D8.2
1st Technical diagram of EQIPD-DWH

EQIPD – GA 777364
IMI2-2016-09-03

European Quality In Preclinical Data

WP8 – Data Management

Lead contributor	Maarten Loos (14 – Sylics) maarten.loos@sylics.com
Other contributors	Tom Van de Casteele (19 – Janssen) tvdcaste@its.jnj.com
	Emily Sena (1 - UEDIN) Emily.Sena@ed.ac.uk
	Jing Liao (1 - UEDIN) jing.liao@ed.ac.uk

Due date	31 Dec 2017
Delivery date	20 December 2017
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	Final	07/08/2017

Document History

Version	Date	Description
V1.0	12 Dec 2017	First Draft
V2.0	16 Dec 2017	Second draft, after review by TvdC, JL and ES
V3.0	19 Dec 2017	Final version

Aim

To generate first diagram of the EQIPD data warehouse (EQIPD-DWH) solution to house the different database structures required for WPs 2 and 4.

Work towards EQIPD-DWH design

Inventory of existing database solutions among partners

Existing database solutions that are available among partners were reviewed and considered for use after modification to allow the various data structures to be stored and retrieved for analyses. These included the CAMARADES database structure (UEDIN) for summary outcome data and according metadata, the MULTI-Part database structure (UEDIN) for individual animal data, AHCODA-DB database structure (Sylics) and the EMIF-platform (used by EFPIA partners). Furthermore the CDISC/SEND-format (used by EFPIA partners; Standard for Exchange of Nonclinical Data) for individual outcome data and according metadata was reviewed as intermediate file format.

Inventory of data structures to be housed in the EQIPD-DWH

Research questions formulated in WP2 were translated into a data inclusion protocol (D8.1). This protocol specifies the data to be included for the analysis of historical data of partners and publically available data. Within the EQIPD consortium, three paradigms were prioritized for which data will be analyzed in WP2, and for which new data will be generated in WP4: Irwin testing, open field activity, and 48 hour EEG monitoring.
Milestone MS36 provided an inventory of the structure of the historical data available in the databases of the EQIPD partners, with respect to these 3 paradigms. In summary, the structure and file format of these historical data is variable among partners. In is expected that newly generated data in WP4 will be similar in structure as the historical data currently available.

1st technical diagram of EQIPD-DWH

The database structures available at partner UEDIN appeared highly suitable as repository for the analyses planned in WP2. Therefore, the EQIPD-DWH solutions will be built around these existing resources. The EQIPD-DWH solution will serve 3 different flows of data:

- Archiving and making accessible of historical data of partners (WP2) and newly generated data on the three selected paradigms (WP4)
- Archiving and making accessible of publically available data, obtained from literature review
- Making EQIPD data accessible beyond the project duration

An overview of the data flow centered around the database structures available at partner UEDIN is depicted in **Figure 1 (in red)**. A more detailed diagram of the database structure available at partner UEDIN for storing of historical individual animal data (WP2) is depicted in **Annex 1**. The Multipart web application may be useful for partners collaborating in the across-site studies in WP4. If WP is going to use this Multipart web application, the database diagram will be similar to the one in **Annex 2**.

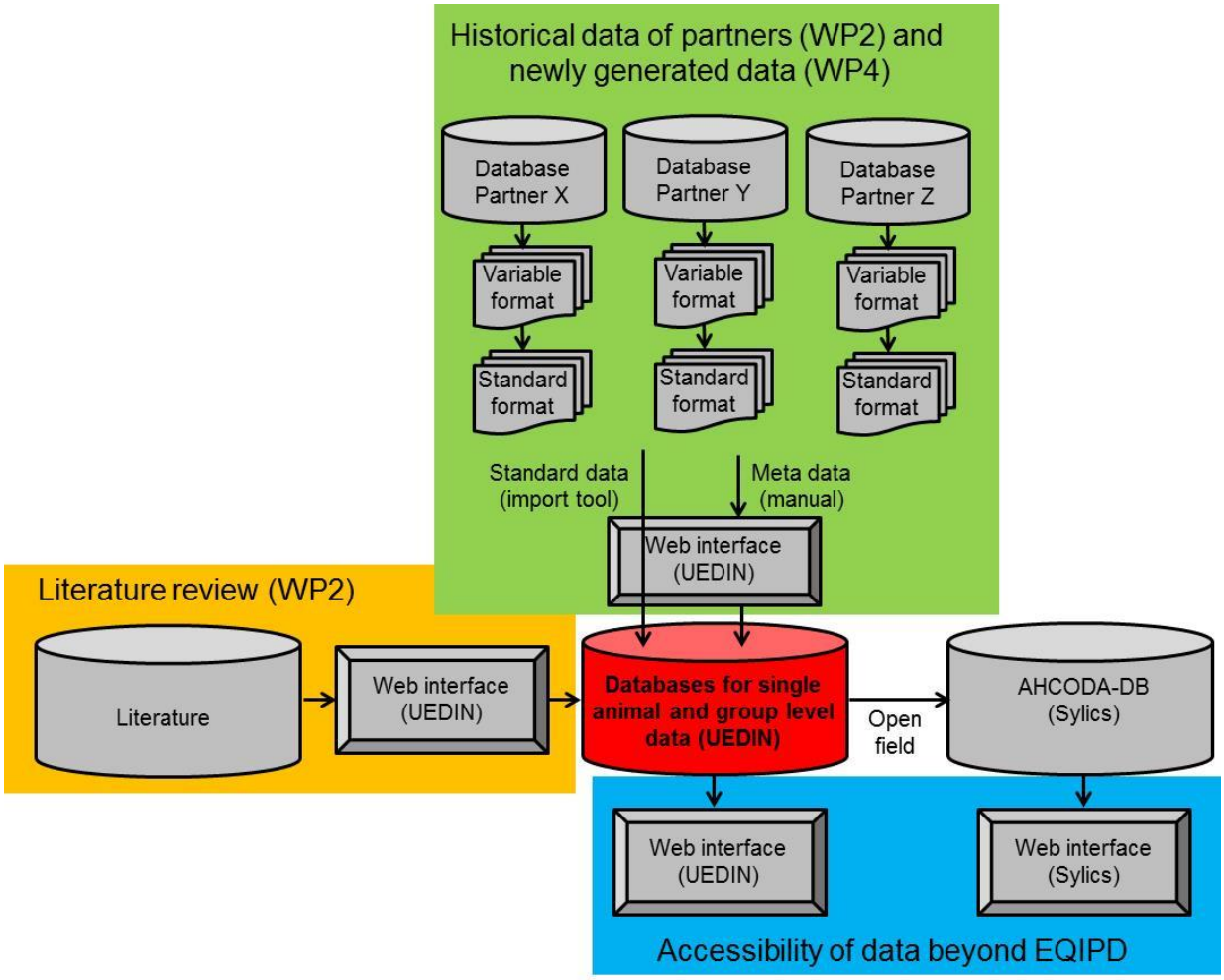
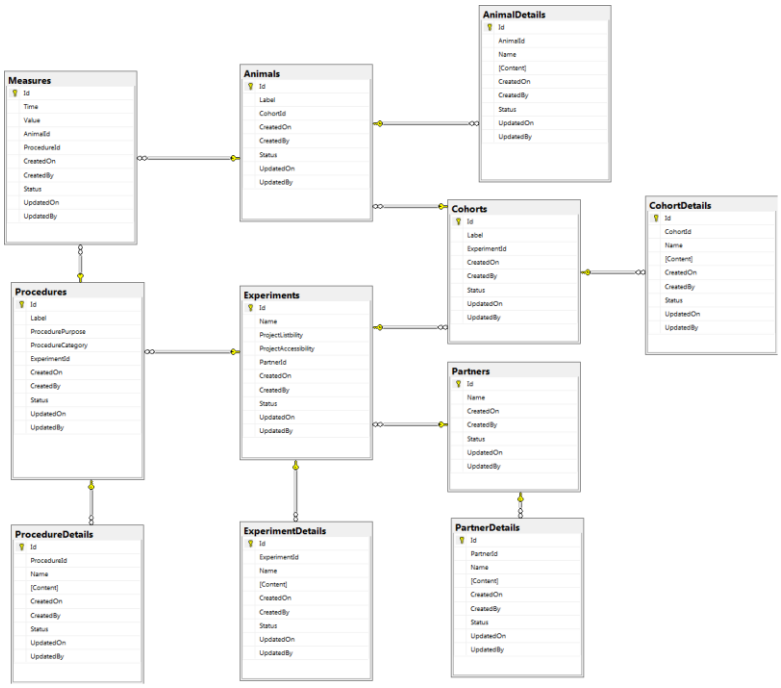


Figure 1 | Global diagram of the EQIPD-DWH. The first flow of data relates to historical and newly generated data of individual animals from partners (green). Among partners, there are different databases solutions in use to archive data of the three selected paradigms. The output of these databases systems of the partners is rich and comprehensive, but the layout/structure is variable among partners. This variable format will be preprocessed to a standardized file format, which will contain data at the level of individual animals. It will be a flat file format, with cases (individual animals) in rows and variables in columns. These variables do not only describe the measures obtained in the three paradigms, but also meta data pertaining to the individual animals (e.g. gender, age, weight). To ensure proper archiving in the database structures available at partner UEDIN (**red**), an import tool will be used import the standardized files. In addition, meta data that relates to the testing protocol will be entered through a web interface. The second flow of data relates to the literature review process (**orange**), which will largely follow previously used procedures by partner UEDIN at Syrf.org.uk. The third flow of data (**blue**) outlines currently available systems to ensure the availability of the EQIPD data beyond the project duration. Additional solutions that will allow storage and sharing of raw data beyond the project duration will be investigated during the next months.

Annex 1 Detailed diagram of database structures at partner UEDIN for historical invididual animal data (WP2)

The diagram below is the first draft for the database for storing the historical data. It shares the core tables from the diagram in Annex 2 (Multipart web application), but because there are less requirements in the front-end, it is simplified.



Annex 2 Detailed diagram of database structures at partner UEDIN for individual animal data in WP4

The Multipart web application may be useful for partners collaborating in the across-site studies in WP4. If WP is going to use this Multipart web application, the database diagram will be similar to the one below.