TransBioLine
Translational Safety Biomarker Pipeline

821283 – TransBioLine

Translational Safety Biomarker Pipeline

WP8 – Data management and analysis

D8.3 Data and Knowledge Management Plan

Lead contributor	Serge Eifes (3 – ITTM)
	serge.eifes@ittm-solutions.com
Other contributors	Andreas Kremer (3, ITTM), Laurie Ben-yair (11, ABX-CRO), Louise Marais (11, ABX-CRO), Anton Behnke (15, Charité), Dana Briesemeister (15, Charité), Alexandra Stege (15, Charité), Bruno Stieger (1, UZH), Michael Merz (1, UZH), Thorsten Hornemann (1, UZH), Mattias Hackl (9, TamiRNA), Kseniya Khamina (9, TamiRNA), Oliver Pötz (10, SIGNATOPE), Stephan Wnendt (10, MLM), Stephan Voswinkel (18, MLM), Maria Jesus Monte Rio (7, USAL)

Due date	26 02 2021
Delivery date	26 02 2021
Deliverable type	R
Dissemination level	CO

Description of Action	Version	Date

Acknowledgement
This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 821283. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA companies.

Contents

Document History

Version	Date	Description
V1.0	16th Sep 2019	Initial DMP - First Draft
V1.1	15th Oct 2019	Initial DMP - Comments
V1.2	25th Oct 2019	Initial DMP - Draft
V1.3	1st Nov 2019	Initial DMP – Final Version
V2.0	20th Mar 2020	First Update DMP – Final Version
V3.0	26 th Feb 2021	Second Update DMP – Final Version

Definitions

Participants of the TransBioLine Consortium are referred to herein according to the following codes:

UZH. University of Zurich, Switzerland

UMA-IBIMA. Universidad de Málaga, Spain

ITTM. Information Technology for Translational Medicine, Luxembourg

Landspitali. Landspitali University Hospital. Iceland

KUM. Klinikum der Universität München. Germany

UNOTT. The University of Nottingham. United Kingdom

USAL. Universidad de Salamanca. Spain

IR-HSCSP-ICCC. Fundació Privada Institut de Recerca de l'Hospital de la Santa Creu i Sant Pau. Spain

TAM. TAmiRNA GmbH. Austria

SIGNATOPE. SIGNATOPE GmbH. Germany

ABX-CRO. ABX-CRO Advanced Pharmaceutical Services Forschungsgesellschaft mbH. Germany

MetaHeps. MetaHeps GmbH. Germany

APHP. Assistance Publique - Hopitaux de Paris. France

SYNAPSE. Synapse Research Management Partners S.L. Spain

Charité. Charité - Universitätsmedizin Berlin. Germany

UNEW. University of Newcastle Upon Tyne. United Kingdom

ULIV. The University of Liverpool. United Kingdom

MLM. MLM Medical Labs GmbH. Germany

UL. University of Leiden. Netherlands

SAS. Servicio Andaluz de Salud. Spain

PFIZER. LTD PFIZER UK

MSD. Merck Sharp & Dohme Corp. USA & France

LLY. Eli Lilly. USA & UK

NOVARTIS. NOVARTIS Pharma AG. Switzerland

ROCHE. F. Hoffman-La Roche Ltd. Switzerland

Janssen. JANSSEN Pharmaceutica NV. Belgium; U.S.A.

SARD. SANOFI-AVENTIS RECHERCHE & DEVELOPPEMENT. France

Grant Agreement. The agreement signed between the beneficiaries and the IMI JU for the undertaking of the TransBioLine project (No 821283).

Project. The sum of all activities carried out in the framework of the Grant Agreement.

Consortium. The TransBioLine Consortium, comprising the above-mentioned legal entities.

Consortium Agreement. Agreement concluded amongst TransBioLine participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

Publishable Summary

The main objective of the TransBioLine project is to generate exploratory and confirmatory data enabling regulatory qualification of new biomarkers that help to optimize drug development and patient safety in the context of DILI, DIKI, DIPI, DIVI and DINI.

The Data Management Plan (DMP) corresponding to deliverable 8.3, as part of WP8 reports the general framework regarding data management, data protection, data ownership, accessibility and sustainability requirements.

As part of making research data Findable, Accessible, Interoperable and Re-usable (FAIR), this document provides an overview on how the FAIR principles have been addressed by the consortium. The DMP gives guidance and provides an oversight of general data management, while it also provides specific data management information for each study.

As mentioned in the Guidelines on Data Management in Horizon 2020[1], the DMP is not a fixed document, but evolves during the lifespan of the project. It addresses the key elements for implementing the FAIR principles on a general and/or dataset by dataset basis and reflects the current status of data management across the consortium. Furthermore, it describes the project's cornerstones for data management including personal data protection as well as related security and ethical aspects. This document therefore provides the general principles how the DMP will be evolving over time.

1. Introduction

The main objective of the TransBioLine project is to generate exploratory and confirmatory data enabling regulatory qualification of new biomarkers that help to optimize drug development and patient safety in the context of drug-induced liver injury (DILI), drug-induced kidney injury (DIKI), drug-induced pancreatic injury (DIPI), drug-induced vascular injury (DIVI) and drug-induced neurologic injury (DINI).

The Consortium Agreement (CA) indicates that a specific DMP will be created. The DMP will abide by the FAIR Data principles and detail the relevant aspects of data management in the TransBioLine project, covering:

- Data description - nature, scope, scale;
- Security, access, sharing;
- Metadata and standards;
- Ethical;
- Formats;
- Storage & back-up;
- Archiving, preservation;
- Compliance with national & EU legislation.

The DMP is an evolving document, not all required information may be available at the moment of writing of this version of the DMP. Therefore, some of the aspects may be described only in a later version of the DMP.

The final version of the DMP is planned to:

- Report on how data will be handled during and after the project
- Describe what data will be gathered, processed or derived
- Describe which methodology and standards were used to generate the data
- Describe whether and how these data will be shared and/or made publicly available
- Describe how the data will be preserved.

2. General principles of the Data and Knowledge Management Plan

This is the second Update DMP for TransBioLine. The DMP is a working document, that will evolve during the project. It will be updated to reflect progress on the project. It is planned to have updates on the document as shown in Table 1 to keep of the progress on the data management. Furthermore, whenever there are changes in the consortium policies or composition, updates on this document will be performed. If necessary additional updates will be performed.

The DMP follows the principles that research data should be FAIR [1].

Name	WP no.	Short name of lead participant	Delivery Date
Initial DMP	8	ITTM	6
First Update DMP	8	ITTM	12
Second Update DMP	8	ITTM	24
Final DMP	8	ITTM	60

Table 1: *Planned updates for the TransBioLine Data and Knowledge Management Plan (DMP)*

The general principles on data and knowledge access rights are defined in the CA (Section 9). Data sharing will be in accordance with the terms and conditions of the Grant Agreement and the CA, in particular the terms and conditions on Human Samples and Personal Data in Appendix 2 and the Ownership and Access Rights conditions as agreed upon in Clauses 6, 7 and 8, to the extent the requested materials and related data constitute results, including prospective in-kind contributed materials and data.

3. Key elements enabling the FAIR principles

3.1. Data set reference and name

We provide here the mapping between the unique identifiers and full names of the data sets to be produced.

3.2. Data set description including standards and metadata

We provide here a description of the generated or collected data set, its origin (in case it is collected), the scale and included data types. Additionally, we provide the standards applied to the data and what metadata has been generated for the data set. Furthermore, we provide information to whom this data set could be useful.

3.3. Data access and sharing

We provide here a description of how the data is made accessible and shared by focusing on access and sharing procedures during and after the project.

3.4 Storage and backup

Here we define the data storage and backup processes during the runtime of the project.

3.5. Archiving and preservation

Here we define the long-term preservation strategy after the project including how long the data (and corresponding metadata) is preserved.

3.6. Discoverable

We provide here the unique identifiers of the individual data sets of TransBioLine and the overview of the related clinical sites including site contacts.

3.7. Useable beyond the original purpose for which it was collected

Can the data produced and/or used in the context of TransBioLine be used by Third parties even long time after collection of the data?

3.8. Interoperable to specific quality standards

We provide here the description what existing standards enabling the interoperability of the data have been used. We focus here specifically on defining the common data model(s) and standard terminologies/ontologies that have been applied to the data.

4. Implementation of the key elements that enable the FAIR principles within TransBioLine

We provide here an overview of the implementation of the FAIR principles in general and across the different TransBioLine data sets as described under Section 3.

Here directly below we provide the generic implementation of principles being common among the different TransBioLine studies.

4.1. Generic implementation of FAIR principles

The first part here below focuses on providing the implementation of the principles that are shared across the TransBioLine studies.

- Principle 3.1: Data set reference and name
  - The corresponding information is study-specific
  - Please refer to section 4.2
- Principle 3.2: Data set description including standards and metadata
  - Where applicable, the standards defined by the Clinical Data Interchange Standards Consortium (CDISC)[2] have been applied.
  - For the corresponding data set description as well as the metadata and the study-specific standards that have been applied, please refer to section 4.2
  - Detailed information on the TransBioLine ID-management process can be found here below under Section 5.3.
- Principle 3.3: Data access and sharing
  - Data access and sharing during the project runtime is implemented as agreed in the CA.
  - Data access and sharing after the project needs separate agreements.
- Principle 3.4: Storage and Backup
  - The storage and backup processes for different TransBioLine software components:
    - tranSMART
      - knowledge management platform for integrative analysis of clinical data
      - cloud-based infrastructure involving daily backups
      - running on server at POST Telecom S.A. (Luxembourg), the service provider for secured cloud infrastructures of ITTM
    - ownCloud
      - data exchange platform for secured transfer of clinical data between partners

- cloud-based file hosting service involving daily backups
- running on server at POST Telecom S.A. (Luxembourg), the service provider for secured cloud infrastructures of ITTM
-

- eCRF tool (Clindex)
  - Clinical data management system for data collection, processing and reporting of clinical data
  - Resides on a server at ABX-CRO in Dresden, Germany. This server will be used as the dedicated data management server and will be housed in a server cabinet in a separate dedicated server room
  - The entire Data Management server is backed up daily. Complete recovery of the database at any time is therefore possible in the event of a system failure.
- Biobank information system (CentraXX)
  - Web and Microsoft SQL Server running on a virtual server farm inside the Charité intranet
  - whole server farm is all time live mirrored on two different locations in Berlin, in case of a fatal accident of one system the other system is there immediately
  - windows server itself gets mirror image backup every day and can be rebuild within 24hours, the backups get stored for 30 days
  - The database itself gets the following backups (storage 30 days):
    - Full back up every Day
    - Transaction Log-Backup every 20 minutes

- Principle 3.5: Archiving and Preservation
  - As for the current DMP version, this is under initial discussion. More detailed information will be provided in the following version(s) of the DMP.
  - The biobank (ZeBanC) is a core facility of the Charité and stores samples and a biobank-specific data set about the end of the study. The basis for the storage period is the Informed Consent Form (ICF) and the TBL Agreement.
- Principle 3.6: Discoverable
  - The corresponding information is study-specific.
  - Please refer to section 4.2
- Principle 3.7: Useable beyond the original purpose for which it was collected
  - The corresponding information is study-specific
    - Depends on the study-specific Informed Consent Form templates.
  - Please refer to Annexure 17, Query_ICF_Sustainability_V1.xlsx
- Principle 3.8:

- CDISC is a global, open, multi-disciplinary, non-profit organization, established to cover study design, data collection, analysis, exchange, submission and other aspects of a series of standards. By following CDISC Clinical Data Acquisition Standards Harmonization (CDASH) ensures Study Data Tabulation Model (SDTM) ready datasets. CDISC core criteria will be applied as shown here below in Table 2.

Standard	Description
Controlled set of terms (CT)	Supports standard vocabulary and coding set CDISC models/standards involved.
Harmonized standards (CDASH) clinical data acquisition	Content for standard case report form the basis of the data collection field.

Table 2: *Application of CDISC core criteria to the TransBioLine clinical data.*

4.2. Implementation of study-specific principles

The second part here below focuses on providing the implementation of the principles that are distinct/not shared across the different TransBioLine studies and have been implemented under Section 4.1. In the current version of the DMP, this section here is not complete due to delayed availability of organ-specific study information.

Here below we introduce a distinction at the level of data collection. This defines three different groups of data:

- Data captured over the electronic Case Report Form (eCRF) mask (called hereafter "Data captured by eCRF mask")
- Data exported from an external database and subsequently imported into the TransBioLine database upload environment (called hereafter "Data originating from database export/import process")
- Data from the IMI Safer and Faster Evidence-based Translation (SAFE-T) project (called hereafter "Legacy SAFE-T data")
- Biomarker Analysis data generated from patient samples (called hereafter "Biomarker data")

4.2.1. Drug-induced Kidney Injury (DIKI)

4.2.1.1. Data captured by eCRF mask

In this section we provide information on the WP1 eCRF template / eCRF database, the ICF templates and Patient Information documents as well as the data dictionary.

For a description of the finalized WP1 eCRF template / eCRF database, we refer here to deliverable D8.2 "Clindex eCRF (software)".

The corresponding ICF templates and Patient Information documents have been finalised. The finalised/ ICF templates and Patient Information documents are available on TransBioLine SharePoint[3].

Please see Annex 10 for the corresponding data dictionary.

4.2.1.2. Legacy SAFE-T data

For the use and integration of the SAFE-T DIKI data in whole or in part into the TransBioLine database. A process has been set up to review patient consent and decide whether the Institut de Recerca - Hospital de la Santa Creu i Sant Pau (IR-HSCSP-ICCC) and ITTM can provide samples and data. Various documents (checklists and application forms) have been prepared for this purpose by ExCom, ITTM and Barcelona to allow a transparent and standardized application and overall approval process.

Based on the request documents and the following checklist (see Annex 19-22), an approval is given by ExCom:

1. Does ITTM agree technically with the request?
2. Is the request supported by BCN-BB?
3. Does the request interfere with ongoing activities in TBL?

4.2.1.3. Biomarker data

Here we make available the corresponding metadata/description for properly documenting the Biomarker data.

Table 3 shows the list of biological parameters that are analysed for WP1.

Biological parameters analysed	Work package	Laboratory	Matrix
Creatinine, urea	WP1-DIKI	MLM	Serum

Creatinine, urea, albumin	WP1-DIKI	MLM	Urine
microRNA NGS/qPCR	WP1-DIKI	TAmiRNA	EDTA-plasma
Protein(s)	WP1-DIKI	Signatope	Urine

Table 3: *Listing of biological parameters analysed by partner/laboratory for WP1-DIKI. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- MLM: Annex 1 and Annex 2
- TAmiRNA: Annex 3
- Signatope: Annex 4

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID and Master sample ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

4.2.2. Drug-induced Liver Injury (DILI)

4.2.2.1. Data originating from database export/import process

In this section we provide information on the WP2 upload Database, the ICF templates and Patient Information documents as well as the data dictionary.

For a description of the finalized WP2 upload Database, we refer here to deliverable D8.2 "Clindex eCRF (software)".

The corresponding ICF templates and Patient Information documents have been finalized. The finalised/ICF templates and Patient Information documents are available on TransBioLine SharePoint[3].

Please see Annex 11 for the corresponding data dictionary.

4.2.2.2. Legacy Safe-T data

The integration of the SAFE-T DILI data in whole or in part into the TransBioLine database. A process has been set up to review patient consent and decide whether the Institut de Recerca - Hospital de la Santa Creu i Sant Pau (IR-HSCSP-ICCC) and ITTM can provide samples and data. Various documents (checklists and application forms) have been prepared for this purpose by ExCom, ITTM and Barcelona to allow a transparent and standardized application and overall approval process.

Based on the request documents and the following checklist (see Annex 19-22), an approval is given by ExCom:

1. Does ITTM agree technically with the request?
2. Is the request supported by BCN-BB?
3. Does the request interfere with ongoing activities in TBL?

4.2.2.3. Biomarker data

Here we make available the corresponding metadata/description for properly documenting the Biomarker data.

Table 4 shows the list of biological parameters that are analysed for WP2.

Biological parameters analysed	Work package	Laboratory	Matrix
Protein(s), albumin, bilirubin	WP2-DILI	MLM	Serum
microRNA NGS/qPCR	WP2-DILI	TAmiRNA	EDTA-plasma
Protein(s)	WP2-DILI	Signatope	Serum
Bile acids	WP2-DILI	USAL	EDTA-plasma
Sphingolipids	WP2-DILI	UZH	EDTA-plasma
Deep Immunophenotyping	WP2-DILI	UNOTT	

Table 4: *Listing of biological parameters analysed by partner/laboratory for WP2-DILI. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- MLM: Annex 1 and Annex 2
- TAmiRNA: Annex 3
- Signatope: Annex 4
- USAL: Annex 5
- UZH: Annex 6

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

4.2.3. Drug-induced Pancreatic Injury (DIPI)

4.2.3.1. Data originating from database export/import process and captured by eCRF mask

In this section we provide information on the WP3 eCRF template / eCRF database, the ICF templates and Patient Information documents as well as the data dictionary.

For a description of the draft WP3 eCRF template / eCRF database, we refer here to deliverable D8.2 "Clindex eCRF (software)".

The corresponding ICF templates and Patient Information documents have been finalized. The finalised/ICF templates and Patient Information documents are available on TransBioLine SharePoint[3].

The corresponding data dictionary will be made available once the corresponding eCRF template has been finalised.

4.2.3.2. Biomarker data

Here we make available the corresponding metadata for properly documenting the Biomarker data and its data management process.

Table 5 shows the list of biological parameters that are analysed for WP3.

Biological parameters analysed	Work package	Laboratory	Matrix
Protein(s)	WP3-DIPI	MLM	Serum
microRNA NGS/qPCR	WP3-DIPI	TAmiRNA	EDTA-plasma
Protein(s)	WP3-DIPI	Signatope	EDTA-plasma

Table 5: *Listing of biological parameters analysed by partner/laboratory for WP3-DIPI. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- MLM: Annex 1 and Annex 2
- TAmiRNA: Annex 3
- Signatope: Annex 4

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

4.2.4. Drug-induced Vascular Injury (DIVI)

4.2.4.1. Data captured by eCRF mask

In this section we provide information on the WP4 eCRF template / eCRF database, the ICF templates and Patient Information documents as well as the data dictionary.

For a description of the finalized WP4 eCRF template / eCRF database, we refer her to deliverable D8.2 "Clindex eCRF (software)".

The corresponding ICF templates and Patient Information documents have been finalised. The finalised ICF templates and Patient Information documents are available on TransBioLine SharePoint[3].

Please see Annex 12 for the corresponding data dictionary.

4.2.4.2. Legacy Safe-T data

The integration of the SAFE-T DIVI data in whole or in part into the TransBioLine database a process has been set up to review patient consent and decide whether the Institut de Recerca - Hospital de la Santa Creu i Sant Pau (IR-HSCSP-ICCC) and ITTM can provide samples and data. Various documents (checklists and application forms) have been prepared for this purpose by ExCom, ITTM and Barcelona to allow a transparent and standardized application and overall approval process.

Based on the request documents and the following checklist (see Annex 19-22), an approval is given by ExCom:

1. Does ITTM agree technically with the request?
2. Is the request supported by BCN-BB?
3. Does the request interfere with ongoing activities in TBL?

4.2.4.3. Biomarker data

Here we make available the corresponding metadata/description for properly documenting the Biomarker data.

Table 6 shows the list of biological parameters that are analysed for WP4.

Biological parameters analysed	Work package	Laboratory	Matrix
Protein(s)	WP4-DIVI	MLM	Serum
Protein(s)	WP4-DIVI	MLM	EDTA-plasma
microRNA NGS/qPCR	WP4-DIVI	TAmiRNA	EDTA-plasma

Protein(s)	WP4-DIVI	Signatope	EDTA-plasma

Table 6: *Listing of biological parameters analysed by partner/laboratory for WP4-DIVI. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- MLM: Annex 1 and Annex 2
- TAmiRNA: Annex 3
- Signatope: Annex 4

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

4.2.5. Drug-induced Neurological Injury (DINI)

4.2.5.1. Data captured by eCRF mask

In this section we provide information on the WP5 eCRF template / eCRF database, the ICF templates and Patient Information documents as well as the data dictionary.

For a description of the finalized WP5 eCRF template / eCRF database, we refer here to deliverable D8.2 "Clindex eCRF (software)".

The corresponding ICF templates and Patient Information documents have been finalised. The finalised ICF templates and Patient Information documents are available on TransBioLine SharePoint[3].

Please see Annex 13 for the corresponding Data dictionary.

4.2.5.2. Biomarker data

Here we make available the corresponding metadata/description for properly documenting the Biomarker data.

Table 7 shows the list of biological parameters that are analysed for WP5.

Biological parameters analysed	Work package	Laboratory	Matrix
microRNA NGS/qPCR	WP5-DINI	TAmiRNA	EDTA-plasma
Protein(s)	WP5-DINI	Signatope/NMI	Serum
Protein(s)	WP5-DINI	Signatope/NMI	CSF

Table 7: *Listing of biological parameters analysed by partner/laboratory for WP5-DINI. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- TAmiRNA: Annex 3
- Signatope: Annex 4

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

4.2.6. Liquid Biopsies

4.2.6.1. Data originating from database export/import process

In this section we provide information on the ICF templates and Patient Information documents as well as the data dictionary.

The corresponding ICF templates and Patient Information documents have been finalized partially as of today (March 10th 2020). The finalised/non-finalised ICF templates and Patient Information documents are/will be made available on TransBioLine SharePoint[3].

Please see Annex 14 for the corresponding data dictionary.

4.2.6.2. Biomarker data

Here we make available the corresponding metadata for properly documenting the Biomarker data and its data management process.

Table 8 shows the list of biological parameters that are analysed for WP5.

Biological parameters analysed	Work package	Laboratory	Matrix
microRNA NHV (NGS platform characterization and references ranges)	WP6-LB	TAmiRNA	EDTA-plasma

Table 8: *Listing of biological parameters analysed by partner/laboratory for WP6-Liquid Biopsies. The corresponding biological matrix of the sample is indicated.*

A detailed description of the file formats and structures for the biomarker analytics results data as provided by the corresponding analytics partners can be found here:

- TAmiRNA: Annex 3

Annex 15 contains the data mapping scheme between ABX-CRO and Charité Biobank (ZeBanC). The export file from the TBL eCRF of ABX-CRO contains the sample relevant data including a minimum data set for the characterization of the patient, the sampling location and the connection to the patient via the specific TBL patient ID. This information is imported into the Biobank information system CentraXX using the import tool implemented by ZeBanC.

5. Protection of personal data

The TransBioLine consortium will conform to the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of

such data, and repealing Directive 95/46/EC (General Data Protection Regulation [GDPR]). Furthermore, it complies to applicable national laws on data protection.

5.1. Informed Patient Consent Form

The ICF template will detail the information on how personal data will be managed. To secure the confidentiality, accuracy, and security of data and data management, the following measures will be taken:

- Patient/subject data is collected and communicated to the TransBioLine project partners in a pseudonymised manner so that no information can be linked in any manner towards potentially identifying the patient/subject
- Patient/subject data is entered/stored into a secured Information Technology (IT) system. Data is processed only for purposes outlined in the patient information and ICF Form templates of the respective studies. Use for other purposes will require explicit patient approval.
- Access to personal data will be granted to partners in non-EU countries for restricted use within the TransBioLine project. Data handling in non-EU countries will be fully conforming to national laws and regulations as well as the GDPR of May 25th 2018. In cases of contradiction, the tighter regulation shall prevail. The necessary and legally adequate measures will be taken to ensure that the data protection standards of the EU shall be complied with. Transfer and subsequent use of TransBioLine data by partners in US will be governed in accordance with federal and state laws.

As mentioned under Section 4.1, Annexure 17: and samples Query_ICF_Sustainability_V1.xlsx was created to support the sustainability of the data and samples collected within the TransBioLine project. ICF templates have been made available in a dedicated directory in the TransBioLine SharePoint for the runtime of the project. Furthermore, a generic ICF template (see Annex 7) has been developed as a support/guidance for the implementation of the organ WP-specific templates.

5.2. De-identification of personal data

As mentioned under Section 5.1, personal data in TransBioLine studies will be transmitted to partners within the consortium only after pseudonymization of the data has taken place by the clinical partner sharing the data.

5.3. ID-Management process for patient and sample ids

The corresponding ID-Management processes are provided by ZeBanC/Charité and ABX-CRO. A detailed description of the structure of patient/master sample ID can be found in Annex 8 and Annex 9. These identifiers are humanly readable and include metadata (e.g. study code and recruitment site) at the level of de-identified subject/patient or sample data.

6. Ethical aspects

The TransBioLine project and its participants are requested to adhering to all relevant international, IMI, and national legislation and guidelines relating to the conduct of clinical studies. The proposed research will be in accordance with defined ethical standards, including those outlined in the Consortium Agreement and the European Code of Conduct for Research integrity.

To achieve the correct balance between research objectives and ethical aspects, TransBioLine is supported by its Ethics Advisory Board (EAB). The EAB consists of three experts with detailed knowledge of ethical policies in the context of clinical research.

The EAB will monitor the progress of the project and ensure a high level of ethical standards in the context of data and knowledge management in TransBioLine. More precisely it will aim at ensuring that all related project activities are ethically sound and compliant with all due rules and regulations, including data privacy considerations.

According to the Consortium Agreement, the responsibilities of the EAB in general and therefore related to data management are as follows:

1. Reviewing the proper application of the ethical rules by the Beneficiaries;
2. Providing advice to the Beneficiaries, the General Assembly and the Steering Committee on ethical issues;
3. Providing advice on the compliance with European ethical laws and regulations and with different guidelines, laws and regulations of countries where studies are being performed; and
4. Providing written ethics periodic reports

7. Data security

According to the CA, the processing of personal data is subject to appropriate security measures (as describe above under Section 5.2) that:

1. are able to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services (e.g., where applicable, pseudonymization of Personal Data);
2. include a process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the Processing.

7.1. Secured IT infrastructure

De-identified personal data will be captured, stored and transferred in the secured TransBioLine IT infrastructure. In this section here below, we focus on defining the following aspects related to the secured IT infrastructure:

- the flow of data and samples across the project
- security aspects of the relevant data management IT components

7.1.1. Overview of data and sample flows

The sample and data flow in the TransBioLine consortium as shown in Figure 1 highlights the flow of samples, (clinical and sample) data and corresponding metadata in TransBioLine.

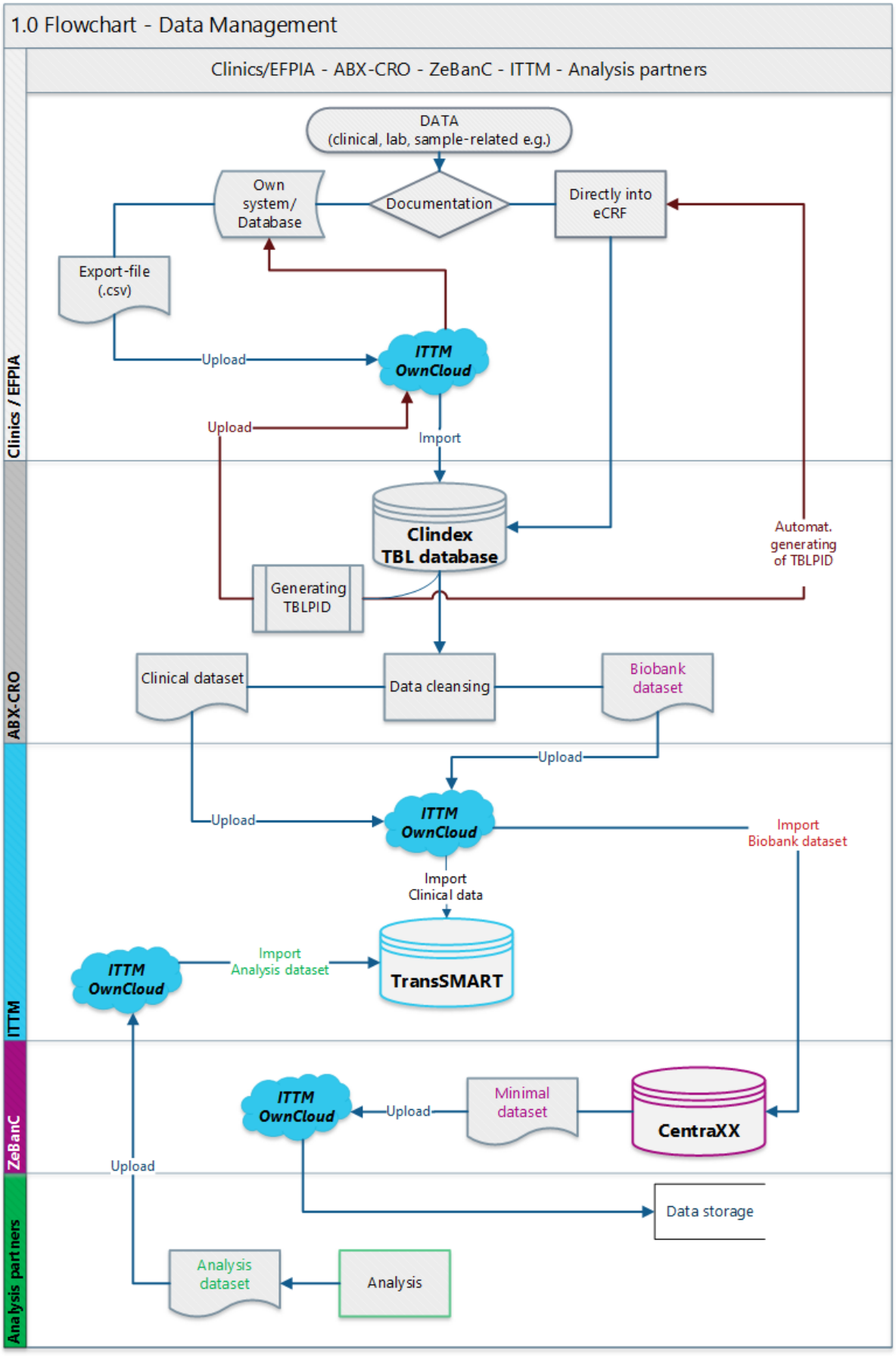

Figure 1: *Sample and data flow diagram in TransBioLine.*

It involves four key software components at the level of data management, namely:

1. Clindex eCRF system: a software system focused on electronic data capture and management for clinical data[4]
2. OwnCloud: an open-source secured file sharing platform[5]
3. CentraXX: Biobank information system focusing on sample data[6]
4. tranSMART: an open-source knowledge data and knowledge management software for translational research data[7].

7.1.2. Security and access rights for the data management-related IT components

To enable the secured usage, storage and transfer of data (and metadata) in the context of TransBioLine, the consortium partners ABX-CRO, ZeBanC/Charité and ITTM have set up a secured IT-infrastructure involving four key software components mentioned her above under section 7.1.1. As can be seen on Figure 2, the software tools Clindex eCRF system, CentraXX Biobank information system and tranSMART implement several security standards and protocols. In this infrastructure, OwnCloud acts as a data exchange component to guarantee the secured data transfer/exchange between partners. The key security aspects at the level of the different software elements (including firewalls and user authentication among others) and the implemented data transfer protocols (Hypertext Transfer Protocol Secure (HTTPS) as well as Rsync via secure socket shell [ssh]) support this secured software environment. Related to the biomarker analytics data, the raw data will be stored locally by the corresponding analytics partner according to state-of-the-art security standards (i.e. data encryption at Rest and access control) in line with the GDPR. Only normalised data will be transferred to OwnCloud for upload to tranSMART.

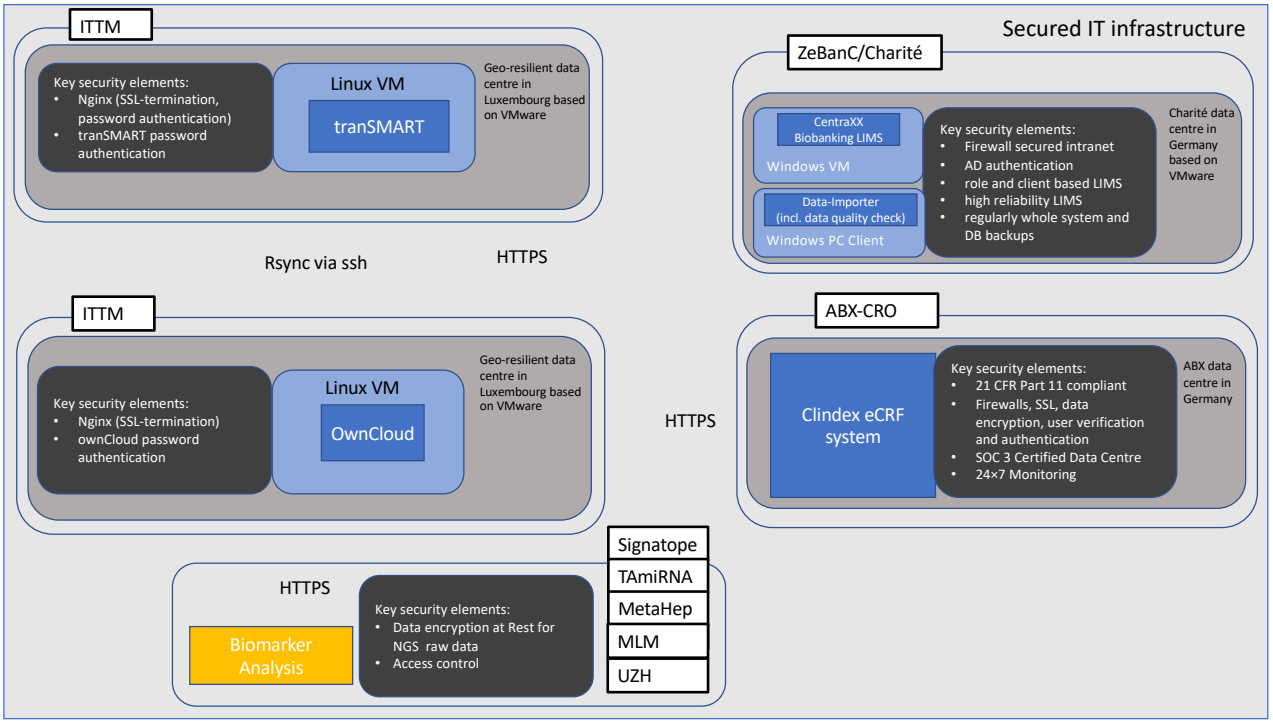

Figure 2: *Secured IT-infrastructure for data management in the context of TransBioLine. This infrastructure involves four key software components including Clindex system, CentraXX, tranSMART as well as OwnCloud. The corresponding key security elements for these components are indicated (see black boxes) as well as the corresponding secured transfer protocols (see text on blue arrows) for data transfer between the key components and with the biomarker-related analytics partners (Biomarker Analysis). Furthermore, the geographical location of the corresponding servers for the software components is indicated in the dark grey boxes. The key security elements for the Biomarker Analysis environments are indicated (see black box).*

As mentioned above, the corresponding data and metadata will be stored and shared in the different IT elements involving restricted access (username and password) to authorized users. Below we provide more detailed information on the security and access right implementation for the different IT components.

7.1.2.1. Clindex eCRF

7.1.2.1.1. Access to the building

Data Managers have access to the database servers via internet from ABX-CRO Data Management offices located in Cape Town, South Africa. The offices are locked (Key access) in a secure Office Park. Security staff is available 12 hours of the day until 6pm. An alarm system is in place, linked to 24 hr security response company.

7.1.2.1.2. Access to the data management server

ABX-CRO servers are located in the server house in an office building in Dresden, Germany. The server house is locked and a video surveillance and alarm system are in place. Only authorised persons have access to the server house via transponder. Any access is logged.

The Data Management server will be stored in a server cabinet in the server room. The door to this room is kept locked at all times. Employees permitted to enter the room, are provided with transponders for this purpose. Each entry to the server room will be logged. Unauthorised personal can only access the server room accompanied by a permitted employee. The server cabinets can be accessed with keys only by permitted employees.

7.1.2.1.3. Access to the database

User-specific usernames and passwords are required to log onto the database. Passwords must consist of at least 6 alpha-numeric characters and are re-set every 12 weeks. All user rights are set by the Database Administrator, who also defines the user groups for each study. Granting and revoking of rights will be documented on a Database Access Form.

The following 2 reports summarize the user access to the database:

- Roles Assigned to Users by Study
- Access to CRFs by Role

7.1.2.2 CentraXX Biobank information system

7.1.2.2.1. Access to the building

Access to Charité buildings and rooms at all are restricted and secured by locks. Only Charité employees have keys or transponders for their rooms or areas depending on their level and application. All three locations of the Charité are guarded by security personal.

7.1.2.2.2. Access to the data management server

Access to the server physically is restricted as well as for every other room inside the Charité. Only registered and special authorized personal is allowed to enter the server areas.

The access to the virtual server is secured by different levels. The first level is the windows authentication, remote to the server or for remote access to the MS-SQL-instance. Usernames and passwords are controlled and steered over the active directory of the Charité domain. The passwords need to be strong enough and have to be changed every three months. If a person is leaving the Charité the account is deactivated automatically. Only registered personal, in this case the administration of the ZeBanC, has access on this level to the database server and component.

7.1.2.2.3. Access to the database

The Database, CentraXX of the company Kairos, has a very finely tuned user and role management system, which makes it possible, to grant or deny access to any person at different levels. User have to login on a client inside the Charité or over a registered VPN. After this step they have to login into the Database where the same user data is used and checked.

The database allows granting access in different levels. At first, each user is granted to one or more projects to which they have access. Inside each project, they get specialized rights like, just as an example, reading or writing. The System consist of more than 100 different rights that can be given or taken based on the role the user has in each project.

7.1.2.3. tranSMART

7.1.2.3.1. Access to the building and data management server

Physical access to the building and servers is controlled according to the procedures of POST Telecom S.A. (Luxembourg), the service provider for secured cloud infrastructures of ITTM.

7.1.2.3.2. Access to the database and web interface

User-specific usernames and passwords are required to log onto the database and web-interface. The passwords are randomly generated and have a minimal length of 12 characters. The user cannot modify the provided password himself. User rights are set by the ITTM software administrator. Access to the underlying database is only granted to ITTM approved system administrators.

7.1.2.4. OwnCloud

7.1.2.4.1. Access to the building and data management server

Physical access to the building and servers is controlled according to the procedures of POST Telecom S.A. (Luxembourg), the service provider for secured cloud infrastructures of ITTM.

7.1.2.4.2. Access to the database and web interface

User-specific usernames and passwords are required to log onto the database and web-interface. The administrator-provided passwords are randomly generated and have a minimal length of 12 characters. User rights are set by the ITTM software administrator. Access to the underlying database is only granted to ITTM approved system administrators.

7.1.2.4.3. A dedicated folder structure as support tool for secured data transfers

The key element of the folder structure, as implemented on TransBioLine OwnCloud, is to enable a secured data exchange between different consortium partners by

supporting various key aspects of the project therefore ensuring secured data transfers of pseudonymised data.

As can be seen on figure 3, a hierarchical folder structure has been implemented enabling for different purposes a secured data exchange between partners:

- Imports from Site
  This folder is used for the data transfer to ABX-CRO from the different sites e.g. WP2, JnJ, Pfizer for import into the TransBioLine Clindex database. This encompasses test as well as actual data files. The test files are used for the initial testing phase and the real patient data will be uploaded to the actual data file.
- Export for Site
  If any sites require an export of their data which includes the TransBioLine identifiers then ABX-CRO will upload the exported files here for them to access.
- Exports for Charité
  This is where ABX-CRO uploads the biobanking data files for Charité and also where sites can load their sample shipping lists (Data sheets).
- Exports to ITTM
  This folder will enable ABX-CRO for the eCRF data as well as the different biomarker analytics partners (SIGNATOPE, MLM, TAM, USAL and UZH) for the sample analytics data to transfer their data files to ITTM for upload to tranSMART.
- Sample status information
  As a support for the stakeholders of the project and to provide an overview of the current status of the samples within the sample and data flow process, TransBioLine WP7 keeps a Sample Status Information File to be accessed on TransBioLine OwnCloud. This document provides a detailed overview on the current status (site, study population, residual amount, shipment date, analysis date, …) of each collected sample within the TransBioLine sample and data flow process.

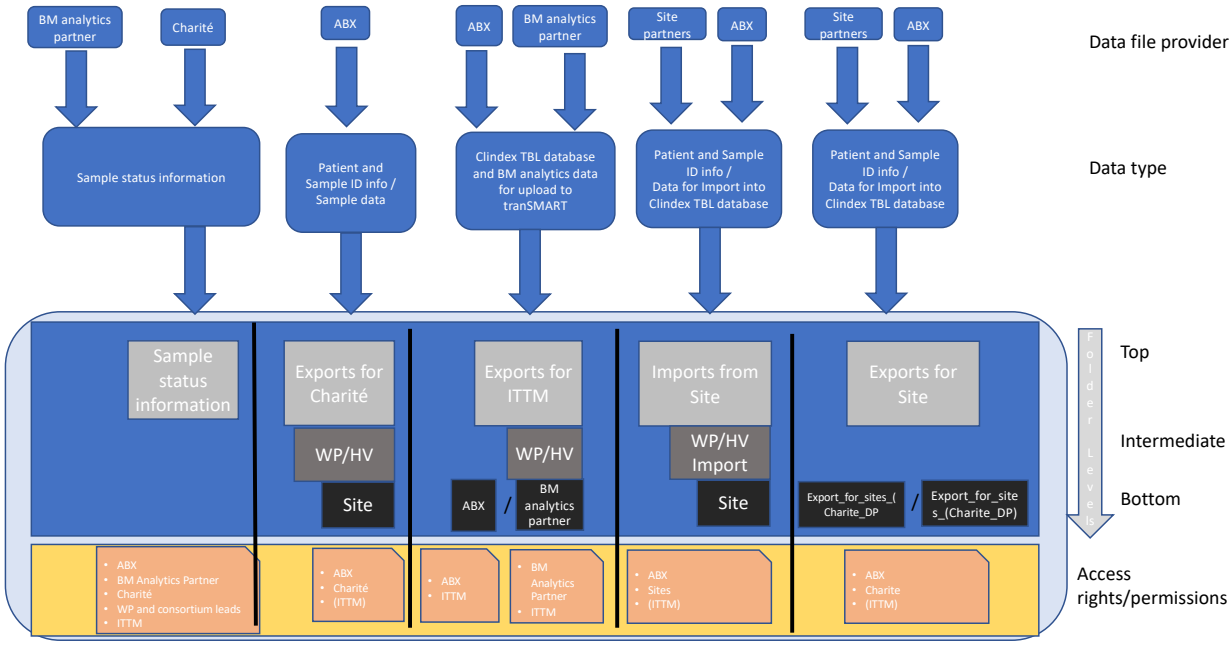

Figure 3: *Folder structure on OwnCloud to enable secured data transfer of patient and biomarker related data. The figure indicates the data file providers, the corresponding data-type as well as the implemented folder structure on OwnCloud and the access rights/permissions for the corresponding folders.*

7.1.2.4.3.1. Sample Status Information file on TransBioLine OwnCloud

As a support for the stakeholders of the project and to provide a summary overview of the current status of the samples in the sample and data flow process, TransBioLine WP7 provides a Sample Status Information File to be accessed on TransBioLine OwnCloud. This document provides a detailed overview on the current status (site, study population, residual amount, shipment date, analysis date, …) of each collected sample within the TransBioLine sample and data flow process.

8. TransBioLine Data Management/Flow "Handbook"

In order to improve the alignment of involved partners on data and sample management processes, TransBioLine WP7 in collaboration with WP8 have established a TransBioLine Data Management/Flow "Handbook".

This handbook focuses on defining the basic principles of data management within TransBioLine as well as providing a general overview of the data and sample flow and the existing processes for secured data transfer within the project.

Its intended audience are partners directly involved in the data and sample management processes. This document (Annex_16) has been made available to the consortium members on TransBioLine SharePoint under the WPX2 folder[9].

9. Conclusion

This second version of the deliverable D8.3 Data and knowledge management plan describes a series of guidelines and processes that have been set up regarding data and knowledge within the context of the TransBioLine project. Future updates of this deliverable will focus on providing further specificity and depth to the foundations for data management practices that have been laid out in this version of the document.

10. References

1. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*.; 2016. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accessed June 11, 2019.
2. CDISC | Clear Data. Clear Impact. https://www.cdisc.org/. Accessed July 3, 2019.
3. TransBioLine DILI - ICFs and Patient Information docs. https://synapsemanagers.sharepoint.com/:f:/r/sites/TransBioLine/wpx2/Shared Documents/Informed_Patient_Consents/Finalised_ICFs_and_PatInfo_docs?csf=1&e=cmR4Wd.

4. Clindex. https://www.fortressmedical.com.

5. OwnCloud. https://owncloud.org/.

6. CentraXX Bio. https://www.kairos.de/en/products/centraxx-bio/.

7. Scheufele E, Aronzon D, Coopersmith R, et al. tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci*. 2014;2014:96-101. http://www.ncbi.nlm.nih.gov/pubmed/25717408. Accessed May 29, 2019.

8. WPX2 Data Management Handbook. https://synapsemanagers.sharepoint.com/:p:/r/sites/TransBioLine/wpx2/Shared%20Documents/WPX2_Data%20Management_Handbook_V0.2_SE.pptx?d=wf9477738a7c8467fb4364b72f80b2811&csf=1&web=1&e=jC6jwu)

10. List of abbreviations

Abbreviation	Full name
CA	Consortium Agreement
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
DIKI	Drug-induced kidney injury
DILI	Drug-induced liver injury
DINI	Drug-induced neurologic injury
DIPI	Drug-induced pancreatic injury
DIVI	Drug-induced vascular injury
DMP	Data management plan
eCRF	electronic Case Report Form
EAB	Ethics Advisory Board
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GDPR	General Data Protection Regulation
HTTPS	Hypertext Transfer Protocol Secure

ICF	Informed patient consent form
IT	Information Technology
LIMS	Laboratory Information management system
SAFE-T	Safer and Faster Evidence-based Translation
SDTM	Study Data Tabulation Model
SSH	Secure socket shell

11. Annexes

Annex 1: MLM Mock Dataset

Annex_1_MLM_Mock_Dataset.xlsx

Annex 2: MLM Dataset Description

Annex_2_MLM_Dataset_Description.docx

Annex 3: TAmiRNA miRNA Data Structure

Annex_3_TAmiRNA_miRNA_Data_Structure.xlsx

Annex 4: Signatope Mock Dataset

Annex_4_Signatope_Mock_Dataset.xlsx

Annex 5: USAL Mock Dataset and Description

Annex_5_USAL_Mock_Dataset_and_Description.xlsx

Annex 6: UZH Lipidomics Dummy Dataset

Annex_6_UZH_Lipidomics_Dummy_Dataset.xlsx

Annex 7: Templates informed consent form and information sheets

Annex_7_TransBioLine - D9.15 Templates of ICF and PIS finalized.pdf

Annex 8: ID Management Process for Patient and Sample Ids

Annex_8_TBL_ID_Management_Process_Version5.4.xlsx

Annex 9: Mapping of codes for institutions and recruitment sites across proposal, id-management and ProEuro DILI

Annex_9_TransBioLine_Mapping_of_CODES_INSTITUTIONS_Version2.xlsx

Annex 10: WP1 Data Dictionary

Annex_10_WP1_Data_Dictionary.xlsx

Annex 11: WP2 Data Dictionary

Annex_11_WP2_Data_Dictionary.xlsx

Annex 12: WP4 Data Dictionary

Annex_12_WP4_Data_Dictionary.xlsx

Annex 13: WP5 Data Dictionary

Annex_13_WP5_Data_Dictionary.xlsx

Annex 14: WP6 Data Dictionary

Annex_14_WP6_Data_Dictionary.xlsx

Annex 15: Mapping ABX to CentraXX

https://synapsemanagers.sharepoint.com/:x:/r/sites/TransBioLine/wp8/Shared%20Documents/Deliverables/D8.3_Second_Update_DMP/Annex_15_Mapping_ABX%20to%20

CentraXX_V5.xlsx?d=wf6ad1408a84848e5872fb175c0183fe5&csf=1&web=1&e=0wHGCc

Annex 16: WPX2 Data Management Handbook

Annex_16: WPX2_Data_Management_Flow_Handbook_V0.2.pptx

Annex 17: Principle 3.7: Useable beyond the original purpose for which it was collected (ICF's)

Annex 17: Query_ICF_Sustainability_V1.xlsx

Annex 19: Safe-T Checklist for Samples

Annex 19: Safe-T Checklist for Samples

Annex 20: Template Safe-T sample request Biobank

Annex 20: Template Safe-T sample request Biobank

Annex 21: Safe-T Checklist for Data

Annex 21: Safe-T Checklist for Data

Annex 22: Template Safe-T patient sample and data request

Annex 22: Template Safe-T patient sample and data request