provisional translation

Data Sharing Policy for the Realization of Genomic Medicine

I. Purpose

The Japan Agency for Medical Research and Development (hereinafter referred to as "AMED") regards the promotion of research and development (R&D) through sharing of genomic research data to which accurate clinical and health checkup information has been added as highly important. This is because these efforts will improve the health of the general public and assist the overcoming of diseases.

Japan's 5th Science and Technology Basic Plan (endorsed by the Cabinet on January 22, 2016) aims to accelerate the production of knowledge through new collaborations between researchers in different organizations, as well as between specialized fields and across borders. This will be achieved by widely utilizing the research outcomes of all manner of users including both those in the academic community, the business world, and private citizens - through the promotion of open science - thereby enabling the production of new sources of value. Furthermore, the plan states that it is also important to maintain an awareness of disciplinary differences in the methods of storing and sharing research data amongst research fields, and to take note of matters such as implementation of intellectual property regimes and open-and-closed strategies that are conscious of the national interest.

The "Interim Report" of the Genomic Medicine Realization Promotion Council (July 2015; Headquarters for Healthcare Policy) also states that it is important to share data such as genomic information connected with clinical and health checkup information amongst different R&D projects for the advancement of research.

Data sharing should be promoted from the perspectives of both efficiency, as it avoids duplication of data measurement and acquisition; and effectiveness, as it enables researchers to attain fresh and important insights by analyzing their data with that of other researchers. However, consideration must also be given to the risk of the appearance of researchers – so-called "research parasites"[1]- who obtain data unilaterally from research and use it in their own research without contributing to the planning or conducting of the original research, as this could seriously impede the motivation of researchers who have worked hard to collect and analyze the original data that they then share and release. Furthermore, sufficient measures also need to be put in place in order to protect the rights of research participants who provide specimens and/or clinical information, including the protection of personal information based on related relevant laws and ordinances as well as ethical guidelines.

[1] Dan L Longo, M.D., and Jeffrey M.D., "Data sharing," New England Journal of Medicine, Vol. 374, No. 3, 276-277 (2016).

For these reasons, AMED has drawn up a data sharing policy for R&D projects funded under the "Japan Genomic Medicine Project," "Project for Psychiatric and Neurological Disorders," "Emerging/Re-emerging Infectious Disease Project of Japan," and "Rare/Intractable Disease Project of Japan," as well as research activities funded under the Research Program on Hepatitis and the Research Program on HIV/AIDS. As a general rule, data sharing is obligatory in the R&D projects to which this policy is applied. This policy[2] provides a framework for achieving balance between protecting the rights of study patients and healthy volunteers as well as researchers who provided data and/or information and advancing research in related fields through data sharing. Consequently, the policy aims to disclose and share genomic data associated with research results and genomic information, including clinical information and analytical/interpretation results in a prompt, extensive, and appropriate manner in order to realize healthcare utilizing genomic information.

II. Definition of Terms and Classification of Data Sharing/Disclosure

1. In this policy, "genomic information" refers to the items listed below, among the information obtained through genomic analyses (e.g., GWAS, SNP array, genomic sequencing, transcriptome analysis, metagenome analysis, epigenomic analysis, and gene expression analysis) of the human or microorganisms that affect to human as well as related information.

(1) Sequence information obtained from germline and somatic DNA.
(2) Genetic polymorphism/variation information present in germline and somatic DNA.
(3) Acquired genomic variations (somatic mutations developing in cancer cells, etc.)
(4) Gene expression profiles, genomic modifications.
(5) Genomic information on DNA derived from microorganisms and others that can affect human health.
(6) Genetic polymorphism/mutational information on DNA derived from microorganisms and others that can affect human health.
(7) Phenotype/clinical information specified by AMED, among related information of its kind.
(8) Other information specified by AMED.

2. In this policy, the "Data Management Plan" refers to a document describing the policy regarding storing, sharing, and disclosing genomic information by R&D project, as prescribed by the Principal Investigator, as presented in IV. 2.

[2] A data sharing policy for information other than genomic information, such as clinical information, is to be developed separately.

3. In this policy, the "database" refers to a data storage system designed to record genomic and accompanying information (stored at the original or representative institution, etc., to be reported to AMED by recording in the Data Management Plan), the AMED Genome Group Sharing Database (AGD),[3] the Medical Genomics Japan Variant Database (MGeND), the Rare Disease Data Registry of Japan, and other public databases designated by AMED, such as the NBDC Human Database (e.g., JGA, NHA, and DRA).

4. The scope of sharing and releasing data under this policy is classified into the three categories listed below, and methods for registering, sharing, and disclosing data will be based on the "Data Management Plan".
   (1) Group Sharing Data: data that can be shared among researchers described in the Data Management Plan, who have registered data in accordance with the Data Management Plan, and researchers whose applications for data access have been approved.[4] As a general rule, data sharing is implemented based on an agreement among the researchers involved. Eligible researchers are shown below. AMED may intervene as needed.
      - Researchers and other interested parties who submit or are expected to submit data that will contribute to the enhancement of the existing data of the research group involved.
      - Researchers who submit, or who are able to submit in the future, data that contributes to the expansion or enhancement of the existing data of the research group involved.
      - Researchers who can contribute to data production or quality improvement or generating extra value.
      - Those who are designated by AMED.
   (2) Controlled-Access Data: data that can be used by data users who have been granted approval to access the data with clarification of the intended purposes and methods of using the data. The data will be available after the registration to the database according to the Data Management Plan.
   (3) Open Data: data that can be used by anyone without any access restrictions. The data will be available after the registration to the database according to the Data Management Plan.

[3] The ADG is a public database built and funded by AMED to operate group sharing data efficiently and effectively, with the cooperation of the National Bioscience Database Center, the Japan Science and Technology Agency (NBDC), and the DNA Data Bank of Japan, the National Institute of Genetics (DDBJ).
[4] Data themselves are undisclosed, but the status of data sharing can be partially disclosed (See IV. 3. (1)).

III. The Scope of R&D Projects Applicable to this Policy

1. This policy applies to R&D projects that fall under items (1) and (2) outlined below from the fiscal 2019 and onward. When this policy is revised, the R&D projects that have already been approved and started before a revision is made will remain applicable to the policy as before the revision.

(1) Of the AMED-funded R&D projects, those generating genomic information.

(2) Research projects for which the application of the data sharing policy is specified in the AMED Application Guidelines.

IV. "Submission Method for Data sharing" and "Preparation of a Data Management Plan", Etc.

1. Submission Method for Data sharing

(1) Group Sharing Data[5] should, as a general rule, be submitted to the designated AGD or stored by the original or a representative research institution[6], and then shared no later than either "two years after the generation of raw data" or "the publication of research results (the date of the research paper selected, patent application publication, etc.; the same applies hereinafter)," whichever occurs first. However, this will not apply when group sharing of the data is difficult for reasons such as ethical considerations, the data contain confidential commercial information, and genomic information on microorganisms that may pose risks to society.

(2) Controlled-Access Data[7] should, as a general rule, be submitted to a public database designated by AMED[8] and disclosed no later than either "two years after the generation of raw data" or "the publication of research results," whichever occurs first. However, this will not apply when it is difficult to implement controlled-access for reasons such as ethical considerations, the data contain confidential commercial information, and genomic information on microorganisms that may pose risks to society.

(3) Open Data[9] should, as a general rule, be submitted to a public database designated by AMED[10] and disclosed no later than either "two years after the generation of raw data" or "the publication of research results," whichever comes first. When a database other than public databases designated by AMED is used, the interested parties will consult with AMED.

---

[5] Data, such as BAM/VCF data for individuals, and data, such as those in disease databases for individuals' clinical information are assumed.

[6] When data are stored at the original or representative research institution, it is necessary to obtain informed consent from participants in the research, including patients and healthy volunteers, regarding the possibility of the data being shared by parties other than the researchers concerned, as prescribed in this policy (See V. 2. below).

[7] Data such as BAM and VCF for individuals is assumed.

[8] Refers to public databases, such as the NBDC Human Database (e.g., JGA, NHA, and DRA).

[9] Such as statistical data for a population, individual members of which are unlikely to be identified.

[10] Refers to public databases, such as the NBDC Human Database (e.g., JGA, NHA, and DRA) and MGeND.

2. Preparation of a Data Management Plan

(1) A Data Management Plan (form designated by AMED[11]) should describe the following items. Part of these items may be omitted.

- Fiscal year of the program (required)
- Program title (required)
- Name and affiliation of the Principal Investigator(required)
- R&D project title (required)
- Title of the project applicable to the Data Management Plan and for which the institution is responsible
- Generic name of the data and data group generated from the research (required)
- R&D data overview (required)
- Repository (data storage site) (required)
- Names, affiliations, and researcher numbers of data scientists[12] (required)
- Person(s) who acquired or collected the R&D data
- R&D data controller
- Distinction between self-controlled data and contractor-specified data
- Expected usages/applications of R&D data
- Methods for acquiring or collecting R&D data
- Policy for using/providing R&D data
- Means for smooth provision (when providing data to others)
  Reason for secrecy and duration (when the institution uses data on its own in secret)
- Expected data volume
- Processing policy
- Other

(2) The Data Management Plan needs to be attached to a form designated by AMED[13] describing the items listed below, by Group Sharing Data, Controlled-Access Data, and Open Data.

- Database in which the data is to be submitted (including databases planned for construction)
- Timing of submission
- Type and scale of data

[11] A form to be separately notified by project. See sample forms in Appendix 1.

[12] In this policy, "data scientists" refer to persons who draw out useful findings from data while securing data quality, rather than simply gathering and processing them, such as researchers responsible for implementing the Data Management Plan, as appointed by the Principal Investigator. The Principal Investigator can also hold the post of "data scientists".

[13] A form to be separately notified by project. See sample forms in Appendix 2.

- Scope of releasing or sharing[14]

(3) The Data Management Plan needs to be attached to the Overall Plan and the Annual Research Plan, to be handled as part of these plans.

(4) When pre-evaluating an open application project, whose scope of evaluation includes descriptions in the Data Management Plan, the R&D Proposal must include the Data Management Plan, as stipulated in the Application Guidelines. Note that adoption conditions of the project evaluation committee, and so on may include, as needed, a request for making modifications to the Data Management Plan.

3. Ascertaining and Evaluating the Implementation Status of the Data Management Plan

(1) After commencing the R&D, the Principal Investigator should submit a report on the status of the preparation and implementation of registering, sharing, and disclosing the data provided in the Data Management Plan (including changes, etc.) in response to a request from AMED and using a designated form.[13] In addition to using the reported content to understand the implementation status, AMED may publish part of a summary of the content. AMED may check the implementation status of the Data Management Plan even after the conclusion of the R&D period.

(2) When it has been found that data have been used, or are likely to be used, for inappropriate purposes or provided to a third party, the Principal Investigator must immediately take measures required to ensure the appropriate control of the data. When it has been found that the Group Sharing Data have been used inappropriately by a person who has access to the data, the Principal Investigator must immediately suspend the person's use of the data and report the situation to AMED.

(3) When an open application project is pre-evaluated, the results of the implementation status of previous Data Management Plans and registration of the existing analysis data are allowed to be evaluated.

(4) Interim and post-implementation evaluations are conducted by adopting the implementation status of the Data Management Plan as an endpoint. This endpoint is intended to evaluate the registration, sharing, and disclosure of the generated data, but not the status of data generation.

V. Protection of Personal Information and Ethical Considerations

1. Research must be conducted while considering the protection of the privacy of study

[14] Also anticipates participation by companies and sets the scope for disclosure/sharing.

participants. To obtain informed consent from them, provisions must be made in the informed consent form explaining the likelihood of the institution conducting a variety of genomic research by sharing and disclosing the data.

2. Sharing/disclosing the data must be implemented in compliance with national laws and regulations, ethical guidelines, and so on.[15]

VI. Intellectual Property

AMED encourages the appropriate protection of intellectual properties while aiming to promote R&D and in the medical field and smoothly put the results into practical applications.[16] Thus, researchers can acquire intellectual property rights for data obtained from research conducted with funding provided by AMED. Furthermore, to increase opportunities for the implementation of secondary research and practical applications of the results of such research, care must be taken to ensure that data sharing is not excessively hampered for the sake of intellectual property rights.

VII. Other

1. This policy is subject to change due to revisions of laws or guidelines.
2. Provisions shall be prescribed separately as necessary regarding commercial usage, contracts amongst the relevant parties, and other matters.
3. AMED is entitled to take measures necessary to promote the use of data registered to the AGD.[17]

Reference: Use of Data

(1) With regard to the use of registry data at when presenting research results, data users shall cite the research paper from which the data was obtained and include an Acknowledgements statement, using the example below for reference.

[15] See "Basic Principles for Human Genome Research" (June 14, 2000; Council for Science and Technology Life Ethics Committee), "Ethical Guidelines for Human Genome/Gene Analysis Research" (March 29, 2001; Ministry of Education, Culture, Sports, Science and Technology; Ministry of Health, Labor and Welfare; Ministry of Economy, Trade and Industry), "NBDC Human Data Sharing Guidelines Ver. 4.0" (August 31, 2018; NBDC), "NBDC Human Data Group Sharing Guidelines ver. 2.0" (August 31, 2018; NBDC), etc.

[16] See AMED Intellectual Property Policy (April 1, 2015; Regulation No. 27 of 2015).

[17] For example, a case in which genome sequence data contained in the AGD can be processed to be searchable, upon obtaining approval of the data provider.

[Example of Acknowledgements]

"The data (some of the data) used in this research was obtained by [Representative Researcher BBBBB] of the Japan Agency for Medical Research and Development [000] Program [AAAA] Research Project and provided via the Japan Science and Technology Agency (JST) National Bioscience Database Center (NBDC) website (http://bioscience dbc.jp/).

(2) Data users may obtain intellectual property rights based on research results obtained through secondary use of open data and controlled-access data.

Revision History

April 2016: Established

October 2017: Revised

• The scope expanded to include "Rare/Intractable Disease Project of Japan."

November 2018: Revised

• The scope expanded to include "Project for Psychiatric and Neurological Disorders," "Emerging/Re-emerging Infectious Disease Project of Japan," Research Program on Hepatitis and the Research Program on HIV/AIDS.

• Addition of sample forms to ascertain the preparation and implementation status of the Data Management Plan.