enhAnced Mobile BiomEtRics

DELIVERABLE: D1.2

Data Management Plan

Contract number:	675087
Project acronym:	AMBER
Project title:	Enhanced Mobile Biometrics
Project duration:	1 January 2017 – 31 December 2020
Coordinator:	Richard Guest, University of Kent, Canterbury, UK

Deliverable Number:	D1.2
Type:	ORDP: Open Research Data Pilot
Dissemination level	Public
Expected submission date	30-06-2017
Date submitted:	14-06-2017

Authors / contributors	K. Jumel, R.Guest, F.Deravi
Contributing partners	UniKent

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant

AMBER

DATA MANAGEMENT PLAN

ADMIN DETAILS

Project Name: AMBER
Grant Title: 675087
Principal Investigator / Researcher: Richard Guest
Description: enhAnced Mobile BiomEtRics Design, implementation and assessment of biometrics on mobile devices
Funder: European Commission (Horizon 2020)
Institution: University of Kent

1. DATA SUMMARY

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

Purpose: Mobile-device biometric assessment using samples collected from human participants. These data will be used to assess novel methods and algorithms for classification and authentication purposes.

Data may be varied. Biometric modalities will include (but not limited to): fingerprint, face, voice, iris, interaction data (such as swipe or movement), device metadata (GPS location, device tilt and movement, cell/wi-fi location etc.). The format of these data may be images, numeric data concerning location or process, or features extracted from these raw data. Where available, publically available datasets will be used. These datasets are typically used widely in the biometric academic research community subject to terms and conditions of use (such as restriction of redistribution of data, publication of images, deletion after completion of study, secure storage of data).

The origin of these data is varied but typically will be made available by other university-based research groups, collected and distributed with Ethics approval. If novel data collection is required then the origin of data will be locally-sourced participants. Collection of data will be governed by local Ethical approval processes. Expected size of data is unknown but may be several Gbs. The data will be useful to the local AMBER consortium and the wider biometric research community (including industrial R & D organisations).

2. FAIR DATA

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Data will be stored at the coordinator's (University of Kent) repository, KAR, and will be kept for 5 years after the end of the project. Where requested, data will be kept for 2 more years. KAR is managed and supported by a team of experts and is free of charge
A naming convention will include a short description of contents, the host institution collecting the data and the month of publication.
Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename
No standards currently exist for biometric data within the Research Data Alliance framework, however we shall use, where possible, ISO/IEC standard formats for data storage which do include provision for metadata.
The real names of participants will NOT be distributed.
Meta data (including demographics) about the subjects will be distributed listed according to a pseudo-random tag assigned to an individual.

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Where possible data will be made available subject to Ethics and participant agreement. However, the personally-identifiable nature of the data collected within AMBER means that in most instances it would be difficult to release collected data. Where data is made available we will do so using the Kent Academic Repository (KAR).
Prior to release, a requesting party will need to contact the Project Coordinator describing their intended use of a dataset. The Project Coordinator will send a terms and conditions document for them to sign and return. Upon return, the dataset will be released. Documentation (and, if available for distribution, software) will be included with the release of the data.

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

As stated, we will adhere to ISO/IEC data interchange formats (19794-X) for the storage of sample and meta data. This will ensure proven interoperability within the biometrics community.

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Due to the sensitive nature of the data they will only be available on application and their use will be restricted to the research use of the licensee and colleagues on a need-to-know basis. This non-commercial licence is renewable after 2 years, data may not be copied or distributed and must be referenced if used in publications. These arrangements will be formalised in a User Access Management licence which describes in detail the permitted use of the data.

3. ALLOCATION OF RESOURCES

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Data will be stored at the coordinator's (University of Kent) repository, KAR, and will be kept for 5 years after the end of the project. Where requested, data will be kept for 2 more years. KAR is managed and supported by a team of experts and is free of charge.

4. DATA SECURITY

Address data recovery as well as secure storage and transfer of sensitive data

Data will stored in Kent Academic Repostiory (KAR) which is managed and supported by a team of experts at the University of Kent and subject to the university's data security measures and backup policies.
Transfer of data is via a Zip process of distribution.
Encryption of sensitive data using shared-key methods.
Password distributed separately.

5. ETHICAL ASPECTS

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

All our work is subject to ethical approval (locally, via an Independent Ethics Advisor and the EC REA). Prior to data collection participants will agree to the terms and conditions outlined in a Participant Information and Consent Form.

6. OTHER

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

None