CoDiet

COMBATTING DIET RELATED NON-COMMUNICABLE DISEASE THROUGH ENHANCED SURVEILLANCE

D3.3 Data Management Plan Update RP1

Deliverable number D3.3

Work Package WP3	Structure of the data
Task 3.1	Data management for data collected in the project
Task Leader	NKUA
Prepared by	Dimitrios Gunopulos (NKUA), Dimitrios Tomaras (NKUA)
Contributors	Dimitrios Tomaras (University of Athens), Sara Arranz Martinez (AZTI), Natalia Zaldua (Microcaya), Orla O' Sullivan (TEAGASC), George Mylonas (Imperial College London), Nieves Embade (CICbioGUNE), Dolores Corella (UVEG)
Version	1.0
Delivery Date	30/06/2024

Funded by the European Union

UK Research and Innovation

Foreword

The work described in this report was developed under the project **CoDiet - Combatting Diet related non-communicable disease through enhanced surveillance** (Grant Agreement number: 101084642; Call: HORIZON-CL6-2022-FARM2FORK-01; Topic: HORIZON-CL6-2022-FARM2FORK-01-10). Any additional information, if needed, should be required to:

Project Coordinator:
Itziar Tueros – itueros@azti.es | AZTI |

WP3 Leader:
Dimitrios Gunopulos – dg@di.uoa.gr | NKUA |

Task Leader:
Dimitrios Gunopulos – dg@di.uoa.gr | NKUA |

Dissemination Level		
PU	Public, fully open	X
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444	
Classified S-UE/EU-S	EU SECRET under the Commission Decision No2015/444	

Funded by the European Union

CoDiet is part of the Horizon Europe programme supported by the European Union.

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Executive summary

There is an enormous amount of scientific research that has already been done about the connections between diet and non-communicable diseases. Understanding what information is already out there is vital to develop new insights. However, reviewing this amount of information manually can be time-consuming and expensive. Artificial intelligence powers a growing number of systems today. We are using a kind of artificial intelligence (AI) called natural language processing (NLP), the same kind used in Chatbots, to help us search for and analyse this information, meaning we can achieve this in a fraction of the time. We aim to create the most comprehensive overview of the relationships between diet, bodily processes, and non-communicable diseases ever produced. In this context, CoDiet is an international research project that aims to combat diet-related diseases through innovative monitoring technologies and personalised nutrition. The project will also develop a tool that will simulate changes in NCDs in response to diet at the population level, with the goal of promoting the uptake of NCD-protective diets. Currently, efforts to tackle these issues on a population scale take a 'one size fits all' approach, and the hope is that personalised dietary advice can lead to more effective results.

The deliverable describes an updated version of the Data Management policy that will be followed in the course of the project, encompassing both data collected in WP2 and data collected previously. In more detail, the report lays out the CoDiet Data management to elaborate; presents an overview of the datasets used by the project; specifies FAIRness to achieve for the methods, tools and data produced by CoDiet; lists the data management activities to take care; and introduces all the necessary principles to comply with General Data Protection Regulation (GDPR), EU AI Act (proposed European law on AI), EU Data Act as well as Laws protecting the privacy of individuals in each member country. The Data Management Plan will be a living document that will be regularly updated by the project partners in order to align with the needs and requirements of the project. As such updated versions of this document will be provided during M18, M36 and M48 (in time with the periodic evaluation/assessment of the project) updating all the necessary sections in light of the needs and findings of the project.

Please check the website of the project (https://www.codiet.eu/) or CORDIS (https://cordis.europa.eu/project/id/101084642) under the deliverables section for additional deliverables and updates.

research activities taking place at Imperial College London and the University of Leicester are supported by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 101084642].

The Beneficiaries of this project are:

No.	Name	Short Name	Country
1	FUNDACION AZTI - AZTI FUNDAZIOA	AZTI	Spain
2	CESKE VYSOKE UCENI TECHNICKE V PRAZE	CVUT	Czech Republic
3	TEAGASC - AGRICULTURE AND FOOD DEVELOPMENT AUTHORITY	TEAGASC	Ireland
4	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTh	Greece
5	Technion – Israel Institute of Technology	Technion	Israel
6	ASOCIACION CENTRO DE INVESTIGACION COOPERATIVA EN BIOCIENCIAS	CIC bioGUNE	Spain
7	National and Kapodistrian University of Athens	NKUA	Greece
8	UNIVERSITAT DE VALENCIA	UVEG	Spain
9	BRUKER BIOSPIN GMBH	BRUKER	Germany
10	MICROCAYA, S.L.	Microcaya	Germany
11	SCIENSANO	Sciensano	Belgium
12	UNIVERSITA DEGLI STUDI DI TRENTO	UNITN	Italy
13	CONSORCIO CENTRO DE INVESTIGACION BIOMEDICA EN RED M.P.	CIBER	Spain
14	ISTITUTO SUPERIORE DI SANITA	ISS	Italy
15	TERVISE ARENGU INSTITUUT	TAI	Estonia
16	IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE	ICL	United Kingdom
17	UNIVERSITY OF LEICESTER	ULEIC	United Kingdom

List of tables, terms and abbreviations

List of tables

List of Terms and Abbreviations

Abbreviation	Definition
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Reproducible
GDPR	General Data Protection Regulation
WP	Work Package
IPR	Intellectual Property Right
GA	Grant Agreement
CA	Consortium Agreement
KB	Kilobytes
MB	Megabytes
GB	Gigabytes
TB	Terrabytes
EU	European Union
AI	Artificial Intelligence
DOI	Digital Object Identifier
VPN	Virtual Private Network
XML	Extensive Markup Language

TABLE OF CONTENTS

Introduction

Purpose and Scope

The purpose of deliverable D3.3: Data Management Plan Update RP1 is to describe an updated version of the data management life cycle for the data to be collected, processed and/or generated by the CoDiet project, encompassing both data collected in WP2 and previously, as explained in WP1. As part of making research data Findable, Accessible, Interoperable and Reusable (FAIR), the project's Data Management Plan (DMP) includes information on the handling of research data during and after the end of the project; what data will be accessed, collected, processed and/or generated; which methodology and standards will be applied; whether data will be shared/made open access; how data will be curated and preserved (including after the end of the project).

The information in this document does not supersede the rules and conditions laid out in the CoDiet Grant Agreement (GA) and those in the CoDiet Consortium Agreement (CA).

Approach for Work Package and relation to other Work Packages and Deliverables

From an organisational point of view, the present D3.3 deliverable is a direct outcome of T3.1: Data Management Plan. However, its scope extends to most major CoDiet activities and more specifically will manage the data that are collected by the consortium, esp. in WP2 and WP5. Its purpose is to ensure that data generated and published within the project are appropriately licensed, openly accessible as much as possible, adequately described, contextualised and documented, and in accordance with all major EU (GDPR, EU AI Act, EU Data Act) and national regulations related to data integrity, security and privacy. As activities of the project evolve, all aforementioned aspects will likely be clarified or modified. To accommodate this reality, the DMP will be a living document, regularly reviewed and revised in order to incorporate and manage currently unforeseen cases and settings of data usage and production in the projects.

Methodology and Structure of the deliverable

The deliverable is structured in accordance with the template and guidelines provided by the EU, and is organised in the following sections: Section 2 provides a data summary addressing issues regarding the purpose of the data access/collection/generation and its relation to the objectives of the project, the types and formats of data the project will generate/collect, the origin of data, the expected size of data, and the data utility (i.e., to whom might it be useful). Section 3 reports on the measures and directions to be adopted to ensure the compliance of CoDiet with FAIR data principles. Section 4 describes the management of other research outputs. Section 5 summarises the allocation of resources and a preliminary cost coverage plan in the context of the project, in order to serve the aforementioned measures. Sections 6 and 7 discuss the main concerns regarding data security, privacy, ethics and the proposed approaches to face them, while Section 8 concludes D3.1.

Data Summary

Pre-existing datasets - Reference Datasets

Comentado [IT1]: a new data set was offered by ICL, nutritech?

UK Biobank

WP3 can develop the methods on pre-existing datasets, such as the UK Biobank. The access to UK Biobank has been budgeted for. UK Biobank holds an unprecedented amount of data on half a million participants aged 40-69 years (with a roughly even number of men and women) recruited between 2006 and 2010 throughout the UK. Showcase (available at http://www.ukbiobank.ac.uk) aims to present the data available for health-related research in a comprehensive and concise way, and to provide technical information for researchers considering applying to use the resource.

nmrshiftdb2

nmrshiftdb2 is a NMR database (web database) for organic structures and their nuclear magnetic resonance (nmr) spectra. It allows for spectrum prediction (^{13}C, ^{1}H and other nuclei) as well as for searching spectra, structures and other properties. The nmrshiftdb2 software is open source, the data is published under an open content license. The core of nmrshiftdb2 are fully assigned spectra with raw data and peak lists. Those datasets are peer reviewed by a board of reviewers. The project is supported by a scientific advisory board. nmrshiftdb2 is part of the NFDI4Chem initiative and will provide a component for a curated repository there. One-dimensional NMR shift data (^{13}C, ^{1}H, ^{15}N, ^{31}P etc.) can be stored as well as two-dimensional spectra of any type. Nuclei and spectrum types can be added as needed. For each spectrum (one-dimensional as well as two-dimensional) peak lists, spectrum images, and raw data can be stored. Together, they form the full characterization of a structure, which is available together with the automatic report and the reviewer opinion for download. The implemented search algorithms can scan the database for example for the following items: substance name, formula, structure, substructures, chemical shifts, and a Hit List of best matches is generated. Furthermore, a shift prediction algorithm for all nuclei has been included.

Data Collected within the project

Data Collection Protocol

We will conduct an observational study with 200 adults at high risk of developing NCD during 8 weeks to validate the efficacy of the selected technologies to monitor food intake and NCD risk in four countries throughout Europe: Ireland, Spain, Greece and the UK (allowing the pilot testing in four different cultures). An eligibility questionnaire considering the inclusion and exclusion criteria will be implemented in each country In Spain volunteers will be recruited from people undergoing their annual medical check-up-cohort (CBG) and in the Preventive Medicine Department (UVEG), in Ireland (Teagasc), in Greece (AUTh) and in the UK (ICL). All participants will sign informed consent declarations prior starting the study. Participants will be assigned to the different sub-cohorts matched for gender and age.

Inclusion criteria: Men and women (50:50) adults (18-65 years), at high risk of developing NCD as assessed using the metabolic syndrome risk score. Overweight or obesity BMI > $25Kg/m2$ plus ~~any two~~one of the following four factors:

Comentado [SAM2]: This criteria has been modified in June

- Raised triglycerides: ≥ 150~~00~~mg/dL ($1.7 mmol/L$)

Comentado [SAM3]: It was a mistake and now it was updated

- Reduced HDL cholesterol: < $40 mg/dL$ ($1.03 mmol/L$) in males or < $50 mg/dL$ ($1.29 mmol/L$) in females.

- Raised blood pressure: systolic BP ≥ 130 or diastolic BP ≥ 85 mm Hg.

- Raised fasting plasma glucose (FPG): ≥ $90 mg/dL$ ($5.0 mmol/L$).

- Current smokers

Exclusion criteria: Outside of specified age and weight range, chronic medical condition including for diabetes and hypertension, a diagnosis of diabetes, cancer, acute infectious disease, renal disease and cardiovascular disease, chronic gastrointestinal condition. Not to use antibiotics during the last 12 weeks before starting the study.

Sample size calculation: G*power software has been used to calculate sample size and actual power considering we would like to assess bivariate correlations between variables. We have selected the "A priori" test of power analysis (based on this article: Behavior Research Methods 2009, 41 (4), 1149-1160, doi:10.3758/BRM.41.4.1149). As input parameters we have selected 2-tailed testing, $\alpha = 0.05$, $power(1-\beta) = 0.8$, correlation $\varrho H0 = 0$ and correlation $\varrho H1$ in the $sample = 0.2$. Then, the corresponding total sample size obtained is 194 and the actual $power = 0.800084$.

At baseline (first visit) and at the end of the study (after 8 weeks) biological samples (blood, urine and faeces) will be collected for the assessment of dietary intake and NCD risk biomarkers together with anthropometric parameters. Moreover, dietary habits and physical activity will me monitorized along 8 weeks.

Demographics and habits: A questionnaire will be designed to collect information on sex, age, income and education level; smoking and drinking habits, medical history at baseline.

Anthropometric measurements: Height, weight, waist, blood pressure. Most complete body composition will be measure with Inbody device (Microcaya).

Written informed consents will be obtained from the participants according the site-specific Ethics Committee requirements.

Genomics data

Genomic DNA will be isolated from blood. Genome-wide genotyping and epigenome-wide methylation determinations will be carried out. Polygenic risk scores (PRS) (only at baseline) and

Epigenetics (DNA methylation + miRNA in a subset) (UVEG) will be computed. Biological samples: 2 samples of 1mL of whole blood will be collected at baseline for genomic DNA isolation and high-density (genome-wide) genotyping as well as for genome-wide DNA methylation. After the intervention, 1 sample of whole blood will be collected for changes in DNA methylation (in a subsample).DNA will be isolated from leukocytes using the Magna Pure 96 device and the corresponding human DNA isolation kits following the manufacturer protocol (Roche). Genome-wide genotyping: After quality control using a Qubit fluorometer and a nanophotometer, DNA samples will be processed using the Illumina Global Screening Array. More than 600.000 sequence variants will be determined. Allele detection and genotype calling will be performed in the Genome Studio genotyping module (Illumina, Inc.). Data cleaning will be performed using standard analysis pipelines implemented in the Phyton programming language combined with PLINK. Additionally, imputation of DNA variants will be undertaken. PRS will be computed for the most relevant phenotypes. For epigenome-wide methylation, after quality control, DNA methylation (from 0 to 100%) in selected CpG sites will be assessed by using the Illumina Human EPIC methylation array. Currently the EPIC v.1 array is not available and we will use the Infinium MethylationEPIC v2.0 BeadChip Kit, a genome-wide methylation screening tool that targets over 935,000 CpG sites in the most biologically significant regions of the human methylome. Bisulfite conversion of DNA will be performed using the Zymo EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, USA) and samples will be hybridized to the Illumina EPIC array, according to the manufacturer's protocol. Microarrays will be scanned with an Illumina HiScan system and ".idat" files will be generated. Quality control procedures will be implemented involving the use of Minfi, Meffil and ewastools R packages, among others.Methylated sites and regions will be analyzed. Beta-values (ranging from 0 to 1 and) will be obtained as metrics to measure methylation levels. Subsequently, beta-values will be converted to M-values as follows: M-value = log2 (beta/(1 − beta)). The advantage of the M-values is the higher homoscedasticity compared with beta-values. Methylation risk scores (MRS) will be computed and constructed. PRS and MRS will be validated and combined in mixed multi-omic scores. In addition, sex-specific, as well as population specific PRS and MRS will be constructed. Exploratory genome-wide association studies (GWASs) for several traits as well as epigenome-wide association studies (EWASs) will be undertaken. In addition, gene set enrichment analysis of the differentially methylated CpG sites will be performed for biological and functional interpretation. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses and Gene ontology (GO) enrichment will be used for the top differentially methylated CpG sites obtained. microRNA profiling will be carried out from RNA isolated from fresh blood (1 mL) in a subset (participants from the same recruitment site: UVEG). After quality control of isolated RNA, selected samples will be processed using a microRNA array profiling the human microRNAome. Quantitative results of microRNA signatures associated to the relevant CoDiet phenotypes will be explored.

New data will be collected and generated both in digital and non-digital format. The research data will exist in a number of states (e.g., raw, cleaned, processed, analysed, archived) and take a number of forms including:

- Results of the quality control from DNA and RNA samples (concentration in ng/mL) ratio of absorbance at 260/280. Results of the RNA quality control (concentration in ng/mL and RIN): in excel and word files.
- Laboratory notebooks (e.g., details of experiments, measurements, observations from fieldwork, etc)

- DNA genotyping results: large datset with the corresponding genotypes: AA, AC, CC for each participant in more than 600.000 genes and the corresponding imputation of additional 1 or several millions of single nucleotide polymorphisms (PLINK files).

- DNA-methylation results: large datset with the corresponding beta and or M-values for methylation in more than 900.000 CpG sites (Idat files).

Final and intermediate files generated from the omic analyses including, genotyping, methylation and microRNA profiles (e.g., cleaning, transformation, normalization, etc.)

- Outputs from bioinformatics and statistical analysis: selected candidate genes, GWASs, PRS, EWAS, selected DNA-methylation sites, MRS, combined PRS and MRS, pathway analyses and other functional analyses both for genotyping and methylation.

- Analytical pipelines

- Several outputs from statistical analyses including summary data for additional analyses.

- Metadata from the corresponding participants (excel files)

Metabolomics data

Microbiome (Teagasc, UNITN). Stools will be collected with validated self-collection kit. One tube containing stool sample with a stabilizing buffer (DNA/RNA Shield Fecal Collection Tube - Zymo Research R1101) will be kept at room temperature until reaching the UNITN lab where it will be stored at –20 °C until the processing step. A Shotgun metagenomic sequencing will be used to profile the gut microbiome. An implemented and standardized experimental and computational framework for the generation of microbiome data will be applied. The sequencing output data will pre-processed including the elimination of the host DNA (human) and only the microbial genetic material from the gut will be used for further analysis. Additionally, NMR metabolomics will be assessed in stools samples containing buffer.

Urine Metabolomics (UCL, CIC BIOGUNE, AUTh): Complementary assays will be applied to provide the most comprehensive description of the metabolome. 8 eppendorfs of 1ml of urine sample will be required to assess the following measures: NMR metabolomics (600 MHz IVDr spectrometers) (ICL and CBG) Urine samples acquisition donnors will be fasting. They will be provided with a urine collector that will be refilled with the first urine of the morning. They will keep the urine in the refrigerator until it is taken to the assigned collection center. Subsequently the urine will be aliquoted into 1.5 mL tubes (1 mL in each tube). Finally all aliquots will be frozen at -80.

Urine process for NMR metabolomics. Urine will be kept at -80°C. After thawing samples, they will be centrifuged at 6000 RPM for 5 minutes at 4°C. 630 µL of the urine supernatant will be transferred to an 1.5 ml Eppendorf with 70 µL of urine buffer containing sodium azide and TSP. After mixing the buffered urine, 600 μL of well mixed sample will be transferred into 5 mm NMR-tube and introduced into the magnet for the acquisition.

600 MHz AVANCE NEO (IVDr): fitted with an automatic cooling sample changer (SampleJet) will be used for urine experiments. All the urine samples will be acquired at 300K. For each sample two experiment were measured: a one-dimensional (1D) 1H spectrum with water presaturation using the nuclear Overhauser effect spectroscopy (NOESY) pulse sequence (noesygppr1d) and a two-dimensional (2D) J-resolved spectroscopy (JRES, jresgpprqf). For a subset of samples, to help with metabolites identification, a 2D 1H total correlation spectroscopy (TOCSY) experiment will be acquired.

CIC bioGUNE provides data in tabular format that has been generated from the cohorts we work with. The cohorts consist of groups from hundreds to thousands of individuals (it is up to the different cohorts we have) who have donated urine and blood samples during their annual medical check-up or another situations. During the donation process, they filled out a questionnaire with anthropometric and medical data, and urine and blood samples were analyzed to extract their biochemical data.

All of this data is included in the cohorts, with descriptive variable names, including the units of measurement. Subsequently, NMR is used to obtain metabolomic data from both urine and serum samples. Specifically, 150 urine metabolites are obtained (expressed in mmol/mol of Creatinine), along with 41 serum metabolites (expressed in mmol/L), and 112 subclasses of lipoproteins (expressed in mg/dL or nmol/L). Since there are metabolites that are found in both urine and blood, suffixes " u" and " b" have been added to differentiate their origin for urine and blood, respectively.

Data Format and Size: All of this data is presented in a tabular format, where different columns contain the individual's code, anthropometric and medical data from the questionnaire, biochemical data, and metabolomic data. For WP1, 10,000 synthetic data have been generated using the synthpop algorithm (Nowok B, Raab GM, Dibben C (2016). "synthpop: Bespoke Creation of Synthetic Data in R." Journal of Statistical Software, 74(11), 1–26. doi:10.18637/jss.v074.i11) with a subset of real data from our Akribea cohort (working population from the Basque Country). The unique code for each synthetic individual is only a sequential number with the prefix "AK". The data will be provided in an Excel file which will occupy approximately 25 MB. This size would be for 10.000 individuals (serum and urine). For UPLC-MS data we also generate excel files in tabular format with all the metabolites quantified expressed in uM or peak intensity.

Tyrosine and tryptophan pathway (AUTh): metabolites will be analysed based on a method currently developed on a triple quadrupole technology. Key intermediates of the two pathways will be mapped, and quantified in both urine and blood.

Untargeted HILIC MS profiling (AUTh): Urine sample (50 μL) will be analysed by HILIC-QTOF MSon a ACQUITY UPLC BEH Amide Column (, 1.7 μm) using a binary mobile phase system consisting of A: 5 mM ammonium formate (pH3) in ACN/H2O (95:5, v/v) and B: 5 mM ammonium formate (pH 3) in H2O/ACN (70:30, v/v).

Untargeted reversed phase LC profiling (AUTh): will be performed on a Reverse Phase system (Acquity BEH C18 column), using a UHPLC-qTOF-MS (Bruker). Elution will be performed using 0.1% (v/v) aqueous formic acid as solvent A and acetonitrile 0.1% formic acid (v/v)) as solvent B. Full scan MS data will be collected over a range of 80 to 1000 m/z as well as MS/MS data both DDA and with no precursor ion selection (bbCID).

Saccharides quantification (AUTh): will be performed by HILIC-MS/MS analysis based on an ACQUITY UPLC BEH Amide Column and a elution system comprised by solvent A: 5 mM ammonium formate (pH3) in ACN/H2O (95:5, v/v) and solvent B: 5 mM ammonium formate (pH 3) in H2O/ACN (70:30, v/v) under gradient elution conditions. MS/MS analysis based on optimised detection of a number of saccharides will be performed on a Triple Quadropole.

Blood metabolome: Serum will be isolated from blood samples to measure: 8 eppendorfs of 500uL of serum.

Serum lipoprotein subfractions: using the Bruker IVDr lipoprotein subclass (ICL). Finally, in the serum analysis an additional report will be obtained: the Bruker IVDr Lipoprotein Subclass Analysis B.I.LISATM. Thanks to this additional report the quantification of 112 lipoproteins parameters is possible. Specifically, it is possible to assess information about the main VLDL, IDL, LDL and HDL classes, six VLDL subclasses (VLDL-1 to VLDL-6) six LDL sub-classes

(LDL-1 to LDL-6), four HDL-subclasses (HDL-1 to HDL-4). Lipoproteins are sorted according to the increasing density and consequent decreasing size.

NMR metabolomics: (ICL and CBG) as described above.

SCFAs quantification Mspect. (ICL) as described above.

Amino acid quantification. (CBG) as described above.

Tyrosine and tryptophan pathway (AUTh) as described above.

Untargeted lipidomics (AUTh) as described in the next subsection.

Acyl carnitines (AUTh): For the analysis of the thirteen (13) acylcarnitines, HILIC-MS/MS will be applied in 50 μL blood Separation will be performed on an ACQUITY BEH Column , 1.7 μm) under isocratic conditions with 90:10 A: B (v/v) (A: MeCN- H2O, 10 mM AF, 95:5 (v/v), pH 3, B: MeCN-H2O, 10 mM AF, 30:70 (v/v), pH 3) at 50 °C column temperature. Detection of analytes will be performed by a XEVO TQD mass spectrometer (Waters, UK) operating in electrospray ionization positive ion mode (+ESI). Detection and quantification will be done by optimised MS/MS transitions which are monitored MRM mode) for each analyte.

Sacharides quantification (AUTh) as described above HILIC-MS/MS on UHPLC-triple quadrupole according to Virgiliou et al 2015 Electrophoresis. Separation on a ACQUITY BEH Column1.7 μm column and detection in MRM mode. Analytes include: Lactose, Arabinose, Arabitol, Fructose, Galactose, Glucose, Inositol, Mannitol, Maltose, Ribose, Sorbitol, Sucrose Xylitol Xylose, Inflammation markers by NMR (CBG and ICL) for the quantitative determination of different biomarkers for systemic inflammation, and consequently, as a potential biomarker for cardiovascular CVD risk assessment, including GlycA, GlycB, SPC , CRP, TNF and targeted pro- and anti-inflammatory cytokines. A J-Edited DIffusional (JEDI) proton NMR spectroscopic approach to selectively augment signals from the inflammatory marker peaks GlycA and GlyCB and a PGPE experiment to quantify SPCs in blood serum NMR spectra, will be also added to the other experiments developed for serum samples.

New data to be collected will be in digital and non-digital format and will be collected as administrative, technical and sequencing/metagenomic data. The research data will exist in a number of states (e.g., raw, cleaned, processed, analysed, archived) and take a number of forms including:

- Results of experiments, trials and studies (in excel files and word files)
- Laboratory notebooks (e.g., details of experiments, measurements, observations from fieldwork, etc)
- Sequence data in the form of nucleotide/protein/metabolomic sequences from:
  - Shotgun metagenomics
  - Metabolomes
  - Static intermediate files generated from the analyses (e.g., contigs from sequence assembly)
- Outputs from bioinformatic analyses
- Analytical pipelines
- Outputs from statistical analyses

- Metadata from clinical studies (excel files)
- Results of Surveys and Questionnaires (excel files)

Lipidomics data

Blood lipidomics: Lipid intake will be assessed by AZTI using targeted lipidomics measuring the fatty profile (including SFA, MUFA, PUFA-omega 3 and PUFA-omega 6 in whole blood by gas chromatography-mass spectrometry (GC-MS) using Dried Blood Spots (DBS) technology that consist of a four drops of whole blood collected on a card, which stabilizes the sample and makes it easier to analyse the fatty acid profile in blood.

Targeted Membrane lipidomics (AZTI) will be applied to study altered lipid metabolism and inflammation. The fatty acid composition of mature RBC membrane phospholipids is obtained from whole blood sample (1 mL) collected in vacutainer tubes containing ethylene diamine tetraacetic acid (EDTA). Samples are centrifuged to remove plasma and collect the pellet with erythrocytes that must be stored in ~~a refrigerator (4ºC) for a maximum of 14 days before being shipped (samples can be accumulated for 14 days and send them together to save shipping costs). Samples cannot be frozen~~at -80ºC until the analysis at the end of the study visits. The samples will be sent to ~~the Lipidomics Laboratory (Bologna, Italy)~~AZTI frozen at -80ºC. ~~at controlled room temperature until arrival (shipment within 24 hours).~~

Comentado [SAM4]: This was changed and approved among WP2 partners.

Untargeted Lipidomic profile will be assed in 50 μL of blood serum by UHPLC-qTOF-MS (AUTh). Pasma samples will be treated with 700 μL of the organic solvent mixture MTBEMeOH (3:1 v/v), followed by 5 min vortex (20 Hz). The samples will be then centrifuged for 30 min at 4 °C and 25,200×g. Supernatants (600 μL) will be transferred to Eppendorf tubes and evaporated to dryness. Reconstitution will be performed with 150 μL H2O:ACN:IPA (1:1:3 v/v). Chromatography will be performed on a UHPLC Elute system (ACQUITY UPLC CSH C18, 1.7 μm column, Waters). The mobile phase system will be consisted of solvent A: ACN:H2O (60:40 v/v), 10 mM ammonium formate, and 0.1% formic acid, and solvent B: IPA:ACN (90:10 v/v) and 0.1% formic acid. Elution will be performed at 55 °C at a flow rate of 0.4 mL/min by applying a gradient profile. Washes with IPA:ACN (90:10 v/v) ACN:H2O (60/40 v/v) will be performed before and after each injection. A 5 and 10 μL extract will be injected in positive and negative mode, respectively. The MS data will be acquired using a TIMS TOF mass spectrometer (Bruker), MS/MS experiments both DDA (data-based acquisition) and with no precursor ion selection will be performed (bbCID). MS/MS data scans will be acquired for QC samples and a randomly selected sample from each group. Calibration with sodium formate, 10 mM) is performed in every injection. A pooled sample (QC) will be prepared by mixing equal volumes of each sample and will be analysed periodically during the batch.

Data Format and Size: Data will be shared in an excel file format and will contain both, membrane lipidomics and DBS data. It will include 42 variables (columns) with fatty acids of interest. The estimated size of the provided data will be around 75KB.

During the first year of the project, AZTI has analysed 45 DBS samples corresponding to samples collected in 5 recruiting sites. Data on fatty acid profile are stored in excel file format. They have a total size of 200Kb. Compared to the initial description, 2 fatty acids have been removed (C17:0 and C20:2) from the list and the excel file includes 21 columns with the following fatty acids:

d_c16_0	d_c16_1w7	d_c18_0	d_c18_1trans9	d_c18_1d_cisw9	d_c18_1vaccenic	d_c18_2d_cis6	d_c18_3w6	d_c18_3w3	d_c20_1w9	d_c20_3w6	d_c20_4w6	d_c20_5w3epa	d_c22_5w3	d_c22_6w3dha	d_c24_1w9	d_sfa	d_mufa	d_pufa	d_o3i	d_w6_w3

Camera devices data

Passive Camera technology (ICL) to quantify the dietary intake, an intelligent wearable camera device (worn on the ear/glasses) will be used to capture eating episodes pervasively and passively providing an objective and more accurate measurement of nutrient intake. Novel computer vision and deep learning techniques will be refined for automatic recognition of food types and estimation of consumed portion sizes. Participant will wear the camera every day during 3 weeks along the 8 weeks of the study (week one, week 4 and week 8).

The camera generates jpg images (which is a standard file format for images), written on an micro-SD card. Image files are collected at a rate of 1 frame per 2 seconds, i.e., 0.5 fps, assuming that we also want to capture brief episodes of snacking during the day. Image filenames are date- and time-stamped. Each image is at HD 1080p resolution 1920x1080 pixels, with a size of a little less than 1MB. As one can imagine, the size of the camera data will probably dwarf any other collected data.

Data Format and Size: Assuming 200 participants wearing the camera daily for 3 weeks, say for 16 hours every day. From a quick calculation, the total size of all images captured per participant would be around 634,178,764,800 bytes = 590 GB. Multiplying this by 200 for all participants, is about 115 TB. These data will also be accompanied with metadata file which will give a thorough description of the type and estimated volume/weight of food consumed, alongside the date- and time-stamped image filename they correspond to. These are text strings and will be provided in CSV format, so will be readable by Excel, for example.

Other devices data

Non-invasive sensors for measuring body composition, stress level, arterial hardening and Advanced Glycation End-products (AGEs) measured by non-invasive sensors by Microcaya. Commercial, non-expensive and portable equipment that can be easily incorporated into health surveillance programs at real environments will be included.

InBody (1 Measurement of 1.5 minutes): Body composition Analyzer. Bioelectrical Impedance Analysis (BIA) is a non- invasive method of measuring impedance by applying alternating electrical currents to a user to measuring their volume of water through impedance values. Low-level electrical currents are sent through the body, and the flow of the current is affected by the amount of water in the body. A healthy balance of body water is critical for good health. With Inbody all segments (arms, trunk and legs) are measured separately. BIA devices measure how this signal is impeded through different types of tissue (muscle has high conductivity but fat slow the signal down). As BIA determines the resistance to flow of the current as it passes through the body, it provides estimates of body water from which body fat is calculated using selected equations. Parameters such as weight, total body water, Fat free mass, skeletal muscle mass and body fat mass are returned. In addition, the ratio of extracellular and total body water of the individual is analysed. Some of these parameters are raw values like resistance and reactance and the phase angle which is a good predictor parameter of the general health of the person.

Data Format and Size: This device data will be provided in excel format with a set of 232 variables describing the data and will have an estimated size of 243KB.

SA3000P (1 Measurement of 3 minutes): Heart Rate Variability & Accelerated Photoplethysmograph Analyzer. The SA3000P is a simple, user-friendly and non-invasive. It provides measurements in 3 minutes using Heart Rate Variability & Accelerated Plethysmography to access overall Cardiovascular and Autonomic Nervous System function. Accelerated Photoplethysmography (APG) is a simple optical technique used to detect volumetric changes in blood in peripheral circulation. It provides the valuable information related to our cardiovascular system. It is widely used in clinical physiological measurement and monitoring the arterial circulation. Heart rate variability or HRV is the physiological phenomenon of the variation in the time interval between consecutive heartbeats in milliseconds. HRV is regulated by the autonomic nervous system (ANS), and its sympathetic and parasympathetic branches, and HRV is commonly accepted as a non-invasive marker of ANS activity. The ANS helps you respond to daily stressors and regulate some of your body's most important systems, including heart rate, respiration and digestion.

Data Format and Size: This device data will be provided in an excel format for the APG data with 15 variables with an estimated size of 127 KB and in an excel format for the HRV data with 40 variables 102 KB. The fields of APG data provide information such as ChartNo, Name, BirthDate, Age, Gender, Exam.Date, Wave Type, AI, AI Status, AE, AE Status, PE, PE Status, HR, HR Status and APGComment. The fields of HRV data contain information such as ChartNo, Name, BirthDate, Age, Gender, Exam.Date, Measure Sensor, ANS Activity, ANS Activity Status, ANS Balance, ANS Balance Status, Stress Resilience, Stress Resilience Status, Stress Index, Stress Index Status, Fatigue Index, Fatigue Index Status, Mean HRT, Mean HRT Status, Electro-cardiac Stability, Electro-cardiac Stability Status, Ectopic Beat, SDNN, PSI, TP, VLF, LF, HF, LFNorm, HfNorm, Lf/Hf, RMSSD, APEN, SRD, TSRD, TP(ln), VLF(ln), LF(ln), HF(ln) and DDRComment. However, fields of the data will be fully anonymized in order to prevent data coupling and also to be in line with the GDPR legislation.

AGE Reader (3 measurements of 12 seconds): Advanced Glycation End products Analyzer. The AGE Reader is a non-invasive monitoring device that uses ultra-violet light to excite autofluorescence in human skin tissue. The autofluorescence is from the level of Advanced Glycation End products (AGEs). The measurement of AGEs provides an immediate cardiovascular risk prediction in 12 seconds. The point-of-care measurement of AGEs offers great opportunities for multiple purposes. AGE Reader applications include: Patient management: early detection of (diabetes) patients at risk of developing cardiovascular complications. Professional health assessment: identify individuals with an increased risk of diabetes and additionally the metabolic syndrome. Each of the analysers include software (local or in the cloud) that allows the management of all measurement information and volunteer's data. These applications offer the functionality to export the data to different files for further analysis with statistical packages or other technologies.

Data Format and Size: This device data will be provided in .csv format with 10 variables and an estimated size of 29KB. The fields of this data provide information such as Patient ID, First Name, Last Name, Age Group, Gender, Date of Birth, Exam Date, Next visit date, AGE, Smoking, Diabetes, LDL cholesterol, HDL cholesterol, Systolic blood pressure, Intima-media thickness, Pulse wave velocity, Physician, Remarks and Sequence number.

Dietary habits data

Food24h recall as a method of collecting nutritional self-reported intake: 3 24h recall records will be collected the days previous to samples collection. Intake24 software will be used (https://intake24.org/info/open-source).

Sociodemographic data

A general questionnaire placed at RedCap system (https://www.project-redcap.org/) will collect sociodemographic information, dietary habits and lifestyle information of 200 volunteers in the study visit 1.

Data Format and Size: These data will be extracted from RedCap system and stored in an excel file that will include 74 variables (columns). The estimated size of the provided data will be around 67KB.

Dataset	Summary	Pre-Existing	Data Format	Collect. By	Est. Size

UKBioBank	UK Biobank holds an unprecedented amount of data on half a million participants aged 4069 years (with a roughly even number of men and women) recruited between 2006 and 2010 throughout the UK.	Yes	NA	UK Biobank	NA
nmrshiftdb2	nmrshiftdb2 is a NMR database (web database) for organic structures and their nuclear magnetic resonance (nmr) spectra. It allows for spectrum prediction (13C, 1H and other nuclei) as well as for searching spectra, structures and other properties	Yes	NA	nmrshiftdb	NA

Table 1: Pre-Existing Datasets Summary

Dataset	Summary	Data Format	Collect. By	Est. Size
Microbiome	New data to be collected will be in digital and non-digital format and will be collected as administrative, technical and sequencing/metagenomic data. The research data will exist in a number of states (e.g., raw, cleaned, processed, analysed, archived) and take a number of forms including: results of experiments, trials and studies, laboratory notebooks, sequence data in the form of nucleotide/protein/metabolomic sequences from shotgun metagenomics, metabolomes, static intermediate files generated from the analyses (e.g., contigs from sequence assembly), outputs from bioinformatic analyses, analytical pipelines, outputs from statistical analyses, metadata from clinical studies, results of Surveys and Questionnaires. CIC bioGUNE provides data in tabular format that has been generated from the cohorts we work with. The cohorts consist of groups from hundreds to thousands of individuals (it is up to the different cohorts we have) who have donated urine and blood samples during their annual medical check-up or another situations	Excel - CSV, Word	AZTI - CIC bioGUNE - UCL - CIC BIOGUNE - AUTh - ICL - CBG	˜25MB
Lipidomics	Data will be shared in an excel file format and will contain both, membrane lipidomics and DBS data. It will include 42 variables (columns) with fatty acids of interest.	Excel - CSV	AZTI	75KB

Table 2: Metabolomics and Lipidomics Datasets Summary

Dataset	Summary	Data Format	Collect. By	Est. Size
Camera Devices	The camera generates jpg images (which is a standard file format for images), written on an micro-SD card. Image filenames are date- and time-stamped. Each image is at HD 1080p resolution 1920x1080 pixels, with a size of a little less than 1MB.	RAW(?)	ICL	115TB
Camera Devices Metadata	These data will also be accompanied with metadata file which will give a thorough description of the type and estimated volume/weight of food consumed, alongside the date- and time-stamped image filename they correspond to.	Excel - CSV	ICL	some KB

Dietary Habits			ICL	

Table 3: Camera device and dietary habits Datasets Summary

Dataset	Summary	Data Format	Collect. By	Est. Size
Other Devices (InBody)	Parameters such as weight, total body water, Fat free mass, skeletal muscle mass and body fat mass are returned. In addition, the ratio of extracellular and total body water of the individual is analyzed. Some of these parameters are raw values like resistance and reactance and the phase angle which is a good predictor parameter of the general health of the person.	Excel-CSV	Microcaya	243KB
Other Devices (SA300P)	This device data will be provided in an excel format for the APG data with 15 variables with an estimated size of 127 KB and in an excel format for the HRV data with 40 variables 102 KB. The fields of APG data provide information such as ChartNo, Name, BirthDate, Age, Gender, Exam.Date, Wave Type, AI, AI Status, AE, AE Status, PE, PE Status, HR, HR Status and APGComment. The fields of HRV data contain information such as ChartNo, Name, BirthDate, Age, Gender, Exam.Date, Measure Sensor, ANS Activity, ANS Activity Status, ANS Balance, ANS Balance Status, Stress Resilience, Stress Resilience Status, Stress Index, Stress Index Status, Fatigue Index, Fatigue Index Status, Mean HRT, Mean HRT Status, Electro-cardiac Stability, Electro-cardiac Stability Status, Ectopic Beat, SDNN, PSI, TP, VLF, LF, HF, LFNorm, HfNorm, Lf/Hf, RMSSD, APEN, SRD, TSRD, TP(ln), VLF(ln), LF(ln), HF(ln) and DDRComment.	Excel - CSV	Microcaya	APG (127KB) - HRV (102KB)
Other Devices (AGE Reader)	This device data will be provided in .csv format with 10 variables and an estimated size of 29KB. The fields of this data provide information such as Patient ID, First Name, Last Name, Age Group, Gender, Date of Birth, Exam Date, Next visit date, AGE, Smoking, Diabetes, LDL cholesterol, HDL cholesterol, Systolic blood pressure, Intima-media thickness, Pulse wave velocity, Physician, Remarks and Sequence number.	Excel - CSV	Microcaya	29KB

Table 4: Other devices Datasets Summary

FAIR Data

Making data findable, including provisions for metadata

Our benchmark data sets will be deposited in permanent repositories (such as Zenodo.org, EGA (https://ega-archive.org), ENA (https://www.ebi.ac.uk/ena/) and MetaboLights (https://www.ebi.ac.uk/metabolights/)) and associated with specific, unique identifiers (such as DOI) unique identifiers (such as DOI) depending on the type of data (e.g., restrictions for the raw genome-wide genotyping data) . and associated with specific, unique identifiers (such as DOI). These data objects will be correlated with the respective metadata, keywords and identifiers so they can be Findable, re-findable and easily searchable. Data access restrictions will be implemented depending on the informed consent signed by the study participant in each site. Metadata will be retrievable by machines and humans, possibly upon the appropriate authentication and authorization processes of the existing platforms, through a well-defined and built protocol of use and in such a way that the metadata can be harvested and indexed. There are no 'Minimal Metadata About ...' (MIA...) standards for our experiments. However, we have a good idea of what metadata is needed to make it possible for others to read and interpret our data in the future.

We will use electronic lab notebooks to make sure that there is good provenance of the data analysis. Data analysis is normally done step-by-step. It is important that for all data the origin and all processing and filtering steps are documented, otherwise results will not be reproducible. Re-users of the data also need this information to decide whether the data can be used for their purpose. In computing, systems like Galaxy and (Jupyter) notebooks often automatically keep provenance information. The provenance will be captured using W3C PROV. We will be keeping the relationships between data clear in the file names. All the metadata in the file names also will be available in the proper metadata.

Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs. Data protection procedures according to the participant signed consent will be carried out.

Making data accessible

Most of the datasets that will be used as part of the project will be offered by the project partners, as well as some pre-existing datasets which will be used as reference datasets, such as NMRShiftDB etc. Some of these datasets are already open to the public, while, at the same time, some of the datasets that will be collected from the consortium partners have high sensitivity. We will be working with the philosophy as open as possible for our data. The data cannot become completely open immediately because of legal reasons. In the cases where private data are aggregated and processed (e.g., as part of a model, functionality of a component or experimental evaluation of models) permission will be requested from the dataset provider prior to making the altered data available. Therefore, concerning the legal reasons and the sensitivity of the data, a data sharing agreement will be required. People can apply to the data access committee that we will set up, prior to getting access to the data, in a formal way (sending a formal request to the data. Access committee).

Regarding the nature of private datasets provided by the project partners, some of the results that will be generated, will be restricted to the partners of the project consortium, while other results will be publicly available in collaboration with the actors involved in each work package. As per our commitment to Ethics and data privacy, all the data access and sharing activities will be rigorously implemented in compliance with the data collection and privacy rules and regulations, as they are applied nationally and in the EU, such as the compliance with GDPR, EU AI Act and EU Data Act.

Depending on their nature and size, datasets that are characterised as openly accessible will be published/stored over zenodo.org, openAIRE and Dataverse.org online repositories. These three platforms ensure that the data will be assigned an identifier. Metadata will be openly available including instructions how to get access to the data. Metadata will available in a form that can be harvested and indexed (managed by the used repository / repositories). Furthermore, we will examine the benefits of sharing computational components developed in the context of the project through Docker-configured containers and possibly publish non-configured versions in the Docker Hub.

Making data interoperable

One of the major challenges to be tackled within the CoDiet project for making the provided components and datasets valuable and usable is interoperability. For this purpose, CoDiet will extensively use semantic web technologies and will incorporate standards and widely used schemas (like XML), ontologies and vocabularies in order to describe and represent all these data assets. If applicable and possible, data will be available through standard-based APIs. If external data sources have to be used, they will also be ingested through their APIs. The file names are very useful as metadata for people involved in the project, but to computers they are just identifiers. To prevent accidents with e.g., renamed files metadata information should always also be available elsewhere in the form of widely used schemas (like XML). For example, metadata about human gut metagenomics studies will be described using "Genomic Standard Consortium Human Gut Minimal Information checklist". This data management plan ensures that all partners agree on the way they describe experiment, studies and samples.

The documentation of the representation of the data will be included and will follow the same approach as the project's source code (via publicly accessible repositories). Dublin Core is a standard documenting domain independent aspects of a resource; including who has created it, audience, function, formatting and licensing. The key to making data citable, searchable and accessible is equipping datasets with metadata – descriptions of and facts and figures about the data – that meets basic standards and adheres to uniform, consistent schema.

We will adhere also to the specifications provided from the DataCite Metadata Standard (https://schema.datacite.org/), as well as, DDI metadata Standard, which is more extensive than Dublin Core and DataCite. It details more of what is in the data and really can help other researchers locate our datasets as an interesting source and reuse them, in order to increase both findability and reusability. DDI-Lifecycle is designed to document and manage data across the entire life cycle, from conceptualization to data publication, analysis and beyond. It encompasses all of the DDI-Codebook specification and extends it. Based on XML Schemas, DDI-Lifecycle is modular and extensible. This version also supports improvements in Classification management (based on GSIM / Neuchatel), non-survey data collection (Measurements), sampling, weighting, questionnaire Design and support for DDI as a Property Graph.

Increase data re-use

We have identified the types of data that will be used during the project. Some types of data (for example "images" , "videos" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG", "H.264", "CSV" [comma separated values], ".XSLX" [Excel Sheet] and ".SQL" [for Tabular Data]) that help to make these data reusable. In CoDiet, we will be using these types of formats to store the data, which allow for ingestion through automated processes (i.e. load them to a database), they are standard data formats widely used by researchers in this field (data analysts and AI experts) and they enable sharing and long term archiving, compared to other formats.

Whenever possible, the datasets will be licensed under an Open Access licence. However, this is dependable on the level of privacy, as well as, the Intellectual Property Rights (IPR) involved in the data primarily from the project partners. A period of embargo will be necessary if the dataset contains specific IPR or other exploitable results that are adequate to justify an embargo. Therefore, the data will be licensed to permit the widest reuse possible when no limitations are identified by the consortium members.

CoDiet's intention is to make as much data as possible available for reuse by third parties. Restrictions will apply whenever privacy, IPR or other exploitation grounds are in play. All datasets will be cleared of bad records, with clear naming conventions, anonymized if necessary, with appropriate metadata and possible readme files to describe their fields.

There are surprisingly many complications that can cause (slight) inconsistencies between results when workflows are run on different compute infrastructures. All datasets generated and collected in CoDiet will undergo a quality check to analyse their plausibility and consistency, ensuring that these datasets can be used to perform assessments and validations of the results produced by the project. Surrounding all tools in the data processing and analysis workflows with the 'boilerplate' code necessary on the computer system someone is using is tedious and error prone. For this purpose, we will also be instrumenting all these tools into pipelines and workflows using automated tools, which e.g., import and analyze the data, such as Python Pandas library and others.

Validation of results without a golden standard is very hard. One way of doing it is to develop two solutions for a problem (two independent workflows or two independently developed tools) to check whether the results are identical or comparable. Therefore, for in CoDiet we will run part of the data set repeatedly to catch unexpected changes in results.

CoDiet will also have a dedicated Virtual organisation within the CESNET infrastructure, provided by the Czech Technical University, isolating it from other projects, even when those other projects have super-user privileges. The data on those systems is properly backed up, thus ensuring and minimizing the risks of data loss. Ideally, all members of the consortium should not carry project data on laptops, USB sticks or other external media. Access to the infrastructure, as well as to all project web services should be addressed via secure http (https://).

All personal data collected such as the genomic and metabolomic data is sufficiently protected. The metabole and lipidomics datasets do not provide sufficient information to identify a person.

We pseudonymize inside the project, only limited people can access the keys (the data collectors and not the consortium on its whole), and therefore data coupling is available only to the data collectors. In case of video data, these will be anonymized in order to for the recording to be non-intrusive.

Other Research outputs

When desirable by the providing party (i.e. CoDiet partners) the source code of CoDiet resources and components will be openly available under appropriate licences (to be selected per individual case). Regarding components developed by CoDiet consortium members, publishable source code and open source software will be accessible via a central repository hosted in popular relevant services such as Github or Gitlab, as well as in the AI4EU platform, to facilitate its use by the wider European AI community, while at the same time there exists the possibility that the consortium may use repositories only accessible to consortium members for development and testing purposes. Finally, as the project evolves, the Data Management Plan will incorporate any additional details regarding the services and repositories used for distributing, documenting and supporting the source code.

Each beneficiary must examine the possibility of protecting its results and must adequately protect them, for an appropriate period and with appropriate territorial coverage, if a) the results can reasonably be expected to be commercially or industrially exploited, and b) protecting them is possible, reasonable and justified (given the circumstances). When deciding on protection, the beneficiary must consider its own legitimate interests and the legitimate interests (especially commercial ones) of the other beneficiaries. Particular emphasis should be given to establish rules and procedures for ownership (and the management of ownership – including protection strategies) of key project results. Results are owned by the beneficiary that generates them. However, due to the strong collaborative work, two or more partners may jointly contribute to an individual result of IP. In these cases, the IP is jointly owned. The joint owners should therefore agree on the terms of the joint ownership through a Joint Ownership Agreement.

Allocation of resources

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible. Making our data FAIR is therefore not a cost that can be separated from the rest of the project.

We will be archiving data (using so-called 'cold storage') for long term preservation after the project but also already during the project. The used data archiving service is budgeted by one or more of the participating institutes. The minimum lifetime of the archive is 5 years. The archival period can be extended – one of the principle investigators involved in the project will decide. The decision whether or not to extend the renewal be based on the actual use of the archived data. Data formats of data in cold storage will not be upgraded over time. Archived data will not be migrated to other storage media over time. None of the used repositories charge for their services. We have a reserved budget for the time and effort it will take to prepare the data for publication.

Regarding any additional costs for making the data and other research outputs FAIR in CoDiet, we provide a high level summary of the costs emerging from the adopted approach for data and knowledge publishing in the following table:

Issue	Action

Costs for making non-patented data and code FAIR	• Fees associated with publication of scientific articles: to be determined • Project website operation: to be determined • Data archiving at zenodo.org, OpenAIRE, Dataverse.org: free of charge • Copyright licensing with Creative Commons/Apache License: free of charge
Partner Responsibilities	Every partner is responsible for the data they produce.
Long-term Preservation	Data preservation of at least 5 years after the project. Long-term preservation of code and open datasets will be provided and associated costs covered by a selected disciplinary repository. Proprietary and sensitive datasets will remain the property of their owners.

Table 5: Allocation of Resources - Issues and Actions

Data Security

Security scope in the context of CoDiet project

CoDiet will use well established platforms that already adopt state-of-the-art security, authentication and authorization mechanisms. In order to gain access to the data in the repositories, partners and third-parties will be given access to these resources via the already existing processes foreseen by these trusted repository services, such as zenodo.org, Dataverse.org and OpenAIRE. These systems have already provisioned for data security procedures since data files and file metadata are stored in multiple online and independent replicas. Besides that, these trusted repository services allow for long-term preservation and curation of the data, since there exist procedures for migrating the data to appropriate and suitable repositories in the unlikely event that these trusted services need to shut down.

We will make use of the data storage infrastructure of an alliance of Czech Institutions, known as CESNET e-infrastructure (https://du.cesnet.cz/). These operate 4 geographically distributed data centers with over 35 PB of capacity, compliant with EN ISO/IEC 27001:2014 security standard. Additionally, regarding the access to the data, the following measures will be put in place to ensure the secure access to the data:

1. Roles & Permissions: For data security, the dataset providers will ensure that access to the data will be granted through an authorised user management system. This system will allow access to the data with different rights depending on the role of the user, so that data minimization compliance from article 5 of GDPR can be achieved (granted access users will have access to only the specific dataset they need to access) and the different access rights ensure the datasets security. This may also apply for the cases when users are granted access to the data through the premises of the dataset providers (e.g., virtual machines with limited administrative roles).

    - Users of the Virtual organisation will be created and managed by the Czech Technical University, upon a verification by a trusted party (i.e., PI of a partner institution in CoDiet). CESNET is entitled to require a proof that the user is entitled to access data storage facilities with respect to Section 1 of these Terms of Service on an annual basis.

2. Encryption: Access to the data will be performed over the latest secure data access mechanisms such as Virtual Private Network(VPN) connections, Secure Shell (SSH) connections, HTTPS (secure web) access to the data and multi-factor authentication mechanisms. CESNET supports the concurrent use of a variety of secure protocols: • NFSv4 protocol is available only for clients using Linux, for Windows users is available only commercial client NFS Maestro. To ensure strong user authentication we are using NFSv4 over Kerberos protocol. NFS is mostly intended for experienced users (because of the complex initial configuration) using Linux.

    - FTPS protocol is FTP which uses TLS protocol to ensure data encryption. FTP is particularly suitable for transferring large files and for the use on MS Windows, e.g., via Total Commander (https://www.ghisler.com/). On the other hand the FTP protocol cannot preserve original ownership and modification time of the files.

- Like the FTPS protocol, the protocols SCP and SFTP are designed for large files. Unlike FTP(S), it hasn't problem with the transmission of other file information

  (permissions, modification times, etc.). As in FTPS, transmission is encrypted. FTP is particularly suitable for transferring large files and for the use on MS Windows, e.g., WinSCP (https://winscp.net/eng/download.php).

- All transfers utilize SSH tunnelling. All users have individual, password protected SSH keys. Users are obliged to use a non-trivial password. Non-trivial password is such that it is impossible to derive from known information about the user, especially is not a person, animal or thing name or a simple combination of those. DS administrators are entitled to perform testing for trivial passwords.

3. Security - Backups: Finally, the dataset providers have already established mechanisms within their premises to ensure the security of the data by creating daily or weekly backups.

   - CoDiet would have a dedicated Virtual organisation within the CESNET infrastructure, isolating it from other projects, even when those other projects have super-user privileges.

   - All data are backed up on tape-based storage (with longer recall times). Other than tape storage, every single file on file system has a storing period of 365 calendar days. CESNET will notify users at least two weeks before exceeding storing period for backup data.

All users are obliged to notify via e-mail address support(at)cesnet.cz if the is aware or suspecting that the infrastructure has been compromised, misused, access passwords have been disclosed, or in any case of other events which can indicate a security incident, such as strange account behaviour, appearance or disappearance of files and so on.

Data protection and GDPR compliance

The processing of any personal data will be conducted solely from CoDiet consortium members and possible Data Sharing Agreements may be signed between each participating member and the data providers before being granted access to the data.

In this context, CoDiet consortium members will pay particular attention to the protection of any personal data during processing, to maximise its utility while complying with applicable legislations of personal data protection. CoDiet project will ensure that no data will be shared and published without ensuring its legal compliance as well as estimating any legal risks. Additionally, the dataset providers have already established the appropriate infrastructure for ensuring the datasets security (backups).

All use case partners have legal departments that will work on the legal framework for the use of the data. The consortium members agree that any Background, Results, Confidential Information and/or any and all data and/or information that is provided, disclosed, or otherwise made available between the Parties during the implementation of the Action and/or for any Exploitation activities (Shared Information), shall not include personal data as defined by Article 2, Section (a) of the Data Protection Directive (95/46/EEC) and applicable local implementing local legislation; or, as from May, 25th 2018, Article 4 of the General Data Protection Regulation(GDPR). That means that data do not include any information relating to an identified or identifiable natural person, where an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier.

Accordingly, each Party will ensure that all data and information contained in Shared Information is anonymized such that it is no longer personal data, prior to providing the Shared Information to such other Parties. Each Party who provides or otherwise makes Shared Information available to any other Party (”Contributor”) represents that, as per applicable Data Protection Legislation: (i) it has the authority to disclose the Shared Information, if any, which it provides to the Parties under this agreement; (ii) where legally required and relevant, it has a legal ground to provide the Shared Information; (iii) there is no restriction in place that would prevent any such other Party from using the Shared Information for the purpose of this Action and the exploitation thereof and iv) takes into account the existence of appropriate safeguards, which may include encryption or anonymisation.

All personal data collected such as the lipidomics and metabole data is sufficiently protected. The metabole and lipidomics datasets do not provide sufficient information to identify a person. For the case of genomics data, three levels of protection are available. Zero level requires full anonymization of the data, whereas the second level allows for sharing pseudoanonymized data. However, for this type of data, we may pseudoanonymize inside the project, only limited people can access the keys (the data collectors and not the consortium on its whole), and therefore data coupling is available only to the data collectors, in line with the GDPR legislations and the contracts signed with the subjects of research.

GDPR Procedure

In the context of CoDiet project, consortium partners have legal departments that work on the legal framework of the data, in such way that appropriate technical and organisational measures, as well as, safeguards, are deployed throughout the personal data lifecycle, ensuring compliance with the General Data Protection Regulation and that any data disclosure between the parties shall not include personal data that allow for person identification and data coupling. A project DPO has been appointed and specific guidance on the data quality requirements and CoDiet’s practices has been circulated. In addition the processing of the data will be conducted using high security mechanisms as aforementioned and upon specific regulated procedure, which we describe in the following section (section 7).

Ethics

Ethics scope within CoDiet Project

Recently, the General Data Protection Regulation has introduced a set of proposals for regulation, namely the European Data Act and the European Act for AI as part of the European Digital Strategy objectives. The implementation of this strategy at any given use case, especially in the domain of research, poses the requirement for a balancing test between the protection of any personal data used and the free movement of such data. Therefore, it is crystal clear that a set of rules, means as well as a set of processes to share data is quite imperative. In this section, we highlight the ethics blueprint of CoDiet, by presenting the wider outline, the wider regulatory sandbox of means, rules and processes, within which the processing of the data is being conducted in order to address indeed the ethics requirements.

Ethics Requirements

CoDiet will comply with the ethical principle throughout the Europe. An Ethics Risk Assessment, which has been filled in by each one of the WP leaders for each one of the deliverables, has been conducted in order to assess the risks derived from the activities of each one of the deliverables. CoDiet will run two human volunteer studies in Ireland, UK, Spain and Greece. The ethical rules of each country will be followed. In addition to collecting health-related data from the participants using questionnaires and various other instruments, data on health incidents will also be collected for the participants (with the periodicity agreed on by the DSMB/Stbe documented based on the participants' clinical records).

Since these are data related to personal health and subject to the legislation on personal data protection, the informed consent of the participant must be obtained, not only to allow the questionnaires and conclusions to be used, but also to give express consent for the participant's clinical records to be included. This point is also specified in the informed consent form available for the general study, which must be completed by all participants from each node. Completion of an additional informed consent form must be requested for the genomic study. The patient must sign the informed consent form(s) in order to participate in the study but can withdraw his or her consent at any time. Each centre is responsible for obtaining authorisation from the Ethics Committee for the participants it recruits, and for safeguarding the signed informed consent forms under appropriate conditions of security and confidentiality.

Since various health questionnaires are also being administered during the study, including the collection of analytical data, data on related illnesses and risk factors, and data on lifestyle variables, it is not possible to make the data generated accessible to the general public. The patient is informed that his or her data will be treated in a confidential manner and that it will be deposited in a database registered with each countries' Data Protection Agency.

Each PI from a participant recruitment centre is responsible for overseeing protection of the data contained in any database or data file registered under his or her name. According to the EU guidance, the participant must be informed regarding the use of his or her data and may then either grant or deny consent for such use. The two new studies that will be undertaken as part of

CoDiet (WP2 and WP5) will have ethical permission granted within each of the country's recruitment takes place in. This will collect genomic, metabolomic and personal data.

All data for analysis will be fully anonymised. The link back to the volunteer will be kept at the unit where the research is been undertaken, The "omic" data, genomic and metabolomic will be used for profiling only and not to assess any risk of disease. All data will be analysed anonymously and will not be linked to the individual data.

The informed consent forms for the CoDiet studies will inform the participant that their data given or generated as part of these studies will be shared with other researchers in CoDiet (a CoDiet researcher is defined as one who is a signatory to the CoDiet a collaboration). Any such data sharing can only take place under conditions where access to the data is controlled, after a researcher has been granted permission by the CoDiet ethical steering Committee. Volunteers be asked to consent to allowing their anonymised data to be stored on research accessible data bases to allow the scientific community access to this rich data sources.

Moreover, CoDiet directly tackles several ethical issues in the use of AI. Technical advances sought and normative/organisational measures suggested will contribute towards improving AI fairness and AI transparency. The partners in the CoDiet project are fully aware of the ethical implications of the proposed research and respect the ethical rules and standards reflected in the Charter of Fundamental Rights of the European Union and in all recent relevant policy, regulation and legal initiatives. The European Union has established a set of principles that should govern all the AI Systems developed through EU-funded research programs. Hereby, we present these principles and discuss how CoDiet project addresses each one of those principal requirements.

Our philosophy is not only to ensure that our community develops best practices and resources for the incorporation of fairness criteria by practitioners of AI and data analytics, but also that it facilitates a better informed, wider engaged, and more equitable participation and impact both for such practitioners and potential users of the AI applications. Our comprehensive understanding of the fairness literature confirms this as a necessity: fairness is not a single focused quantity, but exhibits many potential often in conflict possible means of measurement, and so fairness conscious AI inherently requires transparency in obtaining reliable information on the fairness impact of different design measures, in order to empower the user to make decisions appropriately as based on their own moral values and the surrounding legal requirements.

The EU's Ethical Guidelines for AI systems states seven key dimensions to be evaluated and audited by a cross-disciplinary team, namely (a) human agency and oversight, (b) technical robustness and safety, (c) privacy and data governance, (d) transparency, (e) diversity, non-discrimination and fairness, (f) environmental and societal well-being and, (f) accountability. CoDiet's work directly addresses challenges in oversight, safety, transparency, fairness, and accountability.

- Human Agency and oversight - *AI systems must support human autonomy and decision-making, enabling users to make informed autonomous decisions regarding the AI systems* : The system and techniques developed through CoDiet are not aiming at replacing the human decision-making processes, but to enhance them, allowing for human agency and oversight of the system. CoDiet aims to ensure that the tools that are developed to enhance understanding of NCD risk, dietary monitoring and enhance personalised nutrition will be relevant to all sessions of society across Europe. It is recognised that there are health inequalities based on a range of determinant factors such as socioeconomic status, ethnic and cultural background, access to education, and sex/gender identities.

- Privacy and data governance - *AI systems must guarantee privacy and data protection throughout the system's lifecycle* : The system and techniques developed through CoDiet will use state-of-the-art privacy and security techniques to ensure data security and integrity.

- Transparency - *All datasets and processes associated with AI decisions must be well communicated and appropriately documented* : Any benchmarking datasets along with the respective code of the developed fairness-enhancing AI models will be thoroughly documented.

- Fairness, diversity and non-discrimination - *Best possible efforts should be made to avoid unfair bias* : The AI system built within CoDiet should provide transparency in obtaining any reliable information on the fairness impact of different design measures, in order to empower the user to make decisions appropriately.

- Societal and environmental well-being - *The impact of the developed and/or used AI system/technique on the individual, society and environment must be carefully evaluated and any possible risk of harm must be avoided* : CoDiet project has established a set of precaution measures such as data minimization, data anonymization and purpose limitation steps, to ensure that the data will be used only for the purpose it is intended to. Producing targeted individualized recommendations that are sensitive to drivers of dietary change in vulnerable populations, will lead to improved awareness of the need for a healthier diet.

- Accountability - *Requires that the actors involved in their development or operation take responsibility for the way that these applications function and for the resulting consequences* : CoDiet project has considered a role-and-permission system to ensure that the actors involved in the development process of those systems, as well as, their operation are accountable for the way these applications function and for their resulting consequences.

- Technical Robustness and Safety - *AI systems must maintain their robustness, security and safety throughout their entire existence* : The system and techniques developed through CoDiet, apart from the privacy and security measures considered, will also focus on the robustness of the results provided through the developed system.

The CoDiet consortium confirm that for any applicable ethics issue, the guidance provided in the European Commission Ethics Self-Assessment Guidelines will be rigorously followed. All the members of the CoDiet Consortium do apply a high level of GDPR awareness and performance rooted to the internal policies and safeguards in place. At a higher level of abstraction, the general rules that should be agreed provide the following obligations on behalf of the partners. In processing personal data pursuant to the Consortium Agreement, each member shall:

- process, or permit to be processed, personal data only for the purposes of the performance of this Consortium Agreement and under a solid legal basis (Data Sharing Agreement);

- ensure that personnel are subject to an obligation of confidentiality in respect of the processing of personal data under this Consortium Agreement;

- ensure that appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data;

- not disclose or transfer personal data to any third-party other than where strictly necessary for the purposes of the performance of this Consortium Agreement;

notify the other members of any security incidents, events, weaknesses, data breaches or suspected data breaches impacting or possibly impacting the security of personal data; • be individually responsible for its own processing of personal data pursuant to and in connection with this Consortium Agreement.

Regulatory Sandbox for AI

Hereby, we present the outline of the process required for processing the data for each one of the use case partners, in order to be compliant with the existing EU ethics and the data protection regulations in the context of the CoDiet project.

In the European Act for Artificial Intelligence (Artificial Intelligence Act Proposal), Article 54 states that AI regulatory sandboxes: 'shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time'. For this purpose, since the use case partners are the data providers as well, and the data are being stored into their premises, in this report, the regulatory sandbox for AI needs to be outlined on a per-use case basis.

Involvement of non-EU countries: Since some of the CoDiet consortium members originate from countries outside the EU, the design principle of the procedures for data processing within Regulatory Sandboxes for AI in CoDiet has carefully addressed this information in order to be compliant with the EU GDPR regulations as well as to carefully address how data are transferred (or not) to non-EU countries. The data is only being processed within the justified interests of the project, fairly and lawfully, for a specific and legitimate purpose and only the data necessary to achieve this purpose will be processed, with personal data safety and security measures considered and upon authorization. The GDPR states that ”you do not actually have to 'send' the data to a non-EU country for these provisions to apply; if one of your partners or service providers is located outside the EU and is able to access the personal data you have collected, this amounts to a 'data transfer' in the context of the GDPR. In order to be compliant and lawful, we hereby state that our partners and beneficiaries (ICL and Technion, which originate from United Kingdom and Israel respectively) have an 'adequacy determination' as defined by the European Commission and ensure the same level of data protection as is required under the EU law.

Participants which are established in a non-EU country (if any) undertake to comply with their obligations under the Agreement and:

- to respect general principles (including fundamental rights, values and ethical principles, environmental and labour standards, rules on classified information, intellectual property rights, visibility of funding and protection of personal data)

- for the submission of certificates under Article 24: to use qualified external auditors which are independent and comply with comparable standards as those set out in EU Directive

2006/43/EC (Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts or similar national regulations (OJ L 157, 9.6.2006, p. 87))

- for the controls under Article 25: to allow for checks, reviews, audits and investigations

(including on-the-spot checks, visits and inspections) by the bodies mentioned in that Article (e.g., granting authority, OLAF, Court of Auditors (ECA), etc.). Special rules on dispute settlement apply (See Consortium's Agreement Data Sheet Point 5).

Conclusions

This report describes the data management lifecycle for the data to be processed and generated in the context of the CoDiet project. One of the main objectives of the project is the pursuit of making the research data findable, accessible, interoperable and re-usable (be in line with the FAIR principles). This report focuses on presenting the handling of the research data during and after the end of the project. Regarding the data itself, this report presents the methods of how the data must be processed or/and generated, such as, which methodology and standards should be applied, whether data will be shared through public repositories and being open accessed, , or how data will be protected from the open access, as well as, how the data will be curated and preserved (including after the end of the project). The results will be published in open access papers. However, in the case of clinical studies generated by CoDiet project, access to the data published in these papers should be protected and requested to the corresponding committee by a formal letter.

The current deliverable is the initial version of the CoDiet project Data Management Plan, which will be treated as a live document from hereon. Within the duration of the project and as the project evolves, this report will be revised and updated providing further details and - if possibly needed- amendments to ensure the compliance with all data-related aspects of the CoDiet project and in accordance with all the recent guiding principles of FAIRness, data security and privacy, European Act for Data and European Act for AI.

Annex A - Agreements

Terms and conditions for the access to the CESNET e-infrastructure

(further referred to as the "Terms and conditions")

Preamble

1. **Description and purpose of the CESNET e-infrastructure**
    1. The CESNET e-infrastructure (further in this text referred to as 'the Infrastructure') is a research infrastructure as defined by Act no. 130/2002 Coll., on the Support of Research and Development from the Public Funds, as amended; which provides its users a unique portfolio of information and communication technology facilities.
    2. The Infrastructure is operated by CESNET, association of legal entities, with registered office at Zikova 1903/4, 160 00 Prague 6, Id. No.: 63839172, recorded in the Association Registry maintained by the Municipal Court in Prague under file no. L 58848 (further in the text referred to as the 'Association'). The Association is a 'research organisation in accordance with Act no. 130/2002 Coll., on the Support of Research and Development from the Public Funds and on amendments to some related acts, as amended; or a 'research and knowledge dissemination organisation' or a 'research organisation in accordance with the Communication from the Commission – Framework for State aid for research and development and innovation (2014/C 198/01), section 15 ee).
    3. The Infrastructure is non-public and is not operated primarily to generate profit.
    4. The services of the Infrastructure may only be provided by the Association or an authorised Contractual Partner ('the Partner').

Access Policy (AP)

1. **User access to the Infrastructure**

1. Users of the Infrastructure in particular include entities:

- Focusing primarily on research, experimental development and innovation, including the application of their results in the practice;
- Focusing primarily on education and dissemination of education, culture and prosperity;
- Public administration bodies and local authorities; ▪ Other entities with activities in the public interest.

2. Other entities may only be granted access to the Infrastructure for the purpose of scientific, research, educational and innovation projects. In that

case, the entity concerned shall ensure that the Infrastructure is used in relation to such activities only.

3. Infrastructure users gain access to unique tools including a connection to similar infrastructures abroad.
4. The access to the Infrastructure cannot be legally claimed and the decision of the Association on whether to grant the access is definite.

Acceptable Use Policy (AUP)

1. Rights and obligations of Infrastructure users
    1. Users may use the Infrastructure for activities in compliance with this Acceptable Use Policy, in good manners, respecting the needs of other users and savings the Infrastructure resources. When using the Infrastructure, the users are obliged to adhere to laws and other legal regulations which constitute the Czech legal order. This provision does not affect the right of the Association for damages, nor does it prejudice the Infrastructure users' civil or criminal liability.
    2. No user may use the Infrastructure for activities which:
        1. Constitute an illegal use, interference, change of computer systems, their parts, information or data carriers;
        2. Violate intellectual property rights;
        3. Have an adverse effect operation of the Infrastructure or its constituent facilities, prevent other users from accessing such facilities, threaten the operation of the Infrastructure, or excessively reduce its performance.
    3. Users shall request a prior Association's consent for the Infrastructure to be used by other entities.
    4. Users are obliged to ensure that none of the devices in their competence (owned, leased, borrowed, operated by user, etc.) shall use the Infrastructure for purposes violating this *Terms and conditions* or the contract based on which the uses the Infrastructure services.
    5. Where an user uses the Infrastructure or information about it in violation of this Acceptable Use Policy or the contract based on which the user obtained access to the Infrastructure, or where he help a third party to such an use, both by wilful act or by neglect, the user shall be liable to pay the damages to the Association.
    6. The terms and conditions for the use of individual Infrastructure services are available at https://www.cesnet.cz/services/?lang=en.
2. **Rights and obligations of the Association**
    1. The Association/Partner may restrict/suspend user's access to the Infrastructure if the user has violated his obligations stated above in art. 2 or 3, or other obligations agreed upon in the relevant contract based on which the user had been granted access to the Infrastructure.

2. The Association may discontinue the provision of individual service/services as agreed upon in the relevant contract provided it is not possible to provide it/them as a result of an extraordinary unpredictable and insurmountable obstacle occurred irrespective of Association's will (for instance, but not limited to, force majeure). The

   Association is obliged to notify the users of any such discontinuation. In such cases, however, the Association is not liable for any potential damage incurred.

Final provisions

1. **Final provisions**
   1. The Association reserves the right to change the *Terms and conditions*, in which case the Association shall publish the updated Terms and conditions on the www.cesnet.cz website and shall notify the users of the changes in advance. Users who disagree with the updated *Terms and conditions* may terminate the relevant contract granting the access to the Infrastructure in accordance with the terms and conditions therein stipulated.
   2. This Policy is valid upon its signing by the Association director and effective as of 1 January 2017, fully substituting The Access Policy (AP) from 14 November 2011, including Appendices: 1) Acceptable Use Policy (AUP) and 2) Technical and economic terms and conditions.

In Prague on 7th December 2016

Ing. Jan Gruntorád, CSc. director