HEALTHYCLOUD
Health Research & Innovation Cloud

D2.1 First draft on legal framework for technical safeguards with a focus on cloud usage
Version 0.3

Document Information

Contract Number	965345
Project Website	http://www.healthycloud.eu/
Contractual Deadline	M15, May 2022
Dissemination Level	Public
Nature	Report
Author(s)	Irene Schlünder (BBMRI-ERIC/TMF, WP2), Adrian Thorogood (UNILU, WP2)
Contributor(s)	Harald Wagener (de.NBI-Cloud, WP5) Sina Barysch (de.NBI-Cloud, WP5) Antonios Antonopoulos (EMBL/ELIXIR) Stella Sessi (EMBL/ELIXIR) Salvador Capella (BSC, WP5) Juan González (IACS, WP1)
Reviewer(s)	Ramón Launa (IACS, WP1) Gergely Sipos (EGI, WP5)

Notice: *The HealthyCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement №965345*

	Regina Becker (UNILU, WP2)
Keywords	Legal framework, technical safeguards, organisational safeguards, cloud, usage, privacy by design

Change Log

Version	Author	Date	Description of Change
V0.0	Irene Schlünder	2022/06/22	Table of Contents, Initial suggestions
V0.1	Irene Schlünder, Sina Barysch, Adrian Thorogood, Salvador Capella, Juan Gonzalez, Harald Wagner	2022/07/25	Draft
VO.2	Gergely Sipos, Ramon…, Regina Becker		review
VO.3	all	2022/07/29	submitted

Table of contents

Executive Summary

The ultimate goal of WP2 is to develop a generic data governance model for all FAIR Data Portals making health data available for research. This governance model will describe applicable legal requirements with a focus on the GDPR and outline processes that enable data portals and data providers to share data in a compliant way. This deliverable (D2.1) focuses on identifying legal requirements for privacy and security safeguards that apply while sharing health data for research through a FAIR Data Portal, with a focus on the use of cloud services. It begins by reviewing applicable GDPR principles that must be respected across the data lifecycle, including lawful basis, data minimisation (anonymisation and pseudonymisation), technical safeguards, data subjects' rights, and accountability. The concept of a generic data governance model – applicable across the data lifecycle including submission, storage, access and use – is introduced.

The deliverable then discusses the importance of data security in cloud environments. The cloud offers flexible and scalable storage/compute infrastructure and services to researchers. Cloud environments used in health research are diverse – from hyperscale, to national to local. Clouds must find ways to demonstrate high security standards, e.g., through certifications. Cloud users face the challenge of having to ensure providers have adequate security measures, and must also demonstrate that the measures in place actually correspond to the cloud user's own data protection obligations.

Because many cloud providers are based outside the EU/EEA, cloud users must also consider their GDPR obligations relating to the transfer of personal data to third countries and international organizations. There is currently no reliable transfer mechanism to cloud providers in the US following Schrems II. Standard contractual clauses along with supplementary measures to limit access by foreign governments are one possible approach. A transfer mechanism is also needed for international organisations, though given their independence and internal data protection regulations, these transfers are generally less problematic.

Best practices examples of data governance and security frameworks for the cloud are then provided – including de.NBI Cloud, BIGAN, and the German Medical Informatics Initiative. The deliverable concludes with a list of basic data security requirements including those to be applied while using cloud services. These measures will be further specified by WP5 in Deliverables 5.3 and 5.4. This first draft will be tested in collaboration with WP7 to ensure the align with Healthy Cloud use cases. A second deliverable (D2.2) will follow outlining modular contractual clauses, which will then feed into a final ELSI guideline elaborating on the Generic Governance Model.

1. Background

This deliverable aims to provide a solid overview of GDPR-specific considerations (privacy by design) and requirements to support the selection of specifications necessary for a future cloud environment. Taking the proposed compute solutions by WP5, it will cover considerations applying use of compute resources for data analysis, both in single/central as well as multiple/federated setups and outline appropriate safeguards. This deliverable will also contrast the complementary approaches of data sharing and bringing algorithms to the data, and what requirements either have to ensure controlled data use in a secure environment. As a result of this analysis, a set of guidelines and recommendations are proposed to support the overall architecture. The aim is to expand on the findings on data access (as defined in task 2.2), building on a federated cloud design as specified in WP5 and the use-cases proposed in WP7.

Legal and regulatory compliance of organisational and technical safeguards will be developed in the form of guidelines for performing risk analyses based on qualitative and quantitative metrics. These guidelines will be based on the use cases (WP7) to ensure their applicability to real world scenarios.

2. Definitions

This report relies on definitions from the "Glossary of commonly used terms in the field of health data research" developed by HealthyCloud[1]. Selected terms are reproduced here (in *italics*) for convenience:

- **FAIR Data Portal** *indicates a platform that serves as a contact point and provides a gateway for data providers to grant data users access to data for secondary use in health research. A Fair Data Portal may or may not host data and may or may not decide itself on data access.*
- **Secure Processing Environment** (SPE): *The physical or virtual environment and organisational means to provide the opportunity to re-use data in a manner that allows for the operator of the secure processing environment to determine and supervise all data processing actions, including to display, storage, download, export of the data and calculation of derivative data through computational algorithms[2].*
- **Cloud:** *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models*. [3]
- **Data Controller**: *Under the Data Protection Framework (Regulation (EU) 2018/1725), as well as under the General Data Protection Regulation [GDPR], the data controller is the party that, alone or jointly with others, determines the purposes and means of the processing of personal data. The actual processing may be delegated to another party, called the data processor. The controller is responsible for the lawfulness of the processing, for the protection of the data, and respecting the rights of the data subject,* including but not limited to compliance with the data protection principles

[1] https://zenodo.org/record/5998128#.Yo-mmhPMIsd

[2] Proposal for a regulation of the European Parliament and of the council on European data governance. 2020/0340. ("Data Governance Act") [DGA] https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN)

[3] https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf

of Art. 5.[4],[5]. In the context of sensitive health data, the Data Controller is under some jurisdictions also referred to as Data Custodian and Data Steward, but partly with slightly different meanings.

- **Data processor**: *According to Article 3 (12) of the Data Protection Framework, a processor shall mean "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." The essential element is therefore that the processor only acts "on behalf of the controller" and thus only subject to his instructions.[6] In some cases, the processor may choose not to process the data himself, but may have recourse to a subcontractor who processes the data on his behalf. In practice, this will depend upon the processor agreement entered into with the controller. In the context of this report, where a controller engages an external operator of a Secure Processing Environment, this operator is clearly a "Data Processor", and the term "SPE Operator" also connotes "Data Processor". [Please note that storing data is also a form of data processing.]*
- **Data subject**: *As defined in the GDPR, a data subject is a natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.[7]*
- **Data Provider:** *is the entity who makes data available for secondary use in health research through a FAIR Data Portal.*
- **Data Transfer Agreement (DTA)**: *Legal agreement between a Data Controller and another legal entity providing access to defined data under the control of the Data Controller for purposes defined in the DTA. DTAs have historically been used to enable sharing of sensitive data between organisations (e.g. a hospital providing specific data to a research institution for a research project). There is no standard form of DTA, and with the coming into force of the GDPR, DTAs need to clarify the role of the receiving organisation as either a Data Controller in its own right, or as a Data Processor acting on the instructions of the first Data Controller. (See D2.2 Framework of Modular Contractual Clauses for HRICs for additional guidance and discussion).*

---

[4] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018. Available at: https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32018R1725&from=EN [Data Protection Framework]

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN [GDPR]

[6] [Data Protection Framework] *ibid*

[7] Art. 4(1) [GDPR] *ibid*

3. GDPR Principles to be observed

The GDPR sets the scene for any processing of sensitive data in the EU, and health data usually is considered sensitive. Other EU legislative acts, which may apply in future, are on the horizon, but still not adopted (Data Governance Act proposal, EHDS Regulation proposal, AI Regulation proposal, Data Act proposal, NIS2 Directive proposal). They will, however, not primarily change the landscape in terms of security of data processing (except for elaborating requirements to provide access to data in secure processing environments). Rather these acts aim to clarify the lawful basis of data sharing, to clarify the conditions under which data may be transparently and securely shared, and in some cases to establish an obligation for certain data controllers to make data available for secondary use. As a result, the security principles outlined here will probably not be affected by upcoming legislation. The terms "(secure) cloud usage" or “secure/trusted research environments'' cover diverse contexts and environments, WHICH MAKES IT DIFFICULT TO IDENTIFY WHAT SECURITY MEASURES ARE REQUIRED TO SATISFY LEGAL REQUIREMENTS, as this will always be dependent on context. The problem is analogous to the challenge of defining and designing a compliance programme for a "biobank". Biobanks encompass a structured collection of biosamples and associated data, but also various organizational and governance processes. Identifying legal requirements is not possible without explaining the diverse types of biobanks. Nonetheless, the following provisions will apply to all kinds of platforms and contain key principles:

- Art. 5 GDPR: Principles relating to processing of personal data
- Art. 9 GDPR: Processing of special categories of personal data
- Art. 89 GDPR: Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes
- Art. 25 GDPR: Data protection by design and by default
- Art. 28 GDPR: Processor
- Art. 32 GDPR: Security of processing

For the wording of these provisions see Annex 1.

The GDPR adopts a risk-based approach and so it indicates that the following circumstances need to be considered when the security measures are chosen:

- state of the art of technology;
- the costs of implementation;
- the nature, scope, context and the purposes of processing (see Recital 83 GDPR);
- level of risk of the data processing (see Art. 32 GDPR);
- the sensitivity of the data (e.g., data about health or religion);

- the number of staff at the data controller or processor and the extent of their access to personal data (just a few people can access it, or hundreds; they can read or also edit, delete the data).

Health data are especially protected under the regime of Art. 9 GDPR; they are more sensitive than other data due to their potential of being misused and thus threatening the privacy of patients and citizens. The European Data Protection Supervisor (EDPS) has also conducted a survey of technical and organizational measures required for processing personal data for scientific research throughout numerous countries across the EU[8] . The cornerstones of all secure research environments for processing and sharing health data, however, are more or less clear and can be described as follows:

1. Lawfulness of processing special categories of data

Processing of sensitive data always needs a specific allowance to be lawful. Art. 9 GDPR sets the framework for these special categories of data. The EU landscape of laws providing the appropriate allowance to process data for health research purposes is highly fragmented. Whereas some Member States mainly rely on informed consent when processing health and genetic data[9], others have adopted laws to utilise the derogation clause of Art. 9(2)(j) GDPR also for those data. However, there is a large heterogeneity in the implementation, e.g. sometimes derogations are only applicable to public research stakeholders or under very defined processing contexts.

The proposed Regulation to establish the EHDS aims to provide a Union based derogation for secondary use of health data covering data holders making the data available, health data access bodies responsible for pre-processing and data disclosure and data users pursuing their approved research use. However, it needs to be realised that these derogations, if adopted, require the use of the EHDS mechanisms and will not apply in data sharing and research performed outside the EHDS, which can still take place as before.

In addition, the proposed provisions are not entirely clear yet and will probably be subject to change in the further negotiations between the European Council and European Parliament. The joint Statement of the EDPB and the EDPS on the draft has already pointed to this and other legal uncertainties following from the draft[10].

[8] European Data Protection Supervisor, Study on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research – Final Report (August 2021) https://edpb.europa.eu/system/files/2022-01/legalstudy_on_the_appropriate_safeguards_89.1.pdf

[9] e.g. Germany, see Zenker, S., Strech, D., Ihrig, K., Jahns, R., Müller, G., Schickhardt, C., Schmidt, G., Speer, R., Winkler, E., von Kielmansegg, S.G. und Drepper, J. (2022): Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: Towards a new German national standard. J Biomed Inform 131 S. 104096, https://www.ncbi.nlm.nih.gov/pubmed/35643273

[10] https://edps.europa.eu/data-protection/our-work/publications/edps-edpb-joint-opinions/european-health-data-space_en)

Meanwhile a data provider has to observe the local rules while capturing and delivering data, be it consent or another applicable legal basis in national law.

2. Anonymisation (anonymous use of data?)

Anonymisation of data in the sense of data minimisation is not the rule in research because every anonymisation technique leads to losses in data dimensionality, which makes some research questions impossible to answer. Furthermore, the data can then no longer be linked to a patient or citizen, which excludes follow-up data, so that many questions resulting from intermediate research results cannot be dealt with in a meaningful way without re-collecting an equivalent data set containing the missing links or dimensions. In addition, it becomes impossible to communicate relevant health findings and hence, this could potentially put patients at risk. However, the anonymisation of individual data always becomes important when no legal basis can be found for the further processing of personal data. Targeted anonymisation for specific projects, where very specific values can be dispensed with, still makes sense. However, there are ongoing debates in which cases the anonymisation process as such, which undeniably constitutes data processing, requires its own legal basis. There have even been suggestions made that personal data may be analysed by machine without a specific legal basis, as long as the output results are anonymous (so-called anonymous data access or anonymous data use).

There is also no case law on this topic yet. A final assessment is therefore not yet possible. Any controller should be aware that processing of health and/or genetic data, including anonymisation, may be high risk processing requiring at minimum a data protection impact assessment (DPIA), if not in general a consultation with the supervisory authorities as there are many pitfalls in anonymisation that researchers are not aware of and may therefore misinterpret in their DPIA.

At least the discussion about the permissibility of anonymising data for research has gained momentum. The Federal Commissioner for Data Protection and Freedom of Information in Germany (BfDI) concludes in a position paper from June of this year that, in principle, a compatibility of purpose according to Article 6(4) of the GDPR with the continued validity of the legal basis originally applicable for the collection of the data according to Article 5 (1)(b) in conjunction with Recital 50 of the GDPR is possible for anonymisation processes[11]. In this context, it must be taken into account that anonymisation in the sense of data protection law is only defined by the result, which must meet the criteria of anonymity according to recital 26 of the GDPR. The way in which such a result is achieved, i.e. specifically by counting the

[11] BfDI position paper on anonymisation under the GDPR with special regard to the telecommunications sector. 2020. The Federal Commissioner for Data Protection and Freedom of Information, 29.06.2020, https://www.bfdi.bund.de/DE/Infothek/Transparenz/Konsultationsverfahren/01_Konsulation-Anonymisierung-TK/Positionspapier-Anonymisierung.pdf

data records, another form of aggregation or a coarsening of the individual data in a form that preserves the structure of the data, is not defined in terms of data protection law and is therefore not relevant with regard to the legal evaluation.

3. Other technical safeguards such as pseudonymisation

Pseudonymisation is one well-known measure that can significantly contribute to the safeguards for health data processed for research purposes. This process is not at all new in the design of information systems but gained special attention after the adoption of the GDPR, where pseudonymisation is explicitly referred as a technique which can both promote data protection by design (Article 25 GDPR), as well as the security of personal data processing (Article 32 GDPR). It is also explicitly mentioned in Art. 89 GDPR as default security measure for research.

Pseudonymisation is one of several 'de-identification' techniques (such as aggregation, obfuscation, masking, etc.) intended to remove the association between a set of identifying data and the data subject. Other pseudonymisation definitions such as the ones from ISO, and specifically ISO 25237:2017 Health informatics — Pseudonymization, are built upon this assumption and are similar to the GDPR definition discussed above. Basic pseudonymisation techniques are:

- Counter: Monotonic counter which starts at a certain value and is increased each time a new pseudonym is necessary
- Random number: Random value extracted between a minimum and a maximum boundary each time a new pseudonym is necessary
- Hash function: One-way (non-reversible) cryptographic function transforming input personal data in fixed-length values
- Hash-based message authentication code (HMAC): One-way (non-reversible) cryptographic function adding a key that makes it less predictable than a hash function
- Encryption: Two-way (reversible) cryptographic function transforming an input personal data in values that can be re-transformed in its original format using a key.

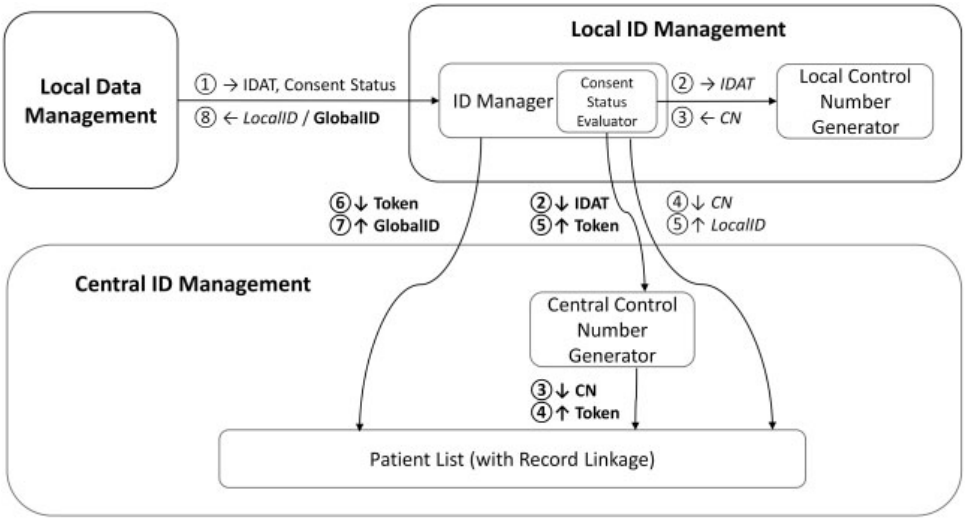
While hash functions can significantly contribute towards data integrity, naive hashing approaches are weak pseudonymisation techniques as they are prone to brute force and dictionary attacks.

A robust approach to generate pseudonyms can be based on the use of keyed hash functions, i.e. hash functions whose output depends not only on the input but also on a secret key (salt).[12] Monotonic counters are easy to generate but may have inherent ordering properties which might weaken their use as a pseudonymisation

[12] For further details see: ENISA: https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques.

technique. Use of random unique identifiers like RFC1422[13] UUIDs may be better suited and can be generated on current systems with little computational overhead. This is in line with the random number approach listed above. Similarly, counters are also considered as weak pseudonymisation technique as they cannot really scale.

Example German Cancer Consortium:



An example pseudonymization flow from the German Cancer Consortium[14]. CN: control number; GlobalID: global ID for network-wide record linkage; IDAT: identifying patient data; LocalID: local ID for one DKTK location. ()

Another issue with pseudonymization flows is that under certain circumstances, re-identification may be required, for example if a research participant or patient revokes his or her consent to the processing. To address these cases, simple one way pseudonymisation is not sufficient, which has given rise to more complex scenarios with intermediaries or data stewardship models, and approaches like EUPID[15], which aims to facilitate linking of data from various sources at the data ingestion point without allowing re-identification from the pseudonymised data. EUPID makes use of hashing and encryption methods to generate the pseudonyms, with a trusted third party holding the encryption keys to facilitate re-identification when needed.

[13] https://datatracker.ietf.org/doc/html/rfc4122

[14] Tremper et al. 2021: MAGICPL: A Generic Process Description Language for Distributed Pseudonymization Scenarios https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0041-1731387

[15] https://eupid.eu/#/home

The EU Commission has launched the pseudonymisation tool "Spider"[16] with interesting functionalities: Record pseudonymisation based on public key infrastructure, linkage of pseudonyms without divulging sensitive personal data, and end-to-end encrypted transfer of pseudonymised data to build cohorts.

It is important to not share the pseudonym widely, otherwise it gets compromised and has then to be changed.

Pseudonymisation is an important step, but not the only security measure to be adopted. The main international Standard on how to manage information security is **ISO/IEC 27001**. The standard was originally published jointly by the International Organisation for Standardisation and the International Electrotechnical Commission in 2005 and then revised in 2013. The certification according to this standard is certainly a seal for trustworthiness. But the certification process is costly and not always necessary for smaller research platforms as long as there are other procedures in place to obtain an independent review of the security measures taken.

For example in Germany, the non-profit organisation TMF e.V.[17] offers a service to the academic health research community to review and discuss data management and protection plans and to give advice on how to improve them. Over the past twenty years, the TMF working group "Data Protection" has consulted more than 100 research projects on the issue of implementing data protection compliance when collecting data and samples. This consultation was based on the generic data protection concepts of the TMF which were coordinated on a national level by the Data Protection Commissioners of the Federation and of the Federal States with the working groups "Science and Research" and "Health and Social Affairs" and published as a TMF book in 2006 (2nd edition 2014). These generic solutions were the first attempt to significantly simplify and accelerate the creation of formally acceptable and nationally implementable data protection concepts for various collaborative research projects as well as the related inspection and coordination process with ethics committees and data protection officers. This concept is intended as a template for researchers' own specific documents, and as an introduction and guide to the complex subject matter.

Another approach are the guidelines of the UK Data Research Alliance[18] on how to establish Trusted Research Environments (TREs), a comprehensive and trustworthy solution that provides the overall services to data discovery, data access request and data analysis, that implies a methodical operation at various levels. The guidelines and principles to set up such a TRE are structured around the "Five Safes"

[16] https://eu-rd-platform.jrc.ec.europa.eu/spider/

[17] https://www.tmf-ev.de/EnglishSite/Home.aspx

[18] https://www.hdruk.ac.uk/news/new-principles-published-to-improve-public-confidence-in-access-and-use-of-data-for-health-research-through-trusted-research-environments/

framework for the access of health data – safe people, safe projects, safe settings, safe data, safe outputs. In the Deliverable 5.1, this framework has been analysed and contextualised to multiple European computational facilities, providing a summarised set of recommendations to set up a "secure cloud" approach (named as SPE in the deliverable), and also including a mapping to the specific technological solutions adopted in such facilities.

Finally, privacy preserving data sharing techniques are gaining ground. Those are especially important in federated infrastructures, where data from different resources are uploaded and/or linked to be analysed. There are different approaches with different trade-offs regarding data privacy on the one hand and usefulness for research on the other hand.[19]

In addition, a **Data Protection Impact Assessment** has to be conducted, wherever a large amount of sensitive data are processed.

Art. 35 GDPR: Data protection impact assessment

1. *[1]Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. [2]A single assessment may address a set of similar processing operations that present similar high risks.*
2. *The controller shall seek the advice of the data protection officer, where designated, when carrying out a data protection impact assessment.*
3. *A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:*
    a. *a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;*
    b. *processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or …*

4. Controlled access

Controlled access is a critical safeguard for sensitive data. It makes sure that only authorised researchers with a reliable background can access the data and that they use the data for a research project that is well defined and has undergone an ethics review. In addition, it ensures the de-identification level of the data, since re-identification depends on context knowledge of a potential attacker. The more

[19] Wirth et al, Privacy-preserving data sharing infrastructures for medical research: systematization and comparison , 2021.

people that have access to de-identified data, the more likely they can be linked to the individual supposedly protected from being identified. For different forms of access governance including access policies, authentication of applicants, and data transfer agreements (DTA) see Milestone Access Governance WP2. A model Data Transfer Agreement will be developed in Deliverable 2.2.

5. Data Subject's Rights and transparency

The GDPR provides the data subject with certain rights, which has to be fulfilled by the data controller(s). These rights do not end with the controller who collects the data but is supposed to be maintained along the processing chain, i.e. by all controllers be they joint controllers or recipients of the data (independent controllers in a row/chain).

These rights are:

- Information according Art. 13, 14, 15, 22 and 34 GDPR, provided in an appropriate manner, *in a concise, transparent, intelligible and easily accessible form, using clear and plain language, and generally in writing or by electronic means. (Art 12 GDPR).*

- Art. 7 Conditions for consent, 7 (3): right to withdraw consent (where it is the legal basis),
- Art. 16 Right to rectification
- Art. 17 Right to erasure ("right to be forgotten")
- Art. 18 Right to restriction of processing
- Art. 19 Notification obligation regarding rectification or erasure of personal data or restriction of processing
- Art. 20 Right to data portability
- Art. 21 Right to object
- Art. 22 Automated individual decision-making, including profiling
- Art. 23 Restrictions

The concrete application of these provisions depends on the legal basis used for collecting the data.

Patient Empowerment should mean more than just observing these rights. Medical ethics requires feedback of incidental findings in certain cases. It has also become a good practice to inform research participants about research results. This is especially important for patients with chronic or rare diseases in order to help them manage their health and wellbeing. The question how to create the transparency for data subjects required by the GDPR has not yet been fully answered. So-called dynamic consent would require that subjects have to be proactively informed about

every data access, but this seems not to be widely preferred or adopted by research projects; in addition consent is not always the appropriate legal basis. Nevertheless, every person naturally has the right to receive information about the use of his or her personal data upon request. For this reason, the MII in Germany for example is establishing a central German " Health Research Data Portal" (FDPG), which bundles the data use requests running through the structures of the MII and generally, without individualised reference, offers information about the data use projects to the data-donating patients at a central point before the data use begins. Extensions of this portal for other projects are currently being discussed as well as interconnections with the National Health Portal of the Federation. Ethical problems arise from technically conceivable individualised information, some of which are currently considered unresolved. The patient or data donor would be indirectly informed about potential health conditions, hereditary predispositions for diseases, etc. – without medical advice and support – through notifications about data use for specific questions. It is still difficult to foresee how such a system would have to be designed if a central data portal were to take over the access decision. In this respect, the transparency requirement of the GDPR speaks more in favour of a decentralised system.

Where citizens can expect personally usable results from the use of their data, the path of feedback should be considered from the outset. This is likely to be the case regularly, especially in the area of genetic analyses, but also in other medical research. Especially in the case of centralised access decisions, the way back must not be excluded from the outset. In the case of consent-based access, the feed-back option must also be included in the consent document.

6. Accountability

Art. 5(2) GDPR: "*The controller is responsible for compliance with paragraph 1 and must be able to demonstrate compliance ("accountability").*"

All obligations under the GDPR therefore apply to the controller, including liability for breaches vis-à-vis the supervisory authorities as well as the data subjects for breaches. In relation to the citizen, the obligations to provide information are certainly just as important as claims for damages. According to Art. 4 No. 7, the controller within the meaning of the GDPR is logically "*the natural or legal person, public authority, agency or other body which alone or jointly with others determines the purposes and means of the processing of personal data*". This sounds simpler than it is, especially if several persons or institutions are involved in the data processing, along the health research data lifecycle, which is likely to be the case in the majority of cases of health research and is especially the case with data FAIR Data Portals, where at least three parties are involved: the data provider, the FAIR Data Portal (or Data Hub) and the data user (see **Figure 1**). Depending on the model, they can take on different roles. The possibilities are briefly summarised here:

Forms of data protection responsibility:

Job processing	**Shared responsibility**	**Transmission** **(Separate responsibility)**
Art. 28 GDPR	**Art. 26**	**Normal case**
Basic dependence on instructions Decision-making power of the person accepting the order possible for the TOMs (Art. 32 GDPR)	Equality: i.e. joint decision-making Own interest in the personal data	Contractual requirements for the handling of the transferred data possible, e.g. limitation of the permitted processing purposes.
The person commissioning the data decides on the purpose and means of data processing	Means and purpose of the data processing are jointly determined	Each responsible person shall determine its purposes and means

The roles in research consortia are still being debated. Different projects have chosen different approaches. Wherever decisions about access are jointly taken, joined controllership is very likely.[20]

Data security is a key obligation under the GDPR for the parties actually processing personal data, with the controller(s) being ultimately responsible. D2.2 Modular Contractual Clauses discusses different relationships in data sharing contexts and the kinds of agreements used to clarify responsibilities, including data security, within these relationships. GDPR roles and responsibilities deserve further consideration in the context of federated systems, which will be explored in the final ELSI guidelines.

It has to be pointed out already here, however, that where a processor is involved in the processing, also the processor has obligations under Art. 32 GDPR and must provide a sufficiently secure processing environment based on the controller's information on the nature of the processing (Art. 28(3)€ GDPR, Art. 32 GDPR, see

[20] Guidelines 07/2020 on the concepts of controller and processor in the GDPR Version 2.0, Adopted on 07 July 2021 https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf; van Veen, E.-B. et al, Joint controllers in large research consortia: a funnel model to distinguish controllers in the sense of the GDPR from other partners in the consortium, 2022, https://open-research-europe.ec.europa.eu/articles/2-80 ;

also EDPB Guidelines 07/2020).[21] It is the controllers' accountability though to demonstrate that they have chosen a processor providing sufficient technical and organisational safeguards (Art. 28(1) GDPR).

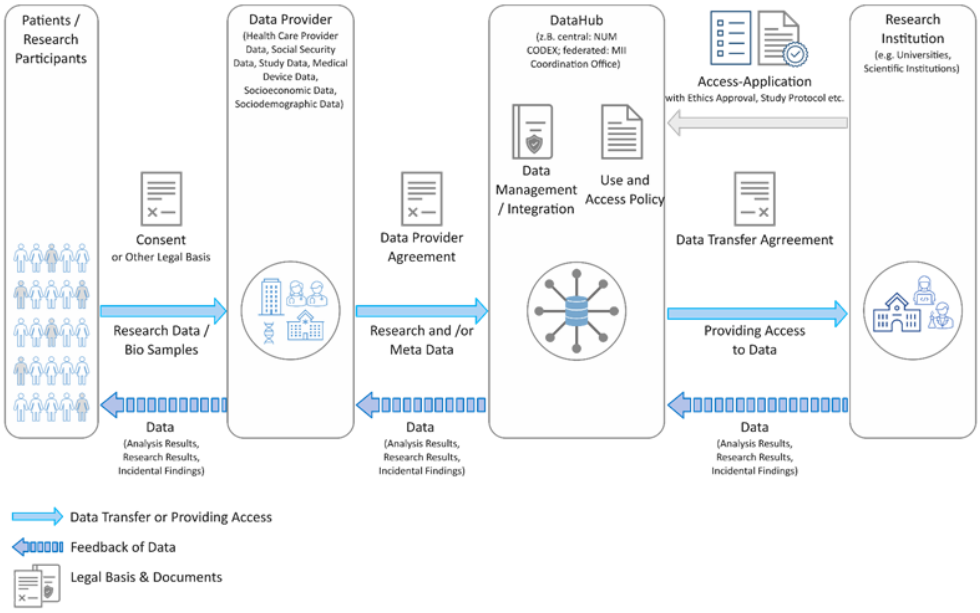
7. Generic Governance Model

These considerations result in a standard model of legal and ethical governance for the sharing of health data for research purposes. This governance applies across the data journey to the three actors mentioned above – the data provider, portal (or hub), and the user (**Figure 1)**. Building a network to make sensitive data available for secondary research use should therefore take into account and follow the organisational measures outlined here. Assumed is the following journey of data in 5 steps or phases:

1. Collecting data from patients/research participants/members of society in compliance with certain rules and with a relevant legal basis (e.g. declaration of consent) (input phase)
2. Transferring the data into a storage system that has specific and appropriate technical and organisational security measures (storage phase)
3. Making data available for research upon request by using a use and access policy including supervisory bodies and procedures to avoid the risk of unauthorised access (provision phase)
4. Using the data for research purposes internally or externally (use phase)
5. In appropriate cases, research results, incidental findings and raw data are returned to the person associated with the state (feed-back phase)

Figure 1. Generic Data Governance Model Across the Health Research Data Lifecycle

[21] There is already case law where processors have been fined for insufficient implementation of Art. 32.



This is important: All steps must be legally congruent, i.e. the legal basis for the data collection must also support the planned processing, further disclosures and uses or there must be another legal basis for those stages. It must be ensured that all other parties involved, whether they are responsible or processing persons, adhere to these rules. This can be done through existing legal requirements (including confidentiality obligations), but also through contracts, as is currently common in research.

This Generic Governance Model comprising the data science lifecycle is the framework for the overall data governance (Milestone on access governance + final ELSI guidelines) + legal security requirements (D2.2) + modular contractual clauses for data processing and use agreements (D2.1).

4. Why apply extensive security measures in clouds?

The use of cloud computing is now seen as a major IT trend, which enables a high degree of flexibility and cost efficiency in the use of computing resources and therefore also has an impact on IT infrastructures used in research. This is particularly true in biomedical research, which in recent years has experienced a sharp increase in storage and computing capacity demands. Technically, Cloud provides IT services via programmatic interfaces. Depending on the abstraction level, (virtualized)

hardware resources such as compute, storage, networking, and machine learning capabilities (Infrastructure as a Service - IaaS), Integrated application development Platforms (Platform as a Service - PaaS) as well as ready-made applications (Software as a Service - SaaS) are offered via network connections and their lifecycle managed via well-defined application programming interfaces.

In general, these offerings hierarchically build on top of each other: To provide Software as a Service via the network, the provider also must have access to the required hardware resources to run the application instances and control software on. The extreme increase in the volume of data processed over the last years, and the corresponding increase in demand for storage and compute is one driver to adopt Cloud Services by third parties. In addition, cloud offerings are convenient because the abstraction provided means that researchers can focus on their fields of research rather than being part time IT administrators and computer scientists, since the academic field is not attractive for IT experts, and there aren't enough idealists to run all the required infrastructures on premises. The industry pushes the trend towards cloud usage by offering cloud oriented tools and technologies, which become a sort of standard and are widely used for training of IT personnel. The increasing amount of genomic data also require centralised and cloud-based solutions in order to help to keep storage and compute costs within reasonable limits. However, it is understandable that the increase in data volumes is also leading to the opposite trend to increasingly leave and process data where it was collected. In accordance with the cloud paradigm, access to and efficient processing and evaluation of the data can then be offered worldwide. In this case, however, the physical location of the resources is a critical parameter and no longer as flexible as the third NIST criterion (see above) actually suggests.

Special features under data protection law, however, only become relevant if the cloud user and the cloud provider are not the same legal entity. Such a constellation can only be completely ruled out in the case of a private cloud, and only if the user is also a provider. In all other constellations, the responsibility for the individual steps of the collection, processing and storage of the data must be clarified. The GDPR links legal responsibility for the content of the decision to handle the data. Accordingly, the legal responsibility remains with the body that primarily has the data at its disposal and may use it in a legally protected manner for research purposes. The prerequisites for this are independent of the use of a cloud. In particular, the following topics still need to be taken into account, irrespective of how secure the cloud processing environment may be.

When discussing Cloud technologies, some differentiation is necessary between the globally present and available Hyperscalers such as AWS, Microsoft Azure, Google Cloud, Alicloud etc.; more regional cloud providers like IONOS SE or OVH / online.net; and local/institutional cloud offerings that work on the principles of cloud computing (service based access to infrastructure / Platforms / Services such as

de.NBI Cloud or usegalaxy.eu) but where locality and regulatory applicability is tightly coupled to clearly delineated locations and research institutions. The usegalaxy.eu case bears special consideration as it makes use of infrastructure-as-a-service not run by the same team, for example the de.NBI Cloud infrastructure in Freiburg.

But even irrespective of the question of how to provision large amounts of genetic data in the cloud, bioinformatics development benefits from the possibility of being able to use larger amounts of computing resources at short notice and at lower cost than is typically possible locally.

This applies not only, but also and especially, to development, testing and validation of new processing and evaluation algorithms for genetic data. For this purpose, the international offers of large and commercial US-American providers are available, among others, even if this does not always appear to be compatible with European data protection standards (see below). Clouds are also used in the medical environment beyond the storage and processing of omics data. One example of this is text mining of large quantities of unstructured clinical data in the cloud for research and quality assurance purposes and for machine learning. There are specific advantages and risks of using clouds for data management in clinical trials. It is clear that for networked medical research, which usually requires a central pooling of data and evaluation options with the necessary regulations and trusts, there is clear potential for the use of clouds. An exciting question here is also whether the academic and publicly funded research world will operate mainly on the user side or will continue to offer cloud services on a larger scale in the future.

Another relevant factor for future development is the increasing integration of patients and their mobile devices. Against the background of today's already strongly cloud-oriented infrastructures in this field in this area, this trend can also lead to an increase in medical research data in the cloud.

As a result, cloud computing allows efficient use of existing resources and enables users to realise a high-availability solution for their application without having to make large investments in their own hardware. This implies, however, additional risks to data protection, as the processing of their data takes place on systems that are not under the direct control of the data controller using cloud services.

Data protection authorities often require extensive control obligations for the cloud user. While it is acknowledged that an on-site audit will not always be possible, a mere assurance by the cloud provider that all data protection requirements are met does not appear to be sufficient. It is therefore proposed that the cloud provider undergo a certification or quality seal procedure[22] on data protection issues at an

[22] https://www.bsi.bund.de/EN/Topics/CloudComputing/CloudCertification/CloudCertification_node.html https://eucoc.cloud/en/home

independent testing body. However, the mere existence of a seal of approval or certificate does not appear to be sufficient; the cloud user must also satisfy themselves that all relevant aspects of their control obligations are covered by a given set of controls covered by the certification in place. These requirements of the data protection authorities of the cloud user apply equally to the auditing of subcontractors.

In the end, users cannot be relieved of their duties under GDPR and other regulations just by using 'secure infrastructure'. Privacy preserving controls are required but not sufficient to allow processing of sensitive medical data in a cloud, i.e. being able to remove - reduce - mitigate risk of privacy or security issues for the organisation running the infrastructure. Encryption at rest / encryption at transit / encryption at processing removes distinct requirements to control that data is not (accidentally or wilfully) exposed to unauthorised third parties.

It is particularly difficult to establish trust with providers that operate internationally, as these providers may be subject to jurisdictions that are not compatible with the GDPR. This environment places high demands on data security and integrity, as attacks or errors can even cause physical damage. Larger cloud providers have on average better and more encompassing security measures in place than most traditional HPC infrastructures and even local/regional cloud offerings. Most of the Hyperscalers (see above) are certified to the highest compliance standards, often multiple, since this is a requirement to do business with governments and financial institutions anyways. But these measures are not sufficient to achieve GDPR compliance if the cloud provider is for example based in the US (see below).

5. Clouds in Third Countries and International Organisations

• Third Countries

Art. 44 GDPR: General principle for transfers

*[1]Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a **third country or to an international organisation** shall take place only if, subject to the other provisions of this Regulation, the conditions laid down in this Chapter are complied with by the controller and processor, including for onward transfers of personal data from the third country or an international organisation to another third country or to another international organisation. [2]All*

provisions in this Chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.

Art. 45 GDPR: Transfers on the basis of an adequacy decision

1. *[1]A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection. [2]Such a transfer shall not require any specific authorisation.*

Art. 46 GDPR: Transfers subject to appropriate safeguards

1. *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.*

As the USA is home to many major providers of IT services and is also an important location for research, the question arises as to the usability of localised IT services. More generally, this is the question of the legal framework for clouds in so-called third countries that are not subject to EU or EEA rules. Until October 2015, US companies had the option of signing up to the US Department of Commerce in accordance with the Safe Harbor Principles. However, in October 2015, the European Court of Justice ruled against the application of this which was based on a decision by the European Commission in 2000 among other things because, it cannot be assumed that the data is sufficiently protected from government agencies such as the NSA.[23] As a successor to Safe Harbor, the EU-US Privacy Shield was negotiated in 2016 between the EU and the United States and was judged by the EU Commission on 12 June 2016 to be adequate with regard to the GDPR level of data protection. This regulation also contains a voluntary commitment component, which can be used by US companies. However, unlike Safe Harbor, this was supplemented by government guarantees and supervisory functions. In 2020, in Schrems II[24], the ECJ again annulled the Commission's decision. **So far, there is no secure basis to transfer data to the US.**

France is currently building a platform to store health data at the national level called French Health Data Hub. The idea is to build a data portal that makes it easier to study rare diseases and use artificial intelligence to improve diagnoses. It is supposed to aggregate data from different sources and make it possible to share

[23] (JUDGMENT OF THE COURT (Grand Chamber), 6 October 2015, in Case C‑362/14, http://curia.europa.eu/juris/document/document.jsf?text=&docid=169195&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=462865).

[24](C-311/18) https://curia.europa.eu/juris/document/document.jsf?text=&docid=228677&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1

some data with public and private institutions for those specific cases. The technical choices have been controversial as the French government originally chose to partner with Microsoft and its cloud platform Microsoft Azure. Microsoft, like many other companies, relies on Standard Contractual Clauses for EU-U.S. data transfers. But the Court of Justice of the EU has made it clear that EU regulators have to intervene if data is being transferred to an unsafe country when it comes to privacy and surveillance. The Conseil d'Etat (**Conseil d'Etat Nr. 444937,** October 2020), decided that an American company could process data in Europe but it would still fall under FISA702 and other surveillance laws. Data would still possibly end up in the hands of American authorities because of extraterritorial application of US laws. In other words, it is being extra careful with health data for now, while Schrems II is still unfolding. The French government is now looking at other solutions for the Health Data Hub. Meanwhile the CNIL, together with the Health Data Hub platform and Microsoft, should try to find technical solutions that make access by the US authorities impossible. In addition, the CNIL should check whether the measures taken to host the data are necessary to fulfil the purpose of the platform. Such technical solutions could be privacy preserving computation such as Trusted Execution Environments.[25]

But many providers of processing environments do not rely on commercial clouds, especially when they are provided by American companies. Instead, Secure Processing Environments are established on Servers based in the EU controlled by the research community itself, one framework being GAIA X funded by the EU.[26]

• International Organisations

In Chapter V, the GDPR treats transfers of personal data to third countries and to international (intergovernmental) organisations in the same manner. This correctly reflects the reality that GDPR does not apply to international organisations, but can also lead to the assumption that there is no difference between controllers or processors in a third country and international organisations.

International organisations are established by an international agreement, have an international legal personality, and a mandate and the powers necessary to fulfil it. International organisations also enjoy privileges and immunities, such as immunity from the jurisdiction of national courts and enforcement and inviolability of premises and archives, in order to ensure their independence.

This independence protects them from interference by their member states, including host countries, and third parties. This means, for example, that law

[25] https://en.wikipedia.org/wiki/Trusted_execution_environment

[26] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

enforcement and intelligence agencies of the host country cannot compel access to the personal data processed by an international organisation.

Moreover, international organisations can adopt internal data protection frameworks – and indeed many have done so – tailored to their status and needs, which provide for data protection principles and the rights of data subjects, and often for internal supervision and redress mechanisms. This means that data transfers to international organisations in most cases would meet the requirements set by the Schrems II CJEU decision.

Nevertheless, controllers subject to GDPR still need to comply with its provisions for a lawful transfer to take place. This does not mean, though, that international organisations must be treated in the same way as a controller or processor in a third country would. However, the options available for such transfers are limited. Administrative arrangements and legally binding and enforceable instruments cannot be used by private actors, while the current Standard Contractual Clauses cannot be used for transfers to international organisations.[27] For this reason, it seems that the (only) appropriate method for transfer of personal data to international organisations is the derogation of 'important reasons of public interest' of Article 49(1)(d) GDPR, the public interest being found in the treaty, as recognised by the EDPB.[28]

6. Best Practice Examples

• de.NBI-Cloud

The de.NBI-Cloud has 8 sites in Germany (Uni Bielefeld, Uni Gießen, Uni Heidelberg, DKFZ Heidelberg, EMBL Heidelberg, Charité Berlin, Uni Tübingen and Uni Freiburg). These multiple SPE sites are using similar technical and operational measures to enable security and privacy. This is intended to enable economies of learning and management, rather than enabling distributed or federated data analysis. Specific project teams are associated with specific individual sites.

For the different Collaboration SPEs, ingestion of external data is an essential preliminary step. Ingestion and anonymisation are the responsibility of project teams, subject to the approved study agreement/project description. de.NBI-Cloud

[27] The European Commission has recently confirmed that the current SCCs are not appropriate for transfers to International Organisations. *See* Questions and Answers for the two sets of Standard Contractual Clauses, Question 25, p. 14.

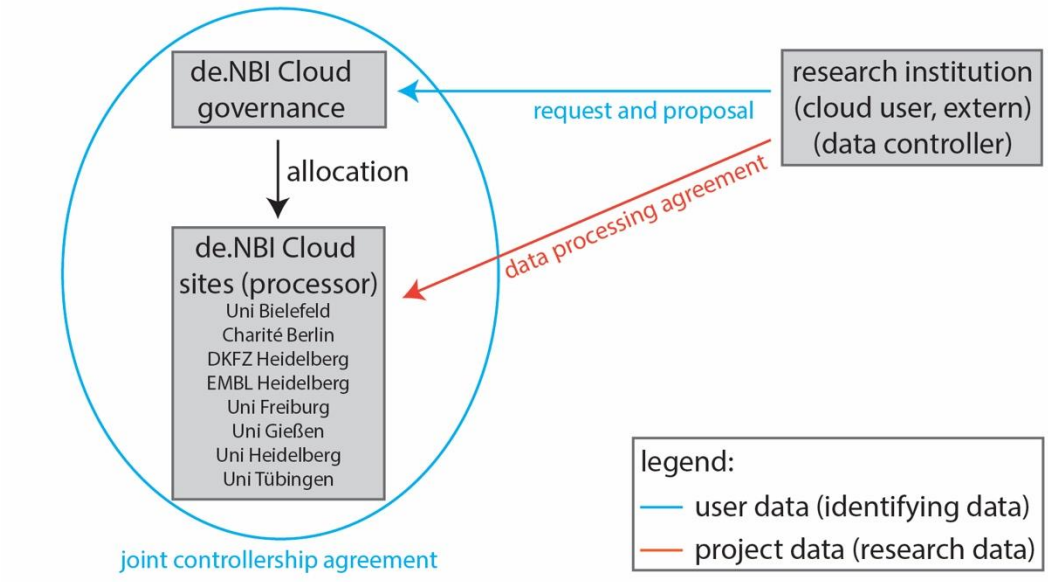
[28] *See* EDPB, Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679, p. 10.

provides infrastructure only, while data and software are the responsibility of the project team at each site. Some sites are certified according to ISO 27001.

Regarding governance, de.NBI-Cloud uses the ELIXIR-AAI (now Life Science Login). Researchers need to apply with their projects at the Cloud main office in Bielefeld. Here, they are reviewed, approved and allocated to one of the 8 sites (taking into account capacities, wishes of the user and certification needs).

Legal documents involved:

- joint controllership agreement (between all de.NBI-Cloud sites; for identifying user data only, as they need to be shared between the de.NBI-Cloud main office in Bielefeld and the processing de.NBI-Cloud site)
- terms of use: https://cloud.denbi.de/documents/3/Nutzungsbestimmungen_de.NBI_Cloud_V1.2_EN.pdf
- Data Processing Agreement (individual for each de.NBI-Cloud site, between processing de.NBI-Cloud site and external cloud user, mainly for sensitive project/research data)
- Portal Privacy Policy (de.NBI-Cloud main office in Bielefeld): https://cloud.denbi.de/about/policies/ (HTML only)
- Cloud Privacy Policy (identical for all sites)



• BIGAN

BIGAN is the platform for secondary use of health data in the Aragón Region, Spain. It is established by executive order (SAN/1355/2018), that defines its legal framework, governance bodies (the BIGAN Oversight Committee, or BIGAN

Committee for short) and technical aspects. IACS Biocomputing unit is responsible for the operational management, development and maintenance of BIGAN infrastructure with the support of IACS staff on the IT, Legal, Ethical, and HR departments and with the assistance of researchers from the Data Science for Health Services and Policy Research Group.

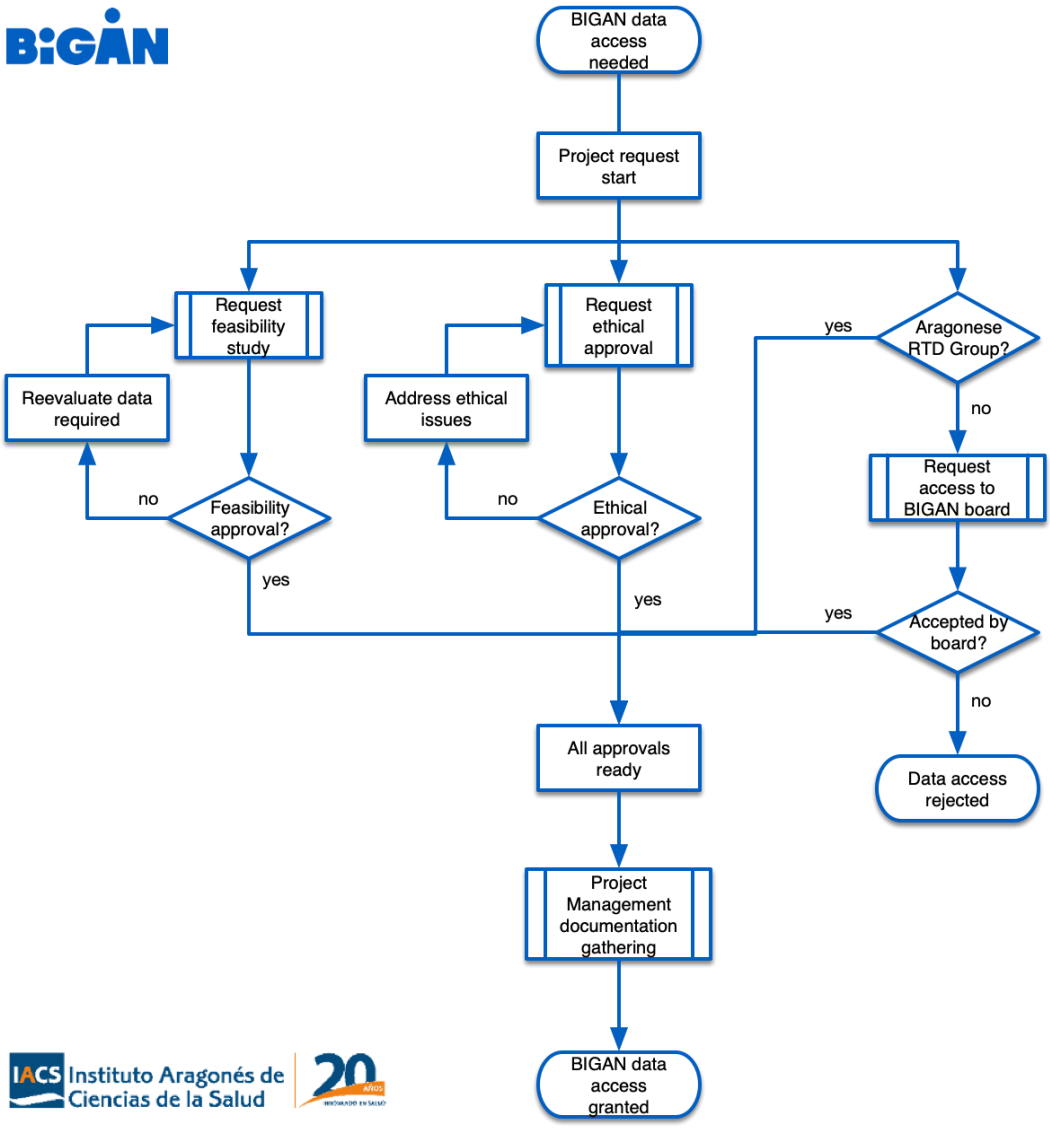
The BIGAN technical solution corresponds to a centralised system that is structured in two elements. The first element is data lake which is divided in two areas: a staging area where ETLs (Extract, Transformation and Load) processes capture and cleanse data from primary information systems, i.e., those used in the Aragonese health system; and an archival area, where the cleaned data is archived for its further reuse. The second element is a data analysis environment, which contains three applications: one application, BIGAN Clinical Management, devoted to compute regular health indicators collection used by Aragonese health community (from clinicians to policy makers). It also includes a *BIGAN labs* section, where advanced indicators are developed and tested; a second application devoted to the research, BIGAN Research, where researchers will analyse the data they have been granted to in a secure manner (conceptually, the SPE side of the platform); and a third application, devoted to capacity building, BIGAN Learning, which is essentially a clone of the research area but for using synthetic data – anonymous data that is artificially created through algorithm from personal health data to emulate its characteristics (and thus not requiring the heavy security constraints). This system currently sits on top of a dedicated four-node cluster attached with a Storage Area Network that holds a variety of data lake area and data processing area databases (Apache Cassandra, PostgreSQL, MongoDB), a Spark cluster instance for high-performance/high-scale data processing and different notebook-like software (Jupyter Notebooks, Zeppelin, RStudio Server, etc.) for the research/capacity application.

In terms of security and data protection, there is a triple pseudonymisation process. First, individual identifiers from primary information systems are pseudonymised prior to its transfer to BIGAN infrastructure using a one-way algorithm, only known by the health services systems. This pseudonym is then encrypted using an encryption key only available in the BIGAN infrastructure. Finally, when data is provided to end users for a given project, the pseudonyms are also encrypted with a key generated for such a project. All in all, this approach tries to avoid possible linkage between multiple extractions and to identify possible data leaks, as each data provision has its own pseudonyms, and it facilitates the reversibility of the pseudonyms, only in the BIGAN environment than then can be transferred to the health service operators in case incidental findings that need patient re-

identification." The system is declared conforming to the Spanish Security Schema for Public Digital Services[29].

The protocol to access data is detailed in the following Figure. To request access to data, the requester requires two approvals: a feasibility study approval, issued by the Biocomputing Unit as experts of the specificities of the data available; and an ethics approval, issued by the Aragonese ethics committee for clinical research (CEICA), or a equivalent Spanish committee based on a mutual recognition principle. In case that the requester group is not part of the Aragonese RTD environment, it will be required the approval of the BIGAN Board. Please note that this protocol applies for secondary use of health data for research and innovation purposes. The secondary use for policy making and regulation restrictions and may require only the mandate of the Aragonese ministry of health when only departments or directorates of health are involved or a discussion in the BIGAN board when other regional ministries are involved.

29 https://administracionelectronica.gob.es/pae_Home/pae_Actualidad/pae_Noticias/Anio2022/Mayo/Noticia-2022-05-04-Publicado-RD-regula-Esquema-Nacional-Seguridad.html



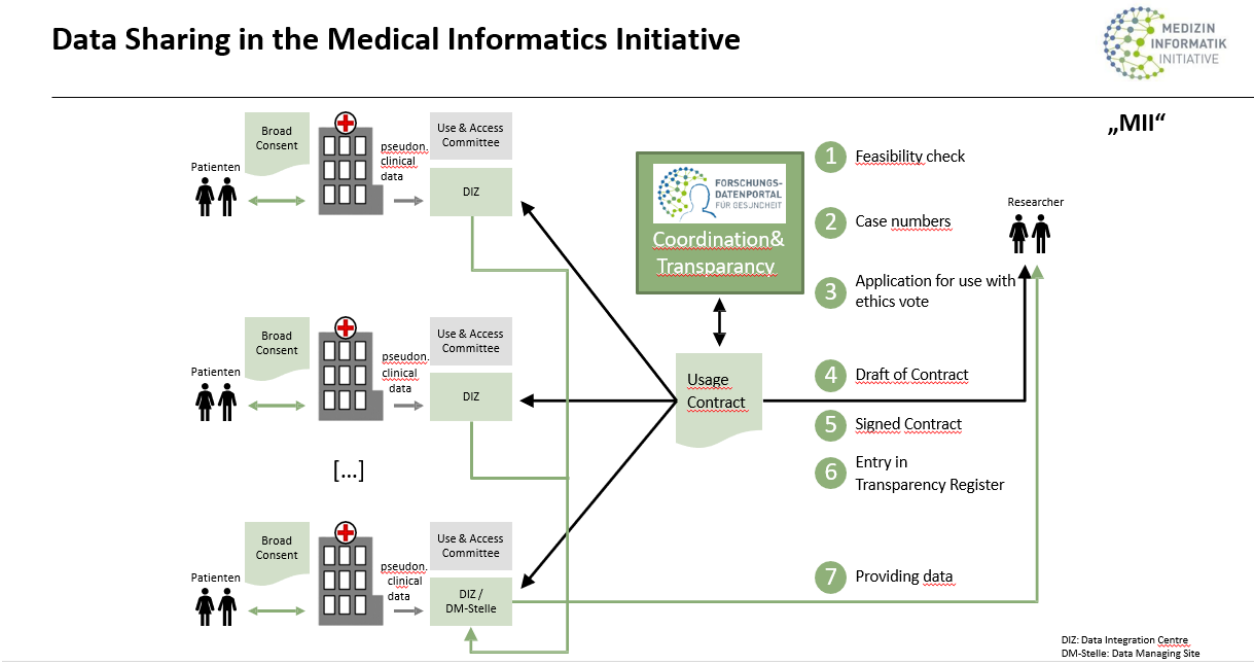
• Medical Informatics Initiative (MII, health care data)

The Medical Informatics Initiative (MII)[30] is an example for implementation of the Generic Governance Model described above. It aims to make the best possible use of the opportunities offered by digitalisation in medicine for care and research through making health care data available for research projects. In a first step, data integration centres (DIZ) are being established and networked at university hospitals and partner institutions. In these centres, the prerequisites will be created to be able to link research and care data across locations. At the same time, innovative IT solutions are being developed for specific medical applications to demonstrate the possibilities of modern digital services and infrastructures in the health sector.

[30] https://www.medizininformatik-initiative.de/en/start

The legal basis for the use of healthcare data within the framework of the MII is the informed consent of the persons to be treated (so-called broad consent), which was coordinated with all data protection supervisory authorities in Germany in an elaborate process over two years.[31]

Access to the MII's data is governed by a use and access policy.[32] There are several modes of data usage: feasibility studies, federated on-site-analysis or transfer of pseudonymised data in certain cases. In order to be allowed to use data for a research project, researchers must first submit an application. This application is reviewed by an independent ethics committee and a "Use and Access Committee" (UAC) at the location of the Data Integration Centre (DIZ). The UAC reviews the applications and approves or rejects them. This procedure rules out any unethical use of the data and ensures a high scientific quality of the data analyses. If a research application is evaluated positively, the scientists are given access to the data after they have concluded a user agreement.[33]



The technical safeguards are outlined in the overall data protection concept[34], which has been discussed and approved by the Data Protection Working Group at

[31] https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung

[32] https://www.medizininformatik-initiative.de/sites/default/files/2020-12/MII_Nutzungsordnung_v1.1.pdf.
[33] https://www.medizininformatik-initiative.de/de/nutzungsvertrag.
[34] https://www.medizininformatik-initiative.de/sites/default/files/2022-03/MII-Datenschutzkonzept_v1.0.pdf.

TMF to comply with the respective Guidelines. In addition, all sites closely collaborate with their local authorities.

The clinical data used for research are pseudonymised before being transferred into the data warehouse of the participating hospitals. It is a special feature of the data protection architecture of the Medical Informatics Initiative that permanently used pseudonyms for the patient data are only generated and processed in the treating facilities or DIZ (Data Integration Center) locations. For all processing across locations, additional pseudonyms are generated and used, which are specific to this processing and are deleted after the processing has been completed.

---

7. Guidelines for risk assessment

Summary of factors to achieve legal compliance of organisational and technical safeguards to support the overall architecture.

1. Determining the data controller (or joint controllers) and processors
2. Governance following the generic model
    - Determine legal basis for sharing data
    - All contracts in place? [Refer to **D2.1 Framework for modular contractual clauses**, in particular relating to data processing agreements and security clauses.]
    - Applying technical and organisational safeguards:
        - Data transfer over encrypted channels.
        - Data minimization (transfer only data necessary for the analysis).
        - Identity verification of users (individual persons) e.g., by communication via institutional email address.
        - Access permissions: formally reviewed on a regular basis (e.g., 6 months).
        - Multi-factor authentication
        - Secure passwords.
        - SSH keys
        - Local computer (accessing use): updated anti-virus/malware. screen lock.
        - Trusted software:
            - Security audits,
            - All security updates made,
            - Access software from trusted source.

        - Ensure deletion of all temporary directories at the end of job or after job failure.
        - Containerization of the analysis.
    - Data security training (staff and users)
    - Security incidents reporting
3. In addition, where clouds are used:
    - Trustworthy cloud provider (risk assessment + having a documented security policy that includes best practices, self-auditing or certification and auditing, ISO) Subcontractors?
    - Cloud provider physically or legally based in a third country?
    - Apply privacy preserving computation techniques
4. Conducting a Data Protection Impact Assessment

Risk analysis with qualitative and quantitative metrics in order to test cloud providers

To be done in collaboration with WP 5, 6 and 7 after D5.3 is submitted.

8. Applying the Guidelines to the use cases

To be done in collaboration with WP 5, 6 and 7 after use cases are clearly defined.

9. Acronyms and Abbreviations

- D – deliverable
- EC – European Commission
- GDPR - General Data Protection Regulation
- HPC – High Performance Computing
- IPR – Intellectual Property Right
- M – Month
- WP – Work Package
- WPL – Work Package Leader

Annex 1

Art. 5 GDPR: Principles relating to processing of personal data

1. *Personal data shall be:*
    a. *processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');*
    b. *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');*
    c. *adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');*
    d. *accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');*
    e. *kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');*
    f. *processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').*
2. *The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').*

Art. 9 GDPR: Processing of special categories of personal data

1. *Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of 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 shall be prohibited.*
2. *Paragraph 1 shall not apply if one of the following applies:*
    a. *the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;*
    b. *…*
    c. *processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;*

d. …

e. processing relates to personal data which are manifestly made public by the data subject;

f. …

g. processing is necessary for reasons of substantial public interest, on the basis of 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;

h. processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;

i. 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;

j. 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.

Art. 89 GDPR: Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

1. *[1]Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject.[2] Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation.[3] Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. [4]Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.*

Art. 25 GDPR: Data protection by design and by default

1. *Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.*

2. *[1]The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. [2]That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. [3]In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.*
3. *An approved certification mechanism pursuant to Article 42 may be used as an element to demonstrate compliance with the requirements set out in paragraphs 1 and 2 of this Article.*

Art. 28 GDPR: Processor

1. *Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject.*
2. *[1]The processor shall not engage another processor without prior specific or general written authorisation of the controller. [2]In the case of general written authorisation, the processor shall inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes.*
3. *[1]Processing by a processor shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller.*

Art. 32 GDPR: Security of processing

1. *Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:*
    1. *the pseudonymisation and encryption of personal data;*
    2. *the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;*
    3. *the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;*
    4. *a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.*
2. *In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.*
3. *Adherence to an approved code of conduct as referred to in Article 40 or an approved certification mechanism as referred to in Article 42 may be used as an element by which to demonstrate compliance with the requirements set out in paragraph 1 of this Article.*
4. *The controller and processor shall take steps to ensure that any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller, unless he or she is required to do so by Union or Member State law.*