adequate documentation of consent to research participation, when data is gathered, regardless of which legal basis is used for the processing/later reuse. Information and promises given to research participant when consenting to research participation may limit the further use of data. Researchers must be honest with participants regarding opportunities to inform of future results where data has gone through a process of anonymization.

[86] See further, article 17 Biomedicine Convention.
[87] The Danish Health Act (Sundhedsloven), § 46, stk.2.
[88] Article 25, Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research.
[89] Ratifications available here www.coe.int/en/web/conventions/full-list/-/conventions/treaty/195/signatures?p_auth=OfELPFij
[90] Heeney and Kerr, 'Balancing the local and the universal in maintaining ethical access to a genomics biobank' BMC Medical Ethics (2017) 18:80 DOI 10.1186/s12910-017-0240-7, p. 4.

4.2.2 Patients' Right to Information

In the case of patients, the Biomedicine Convention recognises the right to know any information collected about one's health, as well the right to not be informed, if wished. In exceptional cases, restrictions may be placed by law on the exercise of these rights.[91] If patients subsequently consent to their data being used for research, this data may become governed as research data, depending on the domestic regulatory regime. It is also possible, in some cases, for patient data to be processed without consent, such as where a research ethics committee gives permission or where provided by law, such as in the case of registries that require data to automatically be transmitted.

4.2.3 Application to harmonisation/integration

The right to access health information in the context of in silico models is de facto subject to the researcher/clinician being able to re-identify and re-contact the data subject. Therefore, the means of data harmonisation/ integration (see table above in section 2.4) can have implications for rights to health information:

1. When data is irreversibly anonymised, it is by definition no longer possible to contact data subjects if their data is harmonised and integrated in in silico modelling.[92] The benefit of this method is that the individual's vulnerability to privacy risks is expected to be very low. Therefore, with a negligible risk for the individual's rights to privacy and data protection, the data could contribute to research and societal progress. For this reason, the EDPB recommended: "*data has to be anonymised where it is possible to perform the scientific research with anonymised data*".[93] At the same time, we submit that it must also be acknowledged that where data is irreversibly anonymised, the data subject's rights to health information are also de facto erased as the individual cannot be re-contacted (as are GDPR rights as anonymised data does not fall under the Regulation). This concern is noted in the explanatory report to the Human Rights and Biomedicine (Protocol):

   "*The ethics committee should be informed if foreseen anonymisation of data would prevent the transmission of [] relevant information*".[94] (Anonymisation in this context refers to GDPR pseudonymisation).

2. Where personal data has been pseudonymised it is traceable to the individual and in theory, the right to findings of significance to one's health, can be realised (although we submit only in limited circumstances as outlined in the first report). As previously noted however, it cannot be assumed that all persons want to receive health information and therefore it is important that the consent process adequately documents information preferences. Also, we note that practical barriers also exist, such as, limited time, resources, out of date contact details or the person may have passed away.

[91] Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine European Treaty Series - No. 164, Oviedo, 4.IV.1997.
[92] As we detailed in the first report, there are questions about the appropriate standard of anonymisation across the EU.
[93] European Data Protection Board, Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak Adopted on 21 April 2020, para 46.
[94] Explanatory Report – CETS 195 – Human Rights and Biomedicine (Protocol), para 57.

Example – Legal barrier to re-contact: Does national law provide a legal basis?

All Danes diagnosed with hepatitis C are registered in a database. In 2018, new treatment became available which can cure hepatitis C. Doctors could see that 3,000 of those registered were not in treatment and wanted to inform them of the new treatment. However, national law did not contain a provision allowing research data, which the data was classified as because it is held in a research database, to be processed for treatment purposes. Eventually, the Minister for Health announced that patients could be lawfully informed based on health professionals' obligation to save lives in an emergency. The example illustrates how the classification of data can have implications for information rights and the need to provide a legal basis.[95]

As of 15 June 2020, under Danish law, personal data from national or regional research registers can be processed to inform data subjects if they suffer from a communicable life threatening or serious illness that can be treated (and processing is necessary for the protection of their vital interests).[96]

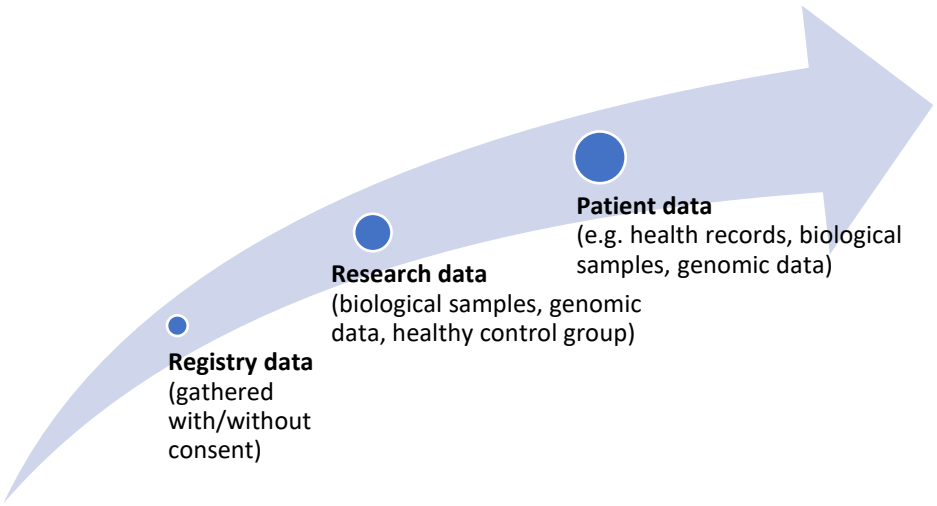
Recommendation 11: When choosing methods of data harmonisation and integration, the research subjects' rights to information should be borne in mind.

The potential for limitations on the individual's health information rights should be borne in mind by researchers when choosing methods of harmonisation and integration, and data protection officers, data controllers and research ethics committees when reviewing applications. The following rights-based factors could also be considered: the terms of the gathering of the data (did the data subject give informed consent to the collection of the data, was a legitimate expectation of re-contact created, was the data subject a child or lacking legal capacity, has a long period passed since the data was collected, what type of data was collected, is it linked to other sources?) Furthermore, provided consent has been given, those who process the data must respect the original consent terms and expectations (see below).

[95] www.dr.dk/nyheder/indland/datalov-blokerer-behandling-til-3000-syge-danskere-minister-laegger-sig-fladt-ned

[96] BKG 2020-06-08 nr 829 Tilbagemelding om væsentlige helbredsmæssige fund fra anmeldelsespligtige sundhedsvidenskabelige og sundhedsdatavidenskabelige forskningsprojekter samt visse registerforskningsprojekter, § 11(1).

We thereby propose the following rights-based checklist when choosing avenues for data integration/ harmonisation:



- ☑ What data sources are being harmonised /integrated?
- ☑ Is the likelihood of **significant** health findings such that recontact could be warranted?
  - If yes, modellers should reflect on the potential for integration/ harmonisation to inhibit re-contact and whether this can be justified based on, for example, the difficulty of obtaining consent or the societal benefit
- ☑ Does any legal/ contractual obligation of re-contact exist under national law?
  - If yes, integration/ harmonisation shall not inhibit re-contacting
- ☑ Is there a contractual/ ethical expectation of recontact, for example, in the original patient consent?
  - If yes, integration/ harmonisation should reflect on re-contacting
- ☑ In the original consent, did the data subject give consent to their data being shared with third parties/ sent to different jurisdictions/ outside EU?
  - Will also shape data integration/ harmonisation opportunities

4.2.3.1 Reflections on Anonymisation of different data sources

In this section, we reflect on the different rights and interests at play when researchers consider anonymising data for in silico modelling purposes, depending on whether the context, such as research and treatment. We suggest other relevant issues include the *type and amount of data, and conditions under which it is stored, such as whether it is linked with other data sources.*

PKU test: new-borns are tested for PKU etc. based on parental consent to the storing and future use for research; left over samples are stored in a biobank.

Societal interests: samples from new-borns can be an important research resource as they represent the whole population, not only those who actively choose to take part in research. **Individual rights**: Consent is given by the child's guardians, who should represent their best interests in childhood. However, for secondary findings emerging later in life, the child's wishes regarding secondary health information are unknown. Provided it is made clear that there is no/ low expectations of re-contact, anonymisation may be a method of integration/ harmonisation that is in line with the data subject's rights given that they do not expect re-contact.

Routine treatment: Patient having a routine operation where tissue/ blood is gathered as part of treatment; leftover material is stored on the basis of consent to treatment and can be used for research provided, for example, reconsent or research ethics approval is obtained.

Societal interests: the samples may provide a valuable resource for researching the condition from which the patient suffers, or other health concerns. In a welfare state context, it could be argued that the reuse of the data for societal research purposes forms part of the societal contract.

Individual rights: if in silico modelling is directly related to the condition that the patient suffers from, this could favour not anonymising the data, at least without informing the research subject as the data subject has a clear interest in receiving information related to their condition. For uses unrelated to the condition, anonymisation may be suitable, provided they have been made aware when consenting that data can be reused for research (and of opportunities to opt out).

Healthy individual donates blood & consents to samples being stored in a biobank for research:

Societal interests: the individuals can function as healthy controls in research projects.

Individual rights: provided the individual has been adequately informed of the possible re-uses of the data and that they should not expect re-contact, it may be in line with the individual's rights to anonymise data.

Data from patient with cancer is automatically stored in a cancer registry:

Societal interests: the research could contribute to new treatments of benefit to patients.

Individual rights: in many countries, this data is automatically stored in registries established by law. We recommend that this is explained to patients, while recognising that during an illness such details may be low on patients' list of concerns. In contrast to the routine treatment case above, registry data may be considered less sensitive as it does not include biological material (which contains genetic data) but merely diagnosis information. Provided that the registry is established by law and the research is designed for the benefit of patients, we consider that anonymisation may be appropriate, so long as the expectation of re-contact is not present.

Healthy individuals consent to take part in a clinical trial of COVID vaccine:

Societal interests: developing new vaccines requires adequate testing.

Individual rights: given the unique COVID emergency setting, one could argue that everyone will benefit from a future covid vaccine, and the individual should not expect future health information. The urgency of

the public health crisis may also justify limits on individual rights. Provided the individual is adequately informed that their data will be used without an expectation of re-contact re secondary findings, data could be anonymised. We expect that studies on serious, long-term side effects could allow for return of results, however.

Terminally ill patient consents to take part in a clinical trial:

Societal interests: patient data may contribute to research and future treatments associated with late-stage illness.

Individual rights: the patient is sadly unlikely to benefit and may be deceased by the time of later modelling. In such cases, anonymisation is unlikely to raise issues relating to the individual's rights to health information. However, genetic relatives may in some cases have an interest in receiving information: however, we regard this interest as limited and only likely to form a legal basis for informing of research findings in rare cases.

As we underscored in the first report, we consider *transparency* when collecting and processing data to be a central tool for protecting data subjects' (including research subjects and patients') rights when harmonising and integrating in silico models. Transparency during the consent process and in healthcare generally can help to clarify expectations regarding later entitlements to information, such as research results and secondary findings. One method can be to revisit consent to remind individuals of the terms of their original consent. We furthermore highlight other methods of increasing transparency, such as reminding data subjects that they have health data stored that can be used for research models. Norwegian law can be of inspiration, which requires that "The patient must have been informed in advance that human biological material can in certain cases be used for research and must have been given the opportunity to opt out of research on human biological material".[97]

Recommendation 12: **The terms of the original consent should guide later data harmonisation/ integration, or new consent should be obtained for uses in conflict with the original consent.**

The individual must have been informed of the potential for further research use and given the opportunity to opt in/ opt out. With increased digitalisation of healthcare and requirements to document consent, determining the terms of the consent should be increasingly accessible. We recommend furthermore that once children reach an age where they can give consent under national law, they be informed that their data is stored in a biobank and can be used for research.

4.3 Transparency

We identify *transparency* as a key legal principle under GDPR and European Health Law (in particular, patients' rights).[98] Machine-learning ('ML') models can entail calculations that are so complex that they are difficult to understand and independently verify. In considering the use of ML in healthcare, we approach the question of transparency from three different scenarios:

1) Solely automated decisions. We suggest these will be unusual in healthcare, as Article 22(4) of the General Data Protection Regulation presents a high bar. However, if solely automatic decisions are

[97] Lov om medisinsk og helsefaglig forskning (helseforskningsloven) LOV-2008-06-20-44, § 28. Unofficial translation.
[98] Miranda Mourby, Katharina Ó Cathaoir and Catherine Bjerre Collin, "Explaining Machine Learning in Healthcare: European Health Law and GDPR" Computer Law & Security Review, Vol 43 (2021), 105611, https://doi.org/10.1016/j.clsr.2021.105611

made (e.g. for inpatient triage) data subjects will have a right to 'meaningful information' about the logic involved.

2) Clinical decisions. These are decisions made ultimately by clinicians—such as diagnosis and suggestions for treatment—and the standard of transparency under the GDPR is lower due to this human mediation.
3) Patient decisions. Decisions about suggested treatments are ultimately taken by the patient or their representative, albeit in dialogue with clinicians. Here, the patient will require a personalised level of medical information, depending on the severity of the risk, and how much they wish to know.

European healthcare law sets a more personalised standard of information requirement than the GDPR. Advice must be tailored to the individual patient according to their needs and priorities, as such there is no monolithic 'explanation' of risk under healthcare law. When giving advice based (even partly) on a ML model, clinicians must have a sufficient grasp of the medically-relevant factors involved in the model output to offer patients this personalised level of medical information. The UK, Ireland, Denmark, Norway and Sweden are examples of European health law jurisdictions which require this personalised transparency to support patients' rights to make informed choices.

In short, in silico models deployed in healthcare must be transparent to *clinicians* in terms of their medical feature-relevance. This enables clinicians to give advice in a way that complies with their patients' rights. This can have the implications for the harmonisation and integration of big data for clinical decision support.

Recommendation 13: When integrating/ harmonising data for development of in silico models to be used in the clinic, transparency standards must be integrated from the start.
Stakeholders integrating/harmonising data for in silico models must be aware that laws must be read in unison and may impose varied obligations/ apply in different contexts. The makers of models should be accountable for ensuring that the quality of data integrated for use in silico models does not inhibit/ undermine the model's ability to provide transparent information to clinicians, as this may infer with the latter's responsibilities to impart information to patients. This would include evaluation of bias in the data, to ensure the features weighted by the model are clinically validated, and would not lead to discrimination if deployed in a healthcare setting. Transparency standards must also be borne in mind when developing research models that later could be used in the clinic. Therefore, transparency standards (particularly regarding feature relevance) should be considered and ultimately integrated into models from the start if the intention is to develop a model to be used in healthcare, in light of patients' strong rights in this area.

We note the European Commission's proposal for an AI Systems Regulation.[99] The Regulation emphasises, inter alia, transparency in line with our approach. The Proposal, if adopted would designate certain AI systems as high-risk, such as, AI systems used to triage patients. The Regulation would introduce a system whereby risk management would be required, as well as data governance that would examine possible biases and data gaps.

[99] European Commission, 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 {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final} Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (COD)

4.4 Patients' Rights to Freedom from Discrimination & Equitable Access to treatment

In the first report, we mapped international legal norms on non-discrimination and equality as they pertain to health research and patient care. We found that under international law, all individuals have a right to enjoy the benefits of scientific progress, and access to healthcare on an equitable basis, free from discrimination. Under international human rights law, discrimination is any distinction, exclusion, restriction or preference which is based on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms. We recommended that *participation*, i.e. individuals' right to be informed, heard and included in healthcare planning and decisions, be guarded as a fundamental principle underpinning these rights.

We also note that EU law prohibits discrimination based on sex, race and ethnicity in healthcare and services. However, EU member states have not passed a broad equality directive, which means that EU law does not prohibit discrimination on other grounds in healthcare.[100] National laws can, and in many cases do, provide protection based on broader grounds.

Since our first report, the Special Rapporteur on contemporary racism has issued (non-binding) recommendations on racial discrimination and emerging technologies.[101] Her recommendations include, "urg[ing] States to adopt an approach to data grounded in human rights, by ensuring disaggregation, self-identification, transparency, privacy, participation and accountability in the collection and storage of data".[102]

The risk of bias has become a reality during the COVID -19 pandemic, where models are being used to support medical decision making and prioritisation, such as, models predicting the course of the infection, models identifying risk of infection and predicting the presence of infection.[103] It has been suggested that these models are at high risk of bias and therefore should not be used routinely, unless properly validated.[104]

For example, one study found that data on ethnicity in patients with COVID-19 is limited.[105] If data is not disaggregated and ethnicity ignored, it can render models developed on one ethnicity less accurate for another. For example, pulse oximeters have been suggested as a tool to estimate when to go to the hospital with COVID infection but these devices are reported to exhibit racial bias.[106] This should cause modellers to reflect on the implications for where the data was gathered, for example, data from hospitals with majority white populations, can be biased when applied to communities of colour.

The authors of another paper on COVID-19 AI bias call for 'transparency in reporting of AI algorithms … to understand intended predictions, target populations, hidden biases, class imbalance problems, and their

[100] Racial Discrimination Directive; Directive 2004/113;
[101] Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Racial discrimination and emerging digital technologies: a human rights analysis, Human Rights Council (18 June 2020, A/HRC/44/57).
[102] Ibid, para 49.
[103] Wynats et al, Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal BMJ 2020; 369 doi: https://doi.org/10.1136/bmj.m1328.
[104] Ibid.
[105] Pan et al, The impact of ethnicity on clinical outcomes in COVID-19: A systematic review, EClinicalMedicine, Volume 23, 2020, https://doi.org/10.1016/j.eclinm.2020.100404.
[106] https://horizon-magazine.eu/article/encoding-same-biases-artificial-intelligence-s-limitations-coronavirus-response.html

ability to generalize emerging technologies across hospital settings and populations'.[107] The authors furthermore support sharing of data to build diverse and representative data sets.[108]

Yet, if the data is not available, developers can do little to ensure data diversity. For example, pregnant women were not included in the original clinical trials of COVID-19 vaccines.[109] Therefore, because the data does not exist, when harmonising efficacy data, women will be underrepresented and pregnant women non-existent in the data set, which can have implications for bias and their right to health. As an example, studies of use of the approved COVID-19 vaccines on pregnanet women, were not initiated until late in the pandemic, and there are still no results from ongoing studies. The lack of data has also served to spread of misinformation that the vaccine has been shown to cause infertility.[110] This is not only problematic from an equity/ non-discrimination standpoint but also practically as women are often front-line health workers in need of the vaccine.

In line with our earlier recommendations, we echo support for transparency in the types of data used for modelling as a precondition for harmonisation and integration. We furthermore recommend reflection on the data sources used for developing models and their implications for all patients. When harmonising and integrating data sources for in silico models, researchers should ensure that datasets are diverse to suit avoid perpetuating bias or discrimination in healthcare. Yet, diversifying data is not enough if the data is inaccurate. In an instructive article, Suresh and Guttag illuminate numerous other forms of bias: historical, measurement, aggregation, evaluation and deployment.[111] We therefore recommend that human rights, including the right to non-discrimination, is included in data impact assessments.[112] The need to tackle bias is noted in the proposed AI regulation, which would require the examination of data biases and gaps, although it is at too early a stage to make conclusions as to whether this provision has teeth.

Recommendation 14: **Developers of in silico models should be accountable for addressing potential bias from the earliest stages of development.**
This is vital also in the gathering and selection of data because if models are tested on biased/ non-representative data, and later used in the clinic, they may be unusable/ harmful when used to predict/ treat persons who are not represented in the data (such as, women/ children/ underrepresented ethnicities). Thus, when harmonising and integrating data for modelling purposes, a relevant reflection is, who is included and excluded and what risks of bias does this create?

[107] Eliane Röösli, Brian Rice, Tina Hernandez-Boussard, Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19, Journal of the American Medical Informatics Association, Volume 28, Issue 1, January 2021, Pages 190–192, https://doi.org/10.1093/jamia/ocaa210
[108] Ibid
[109] Farrell R, Michie M, Pope R. Pregnant Women in Trials of Covid-19: A Critical Time to Consider Ethical Frameworks of Inclusion in Clinical Trials. Ethics Hum Res. 2020 Jul;42(4):17-23. doi: 10.1002/eahr.500060. Epub 2020 Jun 20. PMID: 32562594; PMCID: PMC7323073. The question of whether pregnant women or women of child bearing age can be debated from various perspectives. One perspective is to avoid harm to the foetus, in light of thalidomide, or the woman's potential to carry children. Another perspective is that women are often paternalistically excluded, without adequate reflection as to whether they could be included in a safe manner. This results in men-only clinical trials and treatments not having been tested on women.
[110] In the reverse it has not been tested whether the vaccine can lead to infertility. www.reuters.com/article/uk-factcheck-covid-vaccine-causing-infer-idUSKBN25H20G
[111] Suresh & Guttag, A Framework for Understanding Unintended Consequences of Machine Learning https://arxiv.org/abs/1901.10002
[112] As recommended in the Toronto Declaration (a non-binding human rights declaration by civil society actors).

4.5 Recommendations

This section has made six core recommendations, which are summarised here:

Recommendation 9: **Ensure transparency regarding exemptions from confidentiality**

Recommendation 10: **The rights of research subjects should travel with the data.**

Recommendation 11: **When choosing methods of data harmonisation and integration, the research subjects' rights to information should be borne in mind.**

Recommendation 12: **The terms of the original consent should guide later data harmonisation/ integration, or new consent should be obtained for uses in conflict with the original consent.**

Recommendation 13: **When integrating/ harmonising data for development of in silico models to be used in the clinic, transparency standards must be integrated from the start.**

Recommendation 14: **Developers of in silico models should be accountable for addressing potential bias from the earliest stages of development.**

5. Summary of recommendations

Recommendation 1: Maximise the impact of the DGA consent form

1.1 During the grace period for the application of the Data Governance Act, the European Commission should consult widely among stakeholders across Member States to achieve clarity & consensus as to a standard form of GDPR compliant consent, with consideration given as to whether this consent form could also legitimate transfers to third countries under Article 49 GDPR.

1.2 Data intermediaries and data altruism organisations established under the DGA should explore the use of dynamic consent to ensure that data subjects are kept up to date about research uses of their data, and the identity of any other data controllers who use their information for research.

1.3 Those involved in the designing of the DGA consent form, European Health Data Space and BBMRI-ERIC Code of Conduct for Health Research should be brought into contact to ensure their requirements align and do not perpetuate the uncertainty and confusion within the research community.

Recommendation 2: **Make data collectives as transparent and inclusive as possible**

2.1 Investment should be made at an EU and Member State level to ensure that any Data Cooperatives (or Data Altruism organisations) that contribute to the European Health Data Space do not only serve less vulnerable members of society, but are accessible for a wide range of groups and demographics.

2.2 Data altruism organisations should make it clear that participatnts are not 'data donors', in the conventional sense of donation, and that they retain the right to control how their information is used.

2.3 Data Cooperatives should also be clear with their members about the limitations of their remit, and that members still need to exercise their data subject rights for themselves.

Recommendation 3: **Prioritise the least intrusive data infrastructure**

If multiple infrastructure options are available—as may be the case under the European Health Data Space, for example— care should be taken to select the least intrusive option necessary for each research access. Synthetic data could be seen as the least privacy invasive method of querying information, but this will need to be assessed on a case-by-case basis. In many cases, real-world data is necessary for in silico modelling, in which case remote/technologically controlled access offers more privacy protection than a download with contractual controls.

Recommendation 4: Clearly distinguish between interventional and informational aspects of research as well as different legal grounds for processing personal data for research purposes

We recommend paying close attention to the changes of the pre-GDPR consent framework. In particular, clear distinctions between the interventional and informational aspects of research as well as different legal grounds of data processing should be made. These circumstances should be duly considered by all the different parties, which are supposed to navigate between two parallel regimes of data processing regulations and requirements envisaged in research ethics guidelines.

Recommendation 5: Consider different consent modalities such as broad broad or dynamic consent to fulfill research ethics standards

There are currently few divergent approaches dealing with ethical and legal acceptability of broad consent:

- it might be possible in some situations to accept a broader consent for data processing for research purposes under the GDPR, provided the consent is truly voluntarily (i.e. there is no significant disparity in the power relation between data subjects and controller, or risk of detriment from refusal of processing) and can effectively be revoked).

- another option is to stick to a stricter GDPR based position and to use public interest rather than consent as a legal ground for prospective collection and research on health data and biomaterials. This does not preclude from the need to obtain (broad) consent as required by the research ethics regulations, but in this case consent would be regarded only as an additional safeguard rather than a legal ground for data processing.
- these two opposing options could be reconciled by such consent modality as dynamic consent, which seems to meet both the GDPR and the research ethics requirements.

Recommendation 6: Take public interest as a legal basis for data processing in health research under the 'research condition' scenario

Although there are arguments supporting the re-use of data for scientific research without an identified lawful basis, it is suggested to keep a cautious approach and consider that Articles 6 and 9 GDPR as well as national requirements still need to be satisfied in all cases of further processing for research purposes. In the case of secondary use and waiving of consent, a legal basis for processing data for research use most likely would be public interest.

Recommendation 7: Foster further integration between research ethics and data protection frameworks

We suggest a better integrated framework for processing of data, where research ethics regulations operate along with the GDPR. In such a framework, RECs should have a meaningful role in interpreting GDPR relevant issues such as "public interest", risks and benefits to the research subjects whose data are used for research purposes, especially having in mind that DPOs are not always involved in the development of research protocols and do not directly interact with the researchers. Research ethics regulations could provide a relevant set of criteria to decide whether the justification in terms of public interest could be employed and the waiver of consent justified.

Recommendation 8: In case of research initiatives for data sharing and secondary data analysis based on other legal grounds than consent, ensure that patients are aware of their rights to privacy

Consent of those sharing data for future research helps to maintain trust in research, represents respect of study participants and it also reflects the fundamental principle of research ethics regulations. Therefore, any research initiatives for data sharing and secondary data analysis based on other legal grounds than consent should ensure that patients are aware of their rights to privacy.

Recommendation 9: Ensure transparency regarding exemptions from confidentiality

Clinicians and researchers should be transparent with patients/ research subjects regarding the different possible legal bases that can set aside confidentiality (e.g. when consent will be required or on which basis are available beyond consent, as well as opportunities to opt out).

Recommendation 10: The rights of research subjects should travel with the data.

This requires adequate documentation of consent to research participation, when data is gathered, regardless of which legal basis is used for the processing/later reuse. Information and promises given to research participant when consenting to research participation may limit the further use of data. Researchers must be honest with participants regarding opportunities to inform of future results where data has gone through a process of anonymization.

Recommendation 11: When choosing methods of data harmonisation and integration, the research subjects' rights to information should be borne in mind.

The potential for limitations on the individual's health information rights should be borne in mind by researchers when choosing methods of harmonisation and integration, and data protection officers, data controllers and research ethics committees when reviewing applications. The following rights-based factors could also be considered: the terms of the gathering of the data (did the data subject give informed

consent to the collection of the data, was a legitimate expectation of re-contact created, was the data subject a child or lacking legal capacity, has a long period passed since the data was collected, what type of data was collected, is it linked to other sources?) Furthermore, provided consent has been given, those who process the data must respect the original consent terms and expectations (see below).

Recommendation 12: **The terms of the original consent should guide later data harmonisation/ integration, or new consent should be obtained for uses in conflict with the original consent.**

The individual must have been informed of the potential for further research use and given the opportunity to opt in/ opt out. With increased digitalisation of healthcare and requirements to document consent, determining the terms of the consent should be increasingly accessible. We recommend furthermore that once children reach an age where they can give consent under national law, they be informed that their data is stored in a biobank and can be used for research.

Recommendation 13: **When integrating/ harmonising data for development of in silico models to be used in the clinic, transparency standards must be integrated from the start.**
Stakeholders integrating/harmonising data for in silico models must be aware that laws must be read in unison and may impose varied obligations/ apply in different contexts. The makers of models should be accountable for ensuring that the quality of data integrated for use in silico models does not inhibit/ undermine the model's ability to provide transparent information to clinicians, as this may infer with the latter's responsibilities to impart information to patients. This would include evaluation of bias in the data, to ensure the features weighted by the model are clinically validated, and would not lead to discrimination if deployed in a healthcare setting. Transparency standards must also be borne in mind when developing research models that later could be used in the clinic. Therefore, transparency standards (particularly regarding feature relevance) should be considered and ultimately integrated into models from the start if the intention is to develop a model to be used in healthcare, in light of patients' strong rights in this area.

Recommendation 14: Developers of in silico models should be accountable for addressing potential bias from the earliest stages of development.

This is vital also in the gathering and selection of data because if models are tested on biased/ non-representative data, and later used in the clinic, they may be unusable/ harmful when used to predict/ treat persons who are not represented in the data (such as, women/ children/ underrepresented ethnicities). Thus, when harmonising and integrating data for modelling purposes, a relevant reflection is, who is included and excluded and what risks of bias does this create?

6. Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission Directorate-General for Research and Innovation under Grant Agreement # 825843.

The current review corresponds to Deliverable 3.3 " Recommendations for legally and ethically sustainable transnational data harmonization, integration and in-silico models in personalized medicine" of Work Package 3 "Legal ethics, policy and certification for data-driven *in silico* models in personalized medicine".


EU-STANDS4PM
standards for in silico models
for personalised medicine