IMI2 GA853989 - ERA4TB

European Regimen Accelerator for Tuberculosis

WP1 – Data and Pipeline Management

D1.8 Standardized templates for collection and reporting of clinical and preclinical data available to consortium members – Interim report

Lead contributor	Patrick O'Meara (8 – Critical Path Institute)
Other contributors	
	Ahmad Faizan (8 – Critical Path Institute)
	Pavan Kumar Sudhakar (8 – Critical Path Institute)
	Rick Liwski (8 – Critical Path Institute)
Due date	30/04/2021
Delivery date	23/12/2021
Deliverable type	R
Dissemination level	PU

Document History

Version	Date	Description
0.1	23/10/2020	**Deliverable D1.4:** Initial draft based on current project feedback/updates.
0.2	11/12/2020	**Deliverable D1.4:** Updates based on WP1 partner feedback.
0.3	07/12/2021	**Deliverable D1.8:** • General: o All reference links to documents and websites updated • Section 2: o Information split into subsections 2.1 and 2.2 o Further elaborations on ERA4TB partner requirements to further progress data standards development/progress • Section 3: o In vitro template information completely overhauled based on partners feedback in period January 2021 – November 2021 o Updates file naming convention added based on evolution of procedural requirements in 2021 • Section 4: In vivo template information completely overhauled based on partners feedback in period July 2021 – November 2021 • Sections 5 & 6: Updates on current progress added.

Version	Date	Description
0.4	22/12/2021	**Deliverable D1.8:** • <u>Section 6:</u> o Reference to CMC data/reports now included

Table of Contents

Abstract

This deliverable describes the current updates and developments on the data standards/templates intended being utilised on the ERA4TB project for the reporting and delivery of preclinical and clinical data generated by partners within the ERA4TB project. These standards are to support the ingestion of that data into the ERA4TB Drug Development Information Management (DDIM) platform and to support further downstream use of the data to meet the project objectives.

The information within this version of the document will cover updates/changes to the following topics: (1) In Vitro data reporting/standards developments; (2) In Vivo data reporting/standards requirements; (3) Further elaboration of factors which will influence the progression of standards/templates within the project.

Subject to the influx of data received, and inputs received from the ERA4TB partners, an updated version of data standards will be reported as part of deliverable D1.19 (Final Standardized templates). This document (and subsequently deliverable D1.19) will act as a supplement to the relevant Data Management Plan (deliverables D1.2, D1.11 and D1.16).

1. Project Overview

The European Regime Accelerator for Tuberculosis (ERA4TB) project is a research collaboration funded under the Innovative Medicines Initiative Joint Undertaking 2 (IMI JU2) within the framework of the wider Antimicrobial Resistance Accelerator programme.

The aim of the Antimicrobial Resistance (AMR) accelerator programme is to progress a pipeline of potential medicines, including but not limited to new antibiotics, to treat patients with resistant bacterial infections in Europe and across the globe, or to aid in the prevention of tuberculosis (TB).

The AMR Accelerator programme consist of three pillars under which multiple actions are expected:

- **Pillar A:** Capability Building Network (CBN)

- **Pillar B:** Tuberculosis Drug Development Network (TBDDN)

- **Pillar C:** Company-specific Portfolio Building Networks (PBNs)

The ERA4TB project is a part of Pillar B within the AMR accelerator program. The main objective of this project is to create a European open platform to accelerate the development of new combination regimens for the treatment of TB. The project will start from a collection of anti-TB compounds resulting from different EFPIA drug discovery activities, in varying stages of development, from "late lead" to "investigational new drug", then progress the most promising ones through an ERA4TB 'pipeline' comprised of a variety of in vitro assays, in vivo studies and clinical (Phase I) trials as appropriate for each asset.

The ERA4TB consortium will execute its mission through six specific objectives:

- Implementation of tools and capacities for the evaluation of TB drug candidates to effectively progress compounds from early pre-clinical to clinical development and identify potential new Pan-TB regimens ready for Phase II clinical evaluation.

- Development of modelling and simulation tools, the application of standard and new artificial intelligence (AI) techniques for better characterisation of exposure-effect relationships, optimisation of clinical trial design, prediction of dose and antibacterial effect in humans.

- Management of data generated by the project, integrating also data and knowledge from historical datasets available in reference databases, and from previous/existing consortia and projects, in the context of an ever improving 'learning system' that allows to refine the platform continuously.

- Provide flexible and efficient management able to adapt the capacity and resource allocation level required at each stage of the project, depending on each compound's progression and attrition dynamics and on inherent variables of the multiple combination assays.

- Provide a sustainability plan that incorporates all the synergies and learnings within the project and secures the survival of the platform beyond the life of the project.

- Define and execute an outreach, engagement, dissemination,

- and communication plan in collaboration with regulatory authorities and stakeholders, including patient organizations, to maximise the impact of the project

This document focuses on supporting the third objective listed above, more specifically focusing on the data reporting requirements to deliver data into the grit42 platform within the Drug Development Information Management system (DDIM), herein to be referred as "DDIM-grit42". The core objective for this deliverable is to further develop a set of standards which aspire to meet the following tenets:

- Alignment with FAIR data principles

- Ease of use across the AMR accelerator programs where possible

- Key data elements are reported to meet the end user requirements within the DDIM-grit42 platform

- Alignment with the architectural requirements and/or constraints of the DDIM-grit42 platform

2. Preface on Data Standards

2.1 AMR Data Group

The ongoing progress and development of data standards is partly informed as result of an ongoing collaborations between all pillars within the AMR Accelerator program with the view of creating a set of standards to harmonise both the formatting and terminologies used prior to delivery into the relevant target platform (e.g., DDIM-grit42), data structures that are reusable across the program of AMR accelerator projects and to ensure alignment with FAIR data principles.

The templates, specification documents and associated guidelines borne from the AMR Data Group are controlled and managed within the COMBINE project by Frauenhofer using their OwnCloud platform (https://owncloud.fraunhofer.de/) . Access to these documents is

being managed by the COMBINE team at Frauenhofer, and access can be given to partners involved in the AMR Accelerator program.

2.2 ERA4TB Partner Feedback

During the lifecycle of the ERA4TB project and based upon current asset pipeline progress, an agile approach will likely be adopted to further inform and enhance the data standards culminating with deliverable (D1.19). Factors that can inform future updates/ modifications of these standards include:

- Information on the level of data reported per assessment/experiment received by the relevant partners

- Data consumer requirements to meet downstream needs (e.g., data modelling, pipeline management etc.)

- Changes in regulatory requirements

The success, level of quality, and subsequent finalisation of the aforementioned standards for ERA4TB are dependent on timely feedback and engagement from the relevant ERA4TB partners. Limited, or lack of feedback from the relevant stakeholders may affect or delay the downstream availability of data in order to support the objectives of this project.

Upon identifying a need for update/modification from the AMR data standards, this item will be raised and discussed with members of the AMR Data Group to identify if the current standards are to be modified to meet the needs of the scenario in question with a plan for releasing an updated version of the data standards within the Frauenhofer OwnCloud (if applicable).

Based on project specific use-case requirements, there may be the need in developing a bespoke or project specific template(s), in such situations the templates will likewise be made available to the AMR Data Group whilst ensuring terminology and variable naming conventions are in alignment to support data interoperability. The latest template(s) per WP will be maintained by the Critical Path Institute (C-Path) and are accessible to ERA4TB partners within the Synapse SharePoint.

3. In Vitro Data (Work Package 2)

3.1 Templates/Standards Overview

For all preclinical experiments conducted within **WP2**, data is expected to be reported and maintained in the Raw_Data_Template file (herein to be referred as "WP2 Source Template" - refer to **Section 3.1.1**) for each individual pipeline job as defined within the relevant Asset Progression Plan (APP) and/or Combination Progression Plan (CPP).

In order to integrate the reported data into the DDIM-grit42 platform, the relevant information reported in the WP2 Source Template will be converted into to an adapted version of the AMR Preclinical results template (herein to be referred as "Preclinical Results template") which was initially developed by the AMR Data Group (refer to **Section 3.1.2**) This template was later adapted based on the feedback from the data generating partners within **WP2** and data consuming partners (e.g., **WP5**).

In the event that the relevant asset owner permits the use of all or a subset of the reported data to be integrated the TB Platform for the Aggregation of Preclinical Experiments data (TB-APEX) database for external researcher access. The database in questions is accessible to approved users via the Data Archive platform managed by C-Path.

The permitted data will first be converted into CDISC SEND format prior to integration into TB-APEX. The agreements to facilitate such permissions for integration into TB-APEX are described within the SOP for Data Collaboration Agreements (D1.5) which is accessible to authorised project members only.

3.1.1 Input: WP2 Source Template

As per **Section 3.1**, for all preclinical experiments conducted within **WP2** the data is reported and maintained by the data generators using the WP2 Source Template which is designed to collect and report all required experimental results which includes detailed experimental conditions, compound information and where applicable individual replicate/timepoint level data points/results (e.g., well & plate level data). The data collected within the WP2 Source Template is intended to provide the necessary context for downstream in-depth modelling purposes.

The continued development of source template is being directly informed by the ongoing progress of approved modules within the relevant APP and/or CPP, feedback from the data generating partners (**WP2**) and data consumer requirements (**WP5**). The key focus of these interactions is to minimise the number of source templates needed for each type of experiment or module, whilst also ensuring that all data can be adequately reported/collected for downstream use. Any changes to the source template

will be version controlled (including a revision history) and will be made available to partners within the WP1 Template folder within the Synapse SharePoint area which is accessible to authorised project members only.

Strict adherence to the agreed source template will enable/support:

- Increased automation of data QC and transformation activities

- Improved interoperability across all reported information

- Tabulation of metadata to support downstream activities

- Adequate reporting of provenance information

All results recorded along with the experimental protocol and any other supplementary resources surrounding the experiments should be maintained within an Electronic Laboratory Notebook (ELN) or equivalent platform at site. Subject to the capabilities of the ELN or equivalent platform at site, it is recommended that the WP2 Source Template is incorporated into the platform to support ease of reporting.

NB: To support provenance and traceability, all data generating partners are expected to maintain reported information for a specific APP/CPP job (including reported results) within the directory of their local institution.

3.1.2 Output: Preclinical Results Template

The data from the WP2 Source Template is extracted using programmatic methods and reported into the adapted Preclinical Results Template. A suitable statistical parameter (e.g., mean) is applied to summarise the individual data points. The method and level of summarisation is agreed between the data generator (**WP2**) and data consumer (e.g., **WP5**) ahead of reporting of a specific experiment. The data from the following worksheets in WP2 Source Template are accessed and reported in the Preclinical Results Template using Python programming language by taking information from the following reported tabs:

- <*Experiment Name*>_Summary
- Compounds
- General Info

The outputs from the programmatic transformation into the Preclinical Results Template structure will be loaded into the DDIM-grit42 platform. The latest version of the Preclinical results level template can be accessed from the Synapse SharePoint.

3.2 Template Content

The tables presented int the following pages provides a more detailed overview of the WP2 Source Template (**Tables 1, 2, 3**) and Preclinical Results Template (**Table 4**). Theses tables represents the current iteration of the template (version 4.7) at time of this document has been finalised.

Prior to the reporting of the experimental results, it is recommended that the Synapse SharePoint is accessed to verify whether a partner is working with the latest version of the source template. In the event of doubts, the assigned data management contact(s) in Work Package 1 (**WP1**) can be consulted.

3.2.1 Input: WP2 Source Template Structure

The WP2 Source Template comprises of several worksheets developed based on the requirements from various in-vitro experiments and partner requirements.

3.2.1.1 General Info Worksheet

This worksheet contains metadata information related to the experiments performed (e.g., Experiment ID, Asset involved, Module number, JOB ID etc.). The relevant information collected in this worksheet are directly mapped into their respective columns of the Preclinical Results template. The mapping fields are highlighted in Red and blue respectively to indicate that these fields are mapped into the preclinical results template.

The field "*Experiment Type*" from General Info worksheet is used to hide/display the worksheets specific to a particular experiment, for example on selecting an experiment type as *"TKA"* the worksheets CFU_Calculation and CFU_Summary appears in the source template and hides the worksheets for other experiments (e.g., MIC). This feature is implemented using inbuilt excel macros written within the General_Info worksheet.

Table 1. WP2 Source Template - General Info

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	Experiment ID	Biovia or relevant ELN Identification number	Experiment ID	VARCHAR	R	N	Enter the ELN Identification number generated upon creating an experiment entry.
2	Title of Experiment	Experiment title as per site naming convention	Title of Experiment	VARCHAR	O	N	Recommended to be used to ensure provenance and traceability
3	Type of Experiment	Type of experiment performed	Type of Experiment performed	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss with your data management contact.
4	JOB ID	Job ID as per pipeline specification	Job ID as per pipeline specification document	VARCHAR	R	N	Refer to the APP or CPP for the JOB ID
5	Project	Project	Project under which the experiment is performed	VARCHAR	R	N	e.g., ERA4TB
6	Asset	Asset being experimented	Name of the Asset being experimented	VARCHAR	R	N	Drug asset under investigation
7	APP/CPP No	Asset/Combination Progression Plan number	Asset progression Number	VARCHAR	R	N	APP/CPP number for the relevant job ID.
8	Module Number	Module number	Module number	VARCHAR	R	N	Module number for the relevant job ID.
9	Date of experiment design:	Date of Experiment Design	Date of Experiment Design	DATE	O	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
10	Date of experiment start	Experiment start date	Experiment start date	DATE	R	N	
11	Date of final report	Date of reporting completion	Date of experimental reporting completion	DATE	O	N	
12	Experiment Operator Name:	Experiment Operator Name	Experiment Operator Name	VARCHAR	O	N	
13	Experiment Operator EMAIL	Experiment Operator email.	Experiment Operator email.	VARCHAR	R	N	
14	Experiment Operator Institution	Experiment Operator Institution	Experiment Operator Institution	VARCHAR	R	N	
15	Source Data Location	Source Data Location	Source Data Location	VARCHAR	R	N	Source Data Location at site institute.
16	Approver NAME	Name of the experiment approver	Name of the experiment approver	VARCHAR	O	N	
17	Approver EMAIL	Email of experiment approver	Email of experiment approver	VARCHAR	O	N	
18	Approver Institution	Institution name of the experiment approver.	Institution name of the experiment approver.	VARCHAR	O	N	
19	Background	Background of the experiment	Background of the experiment	VARCHAR	O	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
20	AIM	Aim of the experiment	Aim of the experiment	VARCHAR	O	N	
21	Protocol	Protocol location	Protocol location	VARCHAR	R	N	
22	Human Biological Sample use	Usage of human biological sample	Usage of human biological sample	VARCHAR	O	N	
23	Human Biological Sample use - Comments	Comments on the use of human biological sample	Comments on the use of human biological sample	VARCHAR	O	N	
24	Readout	Readout	Readout	VARCHAR	O	N	
25	Result type	Type of result generated from the experiment	Type of result generated from the experiment	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
26	Replicates	Number of replicates used within the experiment	Number of replicates used within the experiment	VARCHAR	R	N	
27	Biomaterial	Type of Biomaterial used.	Type of Biomaterial used.	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
28	Species	Type of species used.	Type of species used.	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.

3.2.1.2 Compounds Tab

This sheet contains information on the compound, compound concentration, medium, strain, etc. used throughout the course of the experiment. The compounds sheet and <*Experiment-name*>_Summary sheets are merged using GROUP ID variable during transformation of source template to preclinical results template. A new group is created whenever there is a change in any one of the four following parameters; 1. Compound, 2. Compound Concentration, 3. Medium, and 4. Strain.

Table 2. WP2 Source Template - Compounds

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	GROUP_ID	Group Identifier	Group Identifier used within an experiment.	VARCHAR	R	N	Group Identifier used to distinguish different groups within an experiment.
2	INTERNAL_COMPOUND_ID	Name of the compound as referenced at Site.	Name of the compound as referenced at Site.	VARCHAR	R	N	
3	INTERNAL COMPOUND BATCH ID	Batch number of the compound as referenced at Site.	Batch number of the compound as referenced at Site.	VARCHAR	R	N	
4	GRIT COMPOUND ID	Name of the compound as referenced in DDIM-grit42 compound registry.	Name of the compound as referenced in DDIM-grit42 compound registry.	VARCHAR	R	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
5	GRIT BATCH ID	Batch number of the compounded as referenced in DDIM-grit42 compound registry.	Batch number of the compounded as referenced in DDIM-grit42 compound registry.	VARCHAR	R	N	
6	COMPOUND1	Compound 1	Name of the first compound used within an experiment group.	VARCHAR	R	N	
7	COMPOUND1 _CONCENTRATION	Compound1 Concentration	Concentration of the first or only compound used within an experiment group.	VARCHAR/ NUM	R	N	
8	COMPOUND2	Compound 2	Name of the second compound used within an experiment group.	VARCHAR	O	N	
9	COMPOUND2 _CONCENTRATION	Compound2 Concentration	Concentration of the second compound used within an experiment group.	VARCHAR/ NUM	O	N	
10	COMPOUND3	Compound 3	Name of the third compound used within an experiment group.	VARCHAR	O	N	
11	COMPOUND3 _CONCENTRATION	Compound3 Concentration	Concentration of the third compound used within an experiment group.	VARCHAR/ NUM	O	N	
12	COMPOUND4	Compound 4	Name of the fourth compound used within an experiment group.	VARCHAR	O	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
13	COMPOUND4 _CONCENTRATION	Compound4 Concentration	Concentration of the fourth compound used within an experiment group.	VARCHAR/ NUM	O	N	
14	COMPOUND5	Compound 5	Name of the fifth compound used within an experiment group.	VARCHAR	O	N	
15	COMPOUND5 _CONCENTRATION	Compound5 Concentration	Concentration of the fifth compound used within an experiment group.	VARCHAR /NUM	O	N	
16	COMPOUND_CONC ENTRATION_UNIT	Compound(s) concentration Unit	Units for compound concentration values. Please represent all compound(s) concentration using same units within a group.	VARCHAR	R	Y	Units for compound concentration values. Please represent all compound(s) concentration using same units within a group. Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
17	MEDIUM	Medium	Medium used for experiment/assessme nt	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
18	MEDIUM FORM	State of the medium used	State of the medium used	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
19	STRAIN	Bacterial Strain	Bacterial Strain	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
20	Bacterial Strain ATCC REF	Bacterial Strain ATCC REF	Bacterial Strain ATCC REF	VARCHAR	O	N	
21	Bacterial Strain STOCK Nº	Bacterial Strain STOCK Nº	Bacterial Strain STOCK Nº	VARCHAR	O	N	

3.2.1.3   Summary Results Sheet

This sheet contains the actual summarised result values from the experiment in question. These values are either manually entered or populated from <Experiment-Name>_Calculation using pre-populated formulas based on the experiment type selected.

Table 3. WP2 Source Template - Summary Results

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	Sample ID	Identification number assigned to each sample within an experiment.	Identification number assigned to each sample within an experiment.	VARCHAR	O	N	Identification number assigned to each sample within an experiment.
2	GROUP_ID	Group Identifier	Group Identifier used within an experiment.	VARCHAR	R	N	Group Identifier used to distinguish different groups within an experiment.
3	Time-Point (Days)	Timepoint of result collection	Timepoint of result collection	VARCHAR	O	N	Timepoint of result collection. E.g., Day 1, Day 2, etc.
4	Date of Measurement	Date of result measurement.	Date of result measurement.	DATE	R	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
5	Time of Measurement	Time at which the results are collected.	Time at which the results are collected.	TIME (HH:MM)	O	N	Format: 24hr Clock e.g., 13:14, 09:10
6	INTERNAL_COMPOUND_ID	Name of the compounded as referenced at Site.	Name of the compounded as referenced at Site.	VARCHAR	R	N	Name of the compounded as referenced at Site.
7	Statistical Operator	Statistical method	Statistical method used to calculate the result (e.g., Average, Mode, Median)	VARCHAR	R	Y	**STATISTICAL_METHOD** Single value Average Median Mode GeometricMean HarmonicMean #NA (not applicable) NOT IN LIST (SEE COMMENT)
8	Result Operator	Result Operator	Result Operator	VARCHAR	R	Y	**RESULT_OPERATOR** < <= = >= >
9	Result	Result	Experiment Result	VARCHAR/ NUM	R	N	Format to be remain consistent per experiment/assessment

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
10	Result Unit	Unit	Result Unit	VARCHAR	R	Y	Optional for assessment which are qualitative or based on observations (e.g., ABNORMAL/NORMAL) Please refer to full dictionary. If strain not listed, please discuss your data management contact.
11	Log10 Results	Log10 value of the result produced.	Log10 value of the result produced.	VARCHAR	O	N	
12	#NA	Variation	For aggregated values choose a method to report deviations	NUM	O	Y	
13	Optical Density	Optical Density	Optical Density	VARCHAR	O	N	Optical Density
14	VALIDATION	Validation	Validation Status of Result	VARCHAR	R	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
15	CONTROL_GROUP	Control Group	Calculation of the result type based on which type of control	VARCHAR	R	Y	**CONTROL_GROUP** untreated vehicle positive negative plates without compound #NA (not applicable) NOT IN LIST (SEE COMMENT)
16	COMMENTS	Comments	Any additional pertinent information regarding an assessment/experiment which is useful for the interpretation of the data e.g., dosing regimen etc.	VARCHAR	R	N	See also **Section 3.4** regarding "Not in List" scenarios.

3.2.2   Output: Preclinical Results Template

This values in this template are generated by the transformation script which extracts the summarized experimental data from the worksheets mentioned in **Sections 3.2.1.1- 3.2.1.3**. This template structure is configured in the DDIM-grit42 platform for uploading the results for experiments conducted by **WP2**.

Table 4. Output - Preclinical Results

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	STUDYID	Site-specific Study Identifier	Study number	VARCHAR	R	N	Job Id as per the APP/CPP.
2	EXPID	Site-specific Experiment Identifier	Experiment number (e.g., from ELN)	VARCHAR	O	N	Recommended to be used to ensure provenance and traceability
3	CPD_ID	Internal Compound ID	Internal Compound ID	VARCHAR	R	N	System generated identifier (e.g., DDIM-grit42 platform)
4	BATCH_ID	Internal Compound Batch ID	Internal Batch ID	VARCHAR	R	N	System generated identifier (e.g., DDIM-grit42 platform)
5	EXT_CPD_ID	External Compound ID	External Compound ID	VARCHAR	O	N	Compound Identifier per Asset Owner
6	EXT_BATCH_ID	External Batch ID	External Batch ID	VARCHAR	O	N	Compound Identifier per Asset Owner
7	MODULE_NUMBER	Module Number	Module number as per pipeline specification	VARCHAR	R	N	Module number of the respective experiment to be entered. For e.g., M1.3.3
8	APP/CPP	Asset Progression Plan	Asset Progression Plan	VARCHAR	R	N	Asset Progression Plan. For e.g., APP3, CPP1
9	SITE	Site Identifier	Site/Lab where assessment/experiment was conducted	VARCHAR	R	N	Entry should be concise (e.g., partner abbreviations to be used)
10	PROVENANCE	Data Provenance Information	Internal Contact information for additional information on the experiment	VARCHAR	R	N	Contact details to support data provenance (e.g., e-mail address), location of source/raw data files etc.

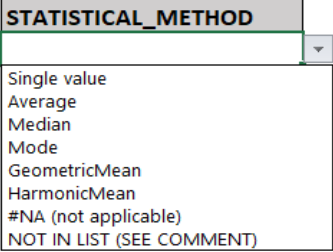| Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
11	EXPERIMENT_DATE	Date of Assessment/ Experiment	Date of Assessment/Experiment	DATE (YYYY-MM-DD)	R	N	ISO 8601 format
12	GROUP_ID	Group Identifier	Group Identifier used within an experiment.	VARCHAR	R	N	Group Identifier used to distinguish different groups within an experiment.
13	BIOMATERIAL	Biomaterial	Biomaterial used for the experiment	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
14	SPECIES_NAME	Species	If animal, cell or tissue was selected for BIOMATERIAL, then specify species	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
15	STRAIN_NAME	Bacterial Strain	Bacterial Strain	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
16	COMPOUND1	Compound 1	Name of the first compound used within an experiment group.	VARCHAR	R	N	
17	COMPOUND1 _CONCENTRATION	Compound1 Concentration	Concentration of the first or only compound used within an experiment group.	VARCHAR/ NUM	R	N	
18	COMPOUND2	Compound 2	Name of the second compound used within an experiment group.	VARCHAR	O	N	
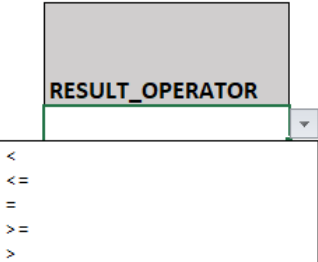
Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
19	COMPOUND2_CONCENTRATION	Compound2 Concentration	Concentration of the second compound used within an experiment group.	VARCHAR/NUM	O	N	
20	COMPOUND3	Compound 3	Name of the third compound used within an experiment group.	VARCHAR	O	N	
21	COMPOUND3_CONCENTRATION	Compound3 Concentration	Concentration of the third compound used within an experiment group.	VARCHAR/NUM	O	N	
22	COMPOUND4	Compound 4	Name of the fourth compound used within an experiment group.	VARCHAR	O	N	
23	COMPOUND4_CONCENTRATION	Compound4 Concentration	Concentration of the fourth compound used within an experiment group.	VARCHAR/NUM	O	N	
24	COMPOUND5	Compound 5	Name of the fifth compound used within an experiment group.	VARCHAR	O	N	
25	COMPOUND5_CONCENTRATION	Compound5 Concentration	Concentration of the fifth compound used within an experiment group.	VARCHAR/NUM	O	N	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
26	COMPOUND_CONCENTRATION_UNIT	Compound(s) concentration Unit	Units for compound concentration values. Please represent all compound(s) concentration using same units within a group.	VARCHAR	R	Y	Units for compound concentration values. Please represent all compound(s) concentration using same units within a group. Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
27	EXPERIMENT	Assessment/Experiment	Name of assessment	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
28	Factor1	Factor column1	Dictionary controlled column names which can be chosen based on the experimental requirements	VARCHAR	O	Y	The column name can be chosen from a list of dictionary values based on the experiment requirements. The list contains the generic experimental parameters like MIC of compound, Temperature, etc.
29	Factor2	Factor column2	Dictionary controlled column names which can be chosen based on the experimental requirements	VARCHAR	O	Y	The column name can be chosen from a list of dictionary values based on the experiment requirements. The list contains the generic experimental parameters like MIC of compound, Temperature, etc.
30	PROTOCOL_NAME	Protocol	Name of protocol (provide as PDF)	VARCHAR	R	N	Document name describing the experimental protocol.
31	NO_OF_REPLICATES	Number of Replicates	Number of replicates used in an experiment	NUM	O	N	Number of replicates used the experiment. Expected to remain constant for a particular experiment.

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
32	RESULT_DATE	Result Collection Date	Date at which the results are observed (YYYY-MM-DD)	DATE (YYYY-MM-DD)	R	N	Date of result collection in YYYY-MM-DD format.
33	RESULT_TIME	Result Time	Time at which the results are observed (HH:MM)	VARCHAR	O	N	Time of result collection in HH:MM 24-hour format.
34	RESULT_TYPE	Result Type	Type of result being generated (e.g., AUC0-t)	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
35	STATISTICAL_METHOD	Statistical method	Statistical method used to calculate the result (e.g., Average, Mode, Median)	VARCHAR	R	Y	STATISTICAL_METHOD Single value Average Median Mode GeometricMean HarmonicMean #NA (not applicable) NOT IN LIST (SEE COMMENT)
36	RESULT_OPERATOR	Result Operator	Result Operator	VARCHAR	R	Y	RESULT_OPERATOR < <= = >= >
37	RESULT_VALUE	Result	Experiment Result	VARCHAR/ NUM	R	N	Format to be remain consistent per experiment/assessment

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
38	RESULT_UNIT	Unit	Result Unit	VARCHAR	R	Y	1. Optional for assessment which are qualitative or based on observations (e.g., ABNORMAL/NORMAL) 2. Please refer to full dictionary. If strain not listed, please discuss your data management contact.
39	See Additional Information	Variation	For aggregated values choose a method to report deviations	NUM	O	Y	#NA SEM StdDev Var Confidence.Norm.Dist Confidence.T.Dist #NA NOT IN LIST (SEE COMMENT)
40	MEDIUM	Medium	Medium used for experiment/assessment	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
41	CONTROL_GROUP	Control Group	Calculation of the result type based on which type of control	VARCHAR	R	Y	CONTROL_GROUP untreated vehicle positive negative plates without compound #NA (not applicable) NOT IN LIST (SEE COMMENT)

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
42	VALIDATION	Validation	Validation Status of Result	VARCHAR	R	Y	**VALIDATION** V (valid) NV (non valid) A (active) NA (not active) NS (no statistical difference) #NA (not applicable)
43	COMMENTS	Comments	Any additional pertinent information regarding an assessment/experiment which is useful for the interpretation of the data e.g., dosing regimen etc.	VARCHAR	R	N	See also **Section 3.4** regarding "Not in List" scenarios.

3.3 WP2 Source Template Terminologies

The dictionaries/terminologies for the fields indicated in **Section 3.2.1** are controlled and maintained by C-Path within the WP2 Source Template itself to support consistency of reporting across all experiments reported.

In a scenario where a field/column with a defined dictionary in the WP2 Source Template does not have an appropriate value in the dictionary list, the relevant partner should raise this item to the assigned data management group within **WP1** for discussion. Based on the discussion, the outcome will be one of the following:

- Update of the data dictionary with new option(s) included ensuring alignment with recognised ontologies (e.g., NCIT) where possible.
- Proceed as per **Section 3.4**

3.4 "Not in List" Scenarios

For fields in the source template which have the dictionary option "NOT IN LIST (SEE COMMENT)" available, this option can be selected in unique cases where an appropriate option does not exist. To ensure clarity of context and ease of parsing out information further downstream. The following entry conventions are recommended.

- Update of the data dictionary with new option(s) included ensuring alignment with recognised ontologies (e.g., NCIT) where possible.
- Scenario 1: Single field affected per record can be written using the convention:

"<variable>:<value>" (see example below)

Type of Experiment	COMMENTS
NOT IN LIST (SEE COMMENT)	Type of Experiment: Test X

- Scenario 2: Multiple fields affected per record can be written using the convention:

"<variable1>:<value1>|<variable2>:<value2>|…. etc" (see example below)

Type of Experiment	RESULT_TYPE	COMMENTS
NOT IN LIST (SEE COMMENT)	NOT IN LIST (SEE COMMENT)	Type of Experiment: Test X |RESULT_TYPE:Type X

- Scenario 3: At least one affected field and additional comments that does not relate to an affected field per record can be written using the convention.

"<variable1>:<value1>|<variable2>:<value2>|Additional Comments…. etc"

(See example below)

Type of Experiment	RESULT_TYPE	COMMENTS
NOT IN LIST (SEE COMMENT)	NOT IN LIST (SEE COMMENT)	EXPERIMENT: Test X |RESULT_TYPE: Type X | Alternative medium was used due to shortage

3.5   File Naming Conventions

For the purpose of consistency and to support findability of all stored data/files generated within the project pipeline, the following conventions will be followed by the partner undertaking a specific experiment.

Table 5. File Naming Conventions

File Name Component	\<Job ID\>_	\<WPx\>_	\<file_name/ Experiment Type\>__	\<Study/ Experiment ID\>_	\<yyyy-mm-dd\>_	\<Sequence Number\>
Description	Unique job identifier for a defined within the relevant APP or CPP	Work Package number under which a specific experiment or activity is conducted under	Experiment Type Name	Unique study or experiment identifier (e.g., ELN ID)	Date of reported file in ISO 8601 format	Optional: Sequence number to be assigned if experiment is split across various source template files
Component Example	Asset1-APP2-M1.1-UNIPV_	WP2_	TKA_	ELN1234_	2021-11-30_	02
Example File Name	Asset1-APP2-M1.1-UNIPV_WP2_ TKA_ ELN1234_2021-11-30_02					

3.6   Future Considerations

Subject to the conditions outlined in **Section 2.2**, if a large degree of divergence in data template requirements has been observed, it may be necessary to convert the data into a CDISC SEND format to harmonise the structure of data prior to ingesting the data into the DDIM. If applicable, such a change in approach will be described in deliverable D1.19 and in the relevant Data Management Plan (deliverables D1.11 and/or D1.16). The possibility to convert such data into CDISC SEND will be subject to data volume, resourcing constraints, and available budget to conduct such activities.

4. In Vivo Data (Work Package 3)

For animal studies, there is a requirement to compare data on an individual animal level rather than just on a study or experiment level. To meet these requirements, a study level data template which also includes an individual animal level data template have been developed by the C-Path. This template is based on a combination of C-Path's in-house In Vivo template, the initial In Vivo data template developed by AMR Data Group and input from **WP5** partners.

4.1 WP3 Template Overview

For all preclinical experiments conducted within **WP3**, data is expected to be reported and maintained in the InVivo_Result_Template file (herein to be referred as "WP3 Source Template").This template is designed to collect individual animal level result including study/experiment level information at source (e.g., ERA4TB laboratory partners) for experiments conducted within **WP3**. Data will be delivered into the DDIM-grit42 platform in the same structure as it is collected at source.

The continued development of source template is being directly informed by the ongoing progress of approved modules within the relevant APP and/or CPP, feedback from the data generating partners (**WP3**) and data consumer requirements (**WP5**). The key focus of these interactions is to minimise the number of source templates needed for each type of experiment or module, whilst also ensuring that all data can be adequately reported/collected for downstream use. Any changes to the source template will be version controlled (including a revision history) and will be made available to partners within the WP1 Template folder within the Synapse SharePoint area which is accessible to authorised project members only.

All results recorded along with the experimental protocol and any other supplementary resources surrounding the experiments should be maintained within an Electronic Laboratory Notebook (ELN) or equivalent platform at site. Subject to the capabilities of the ELN or equivalent platform at site, it is recommended that the data worksheet template is incorporated into the platform to support ease of reporting.

In the event, that the relevant asset owner permits the use of all or a subset of the reported data to be integrated into the TB-APEX database for external researcher access. The permitted data will be converted into CDISC SEND format. The agreements to facilitate such permissions for integration into TB-APEX are described within the SOP for Data Collaboration Agreements (D1.5).

NB: To support provenance and traceability, all data generating partners are expected to maintain reported information for a specific APP/CPP job (including reported results) within the directory of their local institution.

4.2 Template Content

This template consists of four distinct worksheets to support the varying design requirements of the experiment which includes and not but limited to different result types (e.g., PK, PD), results captured at different dosing level and multiple time points. Every sheet contains the STUDY_ID variable to uniquely identify experiment as well as link data from different sheets to the specific study or experiment. In the subsequent pages a more detailed overview of the aforementioned worksheet tabs within the WP3 Source Template will be elaborated upon. These tables represent the current iteration of the template (version 2.0) at the time this document has been finalised.

4.2.1 Trial Summary

Information on this tab provides an overview of the study/experiment in question. This sheet contains information in the form of *KEY* (TSPARM)-*VALUE* (TSVALUE) pair to capture information at study/experiment level which are going to be remain a constant throughout the study or experiment such as Age, Dose Duration, Strain, Site information etc. as well as provenance information to support traceability and FAIR data principles. Full list of required information can be found in **Table 7**.

Table 6. WP3 Source Template - Trial Summary

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	STUDY_ID	Site-specific Study Identifier	Site-specific Study Identifier	VARCHAR	R	N	e.g., Job ID for ERA4TB progression plan activities
2	TSPARM	Parameter Name	Descriptive name of Parameter	VARCHAR	R	N	Pre-defined Parameters along with parameter code to capture study level information which are not going to change throughout the study/experiment. In this sheet TSVALUE needs to be populated only which also includes provenance information to ensure traceability. Refer to **Table 7** for complete list of Parameters with description.
3	TSPARMCD	Parameter Short Name	Short name of Parameter	VARCHAR	R	N	
4	TSVALUE	Parameter Value	Value of Parameter	VARCHAR	R	N	

Table 7. Trial Summary Parameters

TSPARM / Parameter Name	TSPARMCD/ Parameter Short Name	Description	Required or Expected
Asset/Combination Progression Plan Job Identifier	PPREFN	Name, date, and version of Asset o combination progression plan that the pre-clinical study belong to. Only applicable to studies conducted within the ERA4TB project	REQUIRED
Provenance Contact	PRVCNTCT	Contact details of data owner	REQUIRED
Provenance Link	PRVLINK	Location where data owner is storing the study data	REQUIRED
Provence Source Data	PRVDATA	Delimited list of all raw data files used to generate the final data outputs	EXPECTED
Protocol Link	PROTLINK	URL link to Study Protocol	REQUIRED
Age	AGE	Age of subjects planned for the study populated as an NUM. If the planned age of subjects is a range (e.g., "12-14 days"), then populate the age range in the AGETXT variable. Either the AGE or the AGETXT variable should be populated	REQUIRED
Age Text	AGETXT	The age of the subjects at study start, as planned, expressed as a range. If an age NUM value is available, then populate the AGE variable instead. Either the AGE or AGETXT variable should be populated.	REQUIRED
Age Unit	AGEU	Unit of Age, e.g., WEEKS, DAYS	REQUIRED
Dosing Duration	DOSDUR	The longest planned duration from the start of dosing to the first day of planned terminal disposition of the subjects	EXPECTED
Experimental Start Date	EXPSTDTC	Experimental starting date means the date on which the first study specific data are collected	REQUIRED

TSPARM / Parameter Name	TSPARMCD/ Parameter Short Name	Description	Required or Expected
Experimental End Date	EXPENDTC	Experimental completion date means the last date on which data are collected from the study	REQUIRED
Route of Administration	ROUTE	The delivery method by which the Test Article is administered to the subjects	REQUIRED
Species	SPECIES	Common species name of the subject (i.e., test system) under study (e.g., MOUSE, RAT, DOG, MONKEY)	REQUIRED
Subtype of Species	SPECIES_SUBTYPE	e.g., BALB/c	REQUIRED
Sex of Participants	SEX	Planned sex of subjects to participate in the study. e.g., MALE/FEMALE	REQUIRED
Study Category	STCAT	General category of scientific study. (e.g., TOX, MICROBIOLOGY)	REQUIRED
Sponsoring Organization	SSPONSOR	The name of the company (or person) who initiates, supports, or submits the nonclinical study. The parameter contains the name of the specific sponsor	REQUIRED
Strain/Substrain	STRAIN	To identify the vendor-supplied strain/substrain designation for the subject (i.e., test system) under study e.g., ERDMAN	REQUIRED
Route of infection	INFECTION_ROUTE	Route through which Subjects were Infected e.g., AEROSOL	REQUIRED
Lower Limit of Quantitation For PK	LLOQ_PK	Definition of LLOQ for PK Results e.g., 10 ng/mL	REQUIRED
Lower Limit of Quantitation For PD	LLOQ_PD	Definition of LLOQ for PD Results e.g., 0.78 log10 CFU/Lung	REQUIRED
Study Start Date	STSTDTC	The Study Start Date, the date on which the study protocol or plan is approved (signed) by the Study Director. Also known as the study initiation date.	EXPECTED
Study Title	STITLE	The title of the nonclinical study	REQUIRED
Investigational Therapy or Treatment	TRT	The name of the planned Test Article, treatment, therapy administered during the study. Example: Compound Name	REQUIRED
Test Facility Country	TFCNTRY	The country where the Test Facility is located.	EXPECTED
Test Facility Location	TSTFLOC	The full postal address of the Test Facility.	EXPECTED

TSPARM / Parameter Name	TSPARMCD/ Parameter Short Name	Description	Required or Expected
Test Facility Name	TSTFNAM	The name of the Test Facility responsible for the overall conduct of the nonclinical study, or the facility administering the Test Article to Test Subjects	EXPECTED
Treatment Vehicle	TRTV	Vehicle for administration of treatment, such as a liquid in which the treatment drug is dissolved (e.g., Saline).	REQUIRED
End Date/Time of Dose Interval	DOSENDTC	The end date of the dosing interval on the study, as defined by the protocol	EXPECTED
Start Date/Time of Dose Interval	DOSSTDTC	The start date of the dosing interval on the study, as defined by the protocol	EXPECTED
Environmental Temperature	ENVTEMP	The planned environmental temperature for the test subjects. Can be expressed as either a single value (80), or a range (75-80).	EXPECTED
Environmental Temperature Units	ENVTEMPU	The units associated with the environmental temperature. Only "C" or "F" is acceptable.	EXPECTED
Study Length	SLENGTH	The planned length of time for a subject's participation in ISO 8601 format. TSVAL values associated with this TSPARMCD would be "P5M" for a duration of 5 months or "P2W" for a duration of 2 weeks	REQUIRED
Method of Termination	MTHTRM	Describes the planned sacrifice procedure e.g., CO2, ANESTHETIZED CERVICAL DISLOCATION, ANESTHETIZED EXSANGUINATION	EXPECTED
Planned Number of Subjects	SPLANSUB	The planned total number of subjects that will participate in the study.	REQUIRED
Study is Randomized	SRANDOM	Identifies whether the study is randomized. TSVAL values associated with this TSPARMCD would be "Y" or "N."	REQUIRED
Study End Date	STENDTC	The Study End Date: the date on which the final report is approved (signed) by Study Director. Also known as the study completion date.	REQUIRED
Sponsor's Monitor	STMON	The individual responsible for the periodic follow-up regarding the conduct of the nonclinical study. TSVAL values associated with this TSPARMCD would be the specific name of an individual, e.g., "Dr. John Doe."	EXPECTED

4.2.2 Dosing

Data on this tab provides detailed dosing information of animals at all time points as defined by protocol along with other information like weight, infection date, administration route etc. Please note that this worksheet is dedicated for animals that have been given a dosing regimen. If an animal has been infected and no treatment (dose) has been given, such subjects will not be reported in this worksheet..

Table 8. WP3 Source Template - Dosing

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	PROJECT_ID	Project ID	Project Identifier	VARCHAR	R	N	e.g., ERA4TB
2	PROTOCOL	Protocol ID	Name of protocol (provide as PDF)	VARCHAR	R	N	Document name describing the experimental protocol.
3	STUDY_ID	Site-specific Study Identifier	Site-specific Study Identifier	VARCHAR	R	N	e.g., Job ID for ERA4TB progression plan activities
4	EXPID	Site-specific Experiment Identifier	Experiment number (e.g., from ELN)	VARCHAR	R	N	Recommended to be used to ensure provenance and traceability
5	CPD_ID	Internal Compound ID	Internal Compound ID	VARCHAR	R	N	System generated compound identifier (e.g., DDIM-grit42 platform)
6	BATCH_ID	Internal Compound Batch ID	Internal Compound Batch ID	VARCHAR	R	N	System generated batch identifier (e.g., DDIM-grit42 platform)

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
7	EXT_CPD_ID	External Compound ID	External Compound ID	VARCHAR	R	N	Compound Identifier per Asset Owner
8	EXT_BATCH_ID	External Batch ID	External Batch ID	VARCHAR	R	N	Batch Identifier per Asset Owner
9	MODULE	Module Number	Module number as per pipeline specification	VARCHAR	R	N	Module number of the respective experiment to be entered e.g.,M1.3.3
10	APP_CPP	Asset or Combination Progression Plan Number	Asset or Combination Progression Plan Number	VARCHAR	R	N	Asset or Combination Progression Plan. For e.g., APP3 or CPP1
11	SUBJECT_ID	Individual animal identifier	Individual animal identifier	VARCHAR	R	N	Concatenation of ('EXPID','/' ,'ID') to make SUBJECT_ID Unique
12	ID	ID	Animal Identifier	VARCHAR	R	N	One unique value per animal inserted on all records. Consecutive ID number for all animals in particular experiment.

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
13	PMI	Internal Individual Animal Identifier	Internal Individual Animal Identifier	VARCHAR	O	N	To be used when there is also an internal id for Animal e.g., U7836
14	GROUP_ID	Group ID	Group Identifier used within an experiment.	VARCHAR	R	Y	Grouping Variable on the basis of Treatment and dosage level. e.g., TREATMENT-GROUP-1,CONTROL-VEHICLE-GROUP,CTRL-INFECTION-GROUP
15	INFECTION_DATE	Date of infection	Date of infection	DATE (YYYY-MM-DD)	R	N	Date of infection in ISO8601 format
16	CFUINF	CFU Infection	CFU Infection	VARCHAR	R	N	Log10 CFU load used for infection.
17	DOSING_DATE	Date of Dosing	Date of Dosing	DATE (YYYY-MM-DD)	R	N	Date on which treatment dosage were given animal post infection
18	DOSING_TP	Dosing Timepoint	Dosing Timepoint	TIME (HH:MM)	R	N	Dosing timepoints in clock time Format: 24hr Clock e.g., 13:14, 09:10

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
19	SACRIFICE_DATE	Date of Sacrifice	Date of Sacrifice	DATE (YYYY-MM-DD)	R	N	Date on which animal was sacrificed
20	REGIMEN	Regimen	Treatment Regimen as defined by protocol	VARCHAR	R	N	Randomized to regimen (Insert label of each study arm which also should include dose level separated by "-".).
21	COMPOUND_1	Compound 1 Name	Name of Compound Administered	VARCHAR	R	N	Name of the first compound used within a treatment group.
22	COMPOUND_1_DOSE	Dose of Compound 1	Dose of the first or only compound used within a treatment group.	NUM	R	N	Dose of the first or only compound used within a treatment group.
23	DOSE_UNIT	Dose Unit	Unit of the first or only compound used within a treatment group.	VARCHAR	R	Y	Normalized Unit for Dose administered irrespective of compound type e.g., mg/kg
24	FORM	Formulation / vehicle of compound	Formulation / vehicle of compound	VARCHAR	O	N	e.eg. 15% Solutol HS 15 in 50 mM Na-phosphate buffer pH6.5
25	ADMN_ROUTE	Route of administration	Route through which treatment drug was given	VARCHAR	R	N	e.g., oral gavage
26	WEIGHT	Weight	Body weight of animal	VARCHAR	O	N	Individual bodyweight per animal. If weight was not measured, fill with '-99'

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
27	WEIGHT_UNIT	Unit of Weight Measurement	Unit for body weight of animal	VARCHAR	O	Y	
28	AGE	Age	Age of animal	VARCHAR	R	N	Individual age per animal.
29	AGE_UNIT	Unit of Subjects Age	Unite of age	VARCHAR	R	N	e.g., Days, Weeks
30	FEED	FEED	Feed status	VARCHAR	R	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
31	SEVER	Severity or extent of disease.	Severity or extent of disease.	VARCHAR	R	Y	UNAFFECTED, SLIGHTLY AFFECTED (less curious, slower movements, slightly ruffled fur), AFFECTED (prefers to be stationary, ruffled fur), CLEARLY AFFECTED (moves only when manipulated, ruffled fur, half closed eyes), HIGHLY AFFECTED (does not moves when manipulated, ruffled fur, closed eyes, cooler to the touch. Euthanize.)
32	COMMENT	Comments	Any additional pertinent information regarding an assessment/experiment which is useful for the interpretation of the data or to provide value if not present in dictionary	VARCHAR	O	N	See also **Section 3.4** regarding "Not in List" scenarios.

4.2.3 PK

This tab is dedicated for the purpose of capturing Pharmacokinetic (PK) results of various animal speciies over multiple time points and multiple dosing level (regimen) as defined in protocol for the specific study/experiment.

Table 9. WP3 Source Template - PK

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	PROJECT_ID	Project ID	Project Identifier	VARCHAR	R	N	e.g., ERA4TB
2	PROTOCOL	Protocol ID	Name of protocol (provide as PDF)	VARCHAR	R	N	Document name describing the experimental protocol.
3	STUDY_ID	Site-specific Study Identifier	Site-specific Study Identifier	VARCHAR	R	N	e.g., Job ID for ERA4TB progression plan activities
4	EXPID	Site-specific Experiment Identifier	Experiment number (e.g., from ELN)	VARCHAR	R	N	Recommended to be used to ensure provenance and traceability
5	CPD_ID	Internal Compound ID	Internal Compound ID	VARCHAR	R	N	System generated compound identifier (e.g., DDIM- grit42 platform)
6	BATCH_ID	Internal Compound Batch ID	Internal Compound Batch ID	VARCHAR	R	N	System generated batch identifier (e.g., DDIM-grit42 platform)

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
7	EXT_CPD_ID	External Compound ID	External Compound ID	VARCHAR	R	N	Compound Identifier per Asset Owner
8	EXT_BATCH_ID	External Batch ID	External Batch ID	VARCHAR	R	N	Batch Identifier per Asset Owner
9	MODULE	Module Number	Module number as per pipeline specification	VARCHAR	R	N	Module number of the respective experiment to be entered. For e.g.M1.3.3
10	APP_CPP	Asset or Combination Progression Plan Number	Asset or Combination Progression Plan Number	VARCHAR	R	N	Progression Plan. For e.g., APP3, CPP1
11	SUBJECT_ID	Individual Animal Identifier	Individual animal identifier	VARCHAR	R	N	Concatenation of ('EXPID','/' ,'ID') to make SUBJECT_ID Unique
12	ID	ID	Animal Identifier	VARCHAR	R	N	One unique value per animal inserted on all records. Consecutive ID number for all animals in particular experiment.
13	PMI	Internal Individual Animal Identifier	Internal Individual Animal Identifier	VARCHAR	O	N	To be used when there is also an internal id for

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
							Animal e.g., U7836
14	GROUP_ID	Group ID	Group Identifier used within an experiment.	VARCHAR	R	Y	Grouping Variable on the basis of Treatment and dosage level. E.g., TREATMENT-GROUP-1, CONTROL-VEHICLE-GROUP, CTRL-INFECTION-GROUP
15	PK_SAMPLING_DATE	Date of PK sampling	Date of PK sampling	DATE (YYYY-MM-DD)	R	N	Date of PK sampling in ISO8601 format
16	PK_SAMPLING_TP_CT	PK sampling timepoints	PK sampling timepoints	TIME (HH:MM)	R	N	Time of PK Sampling in clock time in format HH:MM
17	PK_SAMPLING_TP_PD	PK sampling timepoints post-dose	Route sampled	VARCHAR	R	N	PK sampling timepoints post-dose (for example: 0.1, 0.5, 1, 2, 4, 7 h post-dose)

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
18	UNIT_SAMPLING_TP	Unit for sampling time	Unit for sampling time	VARCHAR	R	N	e.g., hours.
19	COMPOUND_1	Compound 1 Name	Name of Compound Administered	VARCHAR	R	N	Name of the first compound used within a treatment group.
20	COMPOUND_1_CONC	Concentration of Compound 1	Concentration of Compound 1	NUM		N	For missing information, insert "-99". If below limit of quantification (LLOQ), insert reported value or if unavailable LLOQ.
21	COMPOUND_1_DOSE	Dose of Compound 1	Dose of the first or only compound used within a treatment group.	NUM	R	N	Dose of the first or only compound used within a treatment group.
22	ACTIVE_COMPOUND	Active Compound Name	Active Compound Name	VARCHAR	R	N	e.g., Moxifloxacin
23	ACTIVE_COMPOUND_CONC	Concentration of Active Compound	Concentration of Active Compound	NUM	R	N	For missing information, insert "-99". If below limit of quantification (LLOQ), insert reported value or if unavailable LLOQ.

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
24	CONC_UNIT	Unit for Blood concentration	Normalized Unit for Blood concentration irrespective of blood concentration type e.g., ng/mL	VARCHAR	R	N	Normalized Unit for Blood concentration irrespective of blood concentration type e.g., ng/mL Unit of COMPOUND_1 should not be expressed as ACTIVE_COMPOUND equivalent
25	DOSE_UNIT	Dose Unit	Unit of the first or only compound used within a treatment group.	VARCHAR	R	N	Normalized Unit for Dose administered irrespective of compound type e.g., mg/kg
26	SAMPLING_LOCATION	Route sampled	Route sampled	VARCHAR	R	N	e.g., tail vein, jugular vein
27	SAMPLE_TESTED	Type of sample tested	Type of sample tested	VARCHAR	R	N	e.g., Blood/Plasma/Lung
28	STUDY_DESIGN	Study design type	Study design type	VARCHAR	O	N	e.g., sparse-first dose, rich-steady state
29	REGIMEN	Regimen	Treatment Regimen as defined by protocol	VARCHAR	R	N	Randomized to regimen (Insert label of each study arm which also should include dose level separated by "-".).

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
30	FEED	FEED	Feed status	VARCHAR	R	Y	
31	SEVER	Severity or extent of disease.	Severity or extent of disease.	VARCHAR	R	Y	UNAFFECTED, SLIGHTLY AFFECTED (less curious, slower movements, slightly ruffled fur), AFFECTED (prefers to be stationary, ruffled fur), CLEARLY AFFECTED (moves only when manipulated, ruffled fur, half closed eyes), HIGHLY AFFECTED (does not moves when manipulated, ruffled fur, closed eyes, cooler to the touch. Euthanize.)
32	COMMENT	Comments	Any additional pertinent information regarding an assessment /experiment which is useful for the interpretation of the data	VARCHAR	O	N	See also **Section 3.4** regarding "Not in List" scenarios.

4.2.4 PD

This sheet is used for the purpose of capturing Pharmocodynimac (PD) results (e.g., *CFU* Count) over multiple time points and multiple dosing level (Regimen)) as defined per protocol for the specific study/experiment.

Table 10. WP3 Source Template - PD

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
1	PROJECT_ID	Project ID	Project Identifier	VARCHAR	R	N	e.g., ERA4TB
2	PROTOCOL	Protocol ID	Name of protocol (provide as PDF)	VARCHAR	R	N	Document name describing the experimental protocol.
3	STUDY_ID	Site-specific Study Identifier	Site-specific Study Identifier	VARCHAR	R	N	e.g., Job ID for ERA4TB progression plan activities
4	EXPID	Site-specific Experiment Identifier	Experiment number (e.g., from ELN)	VARCHAR	R	N	Recommended to be used to ensure provenance and traceability
5	CPD_ID	Internal Compound ID	Internal Compound ID	VARCHAR	R	N	System generated compound identifier (e.g., grit42 platform)
6	BATCH_ID	Internal Compound Batch ID	Internal Compound Batch ID	VARCHAR	R	N	System generated batch identifier (e.g., grit42 platform)

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
7	EXT_CPD_ID	External Compound ID	External Compound ID	VARCHAR	R	N	Compound Identifier per Asset Owner
8	EXT_BATCH_ID	External Batch ID	External Batch ID	VARCHAR	R	N	Batch Identifier per Asset Owner
9	MODULE	Module Number	Module number as per pipeline specification	VARCHAR	R	N	Module number of the respective experiment to be entered. For e.g.M1.3.3
10	APP_CPP	Asset or Combination Progression Plan Number	Asset or Combination Progression Plan Number	VARCHAR	R	N	Progression Plan. For e.g., APP3, CPP1
11	SUBJECT_ID	Individual animal identifier	Individual animal identifier	VARCHAR	R	N	Concatenation of ('EXPID','/' ,'ID') to make SUBJECT_ID Unique
12	ID	ID	Animal Identifier	VARCHAR	R	N	One unique value per animal inserted on all records. Consecutive ID number for all animals in particular experiment.
13	PMI	Internal Individual Animal Identifier	Internal Individual Animal Identifier	VARCHAR	O	N	To be used when there is also an internal id for Animal e.g., U7836

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
14	GROUP_ID	Group ID	Group Identifier used within an experiment.	VARCHAR	R	Y	Grouping Variable on the basis of Treatment and dosage level. e.g., TREATMENT-GROUP-1,CONTROL-VEHICLE-GROUP,CTRL-INFECTION-GROUP
15	DATE_PD	PD Date	PD Date	DATE (YYYY-MM-DD)	R	N	PD Date in ISO8601 format
16	PD_Time_CT	Clock-time of euthanizing the animal for CFU determination.	Clock-time of euthanizing the animal for CFU determination.	TIME (HH:MM)	R	N	Clock-time of euthanizing the animal for CFU determination in 24-hour format HH:MM
17	TISSUE	Tissue analysed	Tissue analysed	VARCHAR	R	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
18	CFU_COUNT	CFU Count	Log10 CFU given for each day and replicate.	VARCHAR	R	N	Dependent variable. Log10 CFU given for each day and replicate. For missing information, insert "-99". If below limit of quantification (LLOQ), insert reported value or if unavailable LLOQ.
19	CFU_UNIT	Unit For CFU Count	Unit For CFU Count	VARCHAR	R	N	Unit For CFU Count
20	CFU_STATUS	Sample status for CFU counts	Sample status for CFU count	VARCHAR	Y	Y	
21	TTP_CFU	Time To Positivity for CFU Samples	Time To Positivity for CFU Samples	VARCHAR	R	N	Dependent variable. TTP given for each day and replicate in hours. If negative at day 42, then the sample is negative/No Growth and write 9999

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
22	TTP_CFU_UNIT	Unit of TTP	Unit of Time to Positivity for CFU Samples	VARCHAR	R	N	Unit of Time to Positivity for CFU Samples
23	REPLICATE	Replicate number	Replicate number.	NUM	R	N	Inserted on all records as 1 or 2 and so on
24	REGIMEN	Regimen	Treatment Regimen as defined by protocol	VARCHAR	R	N	Randomized to regimen (Insert label of each study arm which also should include dose level separated by "-".).
25	WEIGHT	Weight	Body weight of animal	VARCHAR	O	N	Individual bodyweight per animal. If weight was not measured, fill with '-99'
26	WEIGHT_UNIT	Unit of Weight Measurement	Unit for body weight of animal	VARCHAR	O	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
27	SEVER	Severity or extent of disease.	Severity or extent of disease.	VARCHAR	R	Y	UNAFFECTED, SLIGHTLY AFFECTED (less curious, slower movements, slightly ruffled fur), AFFECTED (prefers to be stationary, ruffled fur), CLEARLY AFFECTED (moves only when manipulated, ruffled fur, half closed eyes), HIGHLY AFFECTED (does not moves when manipulated, ruffled fur, closed eyes, cooler to the touch. Euthanize.)
28	INCUBATION_TIME	Culture Incubation Time (days)	Culture Incubation Time (days)	VARCHAR	R	N	Culture Incubation Time in number of Days
29	MEDIATYPE	Culture Media Type	Culture Media Type	VARCHAR	R	N	e.g., 7H11 Agar + OADC + PANTA
30	COMMENT	Comments	Any additional pertinent information regarding an assessment/experiment which is useful for the interpretation of the data or to provide value if not present in dictionary	VARCHAR	O	N	See also **Section 3.4** regarding "Not in List" scenarios.

4.3 Terminologies

The dictionaries/terminologies for the fields indicated in **Section 4.2** are maintained by the COMBINE project group within the AMR Accelerator program and can be found within the AMR Knowledge Space.

In a scenario where a field/column with a defined dictionary in the WP3 Source Template does not have an appropriate value in the dictionary list, the relevant partner should raise this item to the assigned data management group within **WP1** for discussion. Based on the discussion, the outcome will be one of the following:

- Update of the data dictionary with new option(s) included ensuring alignment with recognised ontologies (e.g., NCIT) where possible.
- Proceed as per **Section 3.4**

4.4 "Not in List" Scenarios

The conventions for "Not in List" scenarios follow the same recommended approach as per **Section 3.4.**

4.5 File Naming Conventions

The file naming convention for In Vivo data follows the same recommended approach as per **Section 3.5**

4.6 Future Consideration

This data generated for these experiments may require the same approach to be considered as described in **Section 3.6** .

5. Imaging Data (Work Package 4)

Subject to the conditions outlined in **Section 2.2**, further information is required from all WP4 partners to progress the development of templates to support the reporting of imaging experiments. Further details on this topic will be described in deliverable D1.19.

6. GLP Toxicity/Safety Data (Work Package 6)

The data generated under **WP6** is expected to be delivered in a CDISC SEND structure with controlled terminologies in line with CDISC standards. Non-GLP toxicity/safety data and Chemistry, Manufacturing and Controls (CMC) reports are expected to be available as references files in .DOC or .PDF formats, if requested/required by a specific partner within ERA4TB. The requirements to support data provenance will follow the approach described in **Section 7.2**.

In the event that additional data (e.g., non-GLP toxicity/safety data and/or CMC data) is required to be tabulated and integrated into the DDIM-grit42 platform. The development of such templates will be subject to the conditions outlined in Section 2.2. Such developments (if applicable) will be described in deliverable D1.19.

7. Clinical Data (Work Package 7)

7.1 Data Ingestion:

For the ingestion of clinical data into the DDIM-grit42 platform within the DDIM, it is currently the loaded data will follow the conventions as per CDISC SDTM standards with controlled terminologies in line with CDISC standards.

To support the navigation of clinical data within the DDIM-grit42 platform the hierarchy of data will be driven by the following core domains as per **Figure 1** below. This structure may develop over time subject to end user needs and further development of platform capabilities.

Figure   1. DDIM-grit42 Clinical Data Hierarchy
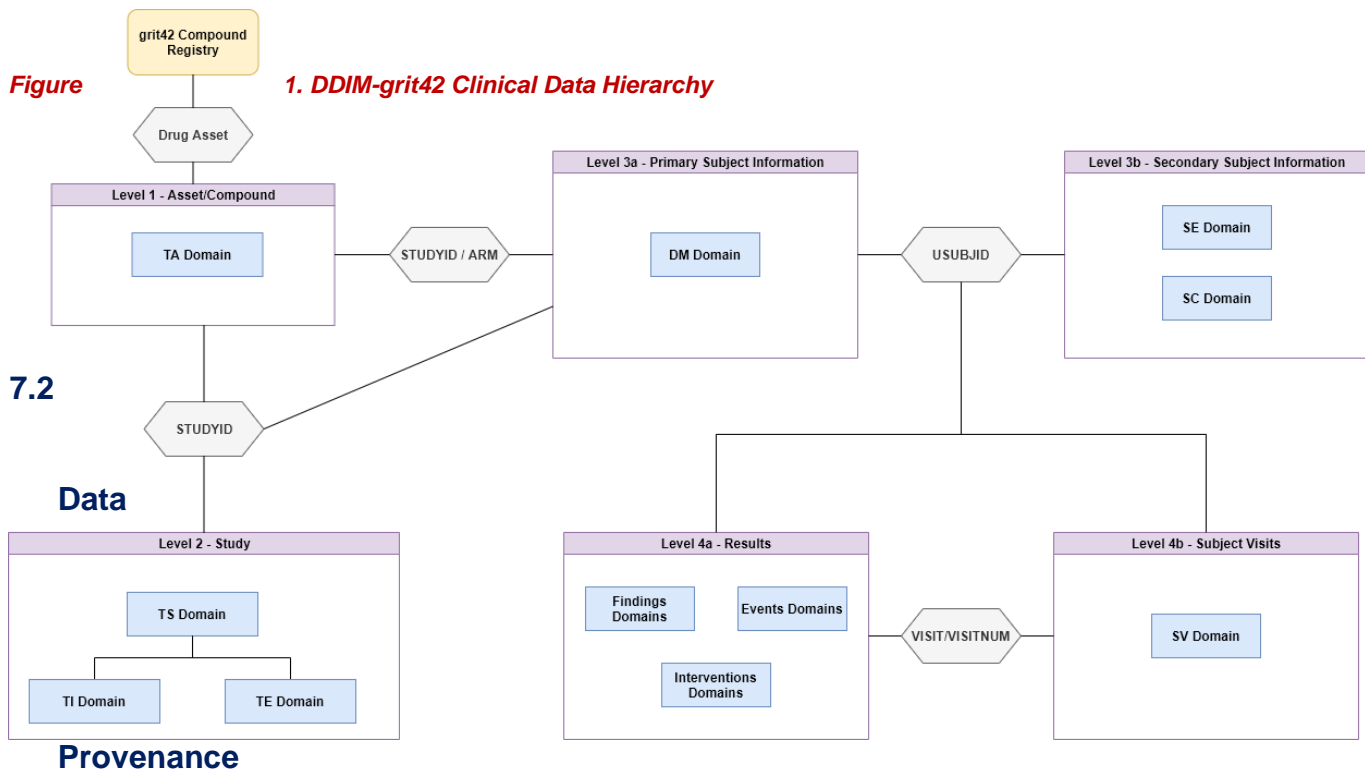
7.2 Data Provenance

In addition to the provision of study, subject and result/event information per treatment arm, further details are required to be reported in order to support data provenance. It is currently planned to utilise the Trial Summary (TS) domain to report such details as a dedicated parameter. The table below provides the list of parameters anticipated to be incorporated into the TS domain to support provenance. This data may be included directly into the TS domain by the data provider or may be merged into the TS Domain using a to be agreed template or flat file

Table 11. Trial Summary Provenance Parameters

Trial Summary Parameter Name [TSPARM]	Trial Summary Parameter Name [TSPARMCD]	NCIT Reference	Description
Progression Plan Reference	PPREF	N/A	Name, date, and version of Asset or Combination progression plan that the clinical study belong too. Only applicable to studies conducted within the ERA4TB project
Module	MODULE	http://purl.obolibrary.org/obo/NCIT_C42721	Module number as per Full Pipeline Specification (D1.1).Only applicable to studies conducted within the ERA4TB project
Provenance Contact	PROVCONTACT		Contact details of data owner
Provenance Link	PROVLINK	http://purl.obolibrary.org/obo/NCIT_C43581	Location where data owner is storing the study data
Provence Source Data	PROVDATA		Delimited list of all raw data files used to generate the final data outputs
Protocol Link	PROTLINK	N/A	URL link to Study Protocol

APPENDIX I – Document and Resource Links

Resource/Documents	Link(s)	Additional Information
CDISC Standards	SDTM (General): http://www.cdisc.org/sdtm SDTM (TB Standards): https://www.cdisc.org/standards/therapeutic-areas/tuberculosis SEND: http://www.cdisc.org/send ADAM: http://www.cdisc.org/adam	
ERA4TB Website	https://era4tb.org/	Public website for ERA4TB
ERA4TB Consortium Agreement	https://synapsemanagers.sharepoint.com/:b:/r/sites/ETB-RA/Shared%20Documents/Consortium%20Agreement%20-%20ERA4TB%20-%20Execution%20version.pdf?csf=1&e=IVnee5	Access managed by Synapse, only accessible to authorised members
ERA4TB Full Pipeline Specification (D1.1)	https://synapsemanagers.sharepoint.com/sites/ETB-RA/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FETB%2DRA%2FShared%20Documents%2FSubmitted%20deliverables%2FD1%2E1%5FFull%20pipeline%20specifications%5FV0%2E4%5FFINAL%2Epdf&parent=%2Fsites%2FETB%2DRA%2FShared%20Documents%2FSubmitted%20deliverables	Access managed by Synapse, only accessible to authorised members
ERA4TB Project Handbook	https://synapsemanagers.sharepoint.com/:b:/r/sites/ETB-RA/wp8/Shared%20Documents/Deliverables/D8.1_Project%20Handbook_v1.2_Final.pdf?csf=1&web=1&e=dhFZt2	Access managed by Synapse, only accessible to authorised members
ERA4TB Data Management Plan (D1.2)	https://synapsemanagers.sharepoint.com/:b:/r/sites/ETB-RA/Shared%20Documents/Submitted%20deliverables/D1.2_Data%20Management%20Plan_v1.0_FINAL.pdf?csf=1&web=1&e=AmRX4V	Access managed by Synapse, only accessible to authorised members
FAIR	https://www.go-fair.org/fair-principles/	Resource on FAIR data principles
NCI Thesaurus	http://www.ontobee.org/ontology/NCIT	NCI Thesaurus OBO Edition