MAESTRIA

Deliverable 5.2
Authorisations for hosting data from related databases, regulatory and ethical aspects

Date: September, 22 2022

HORIZON 2020 – RIA programme
Digital diagnostics – developing tools for supporting clinical decisions by integrating various diagnostic data

Grant Agreement number:	**965286**
Project acronym:	**MAESTRIA**
Contract start date:	**01/03/2021**
Project website address:	**www.maestria-h2020.com**
Due date of deliverable:	**12/2021 - month 10**
Dissemination level:	**Public**

Document properties

Partner responsible	IMT
Author(s)/editor(s)	Luis Pineda, Winnie Chua, Maud Decraene, Nadjia Kachenoura, Waed Khalek, Stéphanie Khoury, Beatriz Lorente, Louise Meyfroit, Laurie Soulat-Dufour, Marcel Tekook, Henry West
Version	1.0

Abstract

MAESTRIA is a European consortium of 18 clinicians, scientists and pharma industrials who are at the forefront of research and medical care of Atrial fibrillation AF and stroke patients. It will create multi-parametric digital tools based on a new generation of biomarkers that integrate artificial intelligence (AI) processing and big data from cutting-edge imaging, electrocardiography and omics technologies. Dedicated core labs will collect and homogenize clinical data. A number of legal and ethical aspects must be considered and respected in order to transfer, host and use clinical data among the partners of the MAESTRIA consortium. This document reports on these aspects at several levels: European, country and partner; discussing also the challenges for complying with those aspects.

Table of Contents

1. Introduction

The MAESTRIA (Machine learning and artificial intelligence for early detection of stroke and atrial fibrillation) project is an 18-partner Research and Innovation action (RIA) with the objective of developing and validating the first integrative diagnostic digital platform for atrial cardiomyopathy diagnosis. This platform will be designed to provide support for improved diagnostic accuracy that increases effectiveness and efficiency of treatments, as well as prevention of the complications of atrial cardiomyopathy, such as atrial fibrillation and stroke.

The WP5 is led by Institut Mines-Télécom (IMT)[1], a French public higher education institution, federating a group of eight engineering and management schools. The objective of WP5 is to coordinate access and sharing of data between partners and to develop strategies of data integration. To this end, the involved partners will develop procedures, methods and computer models to integrate and validate new biological data and then to annotate and homogenise them. WP5 activities will run concurrently with the development of novel algorithms in WP1-3 and the data organised from WP4 as part of an iterative process.

In MAESTRIA, data access and data sharing strategies must align with a number of legal and ethical aspects. All types of personal data processing are regulated by European legislations and entities. Moreover, data transfers between the participating partners require clearance procedures and security guarantees. Additionally, all the medical studies carried out in the context of the project require a prior approval from the corresponding ethics committee. This deliverable reports on these legal an ethical aspects from the general and specific points of view.

1.1. Scope of the Document

This document extends the initial discussions presented in D5.1 “Report on the feasibility of aggregating health data from different countries” regarding legal feasibility of aggregating health data from different countries. As such, the document evokes the regulations mentioned in D5.1, in particular GDPR, but it does not intend to repeat the concepts and articles already reviewed in the previous deliverable. Instead, the document describes how GDPR aspects are addressed by the consortium. This deliverable enriches the general considerations of D5.1 by introducing specific regulations applicable for medical research as well as local regulations for data sharing at each data provider institution.

[1] Institut Mines-Télécom. https://www.imt.fr/en/

In this document, the terms “partners”, “consortium partners” or “consortium members” might be used indistinctly to denote the participants of the MAESTRIA project as listed in the European Commission portal[2].

1.2. Structure of the Document

Section 2 describes how MAESTRIA addresses general regulations concerning data processing, including European (GDPR) and local French law (CNIL). Section 3 presents a set of internal procedures to obtain clearance for data transfer, as described by partner institutions. Section 4 presents an overview of the data transfer agreement prepared by the consortium. Section 5 discusses the challenge of heterogeneity of data sources and transfer procedures. Section 6 summarizes and concludes the report.

[2] CORDIS. EU Research Results. https://cordis.europa.eu/project/id/965286

2. General Hosting Regulations

This section, as most of the content of the document, is oriented towards addressing the regulatory aspects of hosting MAESTRIA's data in its Data Hub. Unless otherwise stated, the analysis presented refers to MAESTRIA's Data Hub, IMT's TeraLab platform as Data Hub responsible, or the process of bringing MAESTRIA's data into the Data Hub. Nevertheless, the objective of hosting the data is a collective effort of all partners carrying out the essential work of enforcing data protection regulation across every stage of data processing.

2.1. GDPR

The General Data Protection Regulation (GDPR)[3] is a regulation on data protection and privacy in the European Union (EU) and the European Economic Area (EEA), published on April 2016.

As a Europe-wide project that revolves around medical data, MAESTRIA is naturally and undoubtedly subject to the GDPR. Deliverable D5.1 has already addressed the main definitions and articles from the regulation that concern the MAESTRIA project, with special focus on personal data, health data, consent, pseudonymisation and security. This subsection focuses rather on the implications of hosting data in compliance with the GDPR.

According to the European Law on Data Protection[4], data processing "covers a *wide range of operations* performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data". Therefore, data hosting, the subject of this deliverable, is considered as data processing; as such, it is subject to GDPR enforcement.

[3] General Data Protection Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[4] What constitutes data processing? https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing_en

IMT's TeraLab platform manages the Data Hub that will host the data from MAESTRIA. TeraLab will follow the GDPR principles applicable to the services it provides, namely storage, organisation, transmission and, eventually, destruction of data. All data hosted in the Data Hub will be previously pseudonymised. On the other hand, TeraLab will not perform any medical data processing, which is out of the professional capabilities of the TeraLab team. Therefore, all medical-related GDPR principles will be enforced by other partners of the project at the different stages, including, but not limited to, collection, alteration, consultation, use, dissemination, and combination.

The EU-funded project GDPR.EU provides a checklist[5] for self-assessment on GDPR compliance, which allows evaluating the actions that MAESTRIA and the Data Hub administrators have or will put in place to comply with the GDPR. This assessment does not constitute a legal binding; however, it sets the foundations for the legal declaration of compliance to be submitted to the French authorities (see Section 2.3). All articles cited in this section correspond to the GDPR.

Lawful basis and transparency

- *Conduct an information audit to determine what information you process and who has access to it.* A data protection impact assessment will be performed in accordance with Article 35, identifying risks to the rights and freedoms of natural persons and proposing mitigation actions.
- *Have a legal justification for your data processing activities.* Following Article 5 on data minimisation (see next subsection), Article 6 on Lawfulness of processing, and Article 9 on processing of special categories of personal data; all data collection and processing will be supported by medical protocols, standard operating procedures and informed consents from patients participating in the studies.
- *Provide clear information about your data processing and legal justification in your privacy policy*. In compliance with Article 12, patients will be informed by the recruiting partners about the purpose of the data collection, the context of the MAESTRIA project and the fact that data will be pseudonymised before it is made available for MAESTRIA medical studies.

Data security

- *Take data protection into account at all times.* Data protection and privacy preservation are core principles in the MAESTRIA project. The partners follow the principles of "Data protection by design and by default".

[5] GDPR Checklist for Data Controllers. https://gdpr.eu/checklist/

- *Encrypt, pseudonymise, or anonymise personal data wherever possible*: all data collected will be pseudonymised following SOP provided by experts coordinating the project's Core Labs. Correct pseudonymisation will be further verified by the Core Labs before the data is made available for use in studies.
- *Create an internal security policy for your team members, and build awareness about data protection.* Security policies in the Data Hub are implemented by TeraLab in consensus with expert medical partners. Data transfer is made using secure protocols, users require password authentication and data is by default not accessible to users, authorisations will be granted per study.
- *Know when to conduct a data protection impact assessment, and have a process in place to carry it out.* A privacy impact assessment (PIA) will be carried out before the data is transferred into the Data Hub. Moreover, PIAs are being implemented by other partners in the context of their internal procedures for data transfer or collection. (See section 3)
- *Have a process in place to notify the authorities and your data subjects in the event of a data breach.* Notification processes will be clearly specified. TeraLab has experience in writing this type of procedures from previous data-sensitive projects. Since TeraLab will hold only pseudonymised data, the particular case of notification to the patients will pass through the recruiting centres.

Accountability and governance

- *Designate someone responsible for ensuring GDPR compliance across your organization.* A Data Protection Officer will be designated for TeraLab; the whole Data Hub administration team will care for GDPR to be enforced. Other partners have also well-established departments ensuring GDPR compliance.
- *Sign a data processing agreement between your organization and any third parties that process personal data on your behalf.* In the context of MAESTRIA, Data Transfer Agreements (DTA) signed between the parties will set the rules for data transfer and processing. A DTA model is included in the project's Consortium Agreement.
- *Appoint a Data Protection Officer (if necessary)* TeraLab is required by law to appoint a Data Protection Officer in order to be able to submit a declaration of compliance to the French authorities. All concerned partners have or will appoint DPOs as well.

Privacy rights

In compliance with Chapter 3 (Art. 12-23) on Rights of the data subject and the French legislation (MR-004, described in section 2.3) on the procedures for exercising the rights of the persons concerned by the research, MAESTRIA will guarantee that, for the data subjects (patients), it is easy to:

- *Request and receive all the information you have about them.*
- *Correct or update inaccurate or incomplete information.*
- *Request to have their personal data deleted.*
- *Ask you to stop processing their data.*
- *Receive a copy of their personal data in a format that can be easily transferred.*
- *Object to you processing their data.*
- *If you make decisions about people based on automated processes, you have a procedure to protect their rights.*

In this case, the data subject should contact their recruiting centre, which shall forward the request to the Data Hub administrators.

2.2. Data Minimisation Principle

As defined by the European Data Protection Supervisor[6], the principle of "data minimisation" means that a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose. They should also retain the data only for as long as is necessary to fulfil that purpose.

The data minimisation principle is recalled in Article 5(1)(c) of the GDPR. The EU Regulation 2018/1725 "on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data" also addresses it on Article 4(1)(c). Both express that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".

Data minimisation is a key concern in MAESTRIA and it is being addressed through several strategies:

- The standard operation procedures for data collection and storage determine the only data that should be collected for the purpose of the MAESTRIA studies, no irrelevant data shall be collected. The patient is informed about the collected data. Moreover, the Core Labs will perform data quality and compliance checks before the data is transferred to the Data Hub.

[6] European Data Protection Supervisor Glossary. https://edps.europa.eu/data-protection/data-protection/glossary/d_en

- Patient data at the Data Hub will be pseudonymised, meaning that personally identifiable information – irrelevant to the study – will not be kept.
- A session on Data Minimisation has been included in the Workshop on data interoperability, held in June 2022 by the data officers from the project. One goal of the discussion was to point out the importance of using, from a technical point of view, only the necessary variables/data from a patient. The main aspect that came out from this session is the importance of minimization and how it not only can help resolve ethical issues but also positively affect the technical outcomes of the project by improving learning models through eliminating over-fitting.
- Data minimisation will be enforced for partners making use of data hosted in the Data Hub. The datasets stored in the Data Hub will not be openly available to (authenticated) members of the project. Instead, the datasets made available to each user shall correspond only to the data required for the studies in which they take part.

On the other end, data minimisation is a beneficial practice from the technical point of view. It allows reducing unnecessary data storage, to monitor and avoid data duplication, and to facilitate data accessibility through an organised and clean data catalogue.

2.3. CNIL / MR-004

The CNIL – French National Commission on Informatics and Liberty ("*Commission Nationale de l'Informatique et des Libertés*") is the French regulatory body in charge of ensuring that data privacy law is applied to the collection, storage, and use of personal data. The CNIL carries out its duties in accordance with the French law n° 78-17[7], known as the French Data Protection Act[8] (*"loi informatique et libertés"*). Following the release of the GDPR in 2016, the Data Protection Act has been brought into compliance and consistency with such European regulation through several decrees[9].

MAESTRIA is subject to the CNIL regulations since the project's Data Hub will be hosted by the IMT's TeraLab platform in servers physically located in France. The CNIL provides reference frameworks for the health sector, including benchmarks (*"référentiels"*) and reference methodologies (*"méthodologies de reference"*) to determine

[7] (In French) Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068624/

[8] English-French glossary on data protection. https://www.cnil.fr/en/english-french-glossary-data-protection

[9] (In French) Entrée en vigueur de la nouvelle loi « Informatique et Libertés » et de son nouveau décret d'application. https://www.cnil.fr/fr/entree-en-vigueur-de-la-nouvelle-loi-informatique-et-libertes-et-de-son-nouveau-decret-dapplication

whether a declaration or a request for authorization from the CNIL is required. One reference methodology applicable to health research studies is the MR-004, to which MAESTRIA intends to adhere, as already do several consortium partners in the framework of similar research studies. The remaining of this section aims at providing a non-exhaustive overview of the MR-004; the objective is to analyse how MAESTRIA is positioned with respect to several aspects of the methodology, rather than present the text in full. The texts for this methodology are published in French; the elements described below include extracts from the French text translated into English by the responsible of this document to the best of their capability.

The reference methodology MR-004 applies to "Research not involving the human person, studies and evaluations in the field of health" (*"Recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé"*); it is detailed in the CNIL website[10] and the official reference text is the deliberation no. 2018-155 of the Official Journal of the French Republic[11]. The MR-004 governs the processing of personal data for study, evaluation or research purposes not involving the human person. More specifically, these studies do not meet the definition of research involving humans [directly], in particular studies involving the reuse of data, which is precisely where MAESTRIA positions itself. The research must be in the public interest, which also relates to MAESTRIA given its multiple public partner institutions. The data controller commits to collect only the data that is strictly necessary and relevant to the objectives of the research (data minimisation principle).

2.3.1. Responsible of the Processing

IMT's TeraLab platform is responsible of MAESTRIA's Data Hub. In the framework of MAESTRIA's health research activities, TeraLab's legal entity *IMT Transfert* must submit to the CNIL a declaration of compliance to the MR-004. It is crucial to distinguish that TeraLab is not in charge nor responsible for any medical analysis or medical processing of data. TeraLab's responsibility is limited to the process of storing the project's data and making it available for the different studies while preserving the principle of data minimisation, i.e. provide only the necessary data for a given study and for the concerned researchers.

TeraLab will appoint a Data Protection Officer in charge of keeping up-to-date the list of studies carried out within the framework of the reference methodology in the register of processing activities. In this case, the records contained in the register should describe which data has been made available for which study and for which users.

[10] (In French) Méthodologie de référence MR-004. https://www.cnil.fr/fr/declaration/mr-004-recherches-nimpliquant-pas-la-personne-humaine-etudes-et-evaluations-dans-le

[11] (In French) Délibération n° 2018-155 du 3 mai 2018. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037187498

2.3.2. Data Collection

In compliance with the data minimisation principle, the data collection must be restricted only to relevant, adequate and limited data to what is necessary with regard to the purpose for which it is processed. The need for the data processing must be scientifically justified in the research protocol. In the MAESTRIA case, the cohorts and clinical studies are listed in the Grant Agreement and the corresponding research protocols will be presented by the responsible partners.

The list of categories of patient data and healthcare professional data that can be processed is exhaustive and listed in the reference methodology official text. The health data of patients, their age, their family situation or their lifestyle, for example, can be collected under MR-004. On the other hand, any data not included in the list provided by the reference methodology, is excluded from the reference methodology. For example, the collection of geocoding data, religious opinions, or law infringement records is excluded.

MAESTRIA data collection will follow the list of categories allowed by MR-004. Data collection procedures have been defined by each of MAESTRIA Core Lab responsible and are documented in D4.3, including the procedure for patient pseudonymisation.

2.3.3. Data Retention

In compliance to MR-004, patient data may be kept for up to two years after the last publication of the research results or, in the absence of publication, until the signature of the final research report. A definitive date cannot currently be determined since the project is ongoing and the last research publication is unknown. MAESTRIA's official end date is 28/02/2026, which may bring the data retention period to, at least, 28/02/2028. Beyond the retention deadline, data is subject to archiving for a period of maximum 20 years or in accordance with the regulations in force. The data-archiving period has not yet been defined for the MAESTRIA data.

The personal data of professionals involved in research may not be kept beyond a period of fifteen years after the end of the last research in which they participated.

2.3.4. Informing People (Patients)

Data collection will be carried out by several MAESTRIA partners and recruiting centres coordinated by partners (AFNET, in the case of the MAESTRIA-AFNET 10 prospective study). Patient's informed consent is at the core of the MAESTRIA studies and all concerned patients are provided with the required information regarding the project and study for which their data will be processed.

In addition, MR-004 allows for data collected not specifically for research to be reused without new individual information being given to the persons concerned:

- When the patient already has the information provided in Articles 13 or 14 of the GDPR; this could, for example, concern several research projects, carried out by the same data controller with identical purposes, identical categories of data and identical recipients;
- When the information provided during the collection of data provides for the possibility of reusing the data, and the patient is provided an information system that they can refer to prior to the implementation of each new data processing; for example, a website on which each research project involving their data is listed. Data reutilisation, however, is subject to each country's legislation; for example, the United Kingdom requires a new patient consent for data reutilisation in other studies.

Currently, data reutilisation beyond the MAESTRIA project is under discussion.

In compliance with Article 13 and 14 of the GDPR, patients must be informed particularly about:

- the identity and contact details of the data controller;
- the contact details of the data protection officer of the controller;
- the purpose of the data processing (presentation of the research project);
- the legal basis for the processing (Article 6 of the GDPR);
- the nature of the information that will be used in the research;
- the recipients or categories of recipients of the data;
- the rights of access, rectification, opposition, erasure, limitation of processing;
- the procedures for exercising these rights;
- the optional nature of participation;
- where applicable, the transfer of personal data outside the European Union and the reference to the appropriate safeguards and the means of obtaining a copy or the place where they have been made available;
- the retention period of the personal data or the criteria used to determine this period.

2.3.5. Security and Confidentiality

The data controller must carry out an impact analysis regarding data protection, which must cover in particular the risks to the rights and freedoms of the persons concerned. It implements the appropriate technical and organizational measures to guarantee a level of security adapted to the identified risks. A single analysis may relate to a set of similar processing operations. The data controller must implement and monitor the application of a security and confidentiality policy.

TeraLab will carry out a risk analysis and detail the procedures for risk mitigation concerning the rights and freedoms of the patients. The risk analysis document will be kept on TeraLab's servers and will be made available to the CNIL upon request. Moreover, TeraLab implements technical security measures to guarantee the data security, including regular backups, secure transfer protocols and transfer traceability. In addition, patient confidentiality is assured by the fact that data hosted on the Data Hub will be guaranteed pseudonymised by the data provider partners.

2.3.6. Transfers of Data outside the European Union

MR-004 authorises data transfers outside of the European Union provided that it corresponds to indirectly identifying data of the persons participating in the research, and the directly (or indirectly) identifying data of the professionals involved in the research. The transfer must be strictly necessary for the implementation of the research or the exploitation of its results, under the conditions provided for by the reference methodology. MAESTRIA data will only be transferred outside of the European Union in one case, when the study requires data to be processed on the premises of the University of Oxford; however, this partner is subject to the United Kingdom's Data Protection Act[12] [13], which is UK's implementation of the GDPR, in spite of the country's departure from the European Union. In all cases, data transfers will be subject to Data Transfer Agreements (see Section 4 of this document) between the involved parties, and will not contain personal (directly identifying) information.

[12] The Data Protection Act. https://www.gov.uk/data-protection
[13] Data Protection Act 2018. https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

3. Regulations and Ethical Aspects from Partners

As discussed in D5.1, Section 3.2, internal regulations for sharing data may apply at each of the data providing partners, in particular for retrospective data, but in some cases also for prospective.

The regulations and approval processes among the data provider partners are heterogeneous. While certain partners require no or little paperwork, others have to go through several stages in order to obtain clearance for data transfers out of their premises or servers. Following this observation, the data provider institutions have been asked to provide an open discussion on the data that they will make available in the context of the MAESTRIA project, describing their own internal procedure(s) for clearance. A number of elements were suggested to steer these discussions:

- Description of the provided data,
- position of the responsible for the authorisation,
- applicable internal regulations,
- process or stages to follow in order to obtain clearance,
- implication and requirements from an ethics committee.

Data transfer authorisations procedures might change along the duration of the project following new internal, national or European regulations. The information contained in the next subsections is valid to the date of submission of this deliverable and has been approved by the partners to be published according to the dissemination level of the document. Some of the discussions reflect the current status of a legal or ethical review that is still ongoing. Given the sensitivity of the internal documents and the public character of this deliverable, the documents themselves are not included as part of the deliverable. The documents supporting the procedures described below may be made available to pertinent authorities of the European Commission upon request and subject to approval of the issuing institution.

3.1. AFNET (Germany)

AFNET (Atrial Fibrillation NETwork) is a German academic research organization integrating an interdisciplinary research network of clinicians, scientists, practices, hospitals and institutes, especially in German speaking countries. AFNET is planning to provide anonymised datasets from the retrospective studies described in Table 1. The Informed Consent Forms (ICFs) from these studies were verified and the Ethics votes to use the anonymised data for MAESTRIA were obtained. The AFNET Managing Director

and Data Protection Officer confirmed by letter that the data sets are in compliance with the study-specific ICFs and the valid data protection laws (GDPR). Finally, AFNET has signed the Data Transfer Agreement with the receiving institution.

In addition, AFNET acts as coordinator partner for the setup of the prospective cohort MAESTRIA-AFNET 10, which is expected to recruit approximately 600 patients from 30 sites in 6 different countries (France, Germany, Greece, The Netherlands, Spain and United Kingdom). D4.3 extensively describes the Standard Operating Procedures for the collection of relevant clinical parameters for Atrial Fibrillation from patients, in particular imaging acquisition, processing and storage: (digital-)ECGs, cardiac CT, MRI, and echocardiograms.

Study	Title	# Patients
ANTIPAF-AFNET 2	A randomized controlled clinical trial to investigate the efficacy of Olmesartane (Angiotensin II receptor blocker) to prevent paroxysmal atrial fibrillation [14]	425
Flec SL-AFNET 3	Short-term pharmacological reversal of atrial fibrillation [15]	635
EAST-AFNET 4	Early treatment of atrial fibrillation for stroke prevention trial [16]	2789
AXAFA-AFNET 5	Anticoagulation using the direct factor Xa inhibitor apixaban during Atrial Fibrillation catheter Ablation: Comparison to vitamin K antagonist therapy [17]	674
NOAH-AFNET 6	Non-vitamin K antagonist Oral anticoagulants in patients with Atrial High rate episodes [18]	2538 (as of 16/08/2022)

Table 1. AFNET retrospective studies

[14] ANTIPAF – AFNET 2 Trial. https://www.kompetenznetz-vorhofflimmern.de/en/research/completed-studies-and-registries/clinical-trials/antipaf

[15] Flec-SL – AFNET 3 Trial. https://www.kompetenznetz-vorhofflimmern.de/en/research/completed-studies-and-registries/clinical-trials/flec-sl

[16] EAST – AFNET 4 Trial. https://www.kompetenznetz-vorhofflimmern.de/en/research/completed-studies-and-projects/completed-clinical-trials/east-afnet-4

[17] AXAFA – AFNET 5 Trial. https://www.kompetenznetz-vorhofflimmern.de/en/research/completed-studies-and-registries/clinical-trials/axafa

[18] NOAH – AFNET 6 Trial. https://www.kompetenznetz-vorhofflimmern.de/en/research/ongoing-trials/noah

3.2. AP-HP (France)

AP-HP, Assistance Publique – Hôpitaux de Paris, is the French public health establishment that exercises the role of University Hospital Centre for Paris and Île-de-France. As such, AP-HP is one of the largest partners of the consortium and, in consequence, data sharing procedures and authorisations are usually more complex than for other partner institutions. Several sub-divisions of the AP-HP participate as data providers of MAESTRIA and each of them follows different procedures. The procedures reported by AP-HP are classified by hospital. In particular, the Pitié-Salpêtrière Hospital contributes with data from several studies, as described below.

3.2.1. Pitié-Salpêtrière Hospital

CATS-AF

The study CATS-AF is an interventional research study (Jardé 1)[19]. This type of research study requires the authorization of the *Agence nationale de sécurité du medicament et des produits de santé (ANSM)* and the *French ethics committee (comités de protection des personnes (CPP))*. The study was submitted and accepted by the ANSM on 25/04/2022 and by the CPP on 01/08/2022.

The content within the CATS-AF protocol and consent form were reviewed by the ICAN project manager, the ICAN legal affairs manager, and the project manager of *l'Unité de Recherche Clinique (URC – Clinic Research Unit)* of Pitié-Salpêtrière hospital before the file was submitted to the ANSM and the CPP. Once the CPP approved the study, we received a letter (signed and dated) with the authorization to begin the patient recruitment. We will have an implementation meeting with the team in September to officially launch the study.

The study protocol and consent form provide clear information on the transfer of data to different entities involved in the study, including the transfer of holter ECG signals to IDOVEN via the *Institut Mines-Télécom* (IMT) Data Hub; both IDOVEN and IMT are part of the MAESTRIA consortium. The pseudonymized holter ECG data will be transferred securely from AP-HP (Pitié-Salpêtrière Hospital and data controller) to IDOVEN (data processor) via MAESTRIA data hub platform managed by IMT located in France. IDOVEN, located in Spain, will only have a pseudonymized version of the holter ECG data and will do their storage and analysis without leaving the EU. The analysis results data will be transferred from IDOVEN to APHP via the MAESTRIA Data Hub platform.

[19] French legislation relative to the research involving human person.
(In French) Loi n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (1)
https://www.legifrance.gouv.fr/loda/id/JORFTEXT000025441587/

CT Scans

AP-HP will also provide CT scan data located in the data warehouse of AP-HP, including:

- ***Imaging data*:** all injected cardiac CT scans and additional abdominal and pelvis CT scans when available for all patients taken care in Pitié-Salpêtrière since 2017 or who came back to Pitié-Salpêtrière since 2017. *2017 being the date of the authorization of the CNIL to process data warehouse of AP-HP.*
- ***Reports*** associated to the CT scans.
- ***Clinical data*** associated to the CT scans in a CSV table (45 data).

A thorough Privacy Impact Assessment (PIA) document has been prepared as required by AP-HP's Data Protection Office (*Bureau de la Protection des Données*). The Data Protection Officer (DPO) of AP-HP is responsible for the final validation of the PIA. The ethics committee of the data warehouse is responsible for the final authorization.

Data will be pseudonymized within the data warehouse by AP-HP teams; this will be followed by an internal quality check made by data scientists from AP-HP and data managers from ICAN.

Two requests were made in the process: first, a proceeding to setup a major project ("*dossier montage grands projets*") was presented to the research team of the Centre of Innovation and Data of the Direction of Information Systems (*"Direction de systèmes d'information, pôle d'innovation, pôle innovation et données, DSI/I&D"*). Second, a request to the scientific and ethic committee ("*dossier de demande au Comité scientifique et éthique*" was submitted to the data warehouse ethics committee. The ethics committee made the requested the following elements:

- MR-004[20] questionnaire to check research do not involve humans (only data)
- protocol of project
- questionnaire specific for access to CT scans
- request of clearance for people involved in data processing
- impact analysis

At the same time, regular meetings were held with technical teams.

[20] Reference methodology of the French National Commission on Informatics and Liberty, discussed earlier in this document.

3.2.2. Saint-Antoine Hospital

Regulatory and ethical measures for the Echocardiography Core Lab data processing.

For the performance of WP1

Assistance Publique - Hôpitaux de Paris (AP-HP) will provide echocardiography data to Owkin for machine learning. As per the French regulation ("loi informatique et libertés" and GDPR), AP-HP (data controller) has registered its commitment to compliance with the CNIL, for health data processing for research purposes. A protocol and notice of information have been established that clearly describe the purpose, methods, data flows (in particular between MAESTRIA Partners) and the right to object for patients, at any time. The protocol and patient information form will be submitted to an internal review board, for ethical approval.

For the performance of WP4 (prospective multicentre cohort)

The Echocardiography Core Lab located in Saint Antoine Hospital (AP-HP, SU) will receive echocardiography exams from all the recruiting centres involved in MAESTRIA-AFNET-10 Cohort. Only pseudonymized echocardiographies will be transferred to the Echocardiography Core Lab. Then the Echocardiography Core Lab will provide comprehensive analysis of the pseudonymised echocardiography (Tomtec, Philips). All the Echocardiography pseudonymized data will be finally stored in the Data Hub, to allow data integration between the different modalities (as per the protocol).

Regulatory approval for the AFNET-10 Cohort will be filed by the sponsor (AFNET) in the different countries recruiting patients. The Echocardiography Core Lab (AP-HP) is identified as a joint-controller in the protocol and in the patient Informed Consent Form (ICF).

As per EU regulation, all data sharing and transfer are governed by a contractual framework, including a consortium agreement and data transfer agreement for each category of data transfer. Contracts are reviewed by each institution's DPO. AP-HP has a data privacy policy in place, available online[21]. (https://www.aphp.fr/patient-public/vos-droits/protection-des-donnees-personnelles-information-patient).

Given the complexity of data sharing in MAESTRIA, regular meetings are organized to ensure transparency towards patients and awareness of data protection. Working groups on data minimization and transfer are also in place.

[21] (In French) Protection des données personnelles – Information PATIENT. https://www.aphp.fr/patient-public/vos-droits/protection-des-donnees-personnelles-information-patient

3.3. Sorbonne University (France)

Sorbonne University (SU) and ICAN Core Lab will transfer data in the context of the AFNET-10 prospective study. AFNET will file for national regulatory approval in each country according to each legal framework. In France, AFNET-10 is a research involving human persons (articles L1221-1 and subsequent of French Public Health Code). This type of research study requires the authorization of the French ethics committee ("*comités de protection des personnes (CPP)*"). The study will be submitted by the sponsor (AFNET) in 2022.

The study protocol and consent forms for each local regulatory filing are harmonized. The study protocol and consent forms provide clear information on the transfer of pseudonymised data to different entities involved in the study, including the transfer of MRI to ICAN/SU (located in France) and, after analysis, from ICAN/SU to IMT Data Hub located in France. ICAN/SU Core Lab will receive pseudonymised MRI data (as checked by the Core Lab upon images download). ICAN and SU data processing is validated by respective DPOs and IT security officers. Storage and analysis will be performed without leaving the European Union. The analysis results data (pseudonymised and tagged with the study patient ID) will be transferred from ICAN/SU to the MAESTRIA Data Hub platform (IMT).

3.4. University of Duisburg-Essen (Germany)

University of Duisburg-Essen, Institute of Pharmacology (UDE) is providing protein expression and corresponding patient background data from own patient cohort in pseudonymised form for WP3. Equivalent measurements will be carried out with subsets of tissue from cohorts of other MAESTRIA participants for WP2. The collection of own patient tissue samples and data is approved by the local ethics committee within the study No.: 12-5268-BO. Written informed consent is obtained for each individual patient participating in the study.

Sensitive information to data of the own tissue samples, allowing patient identification, will NOT be provided to the Data Hub. This is ensured by removal of sensitive information (e.g. name, date of birth) already within the internal process of sample collection by responsible persons of our institution. The data finally provided to the Data Hub will be validated and approved by Professor Dobromir Dobrev, Principal Investigator (PI) of UDE for the MAESTRIA project. Data and material transfer agreements with the receiving or providing institution(s) are to be prepared.

3.5. University of Birmingham (United Kingdom)

The University of Birmingham (UOB) provide data that had been collected as part of a previous EU H2020 project (CATCH ME[22] grant no.633196). In general, this is a combined dataset of studies from five European Countries (United Kingdom, Germany, France, Netherlands and Spain) contributed by partners in the CATCH ME Consortium. The original data were retrospectively recorded and pseudonymised for analytical purposes (Chua, et al. 2019) (DOI: 10.1186/s12872-019-1105-4). UOB is the data controller to ensure that data is used, stored, and processed in accordance with the guidelines of the GDPR. All ethical and regulatory procedures had previously been completed during the lifetime of CATCH ME.

To transfer data, a separate Data Transfer Agreement is required to be signed by UOB and the receiving organisation to define their respective responsibilities according to the GDPR (Consortium Agreement clause 4.4). A template for this agreement is available at Schedule 5 of the Consortium Agreement and the terms of Attachment 5 may be used as a basis by parties involved. Similar procedures would apply for any additional dataset identified for transfer within the lifetime of MAESTRIA.

3.6. University of Oxford (United Kingdom)

The University of Oxford will be contributing pseudonymised patient level data for the MAESTRIA AFNET-10 prospective cohort study. The relevant data for these patients will be pseudonymised onsite by clinicians at the University of Oxford before being transferred to the MAESTRIA Data Hub coordinated by IMT. These data will include patients' basic demographics, medical history, clinical risk factors, pathology results and the results of radiology examinations, as well as wearable data from heart monitors. The actual images from radiology tests such as CT scans, echocardiogram and MRIs will also be pseudonymised and then transferred to the Data Hub. The Director of the University of Oxford Academic Cardiovascular Computed Tomography Core Lab will be responsible for the authorisation of the data transfer.

To authorise this collection and transfer of research data the University of Oxford will apply to a UK Health Research Authority Research Ethics Committee (REC) for study approval so that we can approach patients for consent to participate in the study. This is

[22] Characterizing Atrial fibrillation by Translating its Causes into Health Modifiers in the Elderly https://cordis.europa.eu/project/id/633196

formal ethical approval by an independent panel of experts and community members. This will allow the collection of the data and the transfer of the data from the UK to the EU based Data Hub. Internally within the University of Oxford and the Oxford University Hospitals NHS Foundation Trust (the National Health Service site from where the patients will come from) the 'Data protection by design' process will be followed. A Data Protection and Impact Assessment (DPIA) will be completed and submitted for internal approval by the Head of Information Governance and to the Caldicott Guardian of the Hospital. A Personal Data International Transfer Assessment (PDITA) will also be completed for approval. EEA member states including all EU members and EFTA members are on the UK ICO 'white list' for adequacy of regulations for data transfer, meaning approval for UK to EU transfer of research relevant data is possible. These processes all work to ensure data that is collected is the minimum necessary and that patient rights are protected at all stages of the research. A Third Party Security Assessment may need to be completed for the transfer of imaging data via partners of MAESTRIA such as Caristo Diagnostics for CT data. This will also be approved internally.

Formal notification of receipt from the MAESTRIA Data Hub that the data has been successfully received will be required following transfer; also, formal notification of deletion will be required once the study is over and the data is to be removed.

3.7. Summary of Ethical Documents for Medical Studies

Table 2, which was originally included in the Grant Agreement, summarizes the existing cohorts and the prospective cohorts and clinical studies to build within MAESTRIA and their corresponding ethical approval documents. The table reflects the update on the prospective studies/cohorts that have already started or concluded their approval procedure.

EXISTING COHORTS			
Partner	Name of cohort	Task	Ethical documents provided
P2 AP-HP	FASTRHAC	1.2	• Ethical committee approval 14/04/2014 • French Ministry of research approval 15/04/2015 • CNIL (committee for liberty & informatics) approval 24/11/2015 • Patient information & informed consent form version 04/05/2015
	CDW	5.1	• CNIL (committee for liberty & informatics) approval 19/01/2017 (not provided)
P3 UOXF	OxHVF made from:		
	1) ORFAN	1.1	• Ethical committee approval 04/11/2015
	2) Adipo-RedOx	3.3	• Ethical committee approval 13/06/2011
P5 AFNET	ANTIPAF-AFNET 2	4.1	• Ethical committee vote 27/09/2004 • Patient information & informed consent form version 28/08/2004
	EAST-AFNET 4	4.1	• Ethical committee vote 09/02/2011 • Patient information & informed consent form version 14/04/2011
	AXAFA-AFNET 5	4.1	• Ethical committee vote 20/01/2015 • Patient information & informed consent form version 05/09/2014
	NOAH-AFNET 6	4.1	• Ethical committee vote 20/05/2016 • Patient information & informed consent form version 17/12/2015
P7 MU	RACE V Tissue Bank	2.1 2.3	• Ethical committee approval 26/07/2016
	ISOLATION	2.2	• Ethical committee approval 19/12/2019
	CATCH ME Tissue Bank	2.1	• Catch Me MTA/DSA
P8 UoA	ORFAN	1.1	• Ethical committee approval 04/11/2015 (see UOXF)
MAESTRIA TO-BUILD COHORTS			
P2 AP-HP	CATS-AF	2.2	• ANSM (French National Agency for Medicines and Health Products Safety) approval on 25/04/2022 • CPP (French Ethical committee) approval on 01/08/2022
	MRI-HiRA	5.4	• Not submitted yet
P5 AFNET	MAESTRIA-AFNET 10	4.2	• First submission in Germany on 27/06/2022
P7 MU	MAESTRIA AF-ablation	2.2	• Ethical committee approval 19/12/2019, as part of the ISOLATION study
P12 CRCHUS	Dietary fatty acid imaging	3.3	• Not submitted yet

Table 2. Ethical committees' approvals for MAESTRIA cohorts / clinical studies

4. Data Transfer Agreement

In addition to general regulations and partner-specific procedures, data transfers in the scope of MAESTRIA are regulated by explicit agreements between the consortium members.

4.1. Recipient's Engagements

The Consortium Agreement[23], signed by all the partners, in its Article 4.4 addresses the responsibility of the parties (consortium members) regarding the transfer of equipment and/or data. In particular, the Consortium Agreement mentions the recipient party's engagements so that data:

a) will not be used for other purposes or beyond the time frame for which its needed,
b) will not be disclosed to other parties without prior authorisation of the data supplier,
c) will not be used on human subjects,
d) will be used and stored in accordance with the applicable legal and regulatory provisions,
e) item (e) concerns to equipment, not applicable to data,
f) will be deleted as soon as possible by the recipient if the data subject withdraws consent,
g) will no longer be used and will be returned to the supplier (or destroyed) upon request and/or in the event of the termination of the Consortium Agreement,
h) will be used and stored either in the recipient's premises or under their responsibility.

If necessary, the partners participating in a data transfer should sign a separate agreement. This agreement between the concerned parties shall at least reproduce the applicable conditions of the aforementioned Article 4.4 and be based on the Data Transfer Agreement (DTA) model, included as Attachment 5 in the Consortium Agreement.

[23] The MAESTRIA Consortium Agreement document has been finalised on 10/06/2022, the full text is available on the project's portal of the European Commission, to be consulted by authorised participants.

4.2. Transfer Modalities

The DTA establishes in its section III the different transfer modalities with respect to MAESTRIA's Data Hub; stating that personal data can be processed:

1. directly in a workroom on the Data Hub; or
2. on the data user's (recipient) premise if sent directly by the supplier to the recipient; or
3. on the data user's (recipient) premise if sent by the supplier to the Data Hub and then to the data user; or
4. on the supplier's premises when the data user is accessing the data remotely.

In the first case, a DTA will be concluded by supplier, IMT (as host of the Data Hub) and Data User. In the second and fourth cases, a DTA will be directly concluded between a Data Provider and a Data User. In the third case, there would be either (i) two separate DTA, the first one in between the supplier and IMT; and the second one in between IMT and the data user or (ii) a single agreement with the three concerned parties. Each Party undertakes to respect all national data protection legislation that may apply to them as well as the GDPR.

4.3. Transfer Procedures

The DTA model describes also standard transfer procedures, which, similarly to the rest of the document, might be adapted according to the specific transfer scenario being addressed. The transfer procedures assume that data-related prerequisites such as pseudonymisation, quality verifications or regulatory obligations have already been fulfilled.

The standard transfer procedure requires the use of secure protocols for any transfer and lists the steps followed at the Data Hub to verify data integrity and backup the received data. The transfer procedure should be adjusted in the case of modalities 2 and 4 listed above, where the data does not transit nor is stored in the Data Hub. In these cases, a specific detailed procedure between the provider and recipient should replace the standard steps.

5. Discussion: Heterogeneity of Procedures

One challenge for MAESTRIA is to handle the diversity of studies and data sources involved in the project. This, together with the local regulations at each partner's country, lead to a complex set of procedures to transfer, host and actually be able to use such data, as this document has evidenced. In order to limit such complexity, the MAESTRIA partners coordinating the different tasks and work packages have proposed a number of templates. These templates aim at standardising as much as possible the data collection and transfer procedures.

An example of such a template is the Data Transfer Agreement described in the previous section. An Informed Consent form template has also been prepared by AFNET for the recruitment of patients in the framework of MAESTRIA medical studies.

The proposed templates, however, only set the fundamental elements that these documents must include. The content of the template can -and, likely, should- be adjusted or augmented according to each specific scenario. From early stages of the project, several divergent factors have been identified from one procedure to another, or between institutions. For example:

- GDPR is an EU-wide regulation; however, its enforcement might slightly differ among EU members, according to the local legislations. MAESTRIA's Data Hub must comply with French regulations since it is physically located in France, while patient recruitment and informed consent will comply with the local GDPR-enforcing authorities of the recruiting country.

- Standard Operating Procedures (SOP) have been created for data transfers from recruiting centres to the Core Labs in charge of quality validation. A different procedure had to be written for each Core Lab, since the recruitment, pseudonymisation and transfer procedures differ. In addition, some recruiting centres have requested exceptions on these SOP to comply with their internal legal and security requirements.

- The Data Transfer Agreement template considers general transfer scenarios. Part of this template has been derived from the procedures described in the Data Management Plan. Nevertheless, several data transfer scenarios will happen between partners, which might not align with the initially proposed solution: some data will only transit through TeraLab infrastructure; some data will be transferred

directly from one partner to another, completely bypassing the Data Hub. For these cases, the DTA template should detail the adjusted transfer procedure.

- In a specific scenario on data sharing, the United Kingdom may require patient re-consent if patient data were to be used beyond MAESTRIA, while for other countries a single initial consent would suffice. Currently, the consent form enables the patient to allow or reject further use of their medical data; however, the UK's team preparing the Ethics submission expressed that for new studies the patients need to be asked to re-consent. An adjustment to the consent form may be necessary in this case.

The above list reflects only a few factors observed so far, but the consortium is conscious that the list might grow as the project progresses and the recruitment of patients increases. In anticipation of these situations, a thorough work involving the concerned members has been carried out for the conception and validation of these procedures and templates.

6. Conclusion

This report presented an analysis on how the MAESTRIA project partners address the authorisations for hosting, sharing and processing medical data for the project's clinical studies, from a regulatory and ethical perspective.

First, a GDPR self-assessment allowed describing how MAESTRIA positions with respect to the different categories of GDPR principles. Next, the main elements of the reference methodology MR-004 from the French GDPR-enforcing authorities were presented. Each of these elements was briefly discussed in the context of the MAESTRIA project. MR-004 is the reference methodology that legally applies for the medical studies of the project.

Then, several partners provided a discussion of their internal regulations and clearance procedures in order to be able to share data within the project. The discussion was complemented by a table summarizing the approvals from the corresponding ethics committees for each of the MAESTRIA cohorts/studies.

Finally, the report introduced the existing agreements between partners for data transfer and sharing, as well as their corresponding templates. A discussion regarding how these standard documents help the consortium to address the challenge of heterogeneity of procedures and legislations concluded the report.

As previously stated, the information provided in this report is valid to the date of submission. However, several procedures described in it are evolving as the project progresses. Updates on legal, ethical and technical aspects will be provided on the second delivery of the Data Management Plan (D7.3), expected by M40.

7. Acronyms and References

7.1. Acronyms

Acronym	Meaning
AFNET	Atrial Fibrillation Network
ANSM	Agence Nationale de Sécurité du Médicament et des produit de santé (*French National Agency for Medicines and Health Products Safety*)
AP-HP	Assistance Publique – Hôpitaux de Paris
CNIL	Commission Nationale de l'Informatique et des Libertés (*French National Commission on Informatics and Liberty*)
CPP	Comités de Protection des Personnes (*French Ethics Committee*)
CT	Computed Tomography
DPO	Data Protection Officer
DTA	Data Transfer Agreement
ECG	Electrocardiogram
EEA	European Economic Area
EU	European Union
GDPR	General Data Protection Regulation
ICAN	Institute Of Cardiometabolism and Nutrition
ICF	Informed Consent Form
IMT	Institut Mines-Télécom
MR	CNIL Reference Methodology
MRI	Magnetic Resonance Imaging
PIA	Privacy Impact Assessment
SOP	Standard Operating Procedure
SU	Sorbonne University
URC	Unité de Recherche Clinique (*Clinical Research Unit*)

7.2. References

Chua, Winnie, et al. «Development and external validation of predictive models for prevalent and recurrent atrial fibrillation: a protocol for the analysis of the CATCH ME combined dataset.» *BMC Cardiovascular Disorders*, 2019.