Pistoia Alliance

Collaborate to Innovate

QDatE Best Practice Guidelines for the Storage of Collected Sensor-Generated Data From Remote Monitoring Technologies

These Best Practice Guidelines are created to provide a strategy that derives as much value as possible to the study collecting the data, the participants who supply the data and any external users who are reusing the data outside of its original intended use.

This first iteration of the Quality Data Generation and Ethical Use (QDatE) Best Practice Guidelines is to be used in conjunction with the QDatE Code of Ethics, which are published separately, and will concentrate on preparing sensor-generated data from remote monitoring technologies (SDRM) that have been collected using the FAIR Guiding principles[1] for storage.

There are three additional documents or records that are required to ensure that the data is ready for data storage, and these are discussed in more detail within these Best Practice Guidelines:

1. **Project Data Management Plan or Organizational Data Strategy**
2. **Metadata that is collected and stored alongside the SDRM**
3. **Data Engineering Plan**

Collection of SDRM as per the FAIR Guiding Principles will ensure that the SDRM is Findable, Accessible, Interoperable and Reusable and will save cost, time, and effort in retrospectively making the SDRM data FAIR (something that is not always possible to achieve retrospectively).

In order to ensure the FAIR-ness of the SDRM, this Best Practice Guidelines recommends the use of the FAIR Maturity Indicators Test.[2,3]

INCLUSION AND USE

SDRM includes and is not limited to any digital health related data that is collected via remote monitoring technologies such as:

- **smart devices, e.g., watches or phones**
- **remote medical technological devices, e.g., continuous glucose monitoring devices or accelerometers**
- **remote home monitoring technologies used to collect information, e.g., electric usage monitors as a proxy for insomnia, or pressure flooring to measure gait, walking speed, and falls**

Quality Data Generation and Ethical Use

QDatE

The QDatE Best Practice Guidelines are available for use within all organizations which sponsor, organize, manage, or conduct clinical trials collecting SDRM or ones which design, develop, or market devices that are used to collect SDRM. This includes and is not limited to:

- ❑ **Pharmaceutical companies**
- ❑ **Principal Investigators and other Clinicians**
- ❑ **Clinical Studies Groups**
- ❑ **Clinical Research Organizations**
- ❑ **Academic Researchers**
- ❑ **Healthcare Professionals**
- ❑ **Consumer Sector**

AIM

The aim of the QDatE Best Practice Guidelines is to enable the easy use and reuse of SDRM. This not only provides a cost- and time-saving benefit to the teams running the study but also lessens the burden on the participants who take part in the studies and provides benefits to any additional beneficiaries granted access to the data for analysis outside of the study for which the SDRM was originally collected.

1) Data Management Plan or Data Strategy

A formal project data management plan [4] or organizational level data strategy [5] allows for the automation of the data pipeline management, data cleaning and analytics activities. This will have a number of benefits, including savings in cost and time, as well as making it easier to access the data. This will also enhance the ability of investigators to employ tools developed in the future, such as the addition of Artificial Intelligence.

We encourage that any current fit-for-purpose local project data management plan or organizational level data strategy that is already in place continue to be used, but when new plans are being established, the areas of consideration for a data management plan or data strategy should include and are not limited to:

i) Data type

Data that has been minimally processed for its intended use to enable the easy use and reuse of the data.

ii) For whom are you storing data?

Data must be stored in a way that ensures the quality of the SDRM for its original intended use and for its reuse outside of the original study for which it was collected.

iii) Where to store the data?

The use of any local internal data storage should be utilised in the first instance, but if this does not exist or is not fit-for-purpose there are a number of options in order to store SDRM that range from Enterprise servers, to Enterprise Data Warehouses and Data Lakes, through to Cloud-based systems.

The data should be stored in whichever storage option provides: 1.) the greatest data security and integrity; 2.) adaptability to store large amounts of data from multiple different sources; and 3.) flexibility in enabling the analysis of the data and the scalability for future data collections.

iv) Storage technique

The use of any local internal data storage technique should be utilised in the first instance, but if this is not fit-for-purpose, the storage technique should ensure that the SDRM is stored in a secure, robust, easily interrogated format.

Data Collection

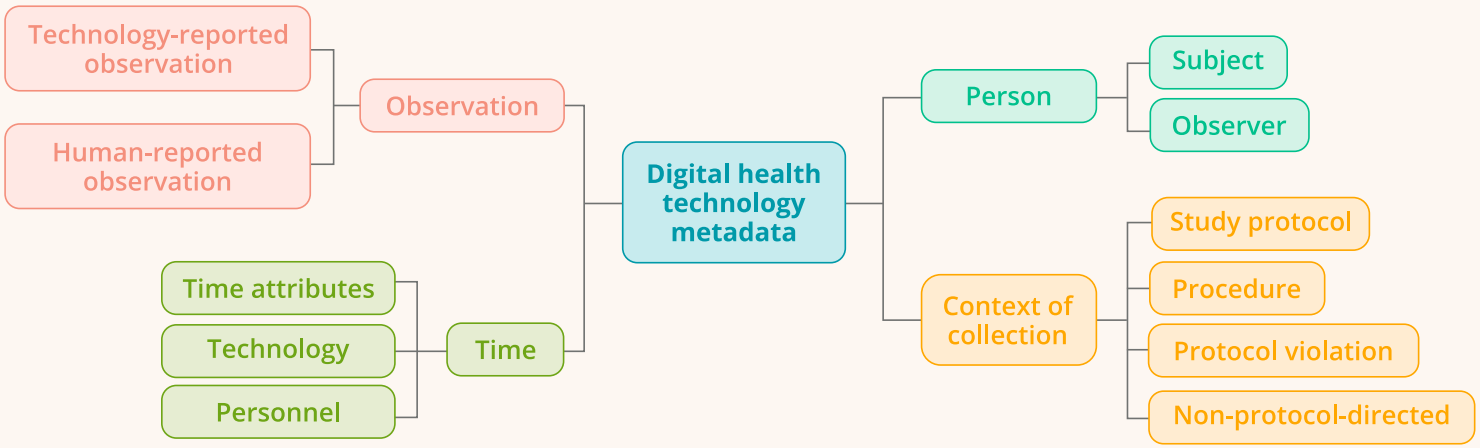
The collection of SDRM should be guided by the FAIR Principles [6] and done so in accordance with the appropriate privacy and security considerations, including those at regional, national, and local levels, and with the required regulatory practices, to ensure the anonymization of the data and preserving user privacy.

2) Metadata to be stored with the SDRM

To ensure the practical and feasible reuse of the SDRM outside of the original study that it has been collected under, rigorous and complete metadata must be collected and stored alongside the SDRM.

A flavour of the minimum metadata required for advancing the use of digital health technologies for clinical research proposed by Badawy et al, 2019 [7], shows the overarching thematic areas of metadata that need to accompany the SDRM tin order to enable easy reuse (see Figure 1).

Figure 1 - Thematic areas of metadata to accompany SDRM



Decrease the risk for inconsistent or incomplete data storage

The ease of use, wearability, and synchronicity of the remote monitoring device for the participant is key. If this is too difficult or burdensome for the participant, there is an increased risk of inconsistent or incomplete data due to a key task that collects or transfers the data not being completed or, in the worst case, the participant withdrawing from the study.

Secure transfer of data

Following the FAIR Guiding Principles enables the SDRM-collected and SDRM-associated metadata to be machine readable. As a result, this Best Practice Guideline recommends an automated data transfer or upload to be done at regular intervals. This helps the checking of data integrity and allows the study to pick up on and resolve any issues that may be arising during the data collection.

Measures are to be in place, such as those within the Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data [8], to protect the transferring data from unauthorised access and undetectable changes that occur during data transfer by ensuring that there are checks in place to flag if the data does not fully transfer to storage or has been compromised during the transfer process.

Data integrity

Throughout the processes of data collection and data transfer, it is vital to ensure that there are well documented checks in place to test for the completeness and integrity of the data. SDRM is often too big for clinical study databases, and this may result in many different derivations - these will need to be fully documented and validated throughout the life course of the study.

3) Data Engineering Plans

Data Engineering Plans are needed to ensure the usability and reusability of the SDRM by documenting the complete data process from start to finish and the steps that are needed to utilize Real World Data (RWD) such as SDRM.

REFERENCES:

1) Wilkinson, M. D. et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship. Data 3, 160018. doi.org/10.1038/sdata.2016.18

2) Wilkinson MD et al 2019 Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci. Data 6, 174. doi.org/10.1038/s41597-019-0184-5

3) https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/metrics

4) https://dmptool.org/

5) Evolving Clinical Data Strategies and Tactics in Response to Digital Transformation, Beth Harper, Michael Wilkinson, Raj Indupuri, Sheila Rocchio, Ken Getz, Therapeutic Innovation & Regulatory Science. doi.org/10.1007/s43441-020-00213-4

6) https://fairtoolkit.pistoiaalliance.org/fair-guiding-principles/

7) Metadata Concepts for Advancing the Use of Digital Health Technologies in Clinical Research, Reham Badawya Farhan Hameed, Lauren Bataille, Max A. Little, Kasper Claes, Suchi Saria, Jesse M. Cedarbaum, Diane Stephenson, Jon Neville, Walter Maetzler, Alberto J. Espay, Bastiaan R. Bloem, Tanya Simuni, Daniel R. Karlin, Digit Biomark 2019;3:116–132, doi: 10.1159/000502951

8) https://edpb.europa.eu/sites/edpb/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf

Pistoia Alliance