945345 – H2O

Health Outcomes Observatory

WP1 – Governance-sustainability-capabilities

# D1.5 H2O information governance and ethics framework


| | |
|---|---|
| Lead contributor | Peter Singleton (8 – The European Institute for Innovation through Health Data) |
| | peter.singleton@chi-group.com |
| Other contributors | Dipak Kalra (8 – i~HD); |
| | Fabian Praßer (2 – Charité) |
| | Dustin Holloway (14 – Takeda); |
| | Nick Bott (14 – Takeda); |


**Document History**

| Version | Date | Description |
|---|---|---|
| V0.10 | 31/05/2021 | First Draft |
| V0.20 | 22/06/2021 | Revisions after internal review |
| V0.21 | 12/07/2021 | Minor corrections |
| VX.X | DD/MM/YYYY | Final Version |

# Table of contents

# Definitions

- **Participants** of the H2O Consortium are referred to herein according to the following codes:
  1. MUW. Medizinische Universitaet Wien.
  2. Charité. Charite – Universitaetmedizin Berlin
  3. EMC. Erasmus Universitair Medisch Centrum Rotterdam
  4. ICS-HUVH. Institut Catala De La Salut – Hospital Universitari Vall d'Hebron
  5. KCL. King's College London
  6. KUL. Katholieke Universiteit Leuven
  7. EPF. Form Europeen des Patients / European Patients' Forum
  8. I-HD. The European Institute for Innovation through Health Data
  9. The Hyve. The Hyve BV
  10. TEAMIT. TEAM IT Research SL
  11. KUH. Karolinska Universitetssjukhuset
  12. UniSR. Universita Vita-Salute San Raffaele
  13. IKNL. De Stichting Integraal Kankercentrum Nederland
  14. TAKEDA. Takeda Pharmaceuticals International AG
  15. NVS. Novartis Pharma AG
  16. ABBVIE. AbbVie INC
  17. Lilly. Ali Lilly and Company Limited
  18. MDT. Medtronic International Trading SARL
  19. Pfizer. Pfizer Limited
  20. ROCHE. F. Hoffman-La Roche Limited
  21. SARD. Sanofi-Aventis Recherche & Development
  22. JDRF. JDRF International
  23. Trial Nation. Trial Nation

- **Grant Agreement**. (Including its annexes and any amendments) The agreement signed between the beneficiaries of the action and the IMI2 JU for the undertaking of the H2O project (Grant Agreement No. 945345).
- **Project**. The sum of all activities carried out in the framework of the Grant Agreement.
- **Consortium**. The H2O Consortium, comprising the above-mentioned legal entities.
- **Consortium Agreement**. Agreement concluded amongst H2O participants for the implementation of the Grant Agreement. The agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

# Abbreviations

| Acronym / Abbreviation | Meaning |
|---|---|
| DMP | Data Management Plan |
| DoA | Description of the Action |
| DPIA | Data Protection Impact Assessment |
| EFPIA | European Federation of Pharmaceutical Industries and Associations |
| EHDEN | European Health Data & Evidence Network |
| EHR | Electronic Health Records |
| GDPR | General Data Protection Regulation |
| H2O | Health Outcomes Observatory (either national or central) |
| PRO | Patient Reported Outcome |
| PROM | Patient Reported Outcome Measure |
| SDC | Statistical Disclosure Control |

# Glossary

A number of technical terms are used that may not be familiar to the average reader:

| Term or phrase | Meaning |
|---|---|
| Aggregate or aggregated data | Data which has been totalled or summarised (e.g. as an average or median value) rather than individual-level data – see micro-data below |
| Data-cube | A multi-dimensional set of aggregate data from which a user may extract specific subsets or produce further summaries |
| Micro-data | Data which relates to an individual or anonymised case – as distinct from aggregate data (see above).<br>Formally (OECD): An observation data collected on an individual object - statistical unit. |
| Statistical Disclosure Control | Recognising that even statistical or aggregate data can be subject to re-identification attacks, additional controls may be applied to further limit access or use of the published statistics. A simple example might be to ensure that all cells in a table have a count of at least 5<br>Formally (OECD): The complex of measures preventing unauthorized access to sensitive statistical information. |
| | |

# Summary

This document describes the Information Governance Framework for the H2O project, including the establishment of the Health Outcomes Observatories. 'Information Governance' is considered to be a broad term covering data protection, confidentiality and ethics as well as high-level elements of information security, which are required when considering the appropriate processing of healthcare data for public benefit.

The document describes a number of instruments which will be needed as part of this 'framework' as well as how these are linked to other project deliverables. It builds on the earlier deliverable, D1.2 Data Management Plan v1.

This deliverable, due for month 9 of the project, will necessarily anticipate products to be developed and refined throughout the whole 5-year project. In some instances, it will lay out broad requirements, in others it may describe the key elements of the final document or template.

The description for this deliverable refers to a 'code of conduct' which is both discussed as a phrase (against a 'code of practice') and elaborated on as an appropriate set of 'principles' and 'standard operating rules' (or 'rules' for short) to be followed within the project and for the operation of the planned observatories. As noted above, these may need to be adjusted or modified to reflect both changing plans and experience gained throughout the project.

# Scope and purpose of this document

This document is one of the key deliverables from Work Package 1 (WP1) due in month 9 of the project (June 2021). It is to be produced under Task 1.3 (T1.3) - Governance, data protection and ethics[1], which is responsible for ensuring robust data protection and GDPR compliance across all H2O data flows.

The actual document is defined quite broadly in the DOA:

| Ref | Title | Description |
|---|---|---|
| D1.5 [M9] | H2O information governance and ethics framework | Portfolio of instruments and codes of practice to ensure GDPR compliance and ethical handling of outcomes data. |

**Note:** This 'framework' document describes the required 'portfolio' of documents required by the H2O project and its eco-system. However, not all of these documents will be immediately available; some will only be defined in outline form at this stage or created as an initial template or model for further development. For example, each Observatory and each consortium member will need to produce DPIAs for GDPR compliance, but can build upon the materials in this document and its exemplar annexes.

[1] Lead i~HD and TAKEDA. Contributors: MUW, Charité, EMC, ICSHUVH, EPF, TEAM-IT, NVS, Lilly, MDT, Pfizer, SARD, Trial Nation, JDRF

# 1. Introduction

## The mission of H2O

The IMI H2O project will deliver patient-centred, ethically and legally sound, extensible and permanent national or regional Health Outcomes Observatories (H2Os), initially in four countries (Tier 1: Austria, Germany, Netherlands, and Spain) for three diseases[2], which will be based on a hybrid model of both federated and centralized data collection, management and analysis. A European-scale network of outcomes-generation pipelines, evidence concentrators and insight comparison and analysis tools will be established to inform all levels of decision makers: to convince and direct them - and the public - towards the most effective and affordable value-based models and to maximise health outcomes. The outcomes data arising from this network, as aggregated data analytics and potentially as anonymised micro datasets, will be useful, and therefore provided as products and services, to a range of industry and public body organisations, as explained in deliverable 6.12.

Patient Reported Outcomes (PROs) reflect the patient perspective in health outcome measurement and are a solid basis for patient empowerment. In a true value-based and patient-centric model, patients not only collect outcomes data, but also control, give access to, and use their health data in a meaningful way. Patients who control their own personal health data, including PROs, will increasingly self-manage their health and take an active role in the interaction with their healthcare providers.

The national and/or regional Observatories will operate under a governance model that will guarantee that data are protected under jurisdictional data protection law. National Observatories will be aligned with a European Observatory to facilitate interoperability, guide reproducibility in other countries, avoid fragmentation using a federated data management approach and promote the benefit of measuring and using outcomes data at regional, national, European and global levels.

The European and national Observatories offer an innovative framework to collect and analyse PRO data accurately and precisely. National Observatories will be connected in a federated way to the pan-European framework. Analytic code will 'travel' to where the data are, and aggregated results will be sent back to the stakeholder who wishes to answer a specific question or to track an outcome. Microdata extractions will be carefully checked using statistical disclosure controls to mitigate reidentification.

We will actively encourage the model to spread to other countries beyond those involved in the current project by creating the governance of the European Observatory in a way that allows new organisations to join. Key prerequisites for this expansion will be the willingness of new partners to ensure harmonisation of outcome standards as well as other outcomes-relevant data and adherence to the ethical and governance model for the management of data.

[2] diabetes, inflammatory bowel disease (IBD), and cancer

# The intended use of health data within H2O

Figure 1 below shows the main data flows feeding into each national Observatory, the way in which the national Observatory may need to partition its processing between identifiable data and pseudonymised or anonymised data, and contribute data and analytics to the European Observatory (darker blue circle on the right).

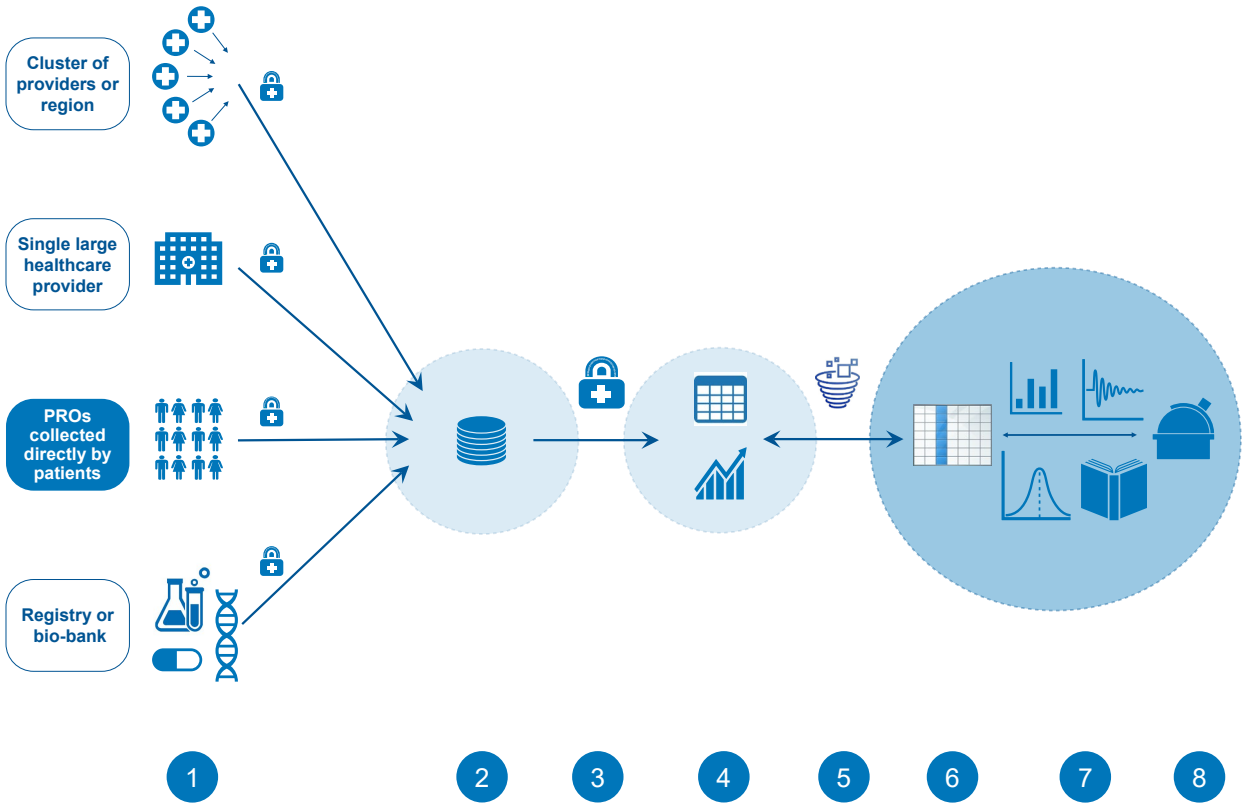


*Figure 1: H2O Data use overview*

This diagram should be interpreted quite broadly in terms of the range of information and the precise dataflows through which data may be gathered. The meaning of the numbered steps in the diagram is as follows:

1. Each national Observatory will always collect PRO data directly from patients, in the disease areas specified in the project. It is hoped that, in the majority of cases, corresponding clinician reported outcomes and complementary outcomes relevant data can be provided by the healthcare provider treating the disease for these patients, on the GDPR legal basis of informed consent, and linked to the PRO data to enrich the dataset per patient. At times other health and care provider sources such as a cluster of primary and secondary care organisations may contribute clinical and other care data. Finally, it is possible that registries and bio-banks may also contribute data that can be linked to individual patients.
2. Each national Observatory will need to hold an identifiable repository of PRO data and other complementary data that can link new data and be provided back to individual patients so that they can access and use their own outcomes data. If consent is provided for this, PROs can also be shared with each patient's clinical team to support joint decision-making and care planning.

3. A separate repository will be required at a national level, whether physical or just-in-time virtual, in order to permit national level data curation and analytics, and to allow each national Observatory to contribute to the European Observatory. A mechanism for providing this is required, and it is to be determined if this function of the national Observatory needs to be a distinct legal entity or can be the same legal entity as is holding the identifiable data repository. The legal entity arrangements, whether a physical second repository is required and if this should be pseudonymised or anonymised might vary between countries depending upon national GDPR derogations and other legislation.
4. Each national Observatory may itself undertake within-country analyses for which it may have public and industry recipients, and it is likely to curate certain anonymised datasets for a broader range of analytic purposes including data sharing with the European Observatory.
5. The European Observatory will run many federated analysis queries on data sets curated by each national Observatory, in order to generate European level outcomes analyses, to enable cross-country comparative analysis, and to generate micro-data sets that can be made available to external parties provided that the risk of reidentification is adequately addressed.
6. The European Observatory will need to collect and combine multiple anonymised data sets and analysis results sets to fulfil its roles.
7. The European Observatory will curate and promote a range of data and analytics products and services to industry and to public authorities, which will be the main income generating activity to sustain the Observatory ecosystem.
8. The European Observatory in collaboration with the National Observatories aims to progressively become the primary recognised resource for health outcomes intelligence across Europe.

## Possible future considerations

There are a number of possible additional flows as well as restrictions on dataflows that may be considered in the future, both within the period of the H2O project and also beyond the project, once the Observatories are established. It may also be that some of these approaches are possible in some member states because of the specific legislative framework or local infrastructure (including national or regional ICT) but not in others.

Some of the aspects to be addressed in future versions of the Data Management Plan and updates to this document are:

- Whether anonymised extracts of micro-data may be shared within strict contractual arrangements (including appropriate security and access controls) for specific purposes by the national Observatories (and with prior agreement of data providers, including patients themselves);
- Whether Observatories may be able to release for open access restricted ‘data cubes’ of aggregated data, possibly subject to Statistical Disclosure Controls (SDC);
- Understand possible interactions with the planned European Health Data Space;
- Understand exactly how patient choices will be applied in terms of interaction through the national Observatory for the purposes of direct care, especially in terms of maintaining adequate records and ensuring that the patient is not put at any additional risk.

These elements should not greatly affect this document, though may require some additional specific rules, as the principles described here should generally be at a higher level of abstraction.

# 2. Overview of the Information Governance Framework (IGF)

This document needs to interface with a number of other formal deliverables, as well as provide input to a number of other H2O project 'products' which will be developed as shown in this diagram:

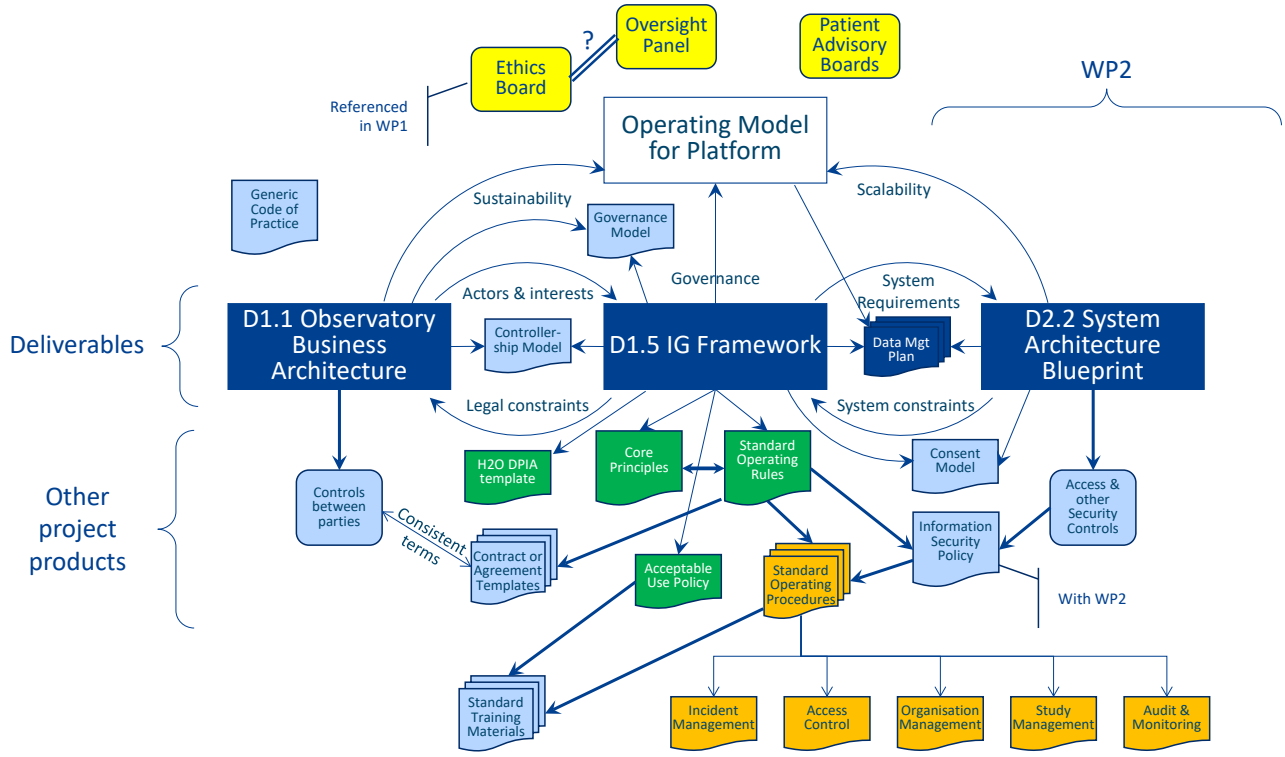


*Figure 2: H2O Information Governance Framework in context*

These 'products', which may be formalised as documents or diagrams, are described in more detail in the following sections.

# Links between IGF and Business Architecture (D1.1)

There is a strong inter-relationship between the IGF and the 'Observatory BusinessArchitecture' with respect to the various legal entities and their roles and responsibilities in terms of finance, sustainability, legal compliance, and ethical oversight. They cannot be developed wholly independently and two aspects of their inter-dependency are described here as a 'governance model' and a 'controllership model'.

## Governance Model

The Governance Model is not a formal product as such; instead, it represents the underlying structures that will be developed in D1.1 Observatory Business Architecture and refined further in D1.8 H2O Business Model. D1.1 has already established 9 Governance Principles for H2O as well as a suggested legal structure for the national and the umbrella observatories.

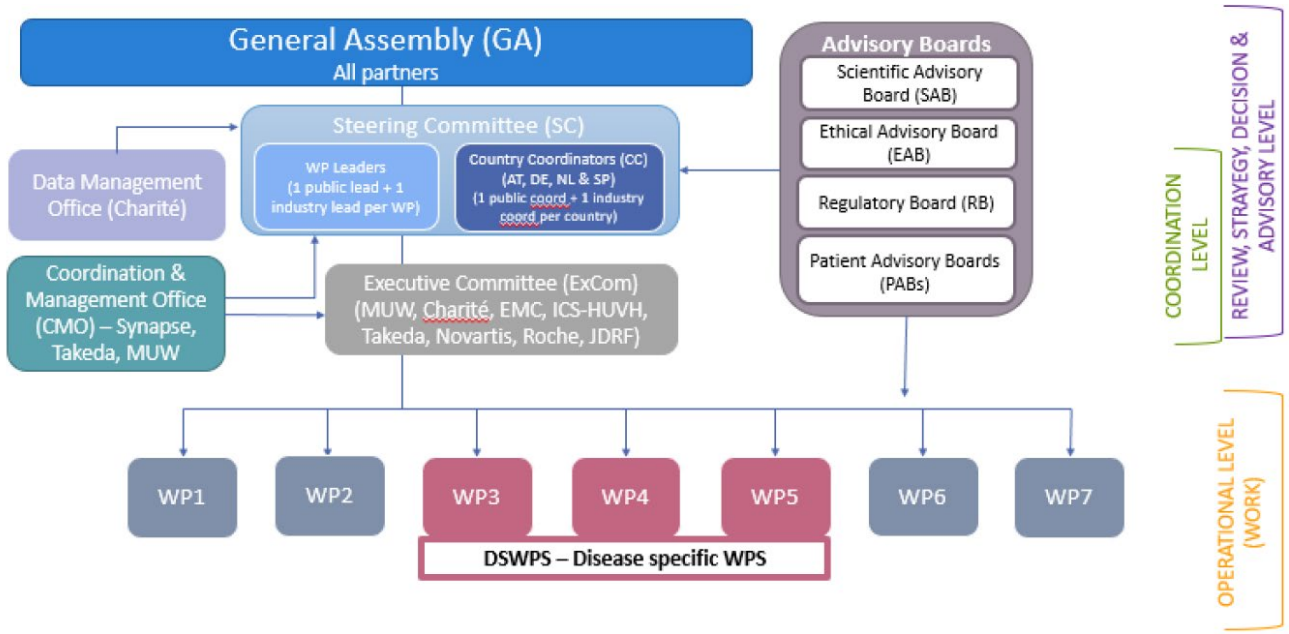


*Figure 3 - Governance diagram from Stage II proposal*

The Governance diagram above (from Stage II proposal) shows how project management and coordination activities relate to the various management, advisory, and oversight boards.

In particular, the Ethics Advisory Board (EAB) is likely to review the overall position with regard to data protection impact assessments as well as reports received concerning legal compliance or potential issues concerning ethical or data protection compliance. It will also have a role in reviewing the compliance of the disease specific work packages (DSWPs) in research ethics approvals.

D1.1 also suggested the following roles within each Observatory:

| Governance aspect | National Observatory role | European Observatory role | European Observatory committee/panel |
|---|---|---|---|
| Management | Managing Director | Managing Director | Managing Board |
| Ethics | Ethics Officer | Chief Ethics Officer | Ethics Board |
| Science | | Chief Scientist | Scientific Advisory Board |
| Analytics | Methods Officer | Chief Methodologist | |
| Patient Liaison | | Chief Patient Officer | Patient Advisory Board |
| Technology/ICT | | Chief Technology Officer | |
| Data Protection | DP Officer* | DP Officer* | |

* legally required

For the IGF, the key aspect here is how the two Data Protection Officers (DPOs) report through to the relevant Observatory management, both operationally (e.g. to Managing Director) or for oversight to the appropriate Management Board to ensure that each Observatory is compliant, both in terms of policies and at an operational level (e.g. handling and reporting of possible incidents or data breaches).

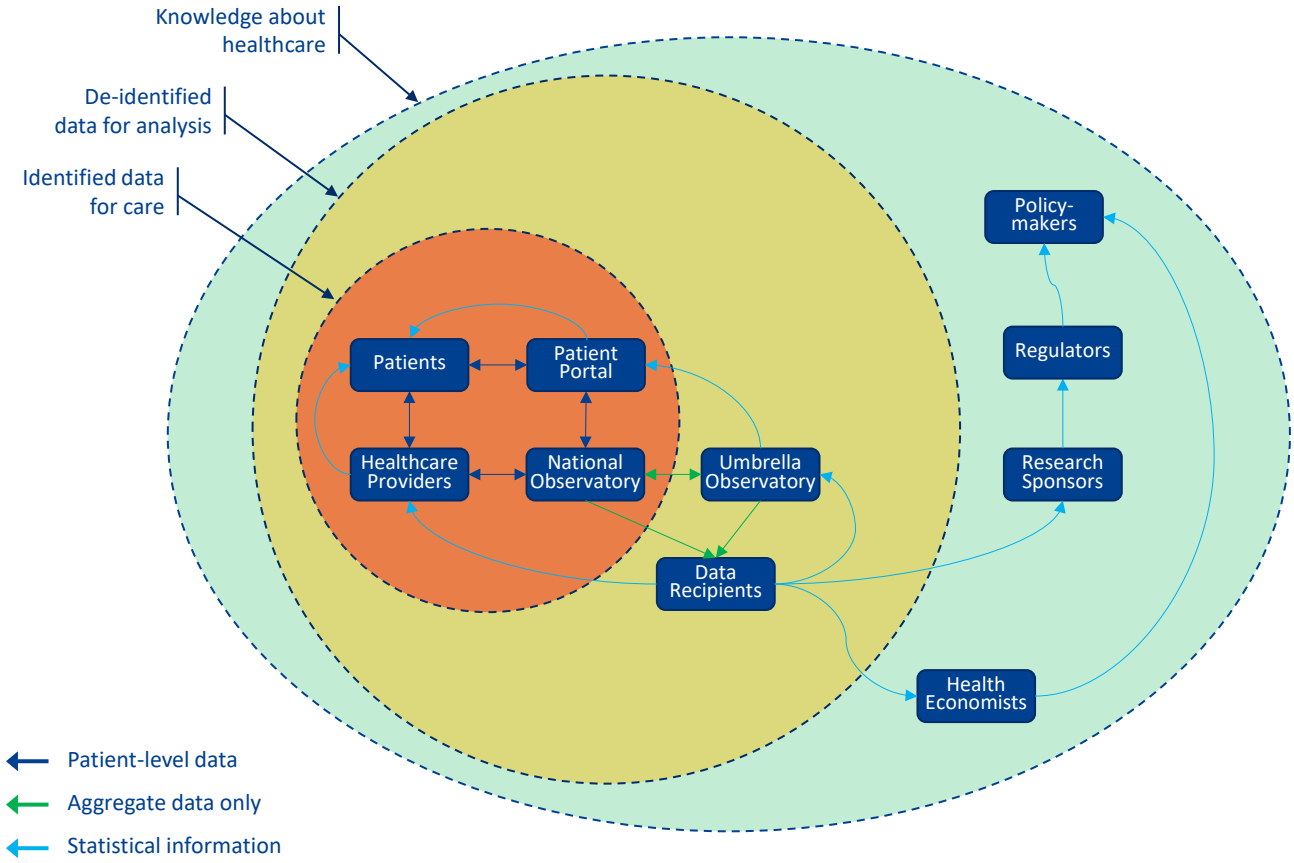


*Figure 4 - Stakeholder Map in terms of information flows*

The IGF and Business Architecture need to consider the different stakeholder groups and how best to interact with them to ensure the sustainability of the eco-system and its compliance with legal requirements and public expectations.

Of particular concern for the IGF are the relationships with stakeholder groups involving exchange of personal data:

- Patients – as providers of outcomes information – have a key concern in establishing identity and authenticating access to potentially sensitive personal data through the Patient Portal
- Healthcare providers – both providing and receiving outcomes data – are concerned with the process of matching up the identity of patients through the Patient Portal and the patient information held by the healthcare provider
- The relevant National Observatory – is invested in the need for effective and secure links (both systems and organisationally) with healthcare providers as well as being transparent with the public and other national bodies, particular national or regional healthcare infrastructure (e.g. patient registration indices)
- Patients – as recipients of health information from the Observatories - both broad information about conditions as well as information which may be personalised to their profiles – needs to trust that this information is accurate, pertinent, and reliable as well as being safe.

## Controllership Model

This 'model' is a term simply for the different participants and how they are responsible (or not) for the various dataflows and holdings across the H2O eco-system. Most legal entities being described will be 'controllers', though whether they act together with others as 'joint controllers' will be part of what the 'Controllership Model' must lay out.

The principal purpose of the Controllership Model (apart from stating which parties are acting jointly and which as sole controller with respect to key processing) will be to help establish what contracts are needed between parties and what the nature of those contracts needs to be.

The 'model' is then chiefly a description of the agreed relationships between the various parties and will both feed into and be informed by the agreements or contracts established between them.

One key aspect will be the decision as to whether an 'observatory' is a single legal entity, handling both identified data from patients and care providers, or two separate entities, one acting as a 'data broker' collating identified data between parties and the other receiving only anonymised data to support health research, developed internally or being shared with other parties.

# Links between IGF and Systems Architecture (D2.2)

Similarly, there must be close liaison between the IGF and the Systems Architecture to ensure that the expected 'rules' can be reliably and efficiently implemented across systems without being rendered ineffective or being so burdensome as to prevent the eco-system being unsustainable or unacceptable as a 'trusted environment' to patients, the public, or regulators.

## Data Management Plan (DMP)

Version 1 of the DMP has already been developed in conjunction with WP2 after liaison with the other work packages and based on their initial intentions and plans. The document will be updated (D1.9 and D1.11) as the project progresses and detailed arrangements develop.

## Consent Model

National Observatories will have to adapt to their home country's health information infrastructure, including existing mechanisms, nationally or regionally, to collect PROMs, so there will be a number of opportunities and constraints on how data can actually flow – and hence what patients may be able to choose. There may also be member state regulations or local healthcare options that need to be respected[3].

However, some of the fundamental choices for any patient would be:

- To simply ignore any PROM initiative – trivial for the patient to do and for H2O to support
- To visit an Observatory website or that of the H2O project
    - And therefore be tracked (anonymously) on the web-site for web-site analysis or other 'marketing' purposes
    - And be offered 'cookie' options, so may or may not create relevant cookies

[3] For example, (though no longer in the EU), England provides a 'national data opt-out' whereby patients can request that their 'confidential patient information' is not used for 'research and planning'

- To register on an Observatory website to investigate further what options are or to receive invitations or newsletters
- (if registered)
    - to see what care providers (integrated with the Observatory) are in the vicinity and
    - to see what PRO surveys or questionnaire might be completed
- to choose to sign-up with the Observatory for PRO purposes, and
    - provide 'proof of identity' information to be sent to linked care providers for identity assurance
        - Level-1 – to allow PROM data from patient to be linked to their EHR
        - Level-2 – to allow private EHR data (in whole or in part) to be passed back to patient
    - provide basic details concerning relevant conditions (to complete PRO forms)
    - select Observatory care providers with whom to share information
    - set broad Observatory data-sharing controls [to be elaborated during the project]
- (if signed-up)
    - To suspend involvement (including data-flows)
    - To stop further involvement
- To make a formal GDPR rights request, including complete deletion or requesting a copy of information held or provided by the patient

This provides a wide range of possible 'consents' to be tracked and supported, both in terms of the detailed options offered at various times and the patient's actual choices of a particular option at a certain time (given that patients must be allowed to change their minds or withdraw entirely).

## Information Security Policy

This document (to be produced with WP2) will define information security requirements at a high-level (ideally, in technology-neutral terms). It will build on the overall dataflows described in the D1.2 Data Management Plan, the business structures and interactions from D1.1 Observatory Business Architecture, and the system architecture described in D2.2 System Architecture Blueprint, as well as the principles and rules from this document. It will also have to support the consent model described here in terms of recording consents against datasets as a whole and possibly against individual records, reflecting patient choices at source, from within national infrastructures, as well as those expressed through the Patient Portal.

The Information Security Policy will describe:

- The Consent Management system – how consents or dissents will be recorded and effected across the H2O eco-system.
- The requirements for Access Control sub-systems at the observatories
- Data Minimisation approaches to be applied across the system architecture, including Pseudonym Management
- Auditing and Monitoring facilities that should be developed within observatory systems to allow effective control of access and early detection of possible misuse or intrusion
- Data Quality checks to be applied at various points (usually at ETL (Extract, Transform, Load) pre-checks)

# Data Protection Impact Assessment (DPIA) template

Consortium partners and Data Access Providers (DAPs) will have their own preferred format for DPIAs, including their own process for assessing the level of risk to the rights and freedoms of data subjects. This template (Milestone 5, due for Month 10) merely lays out the broad detail of the project and how the risks involved might be considered to assist in the development of the actual DPIAs which may be required[4].

Even if not a formal DPIA in a legal sense (as the consortium is not a legal entity) it will still help define and refine possible approaches to information security to be developed in the Information Security Policy to be developed between WP1 and WP2.

This document will necessarily be comparatively high-level about the project as a whole. Each controller within the eco-system will need ensure that there is a suitably detailed DPIA (possibly developed by or in conjunction with other parties) to cover its own processing, reflecting its role within the eco-system. Each national Observatory may produce its own DPIA template for contributing data providers of a particular type (e.g. hospitals) to aid them in participating in the H2O eco-system.

[4] A DPIA is required by Article 35 of the GDPR where a type of processing is likely to result in a high risk to the rights and freedoms of natural persons; in particular, where processing of 'special category' (e.g. health) personal data is to happen then a DPIA is required. This is considered a 'formal' DPIA in this document; however, some form of risk assessment is needed even to establish whether there is any likelihood of a 'high risk' or whether Article 35(3)b would apply (e.g. where data is anonymised) – this might be considered an 'informal' DPIA, which would be considered good practice in any event, though might only appear as a sub-section of a broader document.

# 3. Key Principles for H2O data governance

## GDPR Principles

The General Data Protection Regulation defines 7 key principles for data protection in Article 5:

1. lawfulness, fairness and transparency: [data shall be] processed lawfully, fairly and in a transparent manner in relation to the data subject);
2. purpose limitation: [data shall be] 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;
3. data minimisation: [data shall be] adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;
4. accuracy: [data shall be] 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;
5. storage limitation: [data shall be] 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;
6. integrity and confidentiality: [data shall be] 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;
7. accountability: The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 [the previous 6 principles]).

The first 6 principles build on the principles defined in the Data Protection Regulation, but it is the seventh which introduces additional burdens on the controllers of any 'personal data', such as the need for Data Protection Impact Assessments (DPIAs) which were previously recommended as 'good practice' but are now a legal requirement. However, clear 'accountability' is a key support for public trust in the H2O eco-system, so should be viewed as an objective to be actively sought and promoted rather than treated as a bureaucratic overhead.

The GDPR also includes a further principle of 'privacy by design and by default' (precise requirements are laid out in Article 25). This is the whole purpose of this document – to feed into other work packages and deliverables, such as D2.2, the Architectural Blueprint.

These GDPR principles are critical components of the suggested H2O 'principles' laid out below.

# Recommended H2O Principles

These principles are intended to include aspects of the GDPR principles, but not to simply reproduce them. The principles described here are developed further in the 'Rules' section which provide more detail on how best to meet the 'principles', but will still require organisations to develop processes and procedures to implement both the principles and the 'rules'.

## Operational principles

- Effective oversight – there should be an independent panel[5], including lay representatives, which reviews how patient privacy and legal compliance is preserved across the H2O activities – within the project and across the observatories and their data providers and clients. This may be the H2O Ethics Board initially. The oversight panel should receive regular reports on security assessments and any incidents or near-misses that may occur.
- Risk management – perform DPIAs or other risk assessments as necessary
- Appropriate use – define what is/is not 'appropriate use' for staff and users with access to datasets
- Operational effectiveness – ensure that system controls, while effective, are not overly burdensome and are flexible to support different business models as well as possible changes in legal or cultural aspects of acceptable data processing.

## Organisational principles

- An effective IG structure with assigned responsibilities and policies to ensure that information is used appropriately; clear lines of responsibility – including incident management and reporting structures
- Appropriate IG training for all staff and adequate resourcing to ensure that IG is effectively delivered
- Effective monitoring of access to and use of personal data by staff (and clients, where relevant)
- Routine assessments of legal and other regulatory compliance with reports to the management board
- Effective policies for review of web-site materials and public information to ensure that these cover all aspects of processing, particularly GDPR requirements (Articles 13 & 14)

## System design principles

- Data minimisation – in particular, using minimal identifiers, pseudonyms, or tokens whenever possible
- Retention period restriction – any microdata should only be held (by clients) for a limited period (clearly, observatories need to hold the data indefinitely)
- Clear secure procedures for exchange of datasets – from DAP to observatory; from observatory to client (or other observatory) – datasets encrypted in transit – and at rest
- Strong and ideally certified information security
- Effective access control measures – with access privileges retired or suspended if unused or suspicious activity is discovered
- Adequate and effective audit trials – with tools to support analysis and investigation as well as providing forensic evidence if required

[5] This does not require that all members of the panel are 'independent' only that the overall composition is 'independent' of any particular party or the H2O project itself

## Contractual principles

- Clear data-sharing agreements with conditions that are known to users (see training above) and routinely appraised as being respected
- Where parties receive 'personal data' (or data that is not established as anonymous within the context of GDPR) then the data recipient should have some assured form of security assessment (e.g. ISO27001) or the data provider should have the right to check or audit the data recipient's security arrangements.
- Effective incident reporting procedures (at client and for informing observatory) so that GDPR reporting timelines are met.
- Effective sanctions in case of contractual breach – both that any data breach is contained and further risk of breach eliminated, and that is an effective incentive for clients to abide by the contractual terms
- Clarity over controller/processor statuses, in particular in order to uphold data subject rights. This should include clarity concerning use of sub-processors and responsibilities for resolving any breaches or incidents – all parties should support each other to investigate and find an appropriate resolution
- There should be a clear position over data retention and how data should be thoroughly erased at the end of any project or agreement, including any back-up facilities or other copies of the data received
- Restrictions on use should 'flow down' to staff through employment or confidentiality agreements, supported by appropriate training

It should be noted that contracts may need to make a distinction between the original data received and aggregated or derived data to which any IP or GDPR restrictions would not apply. They may also need to allow for data that has only been trivially manipulated or modified.

# 4. Suggested 'Standard Operating Rules' for Observatory operations

Note: these 'rules' can be linked back to the 'principles' outlined above, but a 'rule' may be attributable to more than one 'principle' (e.g. as supporting 'transparency' and 'security'), so we have not attempted a form of hierarchical numbering from the principles. We have grouped the numbering of the rules by broad categories, loosely related to the principles above.

## A. Oversight

A1 There shall be an independent oversight panel, including lay representatives
A2 The oversight panel shall convene (virtually or in person) at least annually
A3 There shall be a clear 'terms of reference' document for the oversight panel, laying out its scope, range of activities, and composition; this document will be reviewed by the panel at least every three years; variations must be agreed upon by the Observatory management board

## B. Transparency

B1 There shall be clear lay descriptions of the data collected and used across the H2O eco-system on each observatory web-site which should also appear or be linked to from data provider web-sites, so that the public (and particularly possible data subjects) can understand the purposes, risks and benefits of the data processing
B2 The oversight panel should provide an annual report on its activities and any implications for data subjects – with any recommendations for improvement across the H2O eco-system
B3 The 'terms of reference' document and composition of the oversight panel should also appear on Observatory web-sites

## C. Contractual

C1 All staff or contractors will not be granted access to systems with patient data unless they have signed or are subject to confidentiality requirements:
  i) They will not seek to single out or identify the data relating to an individual, unless explicitly authorised to do so
  ii) They will not share or reveal any confidential information unless explicitly authorised to do so
  iii) They will only use the data for the purposes for which they are authorised
  iv) They have read and understood the Acceptable Use Policy (for the system or systems to which they will be granted access)
C2 Joint controllers may decide which party or parties is/are to contract processors on behalf of all the controllers, viz there may be a common contract or just a 'lead' controller
C3 There should be an explicit agreement between any joint controllers
C4 All processors should be subject to contract (as per GDPR Article 28(3))
C5 Any appointment of sub-processors shall be subject to the approval of the contracting controller(s)
C6 All agreements shall state how GDPR data subject rights requests are to be managed and processed
C7 All agreements shall state which parties are responsible for managing potential data breaches (and near-misses) and shall collaborate to resolve such breaches and prevent further breaches

## D. Systems (Data minimisation)

D1 System architecture shall be designed to separate wherever possible personal identifying information (e.g. names and contact details) from other data
D2 Wherever possible individual-level data will be pseudonymised or anonymised

D3 Query systems should be capable of replacing record-keys with a persistent token specific to the recipient organisation (so that different datasets cannot be linked between accounts); this will require a pseudonym management system
D4 Applications to access datasets should select the data attributes required for the purposes of the study or analysis required with a justification; once approved, only these attributes should be available for query selection or extraction
D5 Where possible a sampling mechanism should be applied to minimise the level of attribution possible to a particular record
D6 Query systems should be capable of automatic 'blurring' or grouping of particular data attributes, either as part of an application to access data or where a query produces a small number of records for extraction

## E. Systems (Information Security)

The detailed information security arrangements for Observatories will be developed with WP2 in the Information Security Policy, so these are quite broad 'security' requirements:

E1 Data derived from patient data shall be treated for security purposes as though it were 'personal data' unless aggregated or subject to an assured anonymisation process (e.g. k-anonymisation) to prevent possible re-identification
E2 Access to data and system functionality shall be role-based and limited to the minimum functionality required; exceptional access should be time-limited and provided to an individual rather than added to a generic role
E3 System administrator access should not include access to record-level data and should be subject to at least two-factor authentication; administrator accounts and passwords must be changed from the installation defaults and subject to change whenever key personnel leave
E4 System security policies should ensure that:
  i) User passwords are strong (long & complex, but memorable); forced password refreshes[6] should be avoided unless there is evidence or increased likelihood of compromised accounts
  ii) Accounts are regularly reviewed to identify dormant accounts or exceptional activity
  iii) Accounts are suspended/archived whenever a person changes jobs or roles or leaves an organisation
E5 Users should be trained
  i) How to support effective security
  ii) To respect the confidentiality of the data
  iii) How to detect possible attacks on systems (phishing, etc.) and how to report any possible attacks or misuse
  iv) to use strong passwords
E6 Systems should maintain audit trails of user activity (but avoiding recording any primary data where possible); there should be reporting and analysis systems to allow these audit trails to be used effectively to detect and investigate possible misuse; ideally, the reporting systems should allow the linking of user activity across systems (e.g. from log-in through to level of data accessed)
E7 Systems should maintain a record of queries submitted so that these can be analysed for any evidence of 'jigsaw' attacks or attempts to single-out cases.

[6] Typically where user is required to create a new password every 60 days or similar – this usually causes users to skimp on password strength (or forget their current password) which is counter-productive

## F. GDPR compliance (accountability)

These 'rules' do not seek to cover standard GDPR compliance responsibilities, only adding points of clarity over and above GDPR requirements.

F1 The H2O project [Management Board] shall ensure that all of its members are clear about their position as controllers of any 'personal data' within the H2O project and their resulting legal responsibilities – this is to avoid possible confusion over the role of the H2O project which is not a legal entity.

F2 Each organisation acting as a controller (this includes the observatories, once constituted) shall develop a DPIA to cover their intended operations in respect of patient (or patient-derived) data and, where identified as 'high risk', will consult with their Supervisory Authority

F3 Each organisation acting as a controller will have a clear communications approach to informing data subjects and the general public about its possible use of personal data within the H2O project. This should include transparency about other organisations which may receive patient data as well as information about research outcomes or other benefits from the H2O operations.

# 5. Codes of Practice for Observatory operations

## Code of Conduct versus Code of Practice

In this document, we interpret these two terms as distinct, though this is not always the case in general use.

GDPR Articles 40 & 41 define 'Codes of Conduct' with the implication that these are stronger than a mere 'Code of Practice'[7]. This has led to a greater distinction between the terms – some older 'codes of conduct' might now be termed a 'code of practice' to avoid confusion. For example, the ENCePP 'Code of Conduct' (which provides a set of rules and principles for pharmacoepidemiology and pharmacovigilance studies) might now be termed a 'code of practice', even though it was intended to be prescriptive.

**Code of Conduct:** a binding agreement on rules of behaviour for the members of a group or organisation; members commit to abiding by the code, are registered as such, and are monitored to ensure they comply with the code. This may involve formal certification through a certification authority.

**Code of Practice:** a voluntary set of rules of behaviour for the members of a group or organisation; members may commit to abiding by the code, but there is no register nor certification/validation of compliance. However, failing to follow a code of practice might be grounds for expulsion from the membership of an organisation which sets the code of practice.

There may be some ambiguity within these definitions as to what constitute 'rules of behaviour', but the primary aim of these definitions is to distinguish the terms rather than define them absolutely.

With these definitions, we would describe the suggested H2O 'code of conduct' as described in the DOA as actually a 'code of practice'- at least at this developmental stage. It may develop sufficient rigour to be considered a 'code of conduct' once it is more fully developed and there is an eco-system for certification.

Certification versus accreditation

- Certification represents a written assurance by a third party of the conformity of a product, process or service to specified requirements.
- Accreditation is the formal recognition by an authoritative body of the competence to work to specified standards.

so a certifying body (which assesses organisations' compliance to standards) may need accrediting as assurance that its certificates are worthwhile and reputable.

## Existing Codes of Conduct/Codes of Practice

There are no directly relevant codes of conduct nor codes of practice for the 'observatory' approaches being considered for H2O. However, there are quite a range of existing code or practice which may be pertinent to some aspects of the dataflows envisaged for H2O:

[7] For example, the UK Supervisory Authority, The Information Commissioner's Office (ICO), produced in 2012 some guidance: 'Anonymisation: managing data protection risk code of practice' which was to encourage good practice rather than prescribe exact methods which could be certified.

| Acronym | Organisation | Focus | Notes/comments |
|---|---|---|---|
| ENCePP | encepp.eu | Pharmacoepidemiology and Pharmacovigilance | Part of GVP (Good Pharmacovigilance Practice) |
| ADVANCE | vac4eu.org | Best practice and code of conduct for benefit-risk monitoring vaccines | Designed as an alternative to ENCePP to address conflicts of interest<br>Builds on Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology (ISPE) and the Good Epidemiology Practice (GEP) |
| EMIF | emif.eu | Data catalogues, data access, and distributed querying | Built on ENCePP & EHR4CR SOP (not EHR4CR CoP) |
| ELIXIR ELSI Policy | elixir-europe.org | Omics data re-use | ELIXIR framework for secure archiving, dissemination and analysis of human access-controlled data |
| RD-Connect | rd-connect.eu | Biobanking, including genomics | Produced by RD-Connect GPAP, but adapted from EHR4CR CoP |
| EHR4CR | imi.europa.eu | Secondary Use of Medical Data in Scientific Research Projects | Bahr & Schlünder – developed in parallel with EHR4CR but as part of integration across IMI projects |
| BBMRI | code-of-conduct-for-health-research.eu | | Building on EHR4CR CoP? since 2015<br>No information currently available |
| CORBEL | corbel-project.eu | | Code of Conduct for Health Research (Mayrhofer)= BBMRI above |
| GEANT | geant.net | Data Protection Code of Conduct for identity providers | Pre-GDPR; minimal requirements and relates to identity of end-users |
| ECCRI | Allea.org | | European Code of Conduct for Research Integrity – first published 2011 |
| UKRIO | ukrio.org | | UK Research Integrity Office Published 2009 |
| GLP | oecd.org | Good Laboratory Practice | Chemical safety & testing |
| FAIR | | Research data re-use | Principles: findable, accessible, interoperable and reusable (FAIR) |

# ANNEXES

## Annex 1 – Acceptable Use Policy (AUP) – key points

It is considered that H2O would need to have distinct AUPs for:

- Data analysts receiving or having access to micro-data
- Data analysts receiving only aggregate data or having access to a Trusted |Research Environment (TRE) returning only aggregate results.

Clinical staff receiving identified patient data should already be covered by professional codes of conduct. Technical administrators should also already be covered by employment contracts covering confidentiality and security.

The situation of H2O consortium staff having access to personal data in the form of contact details (other than for patients) is considered out-of-scope for this document as being primarily concerned with the governance of healthcare data. This possibility was covered in D1.2 Data Management Plan v1.

### AUP for aggregate data or statistical results

- H2O aggregate data will be accessible only to *bona fide* researchers in the interest of patients, science and society (be it treatment, research, or innovation that benefits patients and/or society)
- Users must not apply repeated queries to narrow down datasets to a few individuals because the risks of re-identification nor apply reconstruction methods to try to recreate the original data from extracted results
- All users should be aware of relevant data protection and confidentiality legal requirements when using healthcare data in their jurisdiction (which may include derogations from GDPR or additional legislative controls)
- Publications must apply small cell-count checks to ensure that individuals could not be identified by any means from the published data[8]
- Users must be mindful of patient confidentiality and the need to maintain public trust in the H2O eco-system to the benefit of all

### AUP for access to micro-data

- H2O provides data only to bona fide researchers in the interest of patients, science and society only for public healthcare purposes (be it treatment, research, or innovation to benefit patients and/or society).
- Users must not attempt to link H2O datasets to any other datasets (particularly, datasets identifying individuals) because the risks of re-identification and breach of trust
- Users must report any data errors or concerns to H2O without attempting to investigate the root cause because of the risks of singling out and possible re-identifying individuals; H2O will then carry out its own investigations (possibly with the assistance of the user)
- All users must be trained about relevant data protection and confidentiality legal requirements when using healthcare data in their jurisdiction (which may include derogations from GDPR or additional legislative controls), including the specific restrictions when accessing H2O datasets as well as appropriate reporting of possible data issues or breaches

[8] Systems may apply these controls to the actual data provided to users, so this might be unnecessary, but some users (e.g. for regulatory purposes) may receive the direct results of queries when this restrictions would be necessary

- Publications must apply small cell-count checks to ensure that individuals could not be identified by any means from the published data[9]
- Users must be mindful of patient confidentiality and the need to maintain public trust in the H2O eco-system to the benefit of all

[9] Systems may apply these controls to the actual data provided to users, so this might be unnecessary, but some users (e.g. for regulatory purposes) may receive the direct results of queries when this restrictions would be necessary

# Annex 2 – Data-sharing Agreement – Heads of Agreement

'Heads of Agreement' is a term for a business requirement in a possible contract, which may not fully detail the appropriate legal form or precise sub-conditions. This does not attempt to cover what may be 'standard' contract terms, concerning any charges or fees for services or routine matters such as contract variations and notification, except where they have a specific 'information governance' aspect which might otherwise be overlooked. References to GDPR terms, such as 'personal data' are pre-supposed as are any technical definitions concerning datasets or methods for data transfers or systems interfacing.

Note that the actual contractual arrangement might be an overarching 'framework' contract (perhaps for a term of several years) with subsidiary data-sharing agreements for specific datasets, services, or facilities probably for 12 months only, though with the intention of renewal, assuming circumstances do not change.

Caveat: this list of terms is merely advisory – it cannot be taken to be exhaustive and cannot anticipate all of the circumstances to be covered by a particular data-sharing arrangement, so proper legal advice will still be required.

Role of parties

- All parties to the agreement or referenced in the contract shall be identified (where relevant)
- All parties shall be identified as a 'controller' or 'processor' where appropriate
- Where parties are 'joint controllers' then the detailed processing for which they are joint controllers shall be identified and detailed; where each controller maintains individual control of a distinct set of the data, then this shall be detailed (so they might jointly contract a processor for processing services, but otherwise each retain their data exclusively)
- Joint controllers may delegate some contractual matters to a 'lead' controller, but subject to notification of any changes (prior to or immediately after the change as appropriate)

Scope and Purpose of agreement

- The purpose of the agreement will be either of[10]:
    - Direct patient care – where identified data will be shared to a healthcare provider (or providers) to support the care of the specific data subjects
    - Healthcare research – where de-identified, identified, or aggregate data will be shared either through provision of explicit datasets or by provision of a data access facility (possibly a 'trusted research environment')
- The data covered under the agreement would likely be one of
    - Aggregated data (possibly from a TRE)
    - De-identified microdata (probably pseudonymised or even deemed anonymous)
    - De-identified microdata (probably pseudonymised or even deemed anonymous) with a 'linker file' allowing the data to be linked to other de-identified datasets
    - Identified data which may be directly or indirectly linked to other datasets
    - Identified data (though typically only for the purpose of healthcare treatment of the data subjects)
- The agreement must make clear whether data may be further shared by the recipient and under what conditions and subject to what purposes and controls (e.g. that it may only be shared in aggregate form).

[10] There would also be 'service contracts' between a controller or controllers and a processor which must meet the requirements of GDPR Articles 28 and 29 and so are not considered here, except where a 'trusted third-party' is used to provide linkage services to two or more controllers.

- In particular, the data-sharing shall be clear about the territory (or territories) of use of any 'personal data' so that the rights of the data subjects can be properly upheld (e.g. that may only be processed in countries deemed 'adequate' by the EU).
- Alternatively, the agreement may specify requirements for anonymisation (including aggregation) so that territorial restrictions may not be required.

### Data minimisation

- The agreement may specify particular restrictions on the data, such as levels of pseudonymisation or data blurring, publication only of aggregate data, possibly with small cell-count or other statistical disclosure controls. This may involve adherence to particular forms of regulatory guidance including codes or conduct or practice.

### Information security standards

- Ideally, the data recipient should have ISO27001 security certification, but this is not always possible, so an equivalent security assurance mechanism should be required. This may depend on national schemes or standards. Such certification or assurance should be maintained throughout the term of the agreement.
- Any contracted processors would also be required to meet these standards or equivalent. Relevant processors may need to be declared and subject to approval by the data provider[11].
- The data providers should have the right to audit and/or inspect that the recipient is following the agreed procedures to the required standards and can suspend or terminate the agreement if any identified shortcomings are not rectified in a timely manner.

### Data Breach handling

- The parties must report any potential data breach (or near-miss) relating to the data in question to the other party (or parties) within 24 hours in order to be able to meet GDPR Article 33 requirements
- The agreement should make clear who needs to report any data breach to their Supervisory Authority (if not already clear – e.g. in a controller-processor contract)
- All parties must cooperate in investigating whether a data breach has taken place, resolving the breach, and establishing corrective action for the future in a timely manner.
- Where data is expected to be aggregate or anonymised, there should be relevant clauses to allow for the possibility that the data is inadvertently or deliberately re-identified and so become 'personal data' unexpectedly. In particular, what should happen to the data provided (possibly to be deleted entirely or to have such re-identified data purged).

### Intellectual Property Rights

- The agreement should make clear how any IP rights are applied between the parties and between the original data supplied and the results of further data processing.
- The agreement may also require that any publication or further sharing of the data must acknowledge the source of the data and the contribution to the end result (e.g. paper or article). A specific form of words is helpful.

### Data Subject Rights management

- The agreement should make clear how individuals GDPR rights (Articles 15-23) will be managed between the parties – the data subject may apply to any of the controllers involved, but it may be agreed that the actual provision of information or restrictions on processing may be performed by specific parties.
- In particular, it should be clear whether any further recipients of the data (or data derived from it) should be informed of expressed data subject wishes to have further processing stopped or data erased. To this end data recipients should keep records of other parties to which they may release identifiable data.

[11] This measure of control may prove difficult in practice and burdensome for all parties.

### Flow-down of contract terms

- The agreement should require that any sub-licensing or consequent use should be either forbidden or only permitted on terms equivalent to those in the main agreement.
- There may be pre-conditions from 'upstream' agreements with other parties which will need to be included. These may include aspects such as

### Staff/user Training

- The recipient must ensure that staff have the necessary training to meet the obligations imposed by the agreement:
    - Be aware of and understand the processes (e.g. data breach reporting) involved in meeting the terms of the agreement.
    - Understand the confidentiality of the data and the requirements of data protection law (as far as it applies or might apply to the data)

### Termination of agreement

Most agreements have standard terms to cover both the circumstances in which the agreement can be terminated (or will terminate by default at the end of any period specified in the agreement), or the process that must be followed in order to terminate the agreement at the behest on one party alone or by mutual agreement. These 'standard terms' are not covered here.

- The agreement must detail what is to happen to any data shared under the agreement.
    - Where provided for treatment of patients, then data already used in the context of treatment will need to be retained for evidential reasons, but may have been extracted from the delivered dataset into health records, so that the original delivered dataset can be deleted (along with any back-up or other copies)
    - Where provided for research purposes, then any original microdata (and any back-up or other copies) should be deleted with only aggregated analytic datasets retained as evidence for the research.
- Obligations of confidentiality and security of the data must persist after the termination of the agreement.
- The data provider(s) may have the right to suspend or terminate the agreement if the data recipient suffers a significant data breach in respect of personal data or otherwise fails information security requirements.