Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA) - Catalogue of ELSI issues

Éloïse Gennet

Common Infrastructure for National Cohorts in Europe, Canada, and Africa
- CINECA -

Deliverable D7.1
Catalogue of ELSI issues

Work Package:	WP7 - Ethical and legal governance framework for transnational data-sharing
Lead Beneficiary:	European Molecular Biology Laboratory
WP Leader(s):	Emmanuelle Rial-Sebbag (INSERM) Michaela Th. Mayrhofer (BBMRI-ERIC)
Contributing Partner(s):	EMBL-EBI
Contractual Delivery Date:	31 December 2019
Actual Delivery Date:	19 December 2019
Authors of this Deliverable:	Éloïse Gennet & Melanie Goisauf
Reviewed by:	Emmanuelle Rial-Sebbag & Michaela Mayrhofer
Approved by:	Thomas Keane
Dissemination Level:	Public
Grant agreement:	No. 825775 Horizon 2020 (H2020-SC1-BHC-2018-2020)
Type of action:	RIA
Start Date:	1 Jan 2019
Duration:	48 months

Table of contents:

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

Executive Summary

The aim of this deliverable is to give an overview of all the different ethical, legal and societal issues that the CINECA project might be confronted with: public health ethics, personal data protection, ethics of data sharing, protection of consent and vulnerability as well as compliance issues between Canada, Africa and Europe. This deliverable has been elaborated in a bottom up approach, starting from the practical legal and ethical issues encountered notably through Work Package 9 and should thus be read in conjunction with Work Package 9 requirements and deliverables. It will serve as a starting point for our future deliverable 7.2 which will be aimed at identifying and discussing the gaps in the different legislative frameworks and corresponding literature.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

Introduction

The goal of CINECA is to enable the exchange of population scale health data across international borders to allow and promote the reuse of data for health research. The rationale for sharing and reusing data in public health research is deeply rooted in the promotion of a fair distribution of research risks and benefits, and it has become an essential and powerful tool for public health research. This deliverable aims at presenting a catalogue of ELSI issues in CINECA. It has been elaborated in a bottom up approach, starting from the practical legal and ethical issues encountered notably through Work Package 9. It should thus be read in conjunction with Work Package 9 requirements and deliverables. As a basis for the lawful and ethical guarantees for data sharing and reuse within CINECA, all cohorts and consortiums have provided for the copies of their own ethics approvals (Deliverable 9.4), and they are all independently responsible for ensuring researchers accessing data have their own research ethics approval. This deliverable will serve as a starting point for the future deliverable 7.2 which will be aimed at identifying and discussing the gaps in the different legislative or regulatory frameworks and corresponding literature.

As a consequence, this deliverable will be divided into two main parts, the first one focusing on the collective perspectives of international data sharing in public health research (I), the second one examining the opposite perspective of the protection of individual data subjects when their personal data is used for secondary processing (II). Afterwards, future developments will be briefly mentioned (III) before highlighting some of the difficulties encountered in Work Package 9 (IV) and finally listing the references (V).

I. International data sharing in public health research

A. Ethical issues of international data sharing

Data sharing reduces the need for dangerous and burdensome research protocols, avoids unnecessary replication, optimises resources and promotes the gathering of more diverse and rare information (Ohmann et al., 2017; WMA Declaration of Taipei, Recital 5). It is slowly becoming a standard under the pressure of regulatory authorities, journals or funders (Pisani et al., 2016). Although the economical, scientific and ethical advantages of data sharing have been repeatedly passed on (Townend, 2018), the practicalities of how this data sharing should take place is sometimes still unclear.

In this regard, the scientific community developed a set of principles for adequate and fair sharing of data in research. The FAIR principles (Findable, Accessible, Interoperable and Findable) act as a guideline in the context of data sharing and data reuse for research purposes. They describe distinct considerations for contemporary data publishing environments with respect to supporting both manual and automated deposition, exploration, sharing, and reuse. The principles highlight the need to enhance the “ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals” (Wilkinson et al., 2016). The aim of these homogeneous principles, applicable to any data and metadata, is to promote a level of “understandable data” so that they can be processed by any machine. As a prerequisite for adequate data management, data reuse, and

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

data stewardship, the FAIR principles recommend that data and metadata should remain (FORCE11; Boeckhout et al., 2018):

- Findable (F) or discoverable: data and metadata should be uniquely and persistently identifiable. They should be described, identified and registered or indexed in a clear and unequivocal manner. The data shared should thus be free of restrictions (i.e., non-classified information), professional secret and informed consent should be respected, the data should be anonymised or, if pseudonymised, controlled access should be implemented.
- Accessible (A): data and metadata should always be retrievable in a variety of formats that are sensible to humans and machines using persistent identifiers. They should be accessible through a clearly defined access procedure. This implies for instance verifying ethical and legal compliance of access requests, reuse compatibility (compatibility of purposes), and access limitations.
- Interoperable (I): the description of metadata elements should follow community guidelines that use an open, well defined vocabulary. The interoperability procedure must warantee any potential users to use the data and metadata shared (non-discrimination regarding technical capacities of potential users).
- Reusable (R): the description of the essential, recommended, and optional metadata elements should be machine processable and verifiable. Reuse should be easy and data should be citable to sustain data sharing and recognise the value of data". The data must be reused in respect of the right of information and the initial consent given, in respect of any restrictions expressed by data subjects. The lawfulness of the further uses and respect of the licence/MoU must also be warranted.

"Facilitating reuse of data also stands to enhance and raise new risks related to privacy, confidentiality and informational harm" (Boeckhout et al., 2018). If personal data are at stake, an ethics and law review process may be put in place throughout the whole FAIR data process to verify that appropriate technical and organisational actions will be put in place to safeguard the data and metadata shared and the rights of data subjects (Corpas et al., 2018).

Making large scale data exchange across continents possible also offers the opportunity for less marginalisation thanks to more inclusive data and research projects, and thus more equitable access to research benefits[1]. However, this is only an opportunity that has to be seized. Even well intentioned research on data can convey hidden biases and discriminations, from the data itself and/or from the algorithms used to analyse the data (Abiteboul & Stoyanovich, 2019). The Declaration of Taipei of the World Medical Association notably highlights the fact that "The interests and rights of the communities concerned, in particular when vulnerable, must be protected, especially in terms of benefit sharing" (WMA Declaration of Taipei, Recital 17).

One has to beware both statistical biases and societal biases. There can be a statistical bias in the model, in the way that data is presented, summarised, findable etc. However, there can also be a societal bias in the data itself, in the way it has been collected and in the way it represents the real world. Then, algorithms can also be biased themselves, independently from that data, the selection algorithm (that permits to find the data and access it) as well as the research algorithm itself. There is

[1] The CARE Principles for Indigenous Data Governance provide for a good illustration of concerns for marginalisation of minorities and collective benefit of data sharing, https://www.gida-global.org/care (last accessed 18 December 2019).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

still a need to develop mechanisms to ethically validate research methodologies of data research. Unless in order to recall basic principles and fundamental rights, there is still very little legal literature on these issues because of the tight links of such an evaluation with very technical and complex aspects of big data and ethics of algorithms, and the need for close interdisciplinary collaboration and early incorporation of ethical and legal norms.

This is particularly difficult when it includes low and middle income countries (Pisani et al., 2016) as is the case in CINECA. The risks that are raised in the literature is for international data sharing to mostly benefit richer countries and institutions. In that aspect, it can be useful to remember some basic but universal human rights that should be protected from discrimination in access to health benefits from research. In fact, Article 27.1 of the Universal Declaration of Human Rights promotes the right "to share in scientific advancement and its benefits". Besides, Article 2 of the Universal Declaration on Bioethics and Human Rights also aims to "promote equitable access to medical, scientific and technological developments as well as the greatest possible flow and the rapid sharing of knowledge concerning those developments and the sharing of benefits". Finally, the International Bioethics Committee of the UNESCO released a report on big data and health in 2017 that is emphasising on the "responsibilities of all stakeholders" regarding benefit sharing and big data being considered as "a common good for humankind": "advancements and new opportunities provided by science and technology might help reduce and not deepen the inequalities that prevent many human beings from enjoying the highest attainable standard of health" (IBC, 2017, §§73-74).

Within CINECA indeed, the obstacle to data sharing for instance from Europe to Africa is the uncertainty regarding the existence in all of the 34 African countries involved of data protection legislation that would guarantee an equivalent protection of personal (health related and genomic) data than in GDPR. On the contrary, the goal in CINECA and particularly in work package 7 is to permit data sharing in both directions: Canadian and African data to Europe but also from Europe to Canada and Africa. Once the actual legal gaps will be identified (future deliverable 7.2), one of the output of CINECA will be to promote knowledge and expertise on the elaboration of GDPR compliant data access agreements between partners - particularly when transfers are made from Europe to Africa or Canada.

Meanwhile, there are also international ethics instrument that promote general principles and fundamental rights regarding the collection, storage and use of data regarding human biological samples (CIOMS, Guideline 11; WMA Declaration of Taipei, Recital 11), particularly on such concerns as dignity, autonomy, privacy, confidentiality and discrimination (WMA Declaration of Taipei, Recitals 9-12).

B. Considering Societal issues

Beside ethical and legal issues, also societal issues have an impact on the sharing of research and health data across borders. Literature from the field of biobanking has demonstrated that public attitudes towards the sharing of biological samples and health data are divers and shaped by certain assumptions about the meaning and assigned value of these materials and data (Waldby, 2002, Hoeyer, 2008), especially in their cultural and national context (Reardon, 2017). Related questions of, for instance, ownership, benefit-sharing, return of results, collaborations with the health industry and, not least, transnational sharing influence people's readiness to donate to research and the form of informed consent they are willing to give. A major challenge for such collections is that

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

participants are expected to act as “global citizens” in allowing international access to their data for future research without any personal benefit or return (Burton et al., 2010). Taking these considerations into account is crucial to ensure continued public support and the sustainability of platforms and infrastructures for data sharing.

To meet these demands, public and stakeholder engagement activities have become an important element in good biobanking practice. While trust in research institutions and transparency regarding the use, sharing and processing of health data are of major importance for citizens, patients and research participants (Lemke et al., 2010, Goisauf and Durnová, 2018), understandings on how to build trust have changed. Recent stakeholder engagement approaches no longer assume that publics simply need to be educated about science and biotechnology to build trust, but emphasise an increased engagement with publics and advocacy groups in order to co-produce biobank governance (Burgess, 2014, Cambon-Thomsen et al., 2007). Participatory stakeholder engagement exercises in Canada (for example, O’Doherty et al., 2012), Europe (for example, McCormack et al., 2016, Goisauf and Durnová, 2018) and Africa (for example, Folayan et al., 2015, Jao et al., 2015) have shown that collaboration between stakeholders and engagement of participants, communities and patient advocacy organisations are seen as an integral part of the governance structure, furthermore, that information about data access must be provided and that safeguards must be in place to ensure appropriate data use and secure data sharing – the latter especially in terms of privacy protection. The importance of providing information about data sharing with international partners and the need to engage with research and biobank participants also more actively resonates in the attitudes of European biobanks. New regulations, such as the GDPR, have created new demands in terms of transparency in data sharing and data processing, participant engagement and international collaboration, and have highlighted the need to adapt biobank-based research practices (Goisauf et al., 2019). Moreover, the application of IT for secure use and transnational sharing of data is seen as an important issue for the international biobanking community (Harris et al., 2012).

Literature reflecting on previous experiences with sample and data sharing suggest an early involvement of a wide range of stakeholders. Stakeholders comprise professionals in the field of biomedical research and biobanking, regulatory and advisory bodies, research ethics committees and patient advocates. Furthermore, in contributing biomaterials and data, participants become partners of research practices and – as citizens – gatekeepers for further developments (Prainsack, 2014, Rose and Novas, 2004), suggesting to also consider patients and research participants as stakeholders. Therefore, involving a wide range of stakeholders at early stages of the research process and the translation into practice has been raised as an important factor in the implementation process, especially as stakeholders provide important local knowledge and insights into the values and culture that underpin how ELSI and governance are enacted (Gottweis and Kaye, 2012, Kaye, 2011, Murtagh et al., 2017). These considerations are particularly important for international health and research data sharing and the development of an appropriate governance framework.

C. Legal issues of data sharing with regards to GDPR in CINECA

As preliminary clarification, please note that CINECA does not grant any automatic data access nor any shortcut. As developed in the Data Management Plan (deliverable 7.4) and in deliverable 9.9 on the adequate authorisations for exporting data from Europe, all data access will be governed by a

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.

Data Access Agreement between the cohort owner, the CINECA Principal Investigator, and the Principal Investigator's institute.

Although the scope of application of GDPR is very broad (Article 3 GDPR), it is interesting to note GDPR protection will not always be applicable to CINECA data exchanges and processing as not all of the data will be deemed as personal data. In fact, some cohorts, for instance BIOS and Life Lines, have explicitly expressed their wish to limit data access in CINECA to metadata and/or anonymised data, for which GDPR does not apply.

Similarly, when processing does not (or no longer) require identification, Article 11 GDPR provides that "the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation". In that case and where "the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible", and articles 15 to 20 on data subjects' rights do not apply:

- Article 15: Right of access by the data subject
- Article 16: Right to rectification
- Article 17: Right to erasure
- Article 18: Right to restriction of processing
- Article 19: Notification obligation regarding rectification or erasure of personal data or restriction of processing
- Article 20: Right to data portability

When this is not the case and data exchanges do concern personal data, data processing within CINECA will be particularly sensitive and thus be subject to supplementary protection. In fact, sharing and processing of personal data exacerbates the risk of violations of fundamental rights of research subjects and patients like privacy or non-discrimination, all the more so when the data is sensitive as is the case for genomic and health related data (Lamas et al., 2015). In fact, Article 9.1 GDPR prohibits processing of special categories of personal data, notably "genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation".

However, Article 9.2 GDPR provides for several exceptions and legal bases for processing of sensitive data, among which three are relevant for CINECA. The processing of personal data, especially for secondary use, can sometimes overlap between several purposes and thus several legal foundations (thus the difficulties in deliverable 9.8). The European legislator did not specify how to choose the purpose and corresponding legal foundation (nor if combinations are possible) (Rage, 2019; A29WP, 2018). According to the Directorate-General for Health and Food Safety of the European Commission, the data controller, and thus the sponsor or institution of the investigator, is responsible to choose and justify the legal basis for the processing of personal data (DG Health and Food Safety, 2019). In the case of CINECA, three legal bases could thus work.

- Article 9.2.a GDPR: explicit consent
- Article 9.2.i GDPR: public interest in the area of public health
- Article 9.2.j GDPR: scientific research

The use and reuse of such data at the international level is very dependent on each level of normative (legal or ethical) framework: international, national, regional, or even cohort-specific and

This project has received funding from the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.

on different types of normative instruments (binding and non-binding laws and regulations, policies and guidelines from influential stakeholders in the scientific community or funding bodies) (Tassé et al., 2016).

Data sharing between researchers from cohorts like EGCUT (Estonia), BIOS and LifeLines (The Netherlands) or UK Biobank are not hindered by any major legal obstacle as each national law has to comply with the GDPR.

However, GDPR does mention the cases where personal data would be transferred from Europe to third countries or to international organisations. Any transfer has to comply with GDPR (Article 44 GDPR). For the Swiss cohorts, CoLaus and PsyCoLaus, the situation is simple because Switzerland benefits from an adequacy decision from the European Commission (Commission Decision of 26 July 2000), which ensures that there is an adequate level of protection of personal data, and that the transfer does not need a specific authorisation nor additional guarantees.

As for the European Genome-phenome Archive (the EGA, which is part of the international institution of EMBL), H3 Africa or the Canadian cohorts (CHILD, CLSA and CARTaGENE), there is no adequacy decision from the European Commission. The adequacy decision 2002/2/EC about Canada in fact only concerns commercial organisations. The Canadian cohorts are publicly funded and do not qualify as a commercial organisation.

GDPR does provide for a solution to make sure such a transfer offers appropriate safeguards in Article 46 GDPR. “In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available”. Those safeguards can notably be provided for instance by “a legally binding and enforceable instrument between public authorities or bodies” (Article 46.2.a), or by “contractual clauses between the controller or processor and the controller, processor or the recipient of the personal data in the third country or international organisation" (Article 46.3.a).

The role of CINECA’s management team in that regard is to provide support and make sure cohorts’ activities comply with GDPR. As EGA’s activities are tightly linked and embedded with European activities and partners, it has already made sure any data processing is compliant with GDPR and has made the detailed analysis available online.

This is however not the case for Canadian and African cohorts. So far, data exchanges have only been one-sided: from Canada or Africa to Europe. But CINECA partners do aim at exchanging data in the reverse direction, from Europe to Canada and Africa. It will be crucial, in deliverable 7.2, to critically analyse the different legislations in place and compare them the GDPR, for instance the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada and the Protection of Personal Information Act (POPIA) in South Africa.

In order for European data to be accessed and processed by researchers outside of Europe, this Data Access Agreement will have to make sure European data will be processed in compliance with GDPR. It could thus be highly beneficial, in the realm of CINECA, to develop from the practice of the partners exchanging data, a template Data Access Agreement that would both comply with GDPR and be tailored to the specificities of the Canadian and African cohorts in CINECA. This would be in line with Article 50 GDPR on international cooperation for the protection of personal data which promotes

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.

“cooperation mechanisms", "mutual assistance", stakeholder engagement in discussion and activities to develop “the exchange and documentation of personal data protection legislation and practice, including on jurisdictional conflicts with third countries".

II. GDPR legal basis for secondary processing of data in CINECA

According to its Article 3, GDPR applies to any personal data that is processed in the European Union or that has been collected in the European Union even by a controller not established in the EU (Article 3.2.b). GDPR guarantees should thus apply to personal data from European cohorts when processed in Canada or Africa as well as to personal data from Canadian and African cohorts when processed in Europe. Deliverable 9.6 is providing for a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants. The combination of rights and freedoms to be guaranteed will depend on the legal basis for processing: consent, scientific research or public interest.

A. Consent as a legal basis for data reuse for research

“Explicit consent” can be used as a legal basis in GDPR for (primary or secondary) processing of sensitive data such as genomic or health related data (1). In the specific case of secondary processing of personal data for scientific research, GDPR also gives specific provisions raising the question of granularity of consent, which might be particularly relevant for the CINECA project (2) however difficult it might be to interpret. In that regard, the Data Use Ontology (DUO) is a tool that is being used and developed further by CINECA partners which will help promote systematic and reliable interpretations of consent forms regarding secondary processing of personal data for research (3).

1. Consent to sensitive data processing in GDPR

Please note that all cohorts are responsible for respecting the local legal and ethical requirements for obtaining informed consent. All consent forms or templates have been collected in deliverable 9.3, sometimes even more general informed consent guidelines for the cohort or for the consortium.

Processing of personal data can be based on explicit consent Article 9.2.a GDPR (referring to Article 6.1.a GDPR) when it is genomic or health related data as is the case in CINECA. According to Article 4.11 GDPR, consent to data processing must be freely given, specific, informed and unambiguous. This consent can be withdrawn at any time (Article 7.3 GDPR).

As the Article 29 Working Party (A29WP) notes: “The GDPR prescribes that a ‘statement or clear affirmative action’ is a prerequisite for ‘regular’ consent. As the ‘regular’ consent requirement in the GDPR is already raised to a higher standard compared to the consent requirement in Directive 95/46/EC, it needs to be clarified what extra efforts a controller should undertake in order to obtain the ‘explicit’ consent of a data subject in line with the GDPR” (A29WP, 2018). Of course a written statement would be the obvious and ideal way to make sure consent is explicit and most of all to be able to prove it in case of any doubt. However, as the A29WP clarifies, GDPR does not prescribe explicit consent to necessarily be a written and signed statement. Explicit consent can be obtained electronically and even orally, but it would be difficult to prove. The A29WP suggests a two stages verification process making sure that explicit consent is valid: for instance by requiring the data subject to reply “I agree” by email to the controller giving the information about processing and then

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

later confirming agreement by clicking a verification linked or entering a code received per SMS. Another suggestion of best practice when data is kept for a longer period is to refresh consent at appropriate intervals by providing the relevant information again (on the processing as well as on the rights of data subjects) (A29WP, 2018).

2. Consent to secondary processing of personal data for scientific research in GDPR

One main issue observed in CINECA regarding consent is the difficulty to distinguish between the consent to participate in research and the consent to data processing for research. The distinction or lack thereof becomes more apparent for secondary processing of data, which is often not dealt with in informed consent templates, especially in older templates and consent forms as some cohorts' data has sometimes been collected years ago. As noticed in deliverable 9.3 and 9.8, ethics approvals are rather general and do not always permit to clearly differentiate.

And in fact, some cohorts' guidelines on informed consent also promote the use of "broad consent" in order not to hinder secondary use of data for research and permit to still use consent as a legal basis for processing. The issue of secondary processing has even been considered by the European legislator who thought about the granularity of consent (Recital 33 GDPR): "It is often not possible to fully identify the purpose of personal data processing for scientific research at the same time as data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose".

However it is difficult to imagine how consent for secondary processing of personal sensitive data can be both explicit and specific as well as broad. According to Article 29 Working Party (A29WP) on its guideline on consent, Recital 33 GDPR has to be interpreted in a stricter way and with a higher degree of scrutiny when it concerns processing of special categories of data from Article 9 GDPR, as is the case in CINECA. This opinion from the A29WP has been criticised in the literature as it "downplays Recital 33 to the point of becoming nearly non-existent and hence showing an anti-democratic tendency as the legislator inserted that clause in the final text with a purpose" (Van Veen, 2018). This was also stated by the European Parliamentary Research Service, more precisely the Panel for the Future of Science and Technology (STOA, 2019), who confirms that this opinion from the A29WP significantly narrows down the possibility of broad consent for research. Refusing broad consent may discourage researchers to rely on consent and just choose another legal basis. The actual interpretation of this recital may thus require more time and practical experience.

3. Respect of consent instructions and Data Use Ontology (DUO) in CINECA

DUO is a tool that has been developed in GA4GH, which continues to be developed in CINECA and is being used already by several cohorts and consortiums. This Data Use Ontology permits to translate data restrictions into a precise set of "tags" and "terms" that constitute this ontology, in particular the restrictions based on the informed consent of the initial participant for the secondary processing of his or her data. In fact, it is time consuming and difficult for Data Access Committees to have a clear understanding of the different informed consent forms: different language, different expressions and subtleties which are difficult to translate into clear data use conditions. This is what DUO is offering to clarify on a large scale thanks to a unique language system categorising the secondary use conditions and interpret the consent forms in a consistent manner. The latent goal of

This project has received funding from the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.

DUO is also to encourage researchers and data collectors to align to this ontology when creating or adapting their informed consent forms. In fact, semantic interoperability (Liyanage et al., 2015) will be all the more challenging and helpful in the context of CINECA and data sharing between such different consent cultures as Europe, Africa and Canada.

One important stage of the Data Use Ontology is the correct coding of the data restrictions, which is based on current literature on consent codes (Dyke et al., 2016) and metadata profiles of regulatory restrictions (Woolley et al., 2018) and will need further and continuous demonstration of accuracy with ongoing use. DUO aims of course to the automated access to data. However, it is not planned at present for it to be used without human supervision. It is used as a tool to both ease up and speed up the access process. A fully automated process could present risks of violating data subjects' privacy and autonomy and therefore, it is not encouraged to make a fully automated process without gaining more experience and retrospective ethical and legal analysis on these risks.

One underlying question of the use of DUO is also the question of broad consent, or of "bridging consent" (Budin-Ljøsne et al., 2011). This expression refers to the numerous cases in which data sharing is not anticipated in consent forms, especially in disease-specific research settings as mentioned earlier. DUO might offer ways to better identify restrictions or a lack of indications and offer other solutions to interpret consent forms in one way or another.

B. Other legal bases for secondary processing

The goal of CINECA is to promote secondary processing of health and genomic data. Article 6.4 GDPR offers more precise safeguards on the case of secondary processing that is not based on consent:

"Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia:

a. Any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;

b. The context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller;

c. The nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, (...);

d. The possible consequences of the intended further processing for data subjects;

e. The existence of appropriate safeguards, which may include encryption or pseudonymisation".

- Refer to deliverable 9.2 in order to know more about the ethics risks of data processing activities within CINECA, including a Data Protection Impact Assessment.
- Refer to deliverable 9.6 for a description of the technical and organisational measures that will be implemented in CINECA to safeguard the rights and freedoms of the data subjects/research participants.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

- Refer to deliverable 9.7 to get a description of how the data minimisation principle will be observed in CINECA.

1. *Article 9.2.i GDPR: Scientific research*

The exception concerns the cases when “Processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject”. Is this exception only works for secondary use?

If scientific research is the legal basis for processing of sensitive personal data in CINECA, Article 5.1.b and Article 89.1 GDPR apply. The first describes a presumption of compatibility of purposes, only applicable for scientific research, between the primary purpose for initial data collection and later purpose for secondary processing. The second one describes the necessary technical safeguards for rights and freedoms of data subjects, which are described on deliverable 9.6.

Article 89.2 GDPR: Using the legal basis of scientific research for secondary processing permits also to provide for derogations from different data subject rights:

- Article 15: Right of access by the data subject
- Article 16: Right to rectification
- Article 18: Right to restriction of processing
- Article 21: Right to object

2. *Article 9.2.j GDPR: Public interest in the area of public health*

This exception for processing of sensitive personal data could also be relevant for CINECA. It concerns the cases when “Processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy”.

More derogations to fundamental rights of data subjects are possible when using this legal basis. However, the compatibility of purposes for secondary processing will not be automatic as the latter is only applicable in the case of scientific research as explained above.

Nevertheless, using this legal basis for secondary processing allow for more derogations to data subjects rights. In fact, “in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes”, Article 89.3 GDPR permits also to provide for derogations from different data subject rights regarding:

- Article 15: Right of access by the data subject
- Article 16: Right to rectification
- Article 18: Right to restriction of processing
- Article 19: Notification obligation regarding rectification or erasure of personal data or restriction of processing
- Article 20: Right to data portability

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

- Article 21: Right to object

C. Identifying vulnerable groups in CINECA and disentangling ambiguities

CINECA cohorts' data comprehend various groups of people, including vulnerable people and sometimes even exclusively vulnerable people (cohorts focused on children like CHILD, on older adults like CLSA or on data subjects from low resource countries like some of the cohorts from the H3Africa consortium). As such groups often benefit from a stricter protection it is particularly important to identify those vulnerable categories in the main applicable legislations in order to flag potential legal gaps, especially as vulnerable categories might differ from one legal framework to another. Identifying those gaps and finding appropriate solutions without hindering data sharing and thus impeding research is crucial for vulnerable groups not to be marginalised from health research benefits because of an overly cautious and restrictive approach of the reuse of their data.

Regarding vulnerable groups there is an ambiguity because in general, there is a confusion, in biomedical research but not only, between decisional vulnerability and health vulnerability: a vulnerability linked to the ability of protecting one's own interests and make a rational decision versus a rather physical or medical vulnerability (Gennet, 2020). The confusion between both comes from the fact that a health vulnerability in biomedical research can have an impact on cognitive capacities and thus also trigger a decisional vulnerability: this is the case for children or for most mental conditions. However this is not always the case. Pregnant women are able to make rational decisions but are considered vulnerable (CIOMS, 2016, guideline 15). Similarly an older adult can have a physical frailty with any sign of cognitive frailty. Health vulnerability would rather be relevant for actual biomedical research, including more than just data research like for instance clinical trials, and public health research. The ethical issues in that regard would rather concern justice and discrimination issues towards marginalised categories in their access to healthcare and technologies' progress, as mentioned at the beginning of the deliverable, notably regarding recital 17 of the Declaration of Taipei emphasising on the need to promote the interests of vulnerable categories in terms of benefit-sharing.

In this specific part, we will look at two other types of vulnerabilities, decisional vulnerability of course but also vulnerability towards privacy/confidentiality breaches.

1. Vulnerability and "free and informed" consent

The obvious group in GDPR having an explicitly recognised decisional vulnerability are children, as visible in Article 8 GDPR in the specific case of conditions applicable to child's consent in relation to information society services", but also regarding data protection in health research in general (Taylor et al., 2017). In fact, Recital 38 GDPR gives some precisions about the rationale for specific protection: "they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data". Recital 58 GDPR also gives a few precisions on transparency, on the necessity for information to be "concise, easily accessible and easy to understand", notably for children, for which "any information and communication (…) should be in such a clear and plain language that the child can easily understand". As for the actual content and details in this article, there is no need to develop that here as CINECA will not be dealing with information society services.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

However, particular attention should be given so that the consent is actually "freely given" (EDPB, 2019; A29WP, 2018; DG Health and Food Safety, 2019). This implies that the participants should have a "real choice and control". As the EDPB recalls, a clear situation of imbalance of powers between the participant and the sponsor/investigator will imply that the consent is not freely given in the meaning of the GDPR. According to the EDPB: this is for instance the case when a participant belongs to an economically or socially disadvantaged group or in any situation of institutional or hierarchical dependency. As a consequence, as the A29WP notes, consent will not be the appropriate legal basis for most cases. This is why we will present the other legal bases for secondary processing in later developments, although we highly recommend to at least inform data subjects properly even when consent is not the legal basis for processing.

Similarly, there is often a confusion between decisional vulnerability and vulnerability to data/confidentiality breach, independently from any consent issue.

2. Vulnerability and data confidentiality: evaluating the risks of data breaches in CINECA

In that aspect, the particular vulnerabilities, although not explicitly designated as such in the GDPR, actually correspond to all the "sensitive" personal data listed in Article 9 GDPR, including genomic, biometric and health related data. However, even among a particular field of sensitive data, there can be acute vulnerabilities either because the data breach would have more serious consequences, or because the data breach is more probable.

A useful tool to analyse the ethics risks related to data processing and potential data breaches is the Data Protection Impact Assessment (DPIA) in the sense of Article 35 GDPR, which is the object of deliverable 9.2. DPIA could be introducing a tailored process to identify special vulnerabilities regarding data protection in CINECA. In fact, it permits to assess the possible consequences for rights and freedoms of natural persons when the processing is likely to cause specific risks, and in particular when the data is relative to health aspects. This impact assessment, as it obliges to identify the potential risks, could serve as a basis for identifying sources of vulnerabilities for data subjects. First, the controller has to justify the purpose of the processing and the necessity and proportionality of the planned processing regarding the purpose (Article 35.7.a & b GDPR). Second, and this is what is interesting to identify vulnerable people, the controller has to assess the risks to rights and freedoms of natural persons and expose the measures he is planning to put into place in order to mitigate those risks (Article 35.7c & d GDPR). By clearly identifying the risks that are specific to data protection and especially to health-related data protection, the controller will identify the field in which to search for specific vulnerabilities. The A29WP does indeed "encourage the development of sector-specific DPIA frameworks", for instance for specific types of processing or for different types of data (A29WP, 2017, WP248). By searching for ways to mitigate those risks, the controller will have to identify risk-factors, thus further helping the identification of particularly vulnerable people.

The DPIA should not only include the study of the likelihood of risks but also their severity (even if not likely). The likelihood of risks probably refers to technical or organisational factors of vulnerability regarding data protection, i.e. the likelihood of confidentiality breach regarding the security system or type of processing. However, exploring the severity of risks probably implies to explore rather human factors of vulnerability as severity of risks will depend on perception, on different groups of people. And indeed, according to the A29WP, the originality of the GDPR's DPIA in comparison to

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

others relies on the fact that the perspective of data subjects should be taken into account, whereas it is usually limited to technical and organisational aspects of risks.

Two situations obliging the controller to consult the supervisory authority may reflect vulnerability situations. First, particularly vulnerable people could also be identified when the controller thinks that the mitigating measures are not enough to alleviate the risks and that those remain high. If the DPIA shows that the residual risks are still high, the data controller has to consult the supervisory as planned in Article 36 GDPR. "An example of an unacceptable high residual risk includes instances where the data subjects may encounter significant, or even irreversible, consequences, which they may not overcome (e.g.: an illegitimate access to data leading to a threat on the life of the data subjects, a layoff, a financial jeopardy) and/or when it seems obvious that the risk will occur (e.g.: by not being able to reduce the number of people accessing the data because of its sharing, use or distribution modes, or when a well-known vulnerability is not patched)" (A29WP, 2017, WP248). Second, Article 36.5 GDPR obliges data controller to consult the supervisory authority and obtain prior authorisation when the processing takes place as a "task carried out by the controller in the public interest, including processing in relation to social protection and public health", which might be the case in CINECA depending on the legal basis chosen for data processing.

III. Future developments

A. Towards the identification of the legal gaps for deliverable 7.2

This deliverable 7.1 introducing the "catalogue of ELSI issues" for the CINECA project already offers hints of the legal gaps that will be developed in deliverable 7.2 entitled "Catalogue of Canadian, European and African ethical and legal gaps identified". This deliverable will indeed be the occasion to dig deeper into European, Canadian and African legal and ethical frameworks, as well as into the actual governance and practice of the cohorts from the three continents in order to identify the potential differences or even contradictions.

B. Stakeholders engagement on societal issues: preliminary results

Including stakeholders at early stages of the research process and considering ethical, legal and societal issues (ELSI) is key for the establishment of appropriate governance structures. In meeting with CINECAs goal – to develop a federated cloud-enabled infrastructure for making population-scale genomic and biomolecular data accessible across international borders and continents, while keeping with privacy-respecting approaches – engaging with a wide range of stakeholders is a central task in WP7.

Therefore, the aim of the related empirical study in WP7 investigates key requirements for future collaborations, in order to develop a comprehensive health data governance framework that secures privacy, ensures ethical and legal requirements, considers societal matters and ways of engagement. The research design for this study was approved by the INSERM ethics committee (opinion number 19-605). By using the ECOUTER tool (Murtagh et al., 2017), first insights into the expectations, challenges and solutions regarding data generation, usage and sharing have been gained at two events: (1) The perspective of stakeholders from African have been gathered during the 5th African Conference on Emerging Infectious Diseases in Abuja, Nigeria, 7-9.8.2019, and (2) members of the

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

BBMRI-ERIC Stakeholder Forum patient pillar participated in the exercise during the Europe Biobank Week 2019 in Lübeck, Germany, 8.-11.10.2019.

Given the ethical, legal and societal contexts of these first to engagement events, the preliminary findings highlight different, but equally important aspects: (1) African stakeholders highlight the importance of equal partnerships with African countries for fair sharing of data and benefits among all involved stakeholders. Respective mechanisms and safeguards need to be put in place to ensure privacy and to protect patient data from misuse and exploitation, as well as to acknowledge the contributions of the research partners. Solid governance structures, continuous stakeholder engagement and joint development of common standards are needed to build trust and to establish a fair, transparent and functioning data-sharing flow between all partners. (2) Patient advocates from the BBMRI-ERIC stakeholder forum considered the role of regulation (such as the GDPR) for data sharing and the use of personal data. It was discussed how/if legal tools can protect research participants whose (personal) data will be shared and if opt-out and withdrawal is even possible in today's datafied world. (Re)consent was mentioned as an issue and the role of information and understanding of what data actually is for participants to make informed decisions. This was identified as an area of risk together with the question of how data is used and shared by whom for which purposes, and who is in control. Another aspect was the importance of returning results to participants. These considerations have been framed with the more general perspectives/understandings of common sense, pragmatism and the greater good, while trust was put in the centre.

These findings and further engagement exercises throughout CINECA will build the basis to develop accurate and well-grounded ethical and legal recommendations for fair data-sharing between Europe, Africa and Canada.

IV. Work Package 9 requirements' feasibility issues

The added Work Package 9 on compliance with ethical and legal requirements has sometimes been challenging to adapt in the situation of CINECA and to the practice of the cohorts.

Some requirements were confusing as they did not seem to correspond to any existing document in practice, for instance the declaration of compliance with national legislations (deliverable 9.1) or the relevant authorisations for the processing of previously collected data (deliverable 9.8). None of the cohorts were able to provide for such a document and directed us to the initial ethics approval provided with deliverable 9.4 or the informed consent forms provided with deliverable 9.3. One big and recurring challenge in responding to Work Package 9 requirements was also probably coming from the lack of practice and of legal literature regarding GDPR's very recent character, for instance regarding data exchanges to and from outside of Europe (deliverable 9.9) or to reply to the question of whether national legislations were stricter than GDPR regarding health and genomic data (deliverable 9.1). In that aspect, the difficulty could also come from the fact that CINECA involves so many different national legislations and from three different continents: H3 Africa is a consortium involving 34 African countries already.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

V. References

A. Legal documents (ante chronological order)

1. South Africa

Protection of Personal Information Act (POPIA), Act 4 of 2013.

2. Canada

Personal Information Protection and Electronic Documents Act in Canada (PIPEDA)

3. European Union

a. Normative instruments (binding or non-binding)

Resolution of the European Parliament of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement.

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

Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland.

b. Institutional reports

Article 29 Working Party (A29WP)

A29WP, Guidelines on consent under Regulation 2016/679, adopted on 28 November 2017, as last revised and adopted on 10 April 2018, WP259 rev.01.

A29WP, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purpose of Regulation 2016/679, adopted on 4 April 2017, as last revised and adopted on 4 October 2017, WP248 rev.01.

A29WP, Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, adopted on 3 October 2017, WP251.

European Data Protection Board (EDPB)

EDPB, 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)), adopted on 23 January 2019.

European Commission

Independent high-level expert group on artificial intelligence, Ethics guidelines for trustworthy AI, April 2019.

Directorate-General for health and food safety, (revised) Questions and answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation, 10 April 2019.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

European Parliament

Panel for the Future of Science and Technology (STOA), European Parliamentary Research Services. How the General Data Protection Regulation changes the rules for scientific research, PE 634.447, July 2019.

4. *Council of Europe*

a. Normative instruments (binding or non-binding)

Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health related data, adopted on 27 March 2019.

Recommendation CM/Rec(2010)13 of the Committee of Ministers to member States on the protection of individuals with regard to automatic processing of personal data in the context of profiling, adopted on 23 November 2010.

Council of Europe, Convention for the Protection of Individuals with regard to automatic processing of personal data, ETS n°108, 28 January 1981.

b. Institutional reports

Consultative Committee of the Convention for the Protection of individuals with regard to automatic processing of personal data, Guidelines on the protection of individuals with regard to the processing of personal data in the world of big data, Strasbourg, 23 January 2017.

5. *United Nations*

UNESCO, International Bioethics Committee, Report of the IBC on big data and health, Paris, 15 September 2017.

Universal Declaration on Bioethics and Human Rights, adopted by the General Conference of UNESCO on 19 October 2005.

Universal Declaration of Human Rights, General Assembly of the United Nations, Resolution 217 A, adopted in Paris, 10 December 1948.

B. Policy documents

Organisation for Economic Co-operation and Development (OECD), Recommendation of the OECD Council on Health Data Governance, 17 January 2017.

Council for International Organizations of Medical Sciences (CIOMS), International Ethical Guidelines for Health-related Research Involving Humans, in collaboration with the World Health Organization (WHO), Geneva, 2016.

FORCE11, *Guiding principles for findable, accessible, interoperable and re-usable data.* Publishing version B1.0, https://www.force11.org/fairprinciples (last visited: 18/12/2019).

Global Alliance for Genomics and Health (GA4GH), Framework for responsible sharing of genomic and health-related data, 10 September 2014.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

World Medical Association (WMA), Declaration of Taipei on ethical considerations regarding health databases and biobanks, adopted by the 53rd WMA General Assembly Washington DC, USA, in October 2002, and revised by the 67th WMA General Assembly in Taipei, Taiwan, October 2016.

World Medical Association (WMA), Declaration of Helsinki - Ethics principles applicable to medical involving human subjects, adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013.

C. Literature

ABITEBOUL, S. & STOYANOVICH J. 2019. Transparency, Fairness, Data Protection, Neutrality: Data Management Challenges in the Face of New Regulation. J*ournal of Data and Information Quality (JDIQ) - Special Issue on Combating Digital Misinformation and Disinformation and On the Horizon*, 11(3), 15.

BOECKHOUT, M., ZIELHUIS, G.A., BREDENOORD, A.L., 2018. The FAIR guiding principles for data stewardship: fair enough? *European Journal of Human Genetics*, 26:931-936.

BURGESS, M. M. 2014. From ‘trust us’ to participatory governance: Deliberative publics and science policy. *Public Understanding of Science*, 23**,** 48–52.

BUDIN-LJØSNE, I., TASSÉ, A. M., KNOPPERS, B. M. & HARRIS, J. R. 2011. Bridging: from toll bridges to lift bridges? *BMC Medical Genomics*, 4, 69.

BURTON, P. R., FORTIER, I. & KNOPPERS, B. M. 2010. The Global Emergence of Epidemiological Biobanks: Opportunities and Challenges. In: KHOURY, M. J., BEDROSIAN, S. R., GWINN, M., HIGGINGS, J. P. T., IOANNIDIS, J. P. A. & LITTLE, J. (eds.) *Human Genome Epidemiology: Building the Evidence for Using Genetic Information to Improve Health and Prevent Disease*. New York: Oxford University Press.

CAMBON-THOMSEN, A., RIAL-SEBBAG, E. & KNOPPERS, B. M. 2007. Trends in ethical and legal frameworks for the use of human biobanks. *European Respiratory Journal*, 30**,** 373-382.

CORPAS, M., KOVALEVSKAYA, N., McMURRAY, M., NIELSEN, F., 2018. A FAIR guide for data providers to maximise sharing of human genomic data, *PLOS Computational Biology*, 14(3), e1005873.

DEMOTES-MAINARD, J. ET AL. 2019. How the new European data protection regulation affects clinical research and recommendations? *Therapies*, 74,31-42.

DOVE, E. S. ET AL. 2016. Ethics review for international data-intensive research. *Science*, 351(6280), 1399-1400.

DYKE, S. O. M. ET AL. 2016. Consent codes: upholding standard data use conditions. *PLOS Genetics*, 12(1), e1005772.

FOLAYAN, M. O., BROWN, B., HAIRE, B., YAKUBU, A., PETERSON, K. & TEGLI, J. 2015. Stakeholders’ engagement with Ebola therapy research in resource limited settings. *BMC Infectious Diseases*, 15**,** 242.

GENNET, É. 2020. Vulnérabilité et essais cliniques. Réflexions en droit européen. *RGDM*, 74, in press.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

GOISAUF, M. & DURNOVÁ, A. 2018. From Engaging Publics to Engaging Knowledges: Enacting "Appropriateness" in the Austrian Biobank Infrastructure. *Public Understanding of Science*, 28**,** 275-289.

GOISAUF, M., MARTIN, G., BENTZEN, H. B., BUDIN-LJØSNE, I., URSIN, L., DURNOVÁ, A., LEITSALU, L., SMITH, K., CASATI, S., LAVITRANO, M., MASCALZONI, D., BOECKHOUT, M. & MAYRHOFER, M. T. 2019. Data in Question: A Survey of European Biobank Professionals on Ethical, Legal and Societal Challenges of Biobank Research. *PLOS ONE*, 14**,** e0221496.

GOTTWEIS, H. & KAYE, J. 2012. Biobanks for Europe. A challenge for governance. Report of the Expert Group on Dealing with Ethical and Regulatory Challenges of International Biobank Research. Brussels: European Commission Directorate-General for Research and Innovation.

HARRIS, J. R., BURTON, P., KNOPPERS, B. M., LINDPAINTNER, K., BLEDSOE, M., BROOKES, A. J., BUDIN-LJØSNE, I., CHISHOLM, R., COX, D. & DESCHÊNES, M. 2012. Toward a roadmap in global biobanking for health. *European Journal of Human Genetics,* 20**,** 1105-1111.

HOEYER, K. 2008. The ethics of research biobanking: a critical review of the literature. *Biotechnology and Genetic Engineering Reviews*, 25**,** 429-452.

JAO, I., KOMBE, F., MWALUKORE, S., BULL, S., PARKER, M., KAMUYA, D., MOLYNEUX, S. & MARSH, V. 2015. Involving Research Stakeholders in Developing Policy on Sharing Public Health Research Data in Kenya:Views on Fair Process for Informed Consent, Access Oversight, and Community Engagement. *Journal of Empirical Research on Human Research Ethics*, 10**,** 264-277.

KAYE, J. 2011. From single biobanks to international networks: developing e-governance. *Human Genetics*, 130**,** 377.

LAMAS, E., BARH, A., BROWN D., JAULENT M.-C. 2015. Ethical, legal and social issues related to the health data-warehouses: re-using health data in the research and public health research. In: CORNET, R. ET AL. (eds) *Digital Healthcare Empowering Europeans,*EFMI.

LEMKE, A. A., WOLF, W. A., HEBERT-BEIRNE, J. & SMITH, M. E. 2010. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics*, 13**,** 368-377.

LIYANAGE, H., KRAUSE, P. & DE LUSIGNAN S. 2015. Using ontologies to improve semantic interoperability in health data. *Journal of Innovation in Health Informatics*, 22(2), 309-315.

MCCORMACK, P., KOLE, A., GAINOTTI, S., MASCALZONI, D., MOLSTER, C., LOCHMÜLLER, H. & WOODS, S. 2016. 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. *European Journal of Human Genetics*, 24**,** 1403–1408.

MONDSCHEIN, C. F. & MONDA, C. 2019. The EU's General Data Protection Regulation (GDPR) in a research context, In: KUBBEN, P., DUMONTIER, M. & DEKKER, A. (eds), *Fundamentals of clinical data science*, Springer, Cham, 55-71.

MURTAGH, M. J., MINION, J. T., TURNER, A., WILSON, R. C., BLELL, M., OCHIENG, C., MURTAGH, B., ROBERTS, S., BUTTERS, O. W. & BURTON, P. R. 2017. The ECOUTER methodology for stakeholder engagement in translational research. *BMC Medical Ethics*, 18**,** 24.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

O'DOHERTY, K. C., HAWKINS, A. K. & BURGESS, M. M. 2012. Involving citizens in the ethics of biobank research: Informing institutional policy through structured public deliberation. *Social Science & Medicine*, 75**,** 1604-1611.

OHMANN, C. ET AL. 2017. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*, 7, e018647.

PISANI, E. ET AL. 2016. Beyond open data: realising the health benefits of sharing data. *BMJ*, 355,i5295.

PRAINSACK, B. 2014. The Powers of Participatory Medicine. *PLOS Biology*, 12**,** e1001837.

RAGE, V. 2019. La base juridique du traitement des données de santé dans le cadre des recherches impliquant la personne humaine. *Panorama de droit pharmaceutique*, 6, 163-180.

REARDON, J. 2017. *The postgenomic condition: ethics, justice, and knowledge after the genome*, Chicago and London, University of Chicago Press.

RISO, B. ET AL. 2017. Ethical sharing of health data in online platforms – which values should be considered? *Sciences, Society and Policy*, 1, 12.

ROSE, N. & NOVAS, C. 2004. *Biological citizenship,* Blackwell Publishing.

SHABANI, M. & BORRY, P. 2018. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*, 26(2), 149.

TASSÉ, A.-M., BLEDSOE, M. J., GIEPMANS, L. & RAHIMZADEH, V. 2016. Legal and ethical implications of data sharing in international biobanking research: toward a global response. *Biopreservation and biobanking*, 14, 3,

TAYLOR, M. J., DOVE, E. S., LAURIE, G. & TOWNEND D. 2017. When can the child speak for herself? The limits of parental consent in data protection law for health research. *Medical Law Review,* 26(3), 369-391.

TOWNEND, D. 2018. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? *Human Genetics*, 137, 657-664.

TOWNEND, D. ET AL. 2016. Streamlining ethical review for data intensive research. *BMJ,* 354, i4181.

VAN VEEN, E.-B. 2018. Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. *European Journal of Cancer*, 104, 70-80.

VICKERS, A. J. 2016. Sharing raw data from clinical trials: what progress since we first asked "Whose data set is it anyway?", *Trials*, 17, 227.

WALDBY, C. 2002. Stem cells, tissue cultures and the production of biovalue. *Health*, 6**,** 305-323.

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I.J.J., APPLETON, G., AXTON, M., BAAK, A., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3:160018.

WOOLLEY, J. P., ET AL. 2018. Responsible data sharing of biomedical data and biospecimens via the "Automatable Discovery and Access Matrix" (ADA-M), *npj Genomic Medicine*, 3, 17.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

VI. Abbreviations

DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
DUO	Data Use Ontology
EGA	European Genome-phenome Archive
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation

This project has received funding from the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.