1+MG Glossary (v May 26, 2021)

Preface

The 1+MG Glossary aims to provide explanations on the use and specific meaning of data protection, ethics and technical terms relevant to the 1+MG Project. This glossary uses plain language to enable users with different backgrounds to understand these terms. For terms that are defined in the General Data Protection Regulation 2016/679 (GDPR), for example 'personal data', 'controller', and 'processing', the definition in Article 4 GDPR will be legally binding. Explanations of terms based on the GDPR are highlighted in blue, to distinguish them from non-legal (i.e., scientific, technical or ethical terms. This document draws largely on previous efforts and existing definitions (see references).

Please send any feedback or comments to Adrian Thorogood: adrian.thorogood@uni.lu

Data Characterisation

- **Personal data**: data related to a living individual who is likely to be identified by the data directly or in combination with other data (e.g., through a pseudonym). [ref. Art. 4 GDPR]
- **Special categories of data**: types of personal data that are legally recognized as sensitive and whose processing is prohibited unless that requires a special legal justification to be processed (e.g., health, ethnic origin, sexual orientation, and genetic data). [ref. Art. 9(1)(2) GDPR]
- **Anonymous / anonymised data**: data that does not relate to an individual person or that has been irreversibly processed so that the individuals are no longer identifiable by the data controller or by another person. [ref. Rec. 26 GDPR][1]
- **Direct identifiers**: data types that alone are likely to lead to the identity of an individual (e.g. name, identification number, address, online identifier, date of birth).
- **Indirect identifiers**: data types that alone are not likely to lead to the identity of an individual, but are likely to do so in combination with or by comparison to other data.
- **Non-identifiers**: data types that are unlikely to lead directly or indirectly to the identification of an individual, (e.g., basic phenotype data falling within normal ranges, such as gender, age range, or disease status).
- **De-identified data**: data where direct and indirect identifiers have been removed or generalised. The data may still be likely to reveal the individual's identity, depending on the level of de-identification which has taken place.
- **Ethically sensitive data**: data that could have a potentially large impact on the rights and interests of an individual, including data types not legally defined as special categories of data by the GDPR (e.g., some life-style data such as alcohol consumption).
- **Health data**: personal data related to the physical or mental health of an individual independent of its origin (e.g., from the healthcare context, from research, from clinical trials, from the data subject directly, from smart devices, etc.). [ref. Art. 4 GDPR]

[1] An overview and evaluation of anonymisation techniques can be found in the former Working Party 29 opinion 05/2014.

- **Genetic data:** personal data relating to the inherited or acquired genetic characteristics of an individual which give unique information about his/her physiology or health and which result from an analysis of a biological sample from the individual in question. [ref. Art. 4(13) GDPR]
- **Pseudonymised data**: personal data that has been processed in such a manner that it can no longer be attributed to a specific individual without having access to additional information (e.g., a key or additional data), which is kept separately. Pseudonymised data is still considered personal data even if the actor processing the data does not have access to this additional information. [ref. Art. 4 GDPR]
- **Healthcare data**: health data on individuals collected in the healthcare context.
- **Health research data**: health data on individuals collected in the scientific research context.
- **Clinical study data:** health data on individuals collected in clinical studies.
- **Metadata:** a set of data that describes and gives information about other data. Metadata do not include any data that are processed to produce any results of the data use such as health or genetic data. Metadata can be personal data where they contain information on the subject level (e.g. consent decisions of an individual data subject)
- **Data collections:** Data collections are data that come from the same collection context from the same controller(s) and can be characterised with non-personal metadata. A data collection can e.g. refer to data of a cohort or a public registry.
- **Records:** Data related to individual data subjects
- **Data set:** Data that are grouped in a certain context, e.g. for a user's access.
- **Enriched data:** data where additional information is added, e.g. annotations
- **Derived data:** data that have been created by alteration of the original data (e.g. through data curation)

Organizational Roles

- **Data controller:** an entity who determines the purposes and means of the processing of personal data. A controller may not necessarily have direct access to the data. [ref. Art. 4(7) GDPR]
- **Data processor:** an entity who processes personal data on behalf of a controller. [ref. Art. 4(8) GDPR]
- **Joint controllers**: two or more entities who together determine the purposes and means of the processing of personal data, and are therefore jointly responsible for data protection compliance. [Art. 26 GDPR] Responsibilities, rights and obligations among joint controllers are usually defined by agreement.
- **Data custodian**: an organization who collects and uses data and makes initial decisions on data use, sharing, retention and disposal.
- **Data provider**: an entity who decides to make data available.
- **Data recipient:** an entity to whom data is disclosed (receives a physical copy of the data or merely has access of some kind).
- **Data user**: an individual who analyzes or directs the analysis of data with the aim of deriving a result.

- **Data host**: an entity who stores data, often on behalf of another organization or organizations.

Glossary of Terms

Institutional Roles (a single organization may carry out multiple roles)

Initial Controller - the institution responsible for the initial data collection / generation for a primary purpose (e.g., research project and/or healthcare test).

Data Holder - institution who hosts and maintains data in 1+MG (and is usually also the Data Provider).

Data Provider - institution who has the authority to grant or refuse access to data (and is usually also the Data Holder).

Permit Authority - a entity given authority by legislation to grant or refuse access to data (e.g., data held by other organizations)

Data Requester/User - a user who requests access to data and/or uses data.

Technical Infrastructure and Services

Repository - a service for curating and managing data. May be certified by 1+MG.

Secure Processing Environment - a service for remotely analysing and computing on data. May be certified by 1+MG.

Other

1+MG Access Office - a central administrative service to streamline and coordinate access to 1+MG data.

Data Access Committee (DAC) - committee hosted by a Data Provider with a mandate to review access requests according to defined criteria.

Research Ethics Committee (REC) - Group of individuals who undertake the ethical review of research protocols involving humans, applying agreed ethical principles. (World Health Organization, 2009)

Signatory Country - country that has signed the 1+MG Declaration.

1+MG Member Country - country that agrees to be a formal partner in 1+MG.

Processing

- **Processing**: any operation or set of operations performed on personal data, including but not limited to collection, storage, analysis, disclosure, or destruction. [ref. Art. 4(2) GDPR]
- **Cross-border processing**: processing by a controller or processor in the EU/EEA that is carried out in the context of establishments in more than one Member State, or that affects data subjects in more than one Member State. [ref. Art. 4(23) GDPR]
- **International transfer**: the deliberate step to share personal data with an entity under a jurisdiction outside the EU/EEA, which can only be done under the GDPR under certain legal mechanisms.
- **Data disclosure**: to transmit data, disseminate data or otherwise make data available to someone else (e.g., showing data on a screen), whether intentionally or unintentionally.
- **Data access**: the ability, right or permission to act on data in a defined location.
- **Data visiting**: querying or running algorithms on data and receiving the result without having direct access to data.
- **Data hosting**: the storage of data on a stable and accessible platform by one party for one or more other parties.
- **Data analysis**: to process the data with the intent to generate knowledge from the data.
- **Data transmission**: to physically relocate data.
- **Data sharing**: to make data available to another individual, organization, or community for a defined purpose.
- **Data linkage**: bringing together data that relate to the same individual from two or more different sources. This may increase the utility of data (e.g., by revealing new relationships between factors), but may also increase the risk that the individual is re-identified.
- **Further processing**: the processing of personal data for purposes other than those for which the personal data were initially collected by the controller. [ref. Art. 5(1)(b) and 6(4) GDPR]
- **Secondary use**: using data for a different purpose than the one that was driving the collection or generation of the data, whether or not the different purpose was communicated at the time of collection (e.g. using healthcare data for research purposes).
- **Anonymisation**: processing data irreversibly in such a way that they are no longer likely to lead to the identity of an individual. [ref. Rec. 26 GDPR]
- **De-identification**: a process of stripping direct and/or indirect identifiers from data in order to lower the risk the data will lead to the individual's identity.

Modalities of Data Access

- **Open access**: access to data free of charge and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship [ref. Open Data Definition].
- **Controlled access**: a data access model whereby qualified data users apply for data access based on information provided in an application form and their applications are

reviewed by a data access officer or committee. [ref. GA4GH, Data Privacy and Security Policy, 2019]

- **Registered access**: a data access model whereby qualified data users apply for data access to one dataset or multiple datasets at once by providing details of their identity for authentication and agreeing to terms and conditions of data use during the registration process. [ref. GA4GH, Data Privacy and Security Policy, 2019]
- **Data Access Committee** ("**DAC**"): a body that evaluates data access requests and decides, according to defined criteria, who may be granted access to data and for what purposes.

Purposes of Data Processing within the 1+MG

- **Scientific research application**: research undertaken by public or private sector organisations with the aim to gain knowledge on the aetiology, mechanisms, course, treatment and prevention of diseases.
- **Healthcare application**: prevention, diagnosis, or treatment of a disease or health condition in an individual patient.
- **Policy development application**: policy development includes health technology assessment, regulation of medical products and devices, reimbursement strategies, precision medicine policies, as well as in general the planning, management, administration and improvement of healthcare systems.

Other

- **Consent**: the free and informed expression of the will of an individual, or if the individual lacks capacity to consent, the person's legally authorized representative.
- **GDPR Consent to data processing**: the affirmative agreement of a data subject to the processing of his/her personal data. Consent under the GDPR must generally be freely-given, specific, informed, and unambiguous. Consent is a legal basis for processing personal data under the GDPR, and can also legitimate the processing of special categories of data and/or international transfer. [ref. Art. 4(11) GDPR]
- **Consent to research participation:** the free and informed consent of an individual to participate in research as required by good ethical practice and/or law.
- **Consent to healthcare intervention**: the informed consent of an individual to undergo a healthcare intervention as required by good ethical practice and/or law.
- **Opt-in:** permission (e.g. for secondary use of data) that is only valid when actively provided by the individual. Consent as a legal basis under the GDPR generally requires an opt-in. E.g., "please check this box if you agree to make your data available for further research purposes."
- **Opt-out:** active rejection of a default permission (e.g. for secondary use of data) usually accompanied by a reasonable attempt to notify and provide transparent information to the individual. E.g., "your data will be made available for further research purposes unless you contact us and refuse".

- **Technical and organisational measures (TOMs):** Technical measures relate to technological aspects like devices and networks, and include encryption and pseudonymisation. Organisational measures are policies and processes including risk assessments, training, and audits. Parties processing personal data must adopt appropriate technical and organizational measures to ensure security.

References:

BBMRI-ERIC, ELSI Glossary (2019). https://zenodo.org/record/3754255#.Xpha1cgzaUk.

CASRAI, Research Data Management Glossary https://casrai.org/rdm-glossary/

Digital Health Europe, Glossary https://digitalhealtheurope.eu/glossary/

ELIXIR-Europe, Glossary, Abbreviations, and Acronyms
https://elixir-europe.org/about-us/glossary

ELIXIR-Luxembourg Data Glossary (draft).

European Data Protection Supervisor, Data Protection Glossary.
https://edps.europa.eu/data-protection/data-protection/glossary_en

Global Alliance for Genomics and Health (GA4GH), Data Privacy and Security Policy (2019).
https://www.ga4gh.org/wp-content/uploads/GA4GH-Data-Privacy-and-Security-Policy_FINAL-August-2019_wPolicyVersions.pdf

Global Alliance for Genomics and Health (GA4GH), Data Sharing Lexicon (2016).
https://www.ga4gh.org/wp-content/uploads/GA4GH_Data_Sharing_Lexicon_Mar15.pdf

Open Knowledge Foundation, Open Definition. https://opendefinition.org/

Science Europe, Data Glossary http://sedataglossary.shoutwiki.com/wiki/Main_Page

Swiss Personalized Health Network, Glossary, (30 May 2018, v1)
https://sphn.ch/wp-content/uploads/2019/11/Glossary_20180530_SPHN-1.pdf

UK Information Commissioner's Office (ICO), Key Definitions.
https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/

UK Information Commissioner's Office (ICO), Glossary.

https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-fee/glossary/

UK, Understanding Patient Data

Appendix 1: GDPR Principles, Data Subject Rights, and Obligations

These definitions are not currently considered consensus terms as part of the 1+MG Glossary. They may be beyond the intended scope of the Glossary, and may not at this time be comprehensive coverage of the subject matter. They are nonetheless provided as illustrations.

Lawfulness: to process personal data with a valid lawful reason (i.e., legal basis), such as the individual's consent, and in a manner that is generally lawful. [ref. Art. 5(1)(a) GDPR] (ICO)

Fairness: personal data should only be handled in ways that people would reasonably expect and not used in ways that have unjustified adverse effects on them. Whether and how personal data are processed should be carefully considered before beginning. [ref. Art. 5(1)(a) GDPR] (ICO)

Transparency: being clear, open and honest with people from the start about who you are, and how and why you use their personal data. [ref. Art. 5(1)(a) GDPR] (ICO)

Purpose Limitation: the principle that personal data only be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. [ref. Art. 5(1)(b) GDPR]

Data Minimization: personal data are to be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. [ref. Art. 5(1)(c) GDPR]

Accuracy: ensuring the accuracy of personal data and, where necessary, keeping it up to date. [ref. Art. 5(1)(d) GDPR]

Storage Limitation: personal data should generally only be kept in a form which permits identification of data subjects for as long as is necessary for the purposes of processing. [ref. Art. 5(1)(e) GDPR]

Security: ensuring the 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. [ref. Art. 5(1)(f) GDPR]

Data Integrity: data integrity refers to maintaining and assuring the accuracy and consistency of data. [ref. Science Europe Data Glossary]

Right to withdraw consent: where processing is based on GDPR consent, the individual has the right to withdraw that consent at any time. Processing must cease, and the data must be deleted in some circumstances. [ref. Art 7 GDPR]

Right to object: the data subject has the right to object to further processing of personal data under public interest or legitimate interest grounds, unless the controller demonstrates compelling legitimate grounds for processing that override the rights and interests of the data subject. [ref. Art 21 GDPR]

Right of access: an individual's right under the GDPR to obtain information about the processing of his/her personal data, and a copy of the data. This right can be limited through national legislation, in particular when this right impairs the research or renders it impossible. (Art 16 GDPR; Source: EDPS Glossary)

Right of erasure: an individual's right under the GDPR to have his/her personal data deleted in certain circumstances. This right is limited in certain circumstances. (Source: United Kingdom Information Commissioner's Office (ICO), Key Terms)

Right of information: an individual's right to know that their personal data are being processed, the identity of the controller, the purpose(s) of the processing, the recipients, as well as the existence of data subject's rights. The controller has an obligation to provide this information whether or not data have been obtained from the data subject. (Source: EDPS Glossary)

Data protection officer: Under the GDPR, public organisations as well as organisations processing health and genetic data on a large scale need to appoint a data protection officer who is responsible for informing them of and advising them about their data protection obligations and monitoring their compliance with them. [United Kingdom Information Commissioner's Office (ICO), Data Protection Glossary]

Appendix 2. Version History

August 4 2022
- Added definitions around data (metadata; data collections; records etc.)
February 28 2022
- Added definitions for institutional roles, technical and infrastructures services and other actors in 1+MG
May 26 2021
- Color coded legal definitions. - Added definitions for linkage, data custodian, and technical and organisational measures.

Nov 23 2020
- Title changed from 1+MG Data Protection Glossary to 1+MG Glossary as scope not limited to data protection terms. Preamble also clarifies this widened scope. - Replaced the words “Data subject” and “person” throughout with “Individual”. - Applications section definitions no longer make reference to “data access” (this is implied by 1+MG context). - Glossary now uses the terms “direct identifiers”, “indirect identifiers”, and “non-identifiable data” to describe specific data types or attributes (as one may do during a DPIA). - The language “likely to” has been used instead of “reasonably likely” for categories of data, and instead of “easily” for direct identifiers. - “potentially identifying data” term was removed as it overlaps with indirect identifiers. - We decided to keep the term “de-identified data”, though we had comments that it creates confusion with how it differs from non-identifiable data. - We decided to keep illustrative lists of examples for different categories of data, even though they may not always map perfectly to the categories. - “Sensitive data” term is now “ethically sensitive data” to distinguish from legal term. (alternative: high-impact data). - In the organizational roles section, we kept the term “entity” when we feel is colloquially understood to cover both individuals and organizations, though it may have specific meanings under national laws. - Added further processing to complement secondary use definition (which has been more narrowly defined as “using” data). - Consent we now refer to GDPR Consent to Processing, Consent to Research Participation, and Consent to Healthcare Intervention. - Added proposed definitions for opt-in, opt-out. - Other proposed terms not included in this version: Withdrawal of Consent, Data breaches, Data transfer/sharing/access/use agreements; private/DTC genetics/mobile health data. - Appendix 1: GDPR Principles, Data Subject Rights, and Obligations.