PROJECT DELIVERABLE REPORT

DELIVERABLE NUMBER AND TITLE	D2.2.1
TITLE	MINIMAL REQUIREMENTS FOR STUDY DATA
AUTHOR(S)	DR. M. PINART, DR. K. NIMPTSCH, DR. L.O. DRAGSTED, DR. T. PISCHON
WORK PACKAGE	WP 2
TASK	TASK 2.2
WP LEADER	DR. L.O. DRAGSTED
BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE	UCPH, UGENT ,CRA-NUT, ALMA, ISS, UNIROMA, CNR-ISA, IRCCS, UNINA, FEM, TNO, MDC, DIFE, NIHD, ULG, UOA, KU LEUVEN, ILVO, VITO, VUB,TUM
STATUS – VERSION	FINAL - VERSION 1.0
DELIVERY DATE (MONTH)	M18
SUBMISSION DATE	M17
DISSEMINATION LEVEL – SECURITY*	PU
DELIVERABLE TYPE**	R

* Security: PU – *Public*; PP – *Restricted to other programme participants (including JPI Services)*: RE – *Restricted to a group specified by the consortium (including JPI Services)*; CO – *Confidential, only for members of the consortium (including JPI Services)*

** Type: R – *Report*; P – *Prototype*; D – *Demonstrator*; - O - *Other*

CONTENTS

TASK 2.2: MINIMAL REQUIREMENTS FOR STUDY DATA

Minimal requirements include the necessary metadata (i.e. details about what, where, when, why, and how the data were collected, processed, and interpreted) so that the intervention and observational studies identified in task 2.1 (*see deliverable 2.1.1 in task 2.1*) can be uploaded to a central server such as the DASH-IN server at TNO in the Netherlands or at other local sites (so far also UCop). To this end, the studies will fill in a common template (*see **Annex 1***) that will cover all the items related to study design, subject details and measurement data, which are also requested for the specifically tailored quality appraisal tool (*see Deliverable 2.3.1 in task 2.3*) to validate the quality of the participating nutritional studies.

The objective of the task is to identify the minimal requirements for intervention and observational studies so that their data can be shared in future research proposals within the ENPADASI consortium. Because the utility and longevity of data relate directly to how complete and comprehensive the metadata are, we describe in this report the minimal information required to:

- Ensure that the submitted data are sufficient for clear interpretation.
- Improve the ability to compare studies produced by different scientists and organizations.
- Aid the development of reusable data quality metrics.

We have followed the hierarchical structure developed by the ISA Commons (http://www.isacommons.org/), which is a growing community that uses the ISA metadata categories (Investigation, Study and Assay) tracking framework to facilitate standards-compliant collection, curation, management and reuse of datasets in an increasingly diverse set of life science domains. The ISA framework is also part of the COSMOS and Metagenomics Data Infrastructure projects.

- *Investigations* generally describe the overarching aims of the research carried out by a particular project.
- A *Study* is a series of experiments (or assays) which can be combined to answer a particular biological question.
- An *Assay* describes a particular experiment. It allows you to associate data files, SOPs and models together as well as describing the type of assay and any technology required to perform the experiment.

The ISA definition of study to mean a series of experiments and assay to mean both experiment and sample analysis is somewhat unusual and may lead to mistakes. In the following we use the term, assay to mean only a specific sample analytical procedure, questionnaire or test (resulting in data collection), and experiment to mean a specific protocol involving study objects, sampling and assays. A study may comprise several experiments but usually consists of only one so that the two terms become synonymous. With these additional precisions we are using ISA nomenclature. The minimal requirements developed in ENPADASI include:

- Descriptions of how data and files are named, physically structured and stored.
- Details about sampling procedures, analytical methods, and research context.

We have further stratified the requirements as mandatory (M) and optional (O). We will define which of the optionally requested data will be shared prior to or upon acceptance for publication at a later stage.

INVESTIGATION (THE PROJECT CONTEXT)

1. Name of the study and/or experiment in English/Acronym (M)
2. Study or experimental protocol and any protocol deviation/amendments (M)
3. Description of the study aim within the investigation (M)
4. Principal Investigator (name) for the study or experiment described (M)
5. Contact information of the contact person of the study/experiment (M)
6. Consortium involved in the investigation (M)
7. Funding body/bodies for the investigation (M)
8. Study weblink for the investigation or study (URL) (O)
9. Registration number of the study or experiment: i.e. Clinicaltrials.gov (O)
10. Institutional Review Board/Independent Ethics Committee (IRB/IEC) approval number (M)
11. Publications: Type and DOI or File location (M)

STUDY OR EXPERIMENT (A UNIT OF RESEARCH)

1. Study design (M):
   1.1. Observational studies: Cohort, cross-sectional and case-control
   1.2. Intervention studies: randomized clinical trials, cross-over trials
2. Control group (M)
3. Type of controls (M)
4. Randomization (M for intervention/experimental studies)
5. Full randomization method (if any) (M for intervention/experimental studies)
6. Partial randomization method (M for intervention/experimental studies)
7. Other allocation method (not random) (M for intervention/experimental studies)
8. Blinding (yes/no) (M for intervention/experimental studies)
9. Blinding method (M for intervention/experimental studies)
10. Study/experiment terminated (Yes/No/ On-going) (M)
11. Treatments (Number and types of foods, drugs or other treatments compared) (M for intervention/experimental studies)
12. Compliance assessment (for each treatment) (M for intervention/experimental studies)
13. Number of factors (for factorial designs only) (M)
14. Total number of arms (Give number of distinct treatments in the study/experiment for intervention/experimental studies)

15. Explanatory text (Give a textual account of the overall design leading to the total number of groups) or provide the study protocol. (M)
16. Start of recruitment (Start year or date) (M)
17. End of recruitment (End year or date) (M)
18. Follow-up (M)
19. Describe follow-up (i.e. time points and actions taken) (M)
20. Total number of participants recruited in the study (M:xx, F: xx) (M)
    - 20.1. Number of volunteers screened N (M:xx, F: xx)
    - 20.2. Number of volunteers enrolled N (M:xx, F: xx)
    - 20.3. Number of volunteers terminating study N (M:xx, F: xx)
21. Sex of the study participants (Male/Female/Both/Unknown). (M)
22. Age range of the study participants. (M)
    According to the EFSA recommendations for food consumption studies, age can be further stratified in the following categories (O):
    - Infants: 3–11 months
    - Toddlers: 1–2 years
    - Other children: 3–9 years
    - Adolescents: 10–17 years
    - Adults: 18–64 years
    - Elderly: 65–74 years
23. Total number of sample donors (number of individuals with biological samples) stratified by sample type (i.e. whole blood, serum, plasma, urine, etc) (O)
24. Collection of biological samples: (O)
    - 24.1. Sampling identifier
    - 24.2. Organism part (Type of sample): plasma, serum, whole blood, urine, stools etc
    - 24.3. Primary container (PaxGene tubes, 1.5 ml eppendorfs…)
    - 24.4. Storage temperature
    - 24.5. Fasting (if the sample is a fasting sample (crucial when measuring i.e. glucose)) (M)
    - 24.6. Relative time points of sampling event in the study/experiment
    - 24.7. Study or experiment sampling method
    - 24.8. Study or experiment sampling method description (or SOP)

ASSAY (ANALYTICAL MEASUREMENT)

1. Source of data (Questionnaire-based, clinical examination, medical records, sample analyses, tests, etc.) (M)
   The studies should provide information on data sources such as questionnaires, hospital discharge files, abstracts of primary clinical records, electronic medical records, biological measurements, etc. If data is questionnaire-based, it should be stated whether the studies used their own questionnaires or if they used validated or adapted (from validated Q)

questionnaires). If validated questionnaires are used the validation procedure used should be explained.

2. Data on exposures/diets etc. used or recorded in the study/experiment and time points of assessment:
   2.1. Dietary intake (Exposure data)(M)

   Specify method for dietary or nutritional assessment (multiple options are possible):

   2.1.1. Dietary records (innovative alternatives: PDA-technologies, Mobile-phone–based technologies, Camera- and tape-recorder–based technologies)
      2.1.1.1. Estimation of size of the portion: Weighing, Picture book (validated/not validated), Household measures, Known packaging size, Ruler
      2.1.1.2. Number of recorded days
      2.1.1.3. % of record or recall days according to the day of the week (week days, weekend days, unclassified).
      2.1.1.4. % of record or recall days according to the season (Spring, Fall, Winter, Summer, Unclassified)
   2.1.2. 24-Hour Dietary Recall (innovative alternatives: Interactive computer-based technologies, web-based)
   2.1.3. Food Frequency Questionnaire (innovative alternatives: Interactive computer-based technologies, web-based)
   2.1.4. Brief Dietary Assessment Instruments (also known as screeners)
   2.1.5. Diet History consists of Dietary records + 24-hour recall + FFQ (Willett 1998)
   2.1.6. Provided diet, foods or drinks
   2.1.7. Administration method (interview/self-administered/supervised meals/unsupervised meals)

   Indicate what was assessed

   2.1.8. Food Y/N
   2.1.9. Drinks Y/N
   2.1.10. Supplements Y/N

   Nutrient and food intake data submitted

   2.1.11. unadjusted
   2.1.12. adjusted for total energy intake using
   2.1.13. density method
   2.1.14. residual method
3. Physical activity (name of the tool/questionnaire, provider, version, year of the version) (O)
4. Tobacco use (name of the measure, measured/self-reported) (O)
5. Alcohol consumption (name of the measure, measured/self-reported) (O)
6. Anthropometry (name of the measure, measured/self-reported) (O):
   6.1. Weight
   6.2. Height

6.3. Waist circumference
6.4. BMI status (categories)
6.5. %Body fat

7. Socio-demographic information should include according to the EFSA recommendations (O):
   7.1. Marital status
   7.2. Region
   7.3. Rural/urban area
   7.4. Size of the household
   7.5. Household income (country-specific ranges may be applied)
   7.6. Education level (in the case of children, the education level of the parents)
   7.7. Recent employment status
   7.8. Native language (if language minorities in the country)
   7.9. Pregnant or lactating women.
8. Study outcomes and time points of assessment: health outcomes* (O)
   Health outcome OR clinical symptoms of a disease OR health outcome-related biomarkers* (i.e. HbA1c testing for diabetes)
   *This is not a health outcome but an indicator of presence of some disease state. If validated it can be used for disease diagnosis.
   8.1. Declaration of availability of open endpoints in the case of cohort studies with continued central follow-up of all medical incidents, where the follow-up time and end-points may be open.
9. Clear operational definitions of health outcomes*, exposures, and other measured risk factors (M only for outcomes and exposures defined as primary outcomes. Otherwise it should be optional. Dietary intake, alcohol consumption, smoking habits, physical activity, anthropometric and socio-demographic data are M).
10. Experimental metadata: biomarkers, metabolomics, proteomics, genomics, transcriptomics (O)
    10.1. Sampling identifier
    10.2. Assay file name
    10.3. Assay platform
    10.4. Type of assay (physiological, chemical-biological, omics)
    10.5. Measurement (i.e. metabolite profiling)
    10.6. Technology (i.e. mass spectrometry, chromatography)
    10.7. Protocol used (including QC)
    10.8. Alternatively provide reference to the protocol
    10.9. Raw data (Raw spectral data file)
    10.10. Processed data (Data transformation protocol)
    10.11. URL to raw data, if applicable
    10.12. URL to processed data, if applicable

ANNEX 1

See Excel file named “Any study template with QAT 30-06-2016”