IMI2 GA853989 - ERA4TB

European Regimen Accelerator for Tuberculosis

WP1 – Data and Pipeline Management

D1.4 Initial standardized data templates for collection and reporting of clinical and preclinical data available to Consortium members

Lead contributor	Patrick O'Meara (8 – Critical Path Institute)
Other contributors	Rick Liwski (8 – Critical Path Institute)

Due date	31/08/2020
Delivery date	23/12/2020
Deliverable type	R
Dissemination level	PU

Document History

Version	Date	Description
0.1	23/10/2020	Initial draft based on current project feedback/updates.
0.2	11/12/2020	Updates based on WP1 feedback for SC review

Table of Contents

Abstract

This deliverable describes the current development and initial data standards/templates intended to be 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 the following topics; (1) means to support data provenance per data template or standard; (2) preclinical templates for the reporting of In Vitro, In Vivo and Imaging data; (3) overview of the use of CDISC SDTM and SEND for other defined sources of data generated.

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 D1.8. This document (and deliverable D1.8) will act as a supplement to the 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 (IMI JU) 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 initial data reporting requirements to deliver data into the ERA4TB DDIM platform specifically. The core objective for this deliverable is to create a set of standards which aspire to meet the following tenets:

- Alignment with FAIR data principles
- Ease of use across the AMR accelerator programs
- Key data elements are reported to meet the end user requirements within the ERA4TB DDIM
- Alignment with the architectural requirements of the grit42 data capture platform

2. Preface on Data Standards

The development of the initial data standards is a result of an ongoing collaboration 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. 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 are controlled and managed within the COMBINE project by Frauenhofer using their OwnCloud platform (https://owncloud.fraunhofer.de/) . Access to these documents are managed by the COMBINE team at Frauenhofer, and access can be given to partners involved in the AMR Accelerator program.

During the lifecycle of the AMR Accelerator program, and more specifically for this deliverable the ERA4TB project, an agile approach will likely be adopted to further inform and enhance the data standards. Factors that can inform future updates/ modifications of these standards include:

- Information on the level of data reported per experiment received by the relevant partners
- Data consumer requirements to meet downstream needs (e.g. data modelling, pipeline management etc.)
- Changes in regulatory requirements

Upon identifying a need for update/modification, this item will be discussed with members of the AMR Data Group to come to a consensus on how 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 (if applicable). In the event of significant changes being required, these will be documented as part of deliverable D1.8.

3. In Vitro Data (Work Package 2)

The initial focus of the AMR Accelerator program is on aggregated preclinical data sets. “Aggregated” means the combined results of an experiment, not data from an individual experimental subject (e.g. a well or a plate). When providing the results file, the relevant experimental protocol (preferably in PDF format) should also be included prior to delivery to the data management group in ERA4TB.

3.1 Template Overview

This template should be used for the preparation of experimental results (summarized data) for delivery into the grit42 system within the DDIM for ERA4TB. The data worksheet also requires the entry of information to support data provenance. The most current version of the worksheet can be accessed within the AMR Knowledge Space managed by Frauenhofer on behalf of the COMBINE project.

The current template is not purposed for the following:

- Collection of raw data
- Well based data
- Single experimental repeats

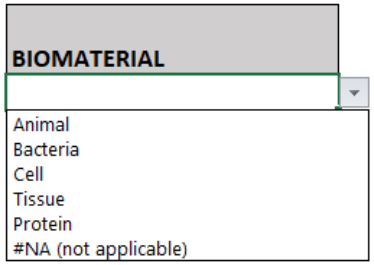
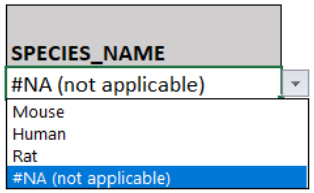
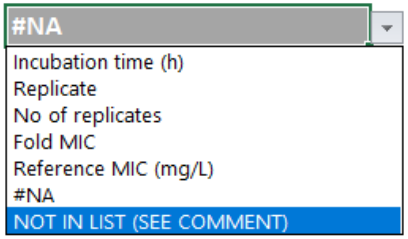
All raw results should be maintained and traceable per experiment identifier (if available) and stored in an electronic laboratory notebook (ELN) or equivalent platform at site. Subject to the capabilities of the ELN or equivalent platform at site, it is recommended that the data worksheet template is incorporated into the platform to support ease of reporting.

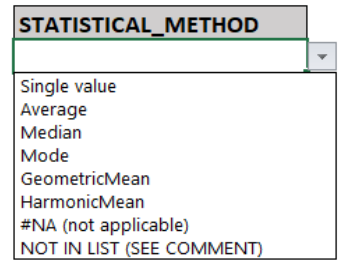
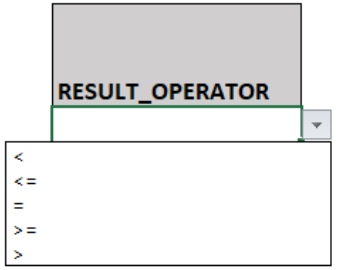
3.2 Template Content

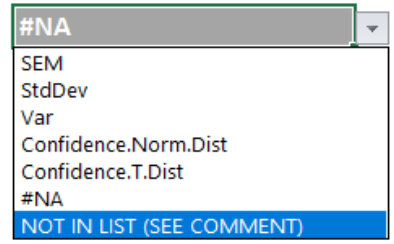
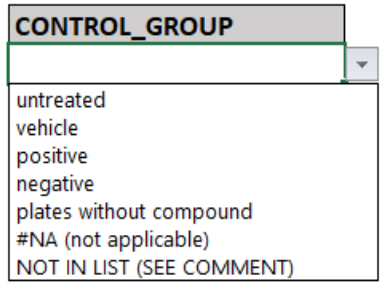
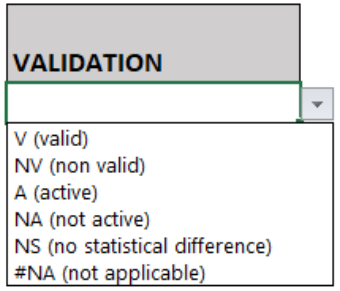
The table in the subsequent pages provides a more detailed overview of the data worksheet tab content within the preclinical template document. The table represents the current iteration of the template (version 4.0).

Prior to reporting experimental results, it is recommended that the Frauenhofer OwnCloud location as described in **Section 3.1** is accessed to verify whether a partner is working with the current template. In the event of doubt, the assigned data management contact(s) in Work Package 1 **(WP1**) can be consulted.

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	
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	Compound Identifier per Asset Owner
4	BATCH_ID	Internal Compound Batch ID	Internal Batch ID	VARCHAR	R	N	Compound Identifier per Asset Owner
5	EXT_CPD_ID	External Compound ID	External Compound ID	VARCHAR	O	N	System generated identifier (e.g. grit42 platform)
6	EXT_BATCH_ID	External Batch ID	External Batch ID	VARCHAR	O	N	System generated identifier (e.g. grit42 platform)
7	SITE	Site Identifier	Site/Lab where assessment/experiment was conducted	VARCHAR	R	N	Entry should be concise (e.g. partner abbreviations to be used)
8	PROVENANCE	Data Provenance Information	Internal Contact information for additional information on the experiment	VARCHAR	R	N	Asset Progression Plan and Module Reference. Contact details to support data provenance (e.g. e mail address), location of source/raw data files etc.
9	EXPERIMENT_DATE	Date of Assessment/Experiment	Date of Assessment/Experiment	DATE (YYYY-MM-DD)	R	N	ISO 8601 format

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
10	BIOMATERIAL	Biomaterial	Biomaterial used for the experiment	VARCHAR	R	Y	
11	SPECIES_NAME	Species	If animal, cell or tissue was selected for BIOMATERIAL, then specify species	VARCHAR	R	Y	
12	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.
13	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.
14	PROTOCOL_NAME	Protocol	Name of protocol (provide as PDF)	VARCHAR	R	N	Document name describing the experimental protocol.
15 - 17	#NA	Factors	These columns can be used to add experimental factors. Please choose a column name from the drop-down list	VARCHAR/NUM	O	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
18	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.
19	STATISTICAL_METHOD	Statistical method	Statistical method used to calculate the result (e.g. Average, Mode, Median)	VARCHAR	R	Y	
20	RESULT_OPERATOR	Result Operator	Result Operator	VARCHAR	R	Y	
21	RESULT_VALUE	Result	Experiment Result	VARCHAR/NUM	R	N	Format to be remain consistent per experiment/assessment
22	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.

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
23	#NA	Variation	For aggregated values choose a method to report deviations	NUM	O	Y	
24	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.
25	CONTROL_GROUP	Control Group	Calculation of the result type based on which type of control	VARCHAR	R	Y	
26	VALIDATION	Validation	Validation Status of Result	VARCHAR	R	Y	

Column #	Variable	Label	Description	Format	Required (R) or Optional (O)	Dictionary (Y/N)?	Additional Information
27	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.

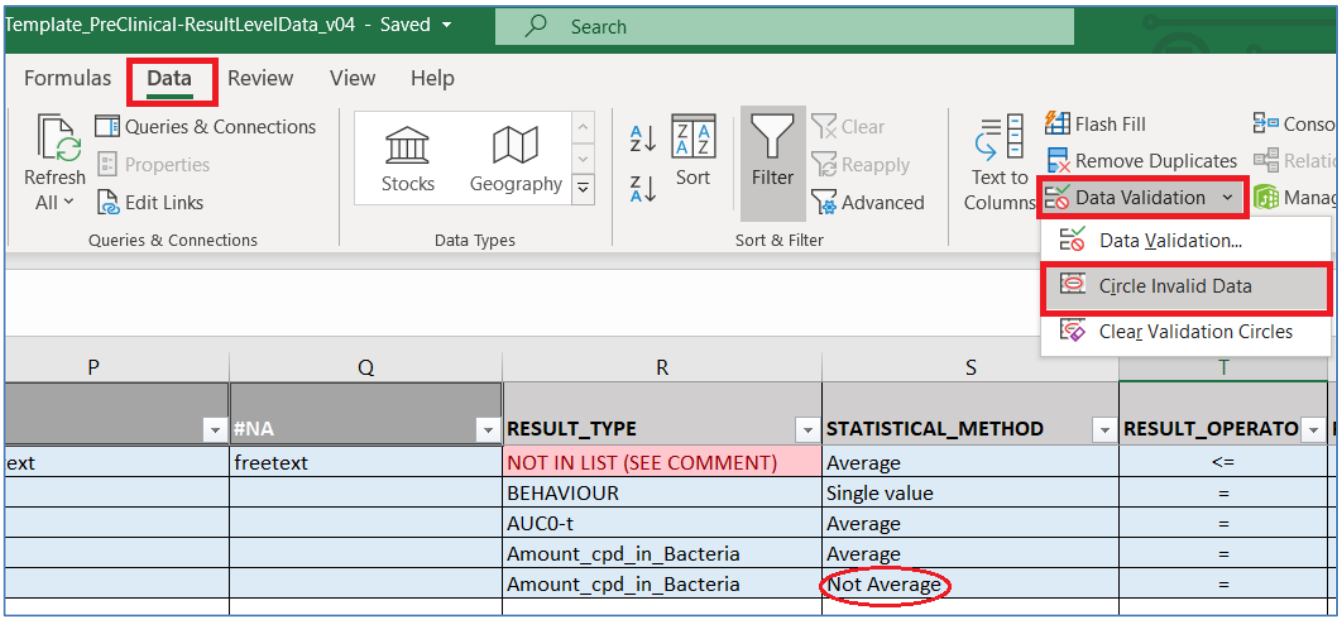
3.3 Terminologies

The dictionaries/terminologies for the fields indicated in **Section 3.2** are maintained by the COMBINE project group within the AMR Accelerator program and can be found within the AMR Knowledge Space .

In the event that a field with a defined dictionary does not have an appropriate option, the relevant partner should raise this item to the assigned data management group for discussion. Subject to this discussion, the outcome will be one of the following:

- Update of the data dictionary with new option(s) included
- Proceed as per **Section 3.4**

To ensure that the reported entries are aligned with the current dictionaries, it is recommended that prior to delivery to the WP1 data management team, the partner preparing the results should select “Circle Invalid Data” and correct any items highlighted as invalid (see example below).



3.4 "Not in List" Scenarios

For fields 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.

- Scenario 1: Single field affected per record can be written using the convention:

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

EXPERIMENT	COMMENTS
NOT IN LIST (SEE COMMENT)	EXPERIMENT: Test X

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

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

EXPERIMENT	RESULT_TYPE	COMMENTS
NOT IN LIST (SEE COMMENT)	NOT IN LIST (SEE COMMENT)	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)

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

As within other IMI projects, naming conventions will help members of ERA4TB to have an overview of existing data and fast access to the data stored on the different repositories (e.g. grit42). It is expected that the partners providing data will keep their own regularly updated inventory listing the studies carried out at their institution and the data supplied to the data management team. The list can then be used for cross referencing purposes.

For the grit42 platform, the assessment/experiment is the essential base-level record which can include multiple documents and data forms/files. The grit42 platform is used to capture in vitro results within the DDIM, the higher levels of study or project may be less relevant than compound ID or regimen, which can be used as connectors between experiments. To ensure consistency and findability, the following filename convention for results generated is recommended:

File Naming Convention	Example
<Project>_ <Partner/Provider>_<Compound/ Regimen>_ <APP #>_ <Experiment ID>_ <Experiment_Type>_ <yyyy-mm-dd>_ <Sequence Number>	ERA4TB_UNIZAR_GSK286_APP1_123-098_TKA_2020-07-01_01

3.6 Future Considerations

Subject to the conditions outlined in Section 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 D1.8 and the Data Management Plan

4. In Vitro Data (Work Package 3)

For animal studies, there is a currently more of a need to compare data on an individual animal level rather than just on a study or experiment level. To meet such a need, a study level data template and an individual animal level data template have been developed by the AMR Data Group.

4.1 Template Overview

This template should be used for the preparation of animal level results and the associated metadata for delivery to the grit42 system within the DDIM for ERA4TB. The most current version of the worksheet can be accessed at the AMR Knowledge Space.

The template is structured into metadata worksheets and a worksheet for animal data which is comprised of a combination study related data (from the metadata worksheets) and manually added individual animal related data (e.g. result data).

As per **Section 3**, it is recommended an ELN or equivalent platform is utilised to trace and maintain all raw data that has supported the generation of the results for a specific experiment.

The current template is not purposed for the following:

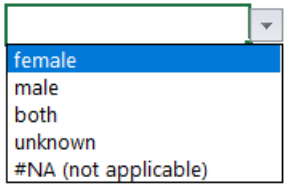
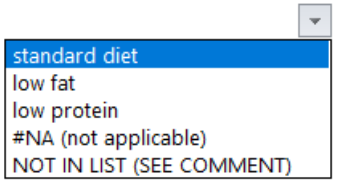
- Raw data
- Clinical Data

4.2 Template Content

4.2.1 Study Info

Information on this tab provides an overview of the study/experiment in question. The inputs reported are subsequently made available within the Animal Data tab of the Animal Level Results template.

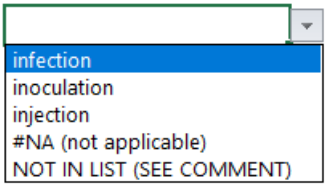
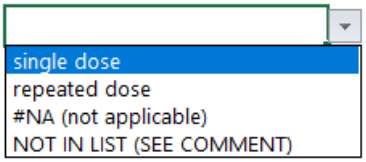
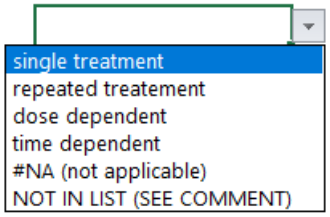
Column #	Category	Variable	Format	Required (R), Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
1	Experiment	project ID	VARCHAR	O	N	
2	Experiment	experiment ID	VARCHAR	R	N	
3	Experiment	protocol ID	VARCHAR	O	N	
4	Experiment	link to protocol		R	N	A written summary of the experimental design including all deviations from the planned setup. A protocol can consist of an SOP/WI & Deviation list
5	Experiment	test drug vendor	VARCHAR	O	N	
6	Experiment	type of experiment	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
7	Experiment	start date	DATE (YYYY-MM-DD)	R	N	ISO 8601 format
8	Experiment	PI/site/contact information	VARCHAR	R	N	Contact information to support data provenance
9	Animal	species	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
10	Animal	Strain	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please

Column #	Category	Variable	Format	Required (R), Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
						discuss your data management contact.
11	Animal	Sex	VARCHAR	R	Y	
12	Animal	age in weeks as a range	VARCHAR	R	N	
13	Animal	individual body weight in gram	NUM	C	N	To be reported if individual animal weights have been taken
14	Animal	mean body weight in gram	NUM	C	N	To be reported if no individual animal weights have been taken
15	Animal	body weight range in gram	NUM	C	N	To be reported if no mean body weight is reported
16	Animal	Vendor	VARCHAR	R	N	
17	Housing	Feed	VARCHAR	R	Y	
18	Housing	supplements	VARCHAR	O	N	
19	Housing	restricted/unrestricted feeding	VARCHAR	C	N	Required for Oral Studies
20	Housing	place of animal experiment	VARCHAR	O	N	

Column #	Category	Variable	Format	Required (R), Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
21	Housing	light/dark cycle	VARCHAR	O	N	
22	Housing	cage size	VARCHAR	O	N	
23	Housing	no animals per cage	NUM	O	N	

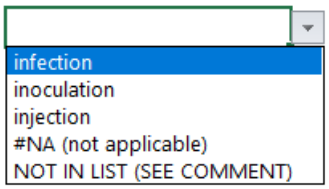
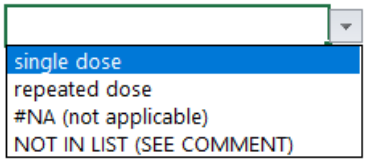
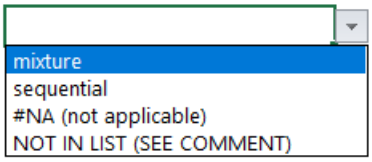
4.2.2 Pre-Treatment

Information on this tab provides details on pre-treatment . The inputs reported are subsequently made available within the Animal Data tab of the Animal Level Results template.

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
1	Pre-Treatment	time-points	VARCHAR	C	N	Required if no duration reported
2	Pre-Treatment	duration or start-end date	VARCHAR	C	N	Required if no timepoint reported
3	Pre-Treatment	type of pre-treatment	VARCHAR	R	Y	
4	Pre-Treatment	dosing information	VARCHAR	R	Y	
5	Pre-Treatment	concentration	NUM	R	N	
6	Pre-Treatment	unit of concentration	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
7	Pre-Treatment	frequency	VARCHAR	R	Y	
8	Pre-Treatment	route of administration	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
9	Pre-Treatment	internal test drug ID	VARCHAR	C	N	Required for non-generic drugs
10	Pre-Treatment	generic test drug ID	VARCHAR	C	N	Required for generic drugs
11	Pre-Treatment	batch/lot	VARCHAR	R	N	
12	Pre-Treatment	pathogen species & strain ID	VARCHAR	C	N	Required for inoculation studies

4.2.3 Treatment

Information on this tab provides details on study treatment . The inputs reported are subsequently made available within the Animal Data tab of the Animal Level Results template.

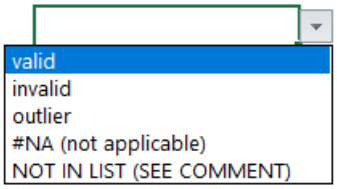
Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
1	Treatment	time-points	VARCHAR	C	N	Required if no duration reported
2	Treatment	duration or start-end date	VARCHAR	C	N	Required if no timepoint reported
3	Treatment	type of treatment	VARCHAR	R	Y	
4	Treatment	dosing information	VARCHAR	R	Y	
5	Treatment	concentration	NUM	R	N	
6	Treatment	unit of concentration	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
7	Treatment	volume applied	NUM	R	N	
8	Treatment	unit volume applied	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
9	Treatment	frequency	VARCHAR	R	Y	
10	Treatment	route of administration	VARCHAR	R	Y	
11	Treatment	regimen of drug combinations	VARCHAR	R	Y	
12	Treatment	internal test drug ID	VARCHAR	C	N	

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
13	Treatment	generic test drug ID	VARCHAR	C	N	
14	Treatment	batch/lot	VARCHAR	R	N	
15	Treatment	pathogen species & strain ID	VARCHAR	C	Y	Required for infection studies
16	Treatment	replicates	NUM	R	N	

4.2.4 Read Out

This tab of the template defines the medium, conditions for sacrifice and expected result measurements per timepoint. The inputs reported are subsequently made available within the Animal Data tab of the Animal Level Results template.

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
1	Termination	sacrifice	VARCHAR	R	Y	Reason for sacrifice or end of experiment
2	Condition	medium	VARCHAR	R	Y	Please refer to full dictionary. If appropriate option not listed, please discuss your data management contact.
3	Readout	Tissue analysed	VARCHAR	R	Y	
4	Readout	Unit 1	VARCHAR	R	Y	Define Unit for experiment readout
5	Readout	time-points 1	VARCHAR	R	Y	Provide list of timepoints per experiment
6	Readout	Value 1	NUM/VARCHAR	R	N	
7	Readout	Unit 2	VARCHAR	C	Y	Required if multiple readouts per experiment

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Information
8	Readout	time-points 2	VARCHAR	C	Y	
9	Readout	Value 2	NUM/VARCHAR	C	N	
10	Readout	Unit 3	VARCHAR	C	Y	Required if multiple readouts per experiment
11	Readout	time-points 3	VARCHAR	C	Y	
12	Readout	Value 3	NUM/VARCHAR	C	N	
13	Readout	Unit 4	VARCHAR	C	Y	Required if multiple readouts per experiment
14	Readout	time-points 4	VARCHAR	C	Y	
15	Readout	Value 4	NUM/VARCHAR	C	N	
16	Readout	Flag valid/invalid	VARCHAR	O	Y	

4.2.5 Animal Data

This tab collates the animal level results of the experiment. Specific inputs into the tabs as described in **Sections 4.2.1 – 4.2.4** are made available for use. It is recommended that this tab should only be completed once all inputs for **Sections 4.2.1 – 4.2.4** have been completed.

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Comments
1	Experiment	experiment ID	VARCHAR	See Additional Comments	N	Data is automatically populated based on inputs from Study Info tab (See **Section 4.2.1**)
2	Experiment	link to protocol	VARCHAR		N	
3	Experiment	type of experiment	VARCHAR		N	
4	Experiment	start date	DATE (YYYY-MM-DD)		N	
5	Experiment	PI/site/contact information	VARCHAR		N	
6	Animal	species	VARCHAR		N	
7	Animal	strain	VARCHAR		N	
8	Animal	sex	VARCHAR		N	
9	Animal	age in weeks as a range	VARCHAR		N	
10	Animal	mean body weight in gram	NUM		N	
11	Animal	body weight range in gram	VARCHAR		N	
12	Animal	vendor	VARCHAR		N	
13	Housing	feed	VARCHAR		N	
14	Housing	restricted/unrestricted feeding	VARCHAR		N	

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Comments
15	Animal	individual animal identifier	VARCHAR	R	N	Derived field. Unique subject identifier generated based on experiment ID and Animal No.
16	Animal	Group	VARCHAR	R	N	
17	Animal	Animal No.	VARCHAR	R	N	
18	Animal	Body weight	NUM	C	N	If not available, then see columns 14 and 15 of the Study Info tab (**Section 4.2.1**)
19	Animal	Cage No.	VARCHAR	O	N	
20	Pre-Treatment	time-points	VARCHAR	See Additional Comments	Y	Dictionary list generated based inputs provided in Pre-Treatment tab (**Section 4.2.2**)
21	Pre-Treatment	duration or start-end date	VARCHAR		Y	
22	Pre-Treatment	type of pre-treatment	VARCHAR		Y	
23	Pre-Treatment	dosing information	VARCHAR		Y	
24	Pre-Treatment	concentration	NUM		Y	
25	Pre-Treatment	unit of concentration	VARCHAR		Y	
26	Pre-Treatment	frequency	VARCHAR		Y	
27	Pre-Treatment	route of administration	VARCHAR		Y	
28	Pre-Treatment	internal test drug ID	VARCHAR		Y	
29	Pre-Treatment	generic test drug ID	VARCHAR		Y	

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Comments
30	Pre-Treatment	batch/lot	VARCHAR	See Additional Comments	Y	Dictionary list generated based inputs provided in Pre-Treatment tab (**Section 4.2.2**)
31	Pre-Treatment	pathogen species & strain ID	VARCHAR		Y	
32	Treatment	time-points	VARCHAR		Y	Dictionary list generated based inputs provided in Treatment tab (**Section 4.2.3**)
33	Treatment	duration or start-end date	DATE (YYYY-MM-DD)		Y	
34	Treatment	type of treatment	VARCHAR		Y	
35	Treatment	dosing information	VARCHAR		Y	
36	Treatment	concentration	NUM		Y	
37	Treatment	unit of concentration	VARCHAR		Y	
38	Treatment	volume applied	NUM		Y	
39	Treatment	unit volume applied	VARCHAR		Y	
40	Treatment	frequency	VARCHAR		Y	
41	Treatment	route of administration	VARCHAR		Y	
42	Treatment	regimen of drug combinations	VARCHAR		Y	
43	Treatment	internal test drug ID	VARCHAR		Y	
44	Treatment	generic test drug ID	VARCHAR		Y	
45	Treatment	batch/lot	VARCHAR		Y	
46	Treatment	pathogen species & strain ID	VARCHAR		Y	
47	Treatment	replicates	VARCHAR		Y	
48	Termination	sacrifice	VARCHAR		Y	Dictionary list generated based inputs provided in Readout tab (**Section 4.2.4**)
49	Condition	medium	VARCHAR		Y	
50	Readout	Tissue analysed	VARCHAR		Y	
51	Readout	Unit 1	VARCHAR		Y	

Column #	Category	Variable	Format	Required (R) or Optional (O) or Conditional (C)	Dictionary (Y/N)?	Additional Comments
52	Readout	time-points 1	VARCHAR	See Additional Comments	Y	Dictionary list generated based inputs provided in Readout tab (**Section 4.2.4**)
53	Readout	Value 1	NUM/VARCHAR		Y	
54	Readout	Unit 2	VARCHAR		Y	
55	Readout	time-points 2	VARCHAR		Y	
56	Readout	Value 2	NUM/VARCHAR		Y	
57	Readout	Unit 3	VARCHAR		Y	
58	Readout	time-points 3	VARCHAR		Y	
59	Readout	Value 3	NUM/VARCHAR		Y	
60	Readout	Unit 4	VARCHAR		Y	
61	Readout	time-points 4	VARCHAR		Y	
62	Readout	Value 4	NUM/VARCHAR		Y	
63	Readout	Unit 5	VARCHAR		Y	
64	Readout	time-points 5	VARCHAR		Y	
65	Readout	Value 5	NUM/VARCHAR		Y	
66	Readout	Flag valid/invalid	VARCHAR		Y	

4.2.6 Scoring

In the event of animal scoring being utilised as a readout, this tab is to be completed to provide metadata defining the local scoring definitions versus the agreed standard definitions (see example in table below)

Score	Local	Systemic
0	Unaffected	Unaffected
1	Slightly affected (swollen thigh)	Slightly affected (less curious, slower movements, slightly ruffled fur)
2	Affected (left hindleg limp or retracted towards abdomen)	Affected (prefers to be stationary, ruffled fur)
3	Highly affected (does not use hind leg). Euthanize.	Clearly affected (moves only when manipulated, ruffled fur, half closed eyes)

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 the event that a field with a defined dictionary does not have an appropriate option, the process as described in **Section 3.3** also applies for In Vivo data.

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**, it is currently planned that the templates described in **Section 3** will be used as a basis for the reporting of imaging results. Further details will be described in deliverable D1.8.

6 GLP Toxicity/Safety Data (Work Package 6)

The data generated under Work Package 6 is expected to be delivered in a CDISC SEND structure with controlled terminologies in line with CDISC standards. Non-GLP toxicity/safety data will only be available as references files in .DOC or .PDF formats, if requested/required by a specific partner within ERA4TB.

Further details requirements to ingest the CDISC SEND data into the grit42 platform within the DDIM will be described deliverable D1.8. The requirements to support data provenance will follow the approach described in **Section 7**.

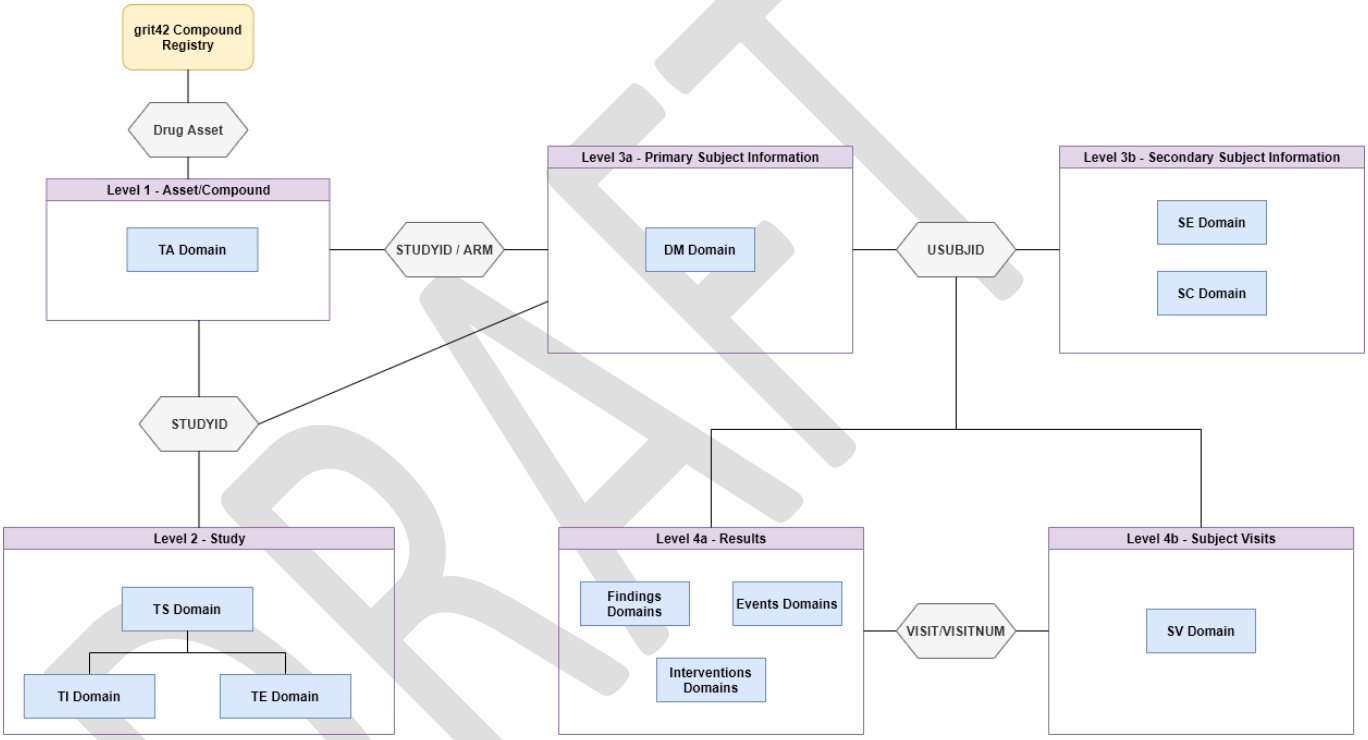
7 Clinical Data (Work Package 7)

7.1 Data Ingestion:

For the ingestion of clinical data into the grit42 platform within the DDIM, it is currently planned that 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 grit42 platform the hierarchy of data will be driven by the following core domains as per **Diagram 1** below. This structure may develop over time subject to end user needs and further development of platform capabilities.

Diagram 1 – 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

Trial Summary Parameter Name [TSPARM]	Trial Summary Parameter Name [TSPARMCD]	NCIT Reference	Description
Asset Progression Plan Reference	APPREF	N/A	Name, date, and version of Asset 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		Location where data owner is storing the study data
Provence Source Data	PROVDATA	http://purl.obolibrary.org/obo/NCIT_C43581	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