Call: H2020-SC1-2018-Single-Stage-RTD
Topic: SC1-BHC-09-2018
Funding Scheme: Research & Innovation Action (RIA)

Deliverable No. 12.2

Data Management Plan (DMP)

Grant Agreement no.:	825828
Project Title:	**EXpanding Platforms for Efficacious mRNA Therapeutics**
Contractual Submission Date:	29/02/2020
Actual Submission Date:	28/02/2020
Responsible partner:	P1 - UMCU

EXPERT has received Funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 825828.

Grant agreement no.	825828
Project full title	EXPERT – EXpanding Platforms for Efficacious mRNA Therapeutics

Deliverable number	**D12.2**
Deliverable title	**Data Management Plan (DMP)**
Type	ORDP (Open Research Data Pilot)
Dissemination level	PU
Work package number	WP12
Work package leader	P1 - UMCU
Author(s)	Prof. Raymond Schiffelers (UMCU) with contributions from other consortium partners
Keywords	Data management, FAIR principles, Open Access, Interoperability, Re-use.

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825828.

Table of Contents

1 Executive Summary 4
2 Introduction 4
3 Data Summary 4
4 FAIR data 14
4.1 Making data findable, including provision for metadata 14
4.2 Making data openly accessible 16
4.3 Making data interoperable 16
4.4 Increase data re-use 18
5 Allocation of resources 18
6 Data security 18
7 Ethical aspects 19
8 Other issues 19
9 Deviations 19
10 References 19

1 Executive Summary

This report is the first deliverable of Task 12 and describes the initial Data Management Plan (DMP) for the EXPERT project, funded by the EU Horizon 2020 Program (825828). The purpose of the DMP is to provide an overview of data types and datasets collected and generated by the project and to define the EXPERT consortium's data management policy.
The EXPERT DMP follows the structure of the Horizon 2020 DMP template[i]. It reflects the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.
This initial version of the DMP defines the general policy and approach to data management in EXPERT that handles data management on the administrative and technical level. This includes for example topics like data and meta-data collection, publication and deposition of open data, the data repository infrastructure and compliance to the Open Access Infrastructure for Research in Europe (OpenAIRE).
It furthermore summarizes the guidelines on FAIR (Findable, Accessible, Interoperable and Reusable) data management[ii]. The DMP will evolve during the lifespan of the project. Next versions will refine and enhance policy aspects and will go into more detail regarding the datasets collected and produced by the EXPERT project.
This document describes the situation at the coordinating institution UMCU. Other consortium partners can use the UMCU cloud infrastructure as it represents state-of-the-art. Future DMP versions will include the actual policy adopted by each individual partner.

2 Introduction

The aim of the Data Management Plan is to ensure that good scientific practice is followed according to the FAIR principles; data should be made 'Findable, Accessible, Interoperable and Re-usable'. The Data Management Plan is an integral part of the research protocol and describes a standardized way how research data are collected, how data are used and stored during research and how data are accessible for others after the research has been completed.
In all honesty, for the majority of data generated in EXPERT we do not, at this time, foresee that reuse will occur. This is because we generate relatively few data points that are influenced by a near endless number of biological variables. We try to capture this variability by working according to SOPs. Still, very rarely, the conditions chosen are relevant for reuse. As a result, it is usually far easier to redo the experiment rather than reuse the data. Nevertheless, we do store the data with relevant descriptive denominators to enable reuse. On an aggregate level the information is suitable for reuse but these are part of the common scientific exchange of information through publications in Open Access journals.
The principle investigator or the researcher is required to use this plan to describe their data management procedures. For each partner institute and type of experiments Annexes are added to this DMP to ensure that we capture all forms of data generation and storage.

3 Data Summary

Within EXPERT we collect data:

- On mRNAs

We perform mRNA synthesis in the facilities of eTheRNA and from there it will be distributed to all partners. This will ensure that all partners work with the same material in their work, and handle and store it in an identical manner for this eTheRNA provides basic characterization of the mRNA

material. The mRNA length, concentration and purity are evaluated with capillary gel electrophoresis. mRNA concentration is measured spectrophotometrically. For the clinical study an additional purification step by HPLC is customary. The generated mRNAs contain a 5' cap and 3' poly A-tail that leads to high RNA stability and increased protein expression in transfected cells. mRNA is stored at −20°C in small aliquots. Data are available to all partners and Advisory Board via the project webpage with continuous consortium annotation.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The data are expected to have limited value for reuse.

- On components of nanomedicines

Targeting ligand. We will screen targeting ligands using the ASSET system or using generated nanobodies/monoclonal antibodies to generate data that describe the physicochemical characterization of the ligand-to-surface coupling. We measure affinity with surface plasmon resonance, orientation of coupling, stability of coupling and ligand density, and the effects of the ligand on performance: cell binding using fluorescence assisted cell-sorting (FACS), internalization (microscopy) and functional mRNA expression in subsets of cells.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The only exception are the microscopy data on intracellular uptake and trafficking. These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. ImageJ is an open source image processing program designed for scientific multidimensional images. ImageJ is highly extensible, with thousands of plugins and scripts for performing a wide variety of tasks, and a large user community[iii]*. The data are expected to have limited value for reuse.*

Nanomedicines. We investigate three types of nanomedicines. The established carriers (LNPs) have been clinically tested with siRNAs, primarily for hepatic and oncology applications. In EXPERT they are adapted for mRNA targeting. The emerging carriers are cell-penetrating peptides. CPP nanoparticles have been widely tested pre-clinically for delivery of various nucleic acid species. Having full control over their chemical synthesis, CPPs offer endless possibilities to add responsive and modulating properties that potentially allows them to outperform established carriers. The exploratory carriers are biologically produced extracellular vesicles. Remarkable pre-clinical results have been obtained with these particles that are at the same time ill-understood. These vesicles also present challenges with regard to reproducible production. The vesicles will be investigated to learn how biological systems transfer mRNA and to feed the critical molecules in this process into the design of synthetic nanocarriers. For the nanomedicines we determine the chemical composition. Loaded with the mRNA the full characterization takes place.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The data are expected to have value for reuse on an aggregate level to facilitate choice for nanomedicines for other projects.

- On the physiochemical characterization of nanomedicines with a mRNA payload

For the nanomedicines with the key physical parameters will include size, size distribution and aggregation propensity, analyzed by batch dynamic light scattering (DLS) and nanoparticle tracking analysis (NTA), and further verified by field flow fractionation and static light scattering (FFF-MALS). Morphology is verified by electron microscopy. Compositional analysis, i.e. targeted quantification of the lipid nanocarrier constituents and the CPPs, will be done by LC-QqQ-MS/MS. Endotoxin analysis will be done both by standard LAL enzymatic assay and a novel LC-MS/MS method. Quantification of protein will be done by enzymatic cleavage and subsequent LC-QqQ-MS/MS analyses as regularly performed in bottom-up proteomics; these proteins include targeting ligands and the target proteins translated from the mRNA payloads, and – if deemed applicable – selected proteins associating with the particle surface. Screening for impurities will be done by LC-DAD-QTOF-MS at all stages, and suspected impurities will be subject for targeted analysis according to regulatory requirements for

pharmaceuticals. Verification of mRNA integrity will be purchased as a single for-fee service by an external supplier (Axolabs) on the final clinical candidate. Overall these measurements provide the full profile for the Investigator's Brochure: particle size, polydispersity index, encapsulation efficiency, zeta potential, pKa, physical stability, chemical stability, morphology, (photo)stability, surface area, resistance to sterilization and mRNA integrity.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The only exception are the electron microscopy data on morphology. These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. The data are expected to have limited value for reuse.

- On the safety aspects of mRNA-nanomedicines in vitro

The SOPs we use within EXPERT on safety of nanomedicines closely follow the Nano Characterization Laboratory guidance[iv] that has been adopted by the European NanoCharacterization Laboratory EUNCL[v]. Cell metabolism will be assessed by the (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) tetrazolium reduction assay that is converted by cells in an energy dependent way. It provides a quick overview of cellular metabolic state. Cell proliferation is assessed by flow cytometry through fluorescent label dilution. The assay detects growth arrest. Cytotoxicity is detected by lactate dehydrogenase release and measures cell membrane integrity loss. Hematocompatibility is tested on blood cells in hemolytic assays, leukocyte proliferation assays, cytokine release assays in whole blood and monocyte/macrophage cultures, FACS analysis of cell surface activation markers. Effects on the coagulation system will be studied in coagulation assays and platelet activation/aggregation assays.
The immune response will be tested in vitro in mouse cell lines (TK-1), primary human PBMCs and in vivo in relevant mouse models and healthy mice. We will probe for cytokine induction using a multiplex ELISA approach, interferon responsive genes, complement activation (looking at the final reactive toxins (C5a, C5b-S)) and for activating lymphocytes (examining early and late activation markers).
Finally, we are actively involved in adoption of High Content Screening technologies, Flow cytometry, and imaging facilities as part of EUNCL
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The only exception are the high content screening microscopy data. These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. The data are expected to have value for reuse on an aggregate level to facilitate choice for nanomedicines for other projects.

- On the performance of mRNA-nanomedicines in vitro

For initial in vitro testing of mRNA delivery efficacy of the different nanocarriers, reporter assays will be utilized. Reporter assays will be based on expression of (1) Cre-protein or (2) Firefly luciferase. Using cells that stably express Cre-responsive reporter constructs, Cre-mRNA delivery and translation ensures a color change from red to green fluorescence can be visualized using fluorescence microscopy. Luciferase emits bioluminescence when it converts its substrate, luciferin. To be able to compare different platforms and ensure standardization, SOPs have been established (D8.1)
To study differences in uptake efficiency, binding and internalization of fluorescently labelled nanocarriers by different target cells will be evaluated using flow cytometry and confocal microscopy at 4°C and 37°C, respectively. Intracellular trafficking will be examined by analysing co-localization with markers for early (EEA), late endosomes/MVBs and lysosomes (Rab7, LAMP1), as a function of time, live cell confocal screening microscopy. As late endosomes are known to exchange materials with the Golgi (C6-NBD-ceramide) and endoplasmatic reticulum (ER) (ER-Tracker), we will also analyse co-localization of nanocarriers with these organelles. To evaluate whether endosomal escape and/or recycling to the extracellular milieu forms a hurdle for mRNA delivery, mRNA delivery experiments will be performed in the presence/absence of endosomolytic agents (e.g. chloroquine) and Niemann Pick C1 protein inhibitors. Internalization mechanisms for nanocarriers will be further

studied using inhibitors of specific endocytic pathways. Both chemical inhibition as well as RNAi approaches will be employed, targeting regulators of e.g. clathrin and caveolae-mediated endocytosis and macropinocytosis. To determine whether internalization route affects transfection efficiency, we will look at effects on both nanocarrier uptake and mRNA delivery activity.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. A significant part of the data are the microscopy data on cell color changes and intracellular trafficking These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. The data are expected to have limited value for reuse.

- On the scale-up and GMP-manufacturability of mRNA-nanomedicines

Once all raw materials have been selected and suppliers contacted, raw materials QC will be defined and established. Those raw materials described in pharmacopeia will be analyzed according to their monography. For those raw materials not included in the pharmacopeia, the QC will be defined according to the Certificate of Analysis (CoA) and quality attributes. In case that the analyses are not validated, instead of fully validated techniques, in some cases, suitability assessment of the analytical method will be performed.
For laboratory experiments, EXPERT has chosen to employ microfluidic manufacture at all its partner research sites. A risk-based method to verify and demonstrate that the process can operate within the predefined specified parameters consistently and produce nanomaterial which meets all its critical quality attributes (CQAs) and control strategy requirements, will be applied. Quality attributes of incoming materials, in-process and finished products will be tested and recorded during laboratory trials, in order to enable continuous process verification.
Starting from laboratory data and design space evaluation, the final manufacturing method for scaling-up will be defined. CQA, in process control (IPC) and range for critical process parameter (CPP) will be stated on the basis of laboratory trials. A dimensional analysis coupled with design of experiments will be implemented for efficient transfer of optimal operating conditions from small to large scale of production. A detailed protocol for demonstration by demo batches will be prepared. According to this data, the scale-up of the manufacturing process will be carried out. CIDETEC's Pilot plant have two different microfluidic reactors, one working at lab scale and the second one working at pilot scale. The sterilization methodology will be developed adequate to the production scale.
QC of the final product will be established and based on a risk analysis, specifications for the release of the batch will be defined. For those tests not available under GMP in the Pilot Plant, specialized pharmaceuticals laboratories will be selected and subcontracted.
For pre-clinical batches will be manufactured. For these Master Batch Record, critical parameters, in-process controls, and testing on the finished products, will be obtained. A final report summarizing data collected during manufacturing, IPC and final test will be produced for the Investigator's Brochure. The relationship between the process inputs (material attributes and process parameters) and the critical quality attributes will be described within defined design space. In this case, the raw materials employed will be those selected for GMP manufacturing. This will ensure good reproducibility in the future GMP production.
GMP production of the selected mRNA-LNPs will preceded by the validation of the sterilization procedure. In case the sterility is carried out by filtration, initially the process will be validated with media fill. Once the process has been already validated sufficient amount of GMP batches will be produced for stability studies and clinical testing.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The data are expected to have value for reuse on an aggregate level to facilitate choice for nanomedicines for other projects.

- On the safety aspects of mRNA-nanomedicines in vivo

RNA nanocarriers that have shown appropriate in vitro and in vivo compatibility will be subsequently evaluated for their tissue distribution profile in vivo to determine organs at risk. In healthy mice, the

tissue distribution profile is assessed by measuring fluorescent carrier and payload in a Pearl® Trilogy/IVIS small animal imaging system. When 95% of the label is cleared form the vasculature, organs are harvested and imaged. This analysis shows the organs that accumulate substantial levels of carrier and or mRNA and are as such most sensitive for side effects. Tissue distribution will also be investigated with the complementary technique of MALDI mass spectrometry imaging (MSI) by MALDI-FT-ICR-MS and – if found suited – MALDI-TOF/TOF-MS, giving label-free distribution data on the nanocarrier components (SINTEF). Primary emphasis will be on the nanobodies and selected nanocarrier matrix components (e.g. lipids), whereas direct imaging of the mRNA payload will be attempted. For absolute quantification of key components, LC-MS/MS will be available.
The GLP-Tox study will be performed on the fixed design of the intratumoral immunostimulatory mRNA-LNP. It is customary to perform these studies at CROs to improve efficiency and objectivity. In vivo toxicology studies are intended to assess the onset, severity, and duration of toxic effects, their dose dependency and degree of reversibility. The GLP toxicology study will mimic the dosing regimen for the clinical study. Several routes of exposure (e.g., oral, intravenous, intramuscular, topical, etc) can be accommodated and multiple species (mice & rabbits) are available. These evaluations include clinical chemistry, hematology, urinalysis, histopathology, and toxicokinetics.
Pigs provide a sensitive and quantitative animal model of non-IgE-mediated Type I (pseudoallergic) hypersensitivity reactions (HSRs) caused by liposomal drugs and many other nano-biopharmaceuticals. The rapidly arising and highly reproducible symptoms, including cardiopulmonary, hemodynamic, hematological, blood chemistry and skin changes closely resemble the clinical picture of severe, life-threatening infusion reactions that represents a major barrier to the clinical use of many state-of-art drugs and drug candidates, e.g., PEGylated aptamer (Pegnivacogin). The pig model is being recommended as a preclinical screening test to identify drugs with anaphylactic potential. Mice and rats can serve as less sensitive species to perform an initial study on tolerability.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. The data are expected to have value for reuse on an aggregate level to facilitate choice for nanomedicines for other projects.

- On the performance of mRNA-nanomedicines in vivo

Antitumor efficacy of mRNA-nanomedicines will be addressed as monotherapy in syngeneic and genetic tumor models of different ontogeny and of increasing resistance to immunotherapeutic interventions to evaluate the general applicability and strength of the approach. A thorough assessment of the quantity and quality of the antitumor immune responses and of the mode of action of lead mRNA formulations will be performed.
Finally, we will address the complementarity of our lead mRNA formulations with two types of immune-modulatory agents.
For syngeneic models, a two-sided tumor inoculation will be performed in which one tumor will be treated with the lead mRNA formulation. Tumor growth of both treated and non-treated tumors will be assessed. This allows one to identify abscopal effects in non-treated tumors, indicative of a systemic antitumor immune response. Mice that reject primary tumors upon treatment will be rechallenged with tumors three months post treatment to address the establishment of long-lived memory.
In the genetic breast cancer model, the first emerging mammary tumor will be inoculated with the mRNA formulation. Impact of treatment on primary tumor growth and on the emergence of metastases will be assessed over time.
Syngeneic models: CT26 (subcutaneous): colon carcinoma; high sensitivity to systemic immunotherapies including ICB's B16 (orthotopic model): melanoma; intermediate sensitivity to immunotherapies; resistant to ICB monotherapy 4T1 (orthotopic model): highly metastatic with low sensitivity to immunotherapies; resistant to ICB monotherapy and most vaccine/ICB combos. Genetic models: MMTV-PyMT mice: model of breast cancer metastasis, in which MMTV-LTR is used to drive the expression of mammary gland specific polyomavirus middle T-antigen, leading to a rapid

development of highly metastatic breast cancer. Humanized mice models: RL follicular lymphoma model: Newborn NSG mice (1-2 days of age) are sublethally irradiated and subsequently receive CD34+ human stem cells isolated from HLA-A2 positive cord blood by injection in the liver. Thirteen weeks after stem cell transfer, 2,5 x 106 human RL follicular lymphoma cells are inoculated s.c. into the mice.
Mechanism of Action. The impact of IT treatment on the quantity and phenotype of tumor infiltrating DCs, macrophages, monocytes, neutrophils, NK cells and T cells will be assessed by flow cytometry. This will be complemented by a thorough NanoString expression profiling of tumors enabling to identify gene signatures correlated with anti-tumor efficacy. Induction of T cells recognizing de novo mutated proteins (neo-epitopes) has been associated with improved therapeutic outcome Earlier studies have identified immunogenic neo-epitopes for CT26, B16-F10 and 4T1 tumors, enabling us to address to which extent our IT mRNA formulations have the capacity to evoke neo-epitope specific T cells by ELISPOT on splenic and intra-tumoral T cells. To address the relative contributions of CD4 T cells, CD8 T cells and NK cells in the observed therapeutic responses, these leukocyte subsets will be depleted by antibodies prior to treatment as reported earlier.
In addition, we explore therapeutic efficacy in cardiac ischemia. Advanced preclinical models of cardiac ischemia-reperfusion are available mimicking the consequences of myocardial infarction and intervention. Mice receive a left coronary artery ligation, followed by reperfusion after 60 minutes by releasing the ligature and removal of tubing. Reflow is confirmed by reversed discoloration of the heart. Intervention can be started immediately or up to 1 month after the procedure by intracardiac injection of the VEGF-mRNA nanomedicine. Ejection volume of the heart is assessed by magnetic resonance imaging. At the end of the experiment animals are euthanized, hearts are excised, dehydrated and fixed after which they are embedded. Serial transverse cryosections of 7 µm are cut, base to apex, for histological and immunohistological stainings.
In the UMCU essentially the same model is also available in pigs. Recently a complete visual model description was provided in JoVE. This provides a unique set-up to test our mRNA nanomedicines in a model of human scale. We expect one lead candidate to be tested in this model as we need both good efficacy as well as exquisite safety due to the sensitivity of the pig to nanomedicine-based interventions
MS-imaging) or the efficacy study with. Furthermore, we also have a 14T MRI for animal work which could be used as fMRI system
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. A significant part of the data are the microscopy data on cell color changes and immunohistochemical stainings. These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. In addition, we have the MRI and MSI images. Also these are vendor-specific formats, but are saved in Digital Imaging and Communications in Medicine (DICOM) format. The data are expected to have limited value for reuse but on an aggregate level can help other projects in a choice for mRNA nanomedicines.

- On clinical safety and performance of the lead mRNA nanomedicines

The data generated within the clinical study represent a separate class of data due to the extensive ethical concerns around sensitive patient data. The EMA recently launched a new policy to ensure that data from clinical trails on new drugs become available[vi].
The mRNA formulation with the most promising anti-tumor activity identified in the preclinical work will be tested in patients with Triple Negative Breast Cancer. The proposed study will investigate the safety, tolerability and efficacy of escalating doses of the selected mRNA formulation (3 dose levels with 8 patients each) compared to placebo (n = 8).
Each patient will receive three administrations of the mRNA formulation prior to start of general treatment (surgery or neoadjuvant chemotherapy) separated by a one week (7 days ± 2 days) interval [at day 0 (start of treatment), one week thereafter on day 7 and on day 14]. The last administration will be performed 2 days preoperatively or before the start of neoadjuvant

chemotherapy. As a surrogate for clinical efficacy, the tumor and peripheral blood samples will be analysed for immunological changes.
Three dose levels of the selected mRNA will be investigated. The starting dose will be calculated based on the results of the preclinical toxicity studies. A Data Safety Monitoring Board (DSMB) will be installed to evaluate the safety and tolerability of the mRNA formulation. The decision to open the next dose level will depend on the recommendation of the DSMB.
The immunological changes in the tumor and peripheral blood samples will provide an indication of the efficacy of the mRNA formulation. Tumor slices will be evaluated using immunohistochemistry to assess the immune cell infiltration and gene expression profiling to characterize the immune cell activation.
The proposed clinical design will give a first indication whether the ultimate aim of the consortium to develop an effective off-the-shelf platform-based nanosized delivery system for mRNA can be successfully implemented in the clinic.
When 50% of the study population is expected to have been recruited a "Midterm recruitment report" will be written. The report shall include an overview of recruited subjects by study site, potential recruiting problems and, if applicable, a detailed description of implemented and planned measures to compensate delays in the study subject recruitment.
At the time of expected results posting or for the last months of the project, whichever comes earlier, the report on status of posting results in the study registry (including timelines when final posting of results is scheduled after end of funding period) will be written.
The immunohistological changes at the tumor site (T-cell infiltration), the changes of the immune gene expression profile, the dynamics of the T cell receptor repertoire, at the tumor site and potentially also the immune response against cancer-germline antigens identified in the immune gene expression profile will be studied. In selected patients, the response to neo-epitopes will be investigated.
These are original data that are generated within the project, they are expressed as numbers/graphs that require limited storage. A significant part of the data are the microscopy data on immunohistochemical stainings and analyses of immune cells in clinical samples. These are high resolution images that are captured in microscope specific formats. For analysis we convert to an image J format. In addition, we have the images obtained from patients scans. Also these are vendor-specific formats, but are saved in Digital Imaging and Communications in Medicine (DICOM) format.
Due to privacy concerns the data on the individual patient level will need to be (pseudo)-anonymized. Since the data are of a very small patient population (n=8) the value of the data is limited. Still they will become an integral part of the larger dataset that is acquired during subsequent studies.

In general, the following table provides a minimal set of data to describe the experimental set-up within EXPERT.

Experimental Description of the minimal information on cell-based mRNA experiments
Adapted from the MIARE guidelines.

The 'experimental description' section provides information about the provenance, hypothesis and key factors in an experiment. We define an experiment as one or more assays that answer a biological question using mRNA reagents.

- Biological, technical, and/or theoretical question/hypothesis.
- Primary contact information
- Publication
- Type of mRNA reagent (CV)
- Number of experimental genes targeted for knock down
- Keywords (CV)
- Target organism genus and species (CV)

Experimental Design

The 'experimental design' section describes factors that apply to the entire experiment. This section enables a user to evaluate the context of an experiment.

- Number of replicates
- Description of replicates (CV)
- Assays being conducted (CV)
- Related information e.g., related publications

Sample Description

The 'sample' section is adapted from the MIAME 1.1 standards. A 'sample' represents the biological material (or biomaterial) for which the data is established.

Bio-source Properties

- Organism genus and species (CV)
- Contact details for sample
- Where descriptors are relevant to the particular sample include
  - Animal/plant strain or line (CV)
  - Sex (CV)
  - Age (CV)
  - Development stage (CV)
  - Organism part (tissue) (CV)
  - Cell type (CV)
  - Cell line (CV)
  - Genetic variation (e.g., gene knockout, transgenic variation) (CV)
  - Individual genetic characteristics (e.g., disease alleles, polymorphisms)
  - Disease state or normal (CV)
  - Clinical information available
  - Individual number (for interrelation of the samples in the experiment)

Biomaterial Manipulations

The 'biomaterial manipulations' section is similar to the "sample" sub section from MIAME v1.1. This section covers manipulations such as cell culture conditions.

- growth conditions/ cell culture conditions
  - Manufacturer name
  - Catalogue number
  - Name
  - Percentage serum

- Separation technique (e.g., none, trimming, microdissection, FACS) (CV)

Pretreatment

The 'pretreatment section' contains a description of the conditions applied prior to the delivery of the mRNA reagent.

• Treatment protocol
• Treatment type (CV)
• Compound (CV)

Assay Plate Description

The 'assay plate description' section contains information on the physical assay plate and the layout (plate map) of the reagents on the assay plate.

• Manufacturer
• Catalogue number
• Assay plate type (CV)
• Plate ID
• Assay plate layout (for each position specify the following)
o Position e.g. row, column
o Reagent group and control type (CV)
o Reagent identifier

mRNA Reagent

The 'mRNA reagent' section describes the mRNA reagent, both experimental and control, used in the study.

• mRNA reagent identifier
• Gene accession number and version
• Official gene symbol
• Entrez Gene ID
• mRNA reagent sequence(s)
• Number of mRNA reagent(s) per well
• mRNA reagent type (CV)
• Modifications to mRNA reagent (CV)
• Reference to vector
o Versioning information/release number
o Manufacturer

Delivery

The 'delivery' section requires information relating to the delivery of the mRNA reagent into a particular biological sample.

• Delivery Type, e.g. nanomedicine composition reverse transfection, electroporation, shooting etc (CV)
• Delivery protocol
o Manufacture of nanomedicine Protocol (if applicable)
o Time of nanomedicine formation (if applicable)
• Delivery reagent
o Delivery reagent type (CV)
o Catalogue number of delivery reagent
o Delivery reagent name
o Manufacturer of delivery reagent
• Final concentration or amount of delivery reagent
• Final concentration of mRNA reagent

• Number of cells per well in delivery plate

AssayPlate

The 'assay plate' section contains information on conditions within the plate where the assaying (data acquisition) take place.

• Number of cells per well
• Time to assay point from delivery of mRNA reagent
• Time of exposure of mRNA reagent (for media changes)
• Media changes
o Media composition
o Time of media change
Post treatment

The 'post treatment' section holds a description of the conditions that are applied after the delivery of the mRNA reagent.

• Treatment protocol
• Treatment type (CV)
• Compound (CV)

Assay

An assay is the process where the effects of mRNA expression are measured (data acquisition). The assay section contains a description of the assay performed and key parameters. Because the effect of mRNA can be measured at various stages, e.g. transcriptome and proteome, this section is technology independent.

• Assay description
• Assay reagents
o Assay reagent name
o Assay reagent catalog number
o Assay reagent manufacturer
• Control definition
• Assay protocol
• Instrument
o Instrument name
o Instrument manufacturer
o Instrument catalogue number
o Type of readout (CV)
o Instrument settings

Data Analysis

The 'data analysis' section requires a definition of the transformation, normalisation and scoring procedures applied to the data produced as described in the assay section.

• Details of filtering of data
• Transformation details
• Reference to analysis script
• Analysis software

o Analysis software name and version
o Analysis software manufacturer
• Normalisation method (CV)
o Normalization parameters
o Controls used to normalize data
• Scoring method (CV)
• Quality controls steps (CV)

Data

The 'data' section describes the required quantitative and or qualitative data.

• Quantitative data
o Description of quantified data
o Unprocessed quantified data
o Normalised quantified data
o Scored data
• Qualitative data
o Description of qualitative data
o Qualitative data

4 FAIR data

4.1 Making data findable, including provision for metadata

As outlined above, we expect that there will be relatively limited direct value for the primary data for reuse. This is because we generate relatively few data points that are influenced by a near endless number of biological variables. We try to capture this variability by working according to SOPs. Very rarely, the conditions chosen are relevant for reuse. As a result, it is usually far easier to redo the experiment rather than reuse the data. Nevertheless, we do store the data with relevant descriptive denominators to enable reuse. Over the past years, several 'minimal information' projects have outlined the metadata that are needed to capture the experimental conditions. Relevant projects are for example the Minimal Information About a Cellular Assay (MIACA)[vii]. Also, such projects can be adapted to fit the needs of EXPERT while still retaining the structure of such recommendations. An example is the MIARE project on RNAi experiments[viii]. Metadata will be created according to the standard ontology adopted in the field according to the EMBL-EBI defined ontology and accessible through their Ontology Lookup Service[ix].
The information will be recorded and organized to be machine-findable when reuse is requested. EMBL-EBI have made tools available that allow automated ontology building using OxO, Zooma and Webulous. The OxO tool[x] is a service for finding mappings (or cross-references) between terms from ontologies, vocabularies and coding standards. OxO imports mappings from a variety of sources including the Ontology Lookup Service and a subset of mappings provided by the UMLS[xi]. In addition, the Zooma tool is a tool for mapping free text annotations to ontology term based on a curated repository of annotation knowledge[xii]. Finally, the Webulous server can be used to create spreadsheets with ontology-based data validations. Webulous can process populated templates and transform them into OWL/RDF using design patterns expressed in Ontology Pre-Processor Language[xiii].
At an aggregate level data will be presented in scientific publications which will become part of repositories. Some EXPERT partners like UMCU already operate their own OpenAIRE-compliant repositories, i.e. Utrecht University Repository. Information stored here is searchable via https://www.narcis.nl/ and findable via search engines like Google Scholar. To ensure that the data of all partners are compliant with the OpenAIRE requirements, the default repository of the EXPERT

project for depositing publications, open data and open source software is Zenodo (http://www.zenodo.org). Zenodo is an EC-co-funded, multidisciplinary repository, for publications and data. A DOI is automatically assigned to all Zenodo files, which can be uploaded in any file format. Zenodo allows researchers to deposit both publications and research data, while providing means to link them. Data is stored in the CERN cloud infrastructure. Zenodo is compliant with the open data requirements of Horizon 2020, the EU Research and Innovation funding programme and OpenAIRE. Furthermore, a CLARITY project page (community) has been set up at https://zenodo.org/communities/expert/ for easy upload of project datasets.

The following Table lists the URLs associated with the EXPERT project

Community URLs
Collection URL:
https://zenodo.org/communities/expert/
Above address links directly to EXPERT community collection.
Upload URL:
https://zenodo.org/deposit/new?c=expert
Above address will automatically ensure people who use it will have their record added to the EXPERT community collection.
Curation URL:
https://zenodo.org/communities/expert/curate/
Above address links to EXPERT's private curation URL. You will find all uploads pending curation.
Harvesting URL:
https://zenodo.org/oai2d?verb=ListRecords&set=user-expert&metadataPrefix=oai_dc
Above address links to a OAI-PMH feed, which can be used by other digital repositories to harvest this community.

Scientific Publications

There are two possibilities for a researcher: publishing in green or in gold open access journals. In case of green open access, Zenodo can be chosen by the researcher as a primary repository for self-archiving. This leaves still the possibility for the additional dissemination of the published publication also at academic social network sites like ResearchGate that do not count as suitable open access repository. For finding suitable green open access publishers, researchers are encouraged to consult RoMEO (http://sherpa.ac.uk/romeo), a searchable database of publisher's policies regarding the self-archiving of journal articles on the web and in Open Access repositories.

In case of gold open access, the scientific publisher's modalities for open access (e.g. embargo periods) must allow the researcher to fulfil the EC's open access obligations. Furthermore, the repository used by the scientific publisher should be OpenAIRE-compliant and issue a DOI. For finding suitable gold open access publishers, researchers are encouraged to consult the Directory of Open Access Journals (https://doaj.org/), a service that indexes high quality, peer-reviewed open access academic journals that use an appropriate quality control system.

4.2 Making data openly accessible

Because of the intellectual property and commercial interests for the data generated within EXPERT we do not make data immediately openly accessible. Rather we offer the possibility of data access via a link at the project website[xiv]. Data access is granted by the EXPERT Project Management Team (PMT) in consultancy with the partner(s) that generated the data. Special attention is given to the data on the clinical study as these additionally raise ethical concerns. Considering the fact that we expect that we will not receive too many requests for data reuse, we think that it is most efficient to make data machine readable for the few instances where this is requested. The FAIRshare organization provides formats to organize this, an example from the MIACA project shows how this can be applied[xv].

To grant access Zenodo offers an opportunity to share data. Metadata for both open, closed, embargoed and restricted records are always publicly available in Zenodo. Data files and data sets for restricted access records are only visible to their owners and to those the owner grants access. Restricted access allows a researcher to upload a dataset and provide the conditions under which he/she grants access to the data. Researchers wishing to request access must provide a justification for how they fulfil these conditions. The owner of the dataset gets notified for each new request and can decide to either accept or reject the request. If the request is accepted, the requester receives a secret link which usually expires within 1-12 months.

4.3 Making data interoperable

We generate data that are stored digitally in multiple file formats. Below is a table with the preferred and acceptable formats for data storage that we will use in EXPERT.

Type	Preferred format(s)	Acceptable format(s)
Text documents	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
Plain text	• Unicode text (.txt)	• Non-Unicode text (.txt)
Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml) • Markdown (.md)
Programming languages	• MATLAB • NetCDF • TextFabric	•
Spreadsheets	• ODS (.ods) • CSV (.csv)	• Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)
Databases	• SQL (.sql) • SIARD (.siard) • CSV (.csv)	• Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Statistical data	• SPSS Portable (.por) • STATA (.dta) • DDI (.xml)	• SPSS (.sav) • SAS (.7dat; .sd2; .tpt)

	• Data and setup (.csv, .txt) • R	
Raster images	• JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2) • DICOM (.dcm)	•
Vector images	• SVG (.svg)	• Adobe Illustrator (.ai) • EPS (.eps) • WMF/EMF (.wmf, .emf) • CDR (.cdr)
Audio	• BWF (.bwf) • MXF (.mxf) • Matroska (.mka) • FLAC (.flac) • OPUS	• WAVE (.wav) • MP3 (.mp3) • AAC (.aac, .m4a) • AIFF (.aif, .aiff) • OGG (.ogg)
Video	• MXF (.mxf) • Matroska (.mkv)	• MPEG-4 (.mp4, .m4a, .m4v) • MPEG-2 (.mpg, .mpeg, .m2v, mpg2) • AVI (.avi) • QuickTime (.mov, .qt)
Computer Aided Design (CAD)	• AutoCAD DXF version R12 (ASCII) (.dxf) • SVG (.svg)	• AutoCAD other versions than R12 (ASCII) (.dwg, .dxf) • DWG (.dwg) • DGN (.dgn)
Geographical Information (GIS)	• GML (.gml) • MIF/MID (.mif/.mid)	• Esri Shapefiles (.shp & related files) • MapInfo (.tab & related files) • KML (.kml) • Esri Geodatabase (.gdb) • Project files/Workspaces (.mxd, .wor, .qgs)
Georeferenced images	• GeoTIFF (.tif, .tiff)	• TIFF World File (.tfw & .tif, possibly with additional files) • JPEG World File (.jgw & .jpg, possibly with additional files) • ERDAS IMAGINE File Format (.img)
Raster GIS	• ASCII GRID (.asc, .txt)	• Esri GRID (.grd & related files) • Surfer Grid (.grd; .srf) • ERDAS IMAGINE File Format (.img)
3D	• WaveFront Object (.obj) • Polygon file format (.ply) • X3D (.x3d) • COLLADA (.dae)	• Autodesk FBX (.fbx) • Blender (.blend) • 3D PDF (.pdf)
RDF	• RDF/XML (.rdf) • Trig (.trig) • Turtle (.ttl)	•

	• NTriples (.nt) • JSON-LD	
Computer Assisted Qualitative Data Analysis (CAQDAS)	• REFI-QDA (Qualitative Data Analysis)	• ATLAS.TI Copy bundle • NVivo Project file

Metadata will be created according to the standard ontology adopted in the field according to the EMBL-EBI defined ontology and accessible through their Ontology Lookup Service.

4.4 Increase data re-use

Open results produced by the project and deposited in Zenodo are usable by third parties after the end of the project. If confidentiality, security, personal data protection obligations or IPR issues related to specific research data that is needed to validate a scientific publication forbid open access, the data will be deposited in a restricted repository and access may be granted upon request and under the conditions of a restricted license.
The open results that are deposited in the Zenodo repository will be available at least 5 years after the conclusion of the project. According to Zenodo's general policies (http://about.zenodo.org/policies/), "items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least".

5 Allocation of resources

Based on experience within the B-SMART project that is now in its 3rd year that has many similarities in the types of data it generates to EXPERT, we anticipate that the raw data are of limited value for reuse. As a result, the allocation of resources is limited and is covered by the institution where the data are generated.
The Project Management Team is responsible for the data management procedures, together with the partner that has generated the data.
For long term preservation of the data, each partner follows their institutional guidelines. For UMCU The Netherlands Code of Conduct for Research Integrity states that research data must be kept for (at least) 10 years. The Utrecht University Policy Framework for Research Data adds that this 10-year period starts after you have published your paper based on the data you are preserving. For medical records, this period is 15 years or longer (WGBO (article 454)) and (patient) data for drug research must be stored for 20 years. The GPDR states that personal data may not be kept longer than is necessary for the purposes for which they were collected or for which they are used. Non-anonymised data may, however, be preserved for historical, statistical or scientific purposes. Within UMCU all data are stored in an electronic laboratory notebook using BIOVIA cloud storage. This service is provided by the department. For sharing data can be uploaded to the Zenodo repository.

6 Data security

All institutions have their own policies to ensure data security. For UMCU the cloud-service electronic laboratory notebook has some built-in features that maximize data security:

Encrypting all traffic to and from UMCU to the BIOVIA Notebook Cloud.

- using Hypertext Transfer Protocol Secure (HTTPS) that encrypts all data sent across the Internet and secures the identification of the BIOVIA Notebook Cloud service via a server certificate.
- HTTPs ensures reasonable protection from “eavesdroppers” and “man-in-the-middle attacks”.

Utilizing Oracle system and object privileges prevent unauthorized access of information.

- using Oracle user repository functionality for username and password authentication.
- securing all pre-defined Oracle accounts, and only our own Operation staff has administrator access to the database.

Integrity: Prevent Unauthorized Data Modification

- BIOVIA Notebook Cloud stores all UMCU data in an Oracle Database in order to provide database integrity
- All data updates are through store procedures that are tested as part of our development process.
- User sessions are logged and the identity of the users who update data is recorded.

Prevent Service Disruptions

- Data Centers and Network Operations Centre (NOC) that are manned 24/7/365.
- Capacity Management process to ensure the availability of all required resources such as bandwidth, data center capacity and utilities (power, cooling, etc.)
- Firewalls are configured with Access Command Lists (ACL) which prevent access to private internal IPs and deny access to all non-Administrative ports.
- Routers configured to prevent Denial of Service (DoS) attacks through the use of anti-spoofing Access Control Listings (ACLs)

7 Ethical aspects

All currently identified ethical parameters to be obeyed are documented in WP1. Those Deliverables have not yet been finalized. Additional details will be reported, as needed, in future versions of the DMP.

The Clinical Annex describes when and where approvals are needed and how to deal with personal data. Further on, it is communicated where informed consent is needed and how to get this. The full clinical trial plan is currently work in progress.

8 Other issues

9 Deviations

10 References

[i] C. Ramjoue and O. Marganne, "TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)," 13 October 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx.

[ii] Directorate-General for Research & Innovation, "Guidelines on FAIR Data Management in Horizon 2020, Version 3.0," EUROPEAN COMMISSION, 26 July 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[iii] https://imagej.net/

[iv] http://ncl.cancer.gov/working-assay-cascade.asp

[v] http://www.euncl.eu/

[vi] https://www.ema.europa.eu/en/news/opening-clinical-data-new-medicines

[vii] FAIRsharing.org: MIACA; Minimal Information About a Cellular Assay; DOI: https://doi.org/10.25504/FAIRsharing.7d0yv9.

[viii] FAIRsharing.org: MIARE-TAB; Minimum Information About an RNAi Experiment Tabular; DOI: https://doi.org/10.25504/FAIRsharing.t8g7yc;

[ix] https://www.ebi.ac.uk/ols/index

[x] https://www.ebi.ac.uk/spot/oxo/

[xi] https://www.nlm.nih.gov/research/umls/

[xii] https://www.ebi.ac.uk/spot/zooma/

[xiii] http://oppl2.sourceforge.net/index.html; Luigi Iannone, Mikel Egaña Aranguren, Alan L. Rector, Robert Stevens: Augmenting the Expressivity of the Ontology Pre-Processor Language. OWLED 2008

[xiv] https://expert-project.eu/

[xv] https://iweb.dl.sourceforge.net/project/miaca/Examples/MIACA_080130/MIACA080130.xml