TransBioLine
Translational Safety Biomarker Pipeline

821283 – TransBioLine

Translational Safety Biomarker Pipeline

WP7 – Assay and Sample Management

D7.03 Data repository for sample data

Lead contributor	Alexandra Stege (15, Charité), Anton Behnke (15, Charité)
Other contributors	Louise Marais (11, ABX-CRO), Laurie Ben-yair (11, ABX-CRO), Serge Eifes (3, ITTM), Andreas Kremer (3, ITTM)

Due date	31.10.2019
Delivery date	16.04.2020
Deliverable type	Websites, patents, filling etc.
Dissemination level	PU

Description of Action	**Version**	**Date**
	V1	02.04.2019

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.

Table of Contents

Document History

Version	Date	Description
V1.0	30 Mar 2020	First Draft
V1.1	01 Apr 2020	Comments
V1.2	06 Apr 2020	Draft
V1.3	16 Apr 2020	Final Version

Definitions

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

01. **UZH**. University of Zurich, Switzerland
02. **UMA-IBIMA**. Universidad de Málaga, Spain
03. **ITTM.** Information Technology for Translational Medicine, Luxembourg
04. **Landspitali**. Landspitali University Hospital, Iceland
05. **KUM**. Klinikum der Universität München, Germany
06. **UNOTT**. The University of Nottingham, United Kingdom
07. **USAL**. Universidad de Salamanca, Spain

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

08. **TAM**. TAmiRNA GmbH, Austria
09. **SIGNATOPE**. SIGNATOPE GmbH, Germany
010. **ABX-CRO**. Advanced Pharmaceutical Services Forschungsgesellschaft mbH, Germany
011. **MetaHeps**. MetaHeps GmbH, Germany
012. **APHP**. Assistance Publique - Hopitaux de Paris, France
013. **SYNAPSE**. Synapse Research Management Partners S.L, Spain
014. **Charité**. Charité - Universitätsmedizin Berlin, Germany
015. **UNEW**. University of Newcastle Upon Tyne, United Kingdom
016. **ULIV**. The University of Liverpool, United Kingdom
017. **MLM**. MLM Medical Labs GmbH, Germany
018. **UL**. University of Leiden, Netherlands
019. **SAS**. Servicio Andaluz de Salud, Spain
020. **PFIZER**. LTD PFIZER, UK
021. **MSD**. Merck Sharp & Dohme Corp, USA & France
022. **LLY**. Eli Lilly, USA & UK
023. **NOVARTIS**. NOVARTIS Pharma AG, Switzerland
024. **ROCHE**. F. Hoffman-La Roche Ltd, Switzerland
025. **Janssen**. JANSSEN Pharmaceutica NV. Belgium, U.S.A.

026. **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 Translational Safety Biomarker Pipeline (TransBioLine, short: TBL) is an international project with 27 partners in 10 countries. The Central Biobank Charité (ZeBanC) carries out the biobanking and ID management for the consortium.

Within the TBL project, ZeBanC is responsible for the coordination of sample-related processes, the provision of sample kits, storage and distribution of the samples to the analytical laboratories. The samples are collected and processed at the recruitment sites and sent periodically to ZeBanC for permanent, quality-assured storage at -80°C. On request, the samples will be sent to the companies and partners for sample analysis.

For biobanking - besides linking the samples with clinical data - the precise characterization of samples and sample-related processes is of great importance. As such, information on the primary container used, by reporting the coagulants, the time between collection and processing start or time of freezing as well as the processing parameters (centrifugation times etc.) are highly relevant in order to decide for which analyses the collected samples are suitable. The aim of this task was the standardized collection of these parameters and their linking to the unique patient pseudonym. Together with the partners ABX-CRO and ITTM, (i) a coordinated ID management, (II) processes for the IT-supported recording of sample-relevant data and (II) harmonized workflows for data transfer in a secure IT infrastructure were established. Our Biobank Information Management System (BIMS) has been adapted so that we can import and search the biobank data record for TBL samples and make it available to our partners as a report at any time.

1. Introduction

The ZeBanC is responsible for quality-assured biobanking based on the requirements of the clinical project partners (clinic, analysis companies, and data processing companies). Besides the definition and standardization of all sample-relevant processes, the storage and provision of samples for analysis, documentation also plays an important role. In the context of biobanking, this mainly refers to information that describes the samples and their quality (pre-processing time span, processing parameters, and storage conditions, etc.) with a link to the patient ID. This is particularly relevant in the context of biomarker analysis, since processing, transport and storage conditions can influence the results of the analysis. To record this information, we have established IT-supported processes and systems wherever possible. Paper documentation or Excel spreadsheets are only used for this purpose in exceptional cases or are used as a vehicle.

The implementation of this objective, which is described in detail below, was only possible in close cooperation with ABX-CRO - eCRF, data source - and ITTM - platform for secure data exchange.

2. Methods

2.1 Data workflow - Overview

In the meetings and discussions with the clinical partners, it quickly became clear that in TBL we have to process and integrate data/information originating from different sources and stored in different formats. A challenging task was and is therefore the standardization of data and data flows between the different partners and systems. For this reason, a close cooperation and intensive information exchange between ABX-CRO, ITTM and ZeBanC has been developed right from the start of the project.

In the discussions/strategies, the requirements including the technical/personnel prerequisites of the clinical partners were always considered leading to solutions which could be integrated into their daily routine.

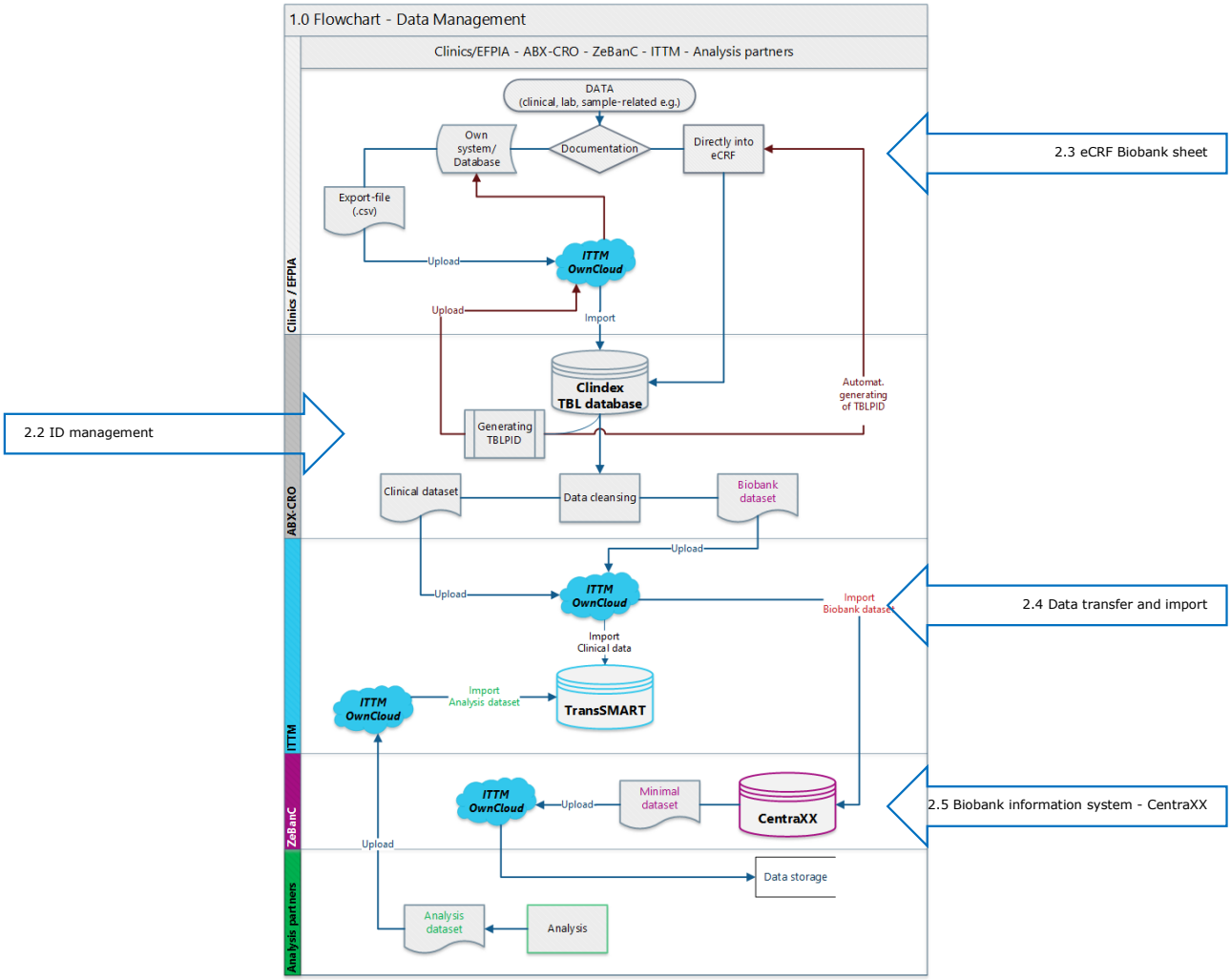


Figure 1 **Data management workflow (Version 20200320) between partners and systems**

2.2 ID Management

Using the established TBL-specific ID management (Appenx 1), patients are pseudonymized, whereby the potential connection of the patient to the clinical and sample-relevant data is always guaranteed. This has been established by ZeBanC in close cooperation with ABX-CRO and ITTM. Either ABX-CRO allocates the unique TBL Patient ID (=MASTER ID) automatically within the eCRF system or manually when the clinical partners provide their data. The ABX database (Clindex) does not store identifying data such as the name. Within the TBL consortium, the patient is managed using only this TBL Patient ID (MASTER ID), so that the privacy of the patient is maintained throughout the data collection, storage, transmission, analysis and reporting process.

Each responsible study physician maintains a secure patient list linking the patient to the TBL Patient ID. Identification of the patient is only possible via this list by the respective clinical partner. All measures to ensure that the list is protected against unauthorized access and an emergency plan are implemented by the recruitment sites.

Figure 2 Visualization of the TBL PID management. Based on the TBL Patient ID, patients can be quickly assigned to a study and a recruitment center.

2.3 eCRF Biobank sheet

The documentation of sample-relevant information as well as the allocation of the sample tubes to the patient is an important part of biobanking in addition to the adequate processing. In order to enable documentation for all partners, regardless of whether a data acquisition device is available in the laboratory or not, each sample kit also includes a sample data sheet. Since it is not practical for biobank staff to enter the information weeks after the collection of the samples, a biobank data sheet was created in the eCRF in cooperation with ABX-CRO. This is tailored to the requirements of each WP. It enables the clinical partner to enter the sample-relevant data directly into the eCRF. This leads to a high degree of standardization. From the database of ABX-CRO (Clindex), this data can then be processed and passed on to the ZeBanC.

BIOBANKING SERUM - VISIT XX		TBL Patient ID XXX	
Sample type	SERUM_SR		
Sampling date [yyyymmdd]		Sampling time [hh:mm]	
Master sample SR ID		Volume master sample SR [ml]	
1. Centrifugation condition	2,000 g 10' 4-8°C	Start 1. Centrifugation [hh:mm]	
Processing SOP conform	Yes / No		
Sample ID SR_1	scan barcode	Sample ID SR_2	scan barcode
Sample ID SR_3	scan barcode	Sample ID SR_4	scan barcode
Sample ID SR_5	scan barcode	Sample ID SR_6	scan barcode
Storage rack ID		Time of Freezing [hh:mm]	
Sample ID SR_7 (5 ml)	scan barcode		
Storage rack ID (5 ml)		Time of Freezing [hh:mm]	
BIOBANKING PLASMA - VISIT XX		TBL Patient ID XXX	
Sample type	PLASMA_EP		
Sampling date [e.g. 01-SEP-2019]		Sampling time [hh:mm]	
Master sample EP ID		Volume master sample EP [ml]	
1. Centrifugation condition	2,000 g 10' 4-8°C	Start 1. Centrifugation [hh:mm]	
Processing SOP conform_PL	Yes / No		
Sample ID EP_1	scan barcode	Sample ID EP_2	scan barcode
Sample ID EP_3	scan barcode	Sample ID EP_4	scan barcode
Sample ID EP_5	scan barcode	Sample ID EP_6	scan barcode
Storage rack ID		Time of Freezing [hh:mm]	
BIOBANKING LIQUID BIOPSY - VISIT XX		TBL Patient ID XXX	
Sample type:	LIQUID BIOPSY_LP		
Sampling date [e.g. 01-SEP-2019]		Sampling time [hh:mm]	
Master sample LB ID		Volume master sample LB [ml]	
1. Centrifugation condition	2,000 g 10' 4-8°C	Start 1. Centrifugation [hh:mm]	
2. Centrifugation condition	10,000 g 10' 4-8°C		
Processing SOP conform	Yes / No		
Sample ID LB_1	scan barcode	Sample ID LB_2	scan barcode
Sample ID LB_3	scan barcode		
Storage rack ID		Time of Freezing [hh:mm]	
BIOBANKING URINE - VISIT XX		TBL Patient ID	
Sample type	URINE_UR		
Sampling date [e.g. 01-SEP-2019]		Sampling time [hh:mm]	
Master sample UR ID		Volume master sample UR [ml]	
Sample ID UR_1	scan barcode	Sample ID UR_2	scan barcode
Sample ID UR_3	scan barcode	Sample ID UR_4	scan barcode
Storage rack ID		Time of Freezing_UR [hh:mm]	

Figure 3 WP4 template for the eCRF biobank sheet

2.4 Data transfer and import

The data exchange in case of the TBL project is file based. The files consist of comma-separated (.csv) data with fixed coordinated rows. To allow importing any kind of data in addition to manual input, our Biobank Information Management System (BIMS) CentraXX uses a Representational State Transfer (REST) Interface. Based on the interface we developed, an import tool that reads the given files, checks its data for correctness and imports the data into the system. The tool checks for errors in the file itself, i.e. if a field is missing, a sample is already known or a data field does not match the correct type, the tool will deny the import. Neither this tool nor the ZeBanC staff are able to correct general errors inside the file!

Sample data sheet	ABX TBL database	CentraXX
Sample data sheet column_desc	**ABX column_desc**	**CentraXX column_desc**
TBL Patient ID (TBLPID)	Master ID	TBL PAT ID
	Study name	Measurement profile TBL study characterization
Site name (SITEID)	Site	Measurement profile TBL study characterization
Patient population (POP)	Study population	Measurement profile TBL study characterization
	Date of consent/last visit (to be used with epoch)	Patient Consent Valid from
	Age unit - Years	Age
	Gender	Sex
	Disease term	ICD Code
Visit (V)	Visit	Visite
Master sample ID	Master sample ID	Proben ID
Sampling date	Sampling date	Extraction date (precision minute)
Sampling date	Sampling time	Extraction date (precision minute)
Sample type	Sample type	Sample kind
Volume master sample	Volume master sample (ml)	Initial amount
Start 1. Centrifugation	Centrifugation start time	Pre-Centrifugation delay (calculated), 1. Centrifugation
Sample ID	Sample ID	TBL Sample ID
Storage location	Lab storage rack ID	Storage location
Time of freezing	Time of freezing	Date of 1st storage
	Processing SOP conform	SOP
		Organization unit
		Primary container
		1. Centrifugation
		2. Centrifugation
		Remaining amount
		Sample container
		Date of receipt

Figure 4 Mapping table sample

A detailed mapping table (see Annex 1) ensures that the data from the ABX-CRO export to CentraXX is assigned correctly.

2.5 Biobanking Information Management System - CentraXX

The standardized collection of relevant data (sample/patient/clinical data) is an important part of biobanking. Within the ZeBanC the Biobank Information Management System (BIMS) CentraXX is used which was developed by the company Kairos. It consists of a MSSQL Database and a Java web frontend accessible from the inside of the Charité intranet and is hosted on the virtual server farm of the Charité. The BIMS is a patient-based system, where each patient is part of one or many organizational units or

clients. The client separates patients and samples from other collections or partners and connects them to the user access management of the system. Thus, it is possible to maintain many different studies or collections in parallel. Based on the rules the administration and coordination of the biobank is able to grant access to the data. In general, the system offers a fixed set of information linked to a patient, sample or other sets. It is possible to extend it by any other data. Another main feature is the ID management. A patient or sample consists of one or many IDs, each is stored in a so called ID Container, which clearly identify this element. These ID Containers guarantee a uniqueness of IDs inside the container, to exclude the same ID on different samples or patients. A Container consist of the set of IDs to fulfill uniqueness checks, a name like "TBL Sample ID" and an ID itself to identify the Container.

Data security, backup and recovery mechanisms

The BIMS is hosted on a virtual server at the Charité intranet, which is managed by the IT experts at the Charité. This includes backups of the entire system so that it can be fully restored in the event of a disaster. In case of a power failure, the system is connected to an emergency power generator. The entire operation is spread over two locations within Berlin with mirrored data at both sites. The database itself is secured by a mechanism on MS SQL level. There is a full backup every Sunday, an incremental backup every day and a transaction log backup every hour. The backup data is stored for 60 days. Every change on the data level is logged by the system and can be tracked by the ZeBanC administration. Users cannot delete entire patients or samples. Deletion requests are made by e-mail with the internal database IDs of the sample or patient. Access and data transfer between the client and server is SSL secured. Every user of the system needs an official Charité domain account. These accounts are validated and are only granted to authenticated persons. The domain account can then be used to log on to the information system.

Sample characterization

CentraXX offers a standardized interface for the acquisition and visualization of patient, disease and sample specific data. To describe the samples, the following data is documented within TBL, if available:

- TBL Patient ID
- Extraction date
- Visit
- Sample kind
- Primary container
- Initial amount
- Remaining amount
- Centrifugation parameters
- Pre-Centrifugation time
- Date of 1st storage

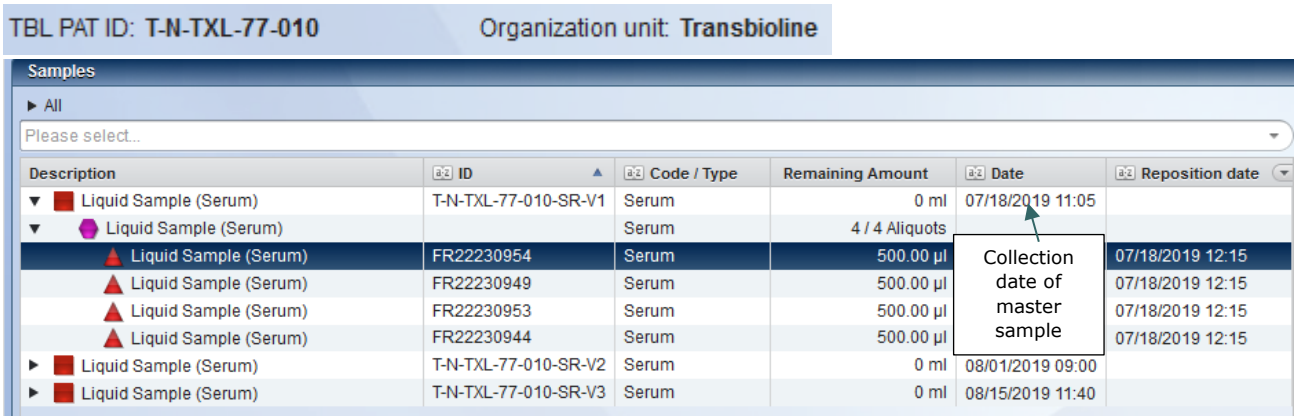


Figure 5 CentraXX screenshot with details regarding available samples and linkage to the TBL Patient ID (MASTERID)

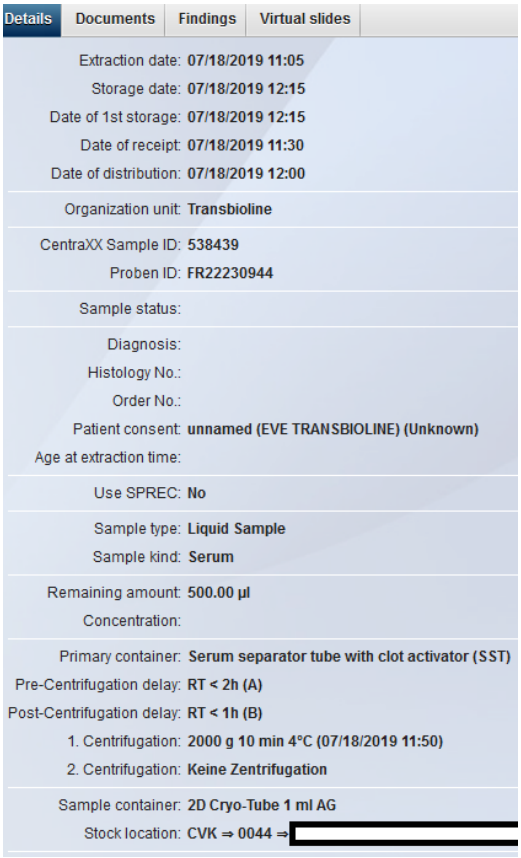


Figure 6 CentraXX screenshot with sample processing details

Apart from the diagnosis, sex and age at the time of sampling, no further patient-specific data is stored in the CentraXX biobank information system.

All other data from the clinical partners (demographic, clinical, laboratory) and from the companies carrying out the biomarker tests are combined in TranSMART and made available to the TBL partner for analysis.

Reporting and Search

CentraXX also offers a search function that enables the biobank staff to compile patient and sample cohorts taking into account certain parameters (sample type, processing type, rounds, etc.). This allows us to quickly create a sample list according to the requirements of the analysis partners, search for samples and provide them with a defined data set (sample related data and TBL Patient ID).

For a better overview regarding the recruitment status, we consider providing a table (optional) containing the following information:

- Work package
- Recruitment location
- Patient cohort
- TBL Patient ID
- Extraction date
- Sample type
- Volume
- Analyzed by … (optional)
- Assay …(optional)

Only samples that are stored in the biobank or that we have transferred to the analysis companies could be listed. The final design has still to be agreed on with the WP leads and the analysis companies.

3. Results

From the ZeBanC perspective, the great challenge within this TBL goal was and is to meet the different requirements of the partners and the studies. The data sets (clinics, laboratory, samples, etc.) of the individual studies differ from each other and the IT and personnel infrastructure at the different locations had to be considered. However, this should not lead to any compromises in terms of standardization and the associated fulfilment of quality requirements.

To achieve this, we established and maintained contact and information exchange with the partners at an early stage in order to jointly define the ID management, data flow and data records. This was the basis for the adjustments in our IT system CentraXX, the customizing of the importer and the definition of the data flows.

All important processes are now defined. Specifications such as ID management and the eCRF biobank sheet have been made available and the ZeBanC IT system has been adapted accordingly. The transfer and implementation of a first biobank specific data set can thus take place at any time. However, experience from other projects has shown that there will be further adaptations and changes afterwards.

4. Conclusion

In close collaboration with ABX-CRO, all data flows and data transfer processes have been defined. With the provision of the secure OwnCloud by ITTM, we have a secured

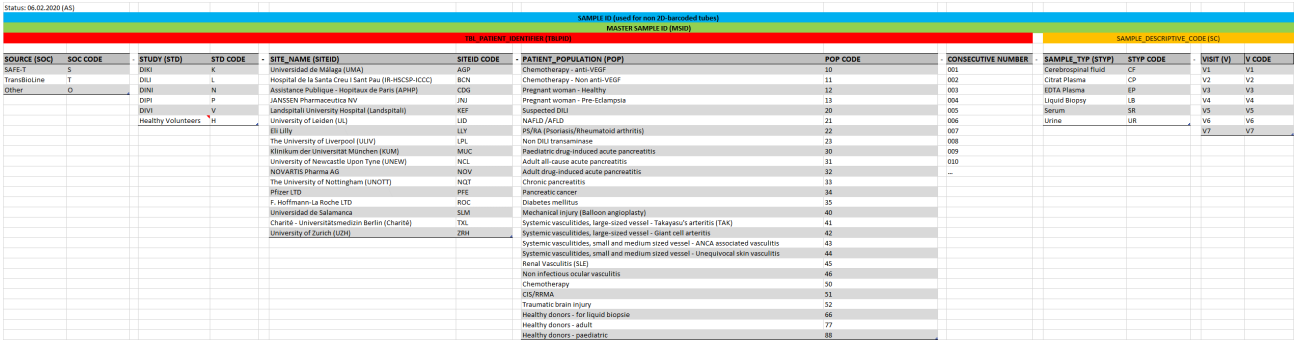
data transfer option. From the IT side, we are prepared in ZeBanC to the extent that we can perform an initial check of the export file from ABX-CRO before uploading it into CentraXX. Afterwards the transfer of the samples to the biobank can take place. This provides the basis for biomarker analysis by companies and laboratories.

Annexes

Annex 1 TBL ID Management

Annex 1_TBL_ID-Management_Version5.5_20200320.xlsx

General:



Status: 06.02.2020 (AS)

SAMPLE ID (used for non 2D-barcoded tubes)
MASTER SAMPLE ID (MSID)
TBL_PATIENT_IDENTIFIER (TBLPID) | SAMPLE_DESCRIPTIVE_CODE (SC)

SOURCE (SOC)	SOC CODE
SAFE-T	S
TransBioLine	T
Other	O

STUDY (STD)	STD CODE
DIKI	K
DILI	L
DINI	N
DIPI	P
DIVI	V
Healthy Volunteers	H

SITE_NAME (SITEID)	SITEID CODE
Universidad de Málaga (UMA)	AGP
Hospital de la Santa Creu i Sant Pau (IR-HSCSP-ICCC)	BCN
Assistance Publique - Hopitaux de Paris (APHP)	CDG
JANSSEN Pharmaceutica NV	JNJ
Landspitali University Hospital (Landspitali)	KEF
University of Leiden (UL)	LID
Eli Lilly	LLY
The University of Liverpool (ULIV)	LPL
Klinikum der Universität München (KUM)	MUC
University of Newcastle Upon Tyne (UNEW)	NCL
NOVARTIS Pharma AG	NOV
The University of Nottingham (UNOTT)	NQT
Pfizer LTD	PFE
F. Hoffmann-La Roche LTD	ROC
Universidad de Salamanca	SLM
Charité - Universitätsmedizin Berlin (Charité)	TXL
University of Zurich (UZH)	ZRH

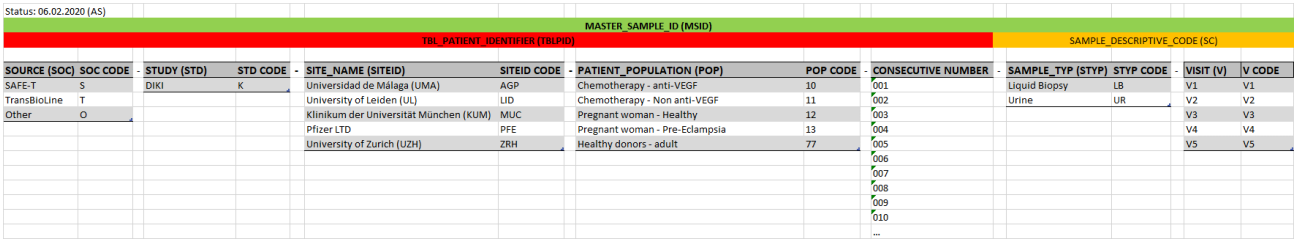
PATIENT_POPULATION (POP)	POP CODE
Chemotherapy - anti-VEGF	10
Chemotherapy - Non anti-VEGF	11
Pregnant woman - Healthy	12
Pregnant woman - Pre-Eclampsia	13
Suspected DILI	20
NAFLD /AFLD	21
PS/RA (Psoriasis/Rheumatoid arthritis)	22
Non DILI transaminase	23
Paediatric drug-induced acute pancreatitis	30
Adult all-cause acute pancreatitis	31
Pregnant woman - Healthy	32
Chronic pancreatitis	33
Pancreatic cancer	34
Diabetes mellitus	35
Mechanical injury (Balloon angioplasty)	40
Systemic vasculitides, large-sized vessel - Takayasu's arteritis (TAK)	41
Systemic vasculitides, large-sized vessel - Giant cell arteritis	42
Systemic vasculitides, small and medium sized vessel - ANCA associated vasculitis	43
Systemic vasculitides, small and medium sized vessel - Unequivocal skin vasculitis	44
Renal Vasculitis (SLE)	45
Non infectious ocular vasculitis	46
Chemotherapy	50
CIS/RRMA	51
Traumatic brain injury	52
Healthy donors - for liquid biopsie	66
Healthy donors - adult	77
Healthy donors - paediatric	88

CONSECUTIVE NUMBER
001
002
003
004
005
006
007
008
009
010
...

SAMPLE_TYP (STYP)	STYP CODE
Cerebrospinal fluid	CF
Citrat Plasma	CP
EDTA Plasma	EP
Liquid Biopsy	LB
Serum	SR
Urine	UR

VISIT (V)	V CODE
V1	V1
V2	V2
V3	V3
V4	V4
V5	V5
V6	V6
V7	V7

WP1 – DIKI:



Status: 06.02.2020 (AS)

MASTER_SAMPLE_ID (MSID)
TBL_PATIENT_IDENTIFIER (TBLPID) | SAMPLE_DESCRIPTIVE_CODE (SC)

SOURCE (SOC)	SOC CODE
SAFE-T	S
TransBioLine	T
Other	O

STUDY (STD)	STD CODE
DIKI	K

SITE_NAME (SITEID)	SITEID CODE
Universidad de Málaga (UMA)	AGP
University of Leiden (UL)	LID
Klinikum der Universität München (KUM)	MUC
Pfizer LTD	PFE
University of Zurich (UZH)	ZRH

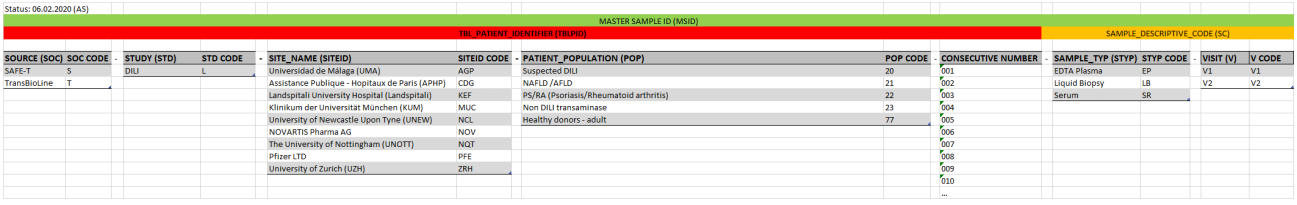
PATIENT_POPULATION (POP)	POP CODE
Chemotherapy - anti-VEGF	10
Chemotherapy - Non anti-VEGF	11
Pregnant woman - Healthy	12
Pregnant woman - Pre-Eclampsia	13
Healthy donors - adult	77

CONSECUTIVE NUMBER
001
002
003
004
005
006
007
008
009
010
...

SAMPLE_TYP (STYP)	STYP CODE
Liquid Biopsy	LB
Urine	UR

VISIT (V)	V CODE
V1	V1
V2	V2
V3	V3
V4	V4
V5	V5

WP2 – DILI:



Status: 06.02.2020 (AS)

MASTER SAMPLE ID (MSID)
TBL_PATIENT_IDENTIFIER (TBLPID) | SAMPLE_DESCRIPTIVE_CODE (SC)

SOURCE (SOC)	SOC CODE
SAFE-T	S
TransBioLine	T

STUDY (STD)	STD CODE
DILI	L

SITE_NAME (SITEID)	SITEID CODE
Universidad de Málaga (UMA)	AGP
Assistance Publique - Hopitaux de Paris (APHP)	CDG
Landspitali University Hospital (Landspitali)	KEF
Klinikum der Universität München (KUM)	MUC
University of Newcastle Upon Tyne (UNEW)	NCL
NOVARTIS Pharma AG	NOV
The University of Nottingham (UNOTT)	NQT
Pfizer LTD	PFE
University of Zurich (UZH)	ZRH

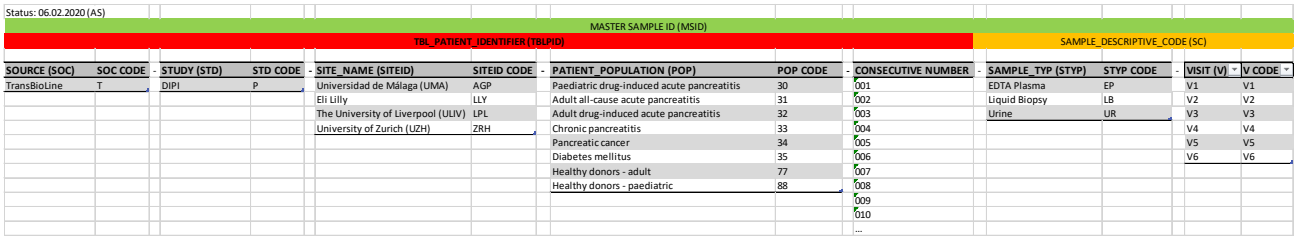
PATIENT_POPULATION (POP)	POP CODE
Suspected DILI	20
NAFLD /AFLD	21
PS/RA (Psoriasis/Rheumatoid arthritis)	22
Non DILI transaminase	23
Healthy donors - adult	77

CONSECUTIVE NUMBER
001
002
003
004
005
006
007
008
009
010
...

SAMPLE_TYP (STYP)	STYP CODE
EDTA Plasma	EP
Liquid Biopsy	LB
Serum	SR

VISIT (V)	V CODE
V1	V1
V2	V2

WP3 – DIPI:



Status: 06.02.2020 (AS)

MASTER SAMPLE ID (MSID)
TBL_PATIENT_IDENTIFIER (TBLPID) | SAMPLE_DESCRIPTIVE_CODE (SC)

SOURCE (SOC)	SOC CODE
TransBioLine	T

STUDY (STD)	STD CODE
DIPI	P

SITE_NAME (SITEID)	SITEID CODE
Universidad de Málaga (UMA)	AGP
Eli Lilly	LLY
The University of Liverpool (ULIV)	LPL
University of Zurich (UZH)	ZRH

PATIENT_POPULATION (POP)	POP CODE
Paediatric drug-induced acute pancreatitis	30
Adult all-cause acute pancreatitis	31
Adult drug-induced acute pancreatitis	32
Chronic pancreatitis	33
Pancreatic cancer	34
Diabetes mellitus	35
Healthy donors - adult	77
Healthy donors - paediatric	88

CONSECUTIVE NUMBER
001
002
003
004
005
006
007
008
009
010
...

SAMPLE_TYP (STYP)	STYP CODE
EDTA Plasma	EP
Liquid Biopsy	LB
Urine	UR

VISIT (V)	V CODE
V1	V1
V2	V2
V3	V3
V4	V4
V5	V5
V6	V6

WP4 – DIVI:

Status: 06.02.2020 (AS)

MASTER SAMPLE ID (MSID)																			
TBL_PATIENT_IDENTIFIER (TBLPID)																SAMPLE_DESCRIPTIVE_CODE (SC)			
SOURCE (SOC)	SOC CODE	-	STUDY (STD)	STD CODE	-	SITE_NAME (SITEID)	SITEID CODE	-	PATIENT_POPULATION (POP)	POP CODE	-	CONSECUTIVE NUMBER	-	SAMPLE_TYP (STYP)	STYP CODE	-	VISIT (V)	V CODE
SAFE-T	S		DIVI	V		Universidad de Málaga (UMA)	AGP		Mechanical injury (Balloon angioplasty)	40		001		EDTA Plasma	EP		V1	V1
TransBioLine	T					Hospital de la Santa Creu I Sant Pau (IR-HSCSP-ICCC)	BCN		Systemic vasculitides, large-sized vessel - Takayasu's arteritis (TAK)	41		002		Liquid Biopsy	LB		V2	V2
						Pfizer LTD	PFE		Systemic vasculitides, large-sized vessel - Giant cell arteritis	42		003		Serum	SR		V3	V3
						F. Hoffmann-La Roche LTD	ROC		Systemic vasculitides, small and medium sized vessel - ANCA associated vasculitis	43		004		Urine	UR			
						University of Zurich (UZH)	ZRH		Systemic vasculitides, small and medium sized vessel - Unequivocal skin vasculitis	44		005						
									Renal Vasculitis (SLE)	45		006						
									Non infectious ocular vasculitis	46		007						
									Healthy donors - adult	77		008						
												009						
												010						
												...						

WP5 – DINI:

Status: 06.02.2020 (AS)

MASTER SAMPLE ID (MSID)																			
TBL_PATIENT_IDENTIFIER (TBLPID)																SAMPLE_DESCRIPTIVE_CODE (SC)			
SOURCE (SOC)	SOC CODE	-	STUDY (STD)	STD CODE	-	SITE_NAME (SITEID)	SITEID CODE	-	PATIENT_POPULATION (POP)	POP CODE	-	CONSECUTIVE NUMBER	-	SAMPLE_TYP (STYP)	STYP CODE	-	VISIT (V)	V CODE
TransBioLine	T		DINI	N		Pfizer LTD	PFE		Chemotherapy	50		001		Cerebrospinal fluid	CF		V1	V1
Other	O					Charité - Universitätsmedizin Berlin (Charité)	TXL		CIS/RRMS	51		002		Serum	SR		V2	V2
									Traumatic brain injury	52		003					V3	V3
									Healthy donors - adult	77		004						
												005						
												006						
												007						
												008						
												009						
												010						
												...						

ID Visualization:

- TBL Patient ID (TBLPID)

SOURCE-STUDY-SITE_NAME-PATIENT_POPULATION-CONSECUTIVE_NUMBER

Example: TBL-DIKI-University of Leiden-Pregnant woman with Preeclampsia-001

Code: T-K-LID-12-001

- Master sample ID (MSID=Patienten-ID plus sample descriptive code)

SOURCE-STUDY-SITE_NAME-PATIENT_POPULATION-CONSECUTIVE_NUMBER-SAMPLE_TYP-VISIT

Example: T-K-LID-10-001-EDTA Plasma-Visit 1

Code: T-K-LID-12-001-EP-V1

Annex 2 Mapping Table

Annex 2_Mapping_Sample data sheet to ABX to CentraXX V1.xlsx

Status: Version 1 / 27.03.2020 AS					
Sample data sheet	ABX TBL database				CentraXX
Sample data sheet column_desc	**ABX column_name**	**ABX column_desc**	**ABX data_type**	**ABX data_format / description**	**CentraXX column_desc**
TBL Patient ID (TBLPID)	MASTERID	Master ID	CHAR	Generated by ABX-CRO or automatically within the eCRF: e.g. T-K-LID-06-001	TBL PAT ID
	STUDYID	Study name	LONG	DIKI, DILI, DINI, DIPI, DIVI	Measurement profile TBL study characterization
Site name (SITEID)	SITEID	Site	LONG	CODELIST ID MANGEMENT	Measurement profile TBL study characterization
Patient population (POP)	POP	Study population	LONG	CODELIST ID MANGEMENT	Measurement profile TBL study characterization
	DSSTDAT	Date of consent/last visit (to be used with epoch)	CHAR	ddmmmyyyy	Patient Consent Valid from
	AGEU	Age unit - Years	DATE		Age
	SEX	Gender	CHAR	1=male, 2 = female	Sex
	MHTERM	Disease term	CHAR	Diagnosis	ICD Code
Visit (V)	VISIT	Visit	LONG	CODELIST ID MANGEMENT	Visite
Master sample ID	MSID	Master sample ID	CHAR	Generated by ABX-CRO or automatically within the eCRF: e.g. T-K-LID-06-001-EP-1A	Proben ID
Sampling date	LBDAT	Sampling date	DATE	ddmmmyyyy	Extraction date (precision minute)
Sampling date	LBTIM	Sampling time	TIME	hh:mm	Extraction date (precision minute)
Sample type	LBSPEC	Sample type	CHAR		Sample kind
Volume master sample	LBVOL	Volume master sample (ml)	DECIMAL		Initial amount
Start 1. Centrifugation	LBCENTIM	Centrifugation start time	TIME	hh:mm	Pre-Centrifugation delay (calculated), 1. Centrifugation
Sample ID	LBREF	Sample ID	CHAR	e.g. FD22222222	TBL Sample ID
Storage location	LBSTORID	Lab storage rack ID	CHAR		Storage location
Time of freezing	LBFRETIM	Time of freezing	TIME	hh:mm	Date of 1st storage
	LBSPCCND	Processing SOP conform	LONG	CODELIST ID MANGEMENT	SOP
					Organization unit
					Primary container
					1. Centrifugation
					2. Centrifugation
					Remaining amount
					Sample container
					Date of receipt

Annex 3 Data and Knowledge Management Plan

Annex 3_D8.3 Data and Knowledge Management Plan.docx

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	31 07 2019
Delivery date	01 11 2019
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.

Table of 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	12th Mar 2020	First Update DMP – First Draft

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 Initial 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 LIMS (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 backup 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 will also be able to store samples and data about the end of the study. The basis for the storage period is the Informed Patient Consent Form 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 Patient Consent Form templates.
    - This information will be provided in the following version of the DMP for WP1 (DIKI), WP3 (DIPI), WP4 (DIVI) and WP5 (DINI) as currently the finalized Informed Patient Consent Form template is currently only available for WP2 (DILI).

- Please refer to section 4.2
- 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 eCRF database 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 her to deliverable D8.2 "Clindex eCRF (software)".

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 10 for the corresponding data dictionary.

4.2.1.2. Legacy SAFE-T data

The integration of the SAFE-T DIKI data in whole or in part into the TransBioLine database takes place after all ethical and patient-specific concerns have been evaluated. The decision on inclusion will be taken by the Executive Committee after obtaining the necessary support from specific TransBioLine consortium partners and the available and volunteer advise from the Ethical Advisory Board.

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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

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 her to deliverable D8.2 "Clindex eCRF (software)".

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 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 takes place after all ethical and patient-specific concerns have been evaluated. The decision on inclusion will be taken by the Executive Committee after obtaining the necessary support from specific TransBioLine consortium partners and the available and volunteer advise from the Ethical Advisory Board.

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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

4.2.3. Drug-induced Pancreatic Injury (DIPI)

4.2.3.1. Data originating from database export/import process

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 her to deliverable D8.2 "Clindex eCRF (software)".

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].

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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

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 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 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 takes place after all ethical and patient-specific concerns have been evaluated. The decision on inclusion will be taken by the Executive Committee after obtaining the necessary support from specific TransBioLine consortium partners and the available and volunteer advise from the Ethical Advisory Board.

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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

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 her to deliverable D8.2 "Clindex eCRF (software)".

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 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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

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

Please see Annex 15 for the biological sample-specific data dictionary. This dictionary maps to the biological sample metadata associated with the Biomarker data. The corresponding document shows the export format for the data transfer from ABX-CRO to ZeBanC/Charité (see spreadsheet "Export data for ZeBanC") as well as the template containing the sample information as provided to the analytics partners (see spreadsheet "Template for partners").

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, plans to support the sustainability of the TransBioLine data including the corresponding strategy for the ICF templates are currently under discussion and will be integrated into this document once they are available. Considering the current status, it is planned to make the ICF templates 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. De-

identification of personal data will be done on clinical partner side that collect the data before sharing data with other partners.

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 human readable and include metadata (e.g. study code and recruitment site) at the level of de-identified subject or samples.

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.

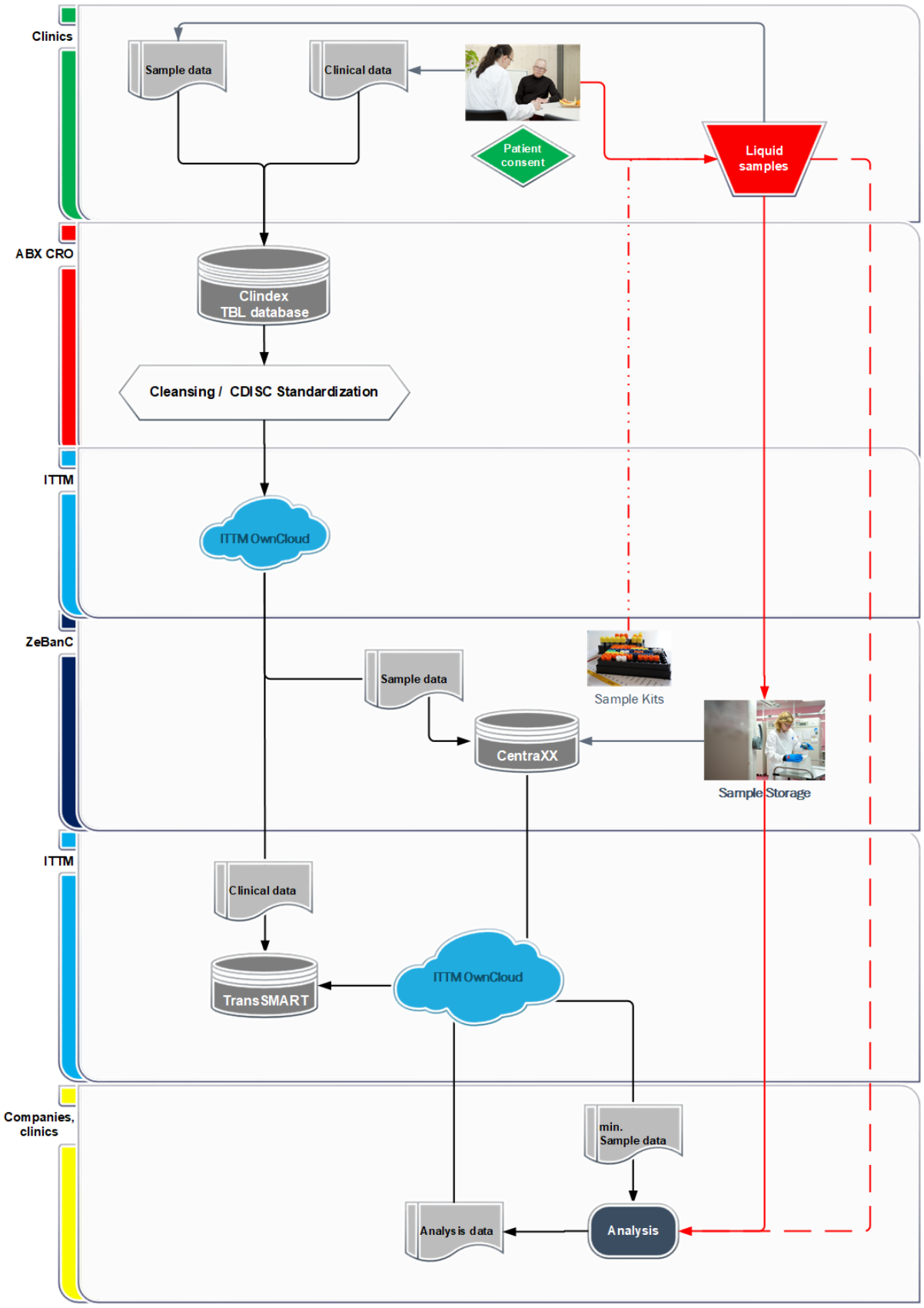


Figure 1: *Sample and data flow diagram in TransBioLine.*

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 Biobanking LIMS: a laboratory information management system (LIMS) 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 Biobanking LIMS 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.

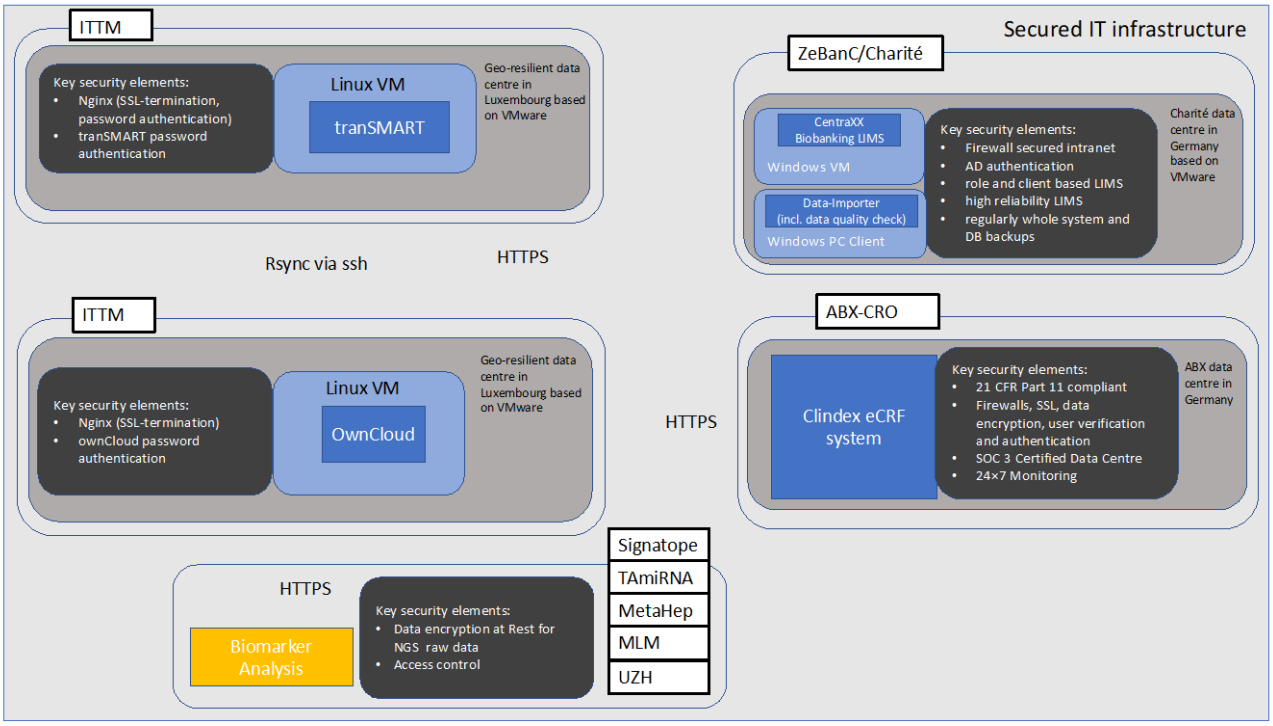


Figure 2: *Secured IT-infrastructure for data management in the context of TransBioLine. This infrastructure involves four key software components including Clindex system, CentraXX Biobanking LIMS, tranSMART as well as OwnCloud. The corresponding key security elements for these components are indicated (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 security 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 Biobank LIMS (CentraXX)

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 allow 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.

8. Conclusion

This first 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.

9. 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.

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

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.

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: Generic Informed Patient Consent Form

Instructions to site: This template is available to use for the TransBioLine study. Update all blue text and delete all red help text before submitting / using.

WP1-DIKI / WP2-DILI / WP3-DIPI / WP4-DIVI / WP5-DINI

TransBioLine Patient Information Leaflet

For informed consent concerning the donation, storage, and utilization of biological materials and supporting data as well as the

collecting, processing and usage for scientific research purposes

Dear patient,

You are currently being treated as a patient at the_____ *clinic/hospital or participate* in the_____ *study* and therefore are invited to take part in this 'Translational Safety Biomarker Pipeline (TransBioLine)' study.

This research study as well as the present patient information and Informed Consent Form (ICF) have received a favourable opinion from the independent Ethics Committee.

The study is funded by the Innovative Medicines Initiative (IMI). IMI is the world's biggest public-private partnership (PPP) in life sciences. It is a partnership between the European Union (represented by the European Commission) and the European pharmaceutical industry.

Your participation in this study is voluntary. You can decide anytime to withdraw your consent to the study without indicating any reasons. The decision not to participate or the premature withdrawal of your consent will not have any negative impact on your further medical care.

Please read the following information carefully as an addition to the oral consultation with your study doctor. Please do not hesitate to ask questions. In addition, please feel free to discuss the information given to you with your family or family doctor before making any decision.

Please sign this consent only:

- If you fully understand the course of this research study,
- If you are willing to give your consent to participate,
- If you are fully aware of your rights as a patient before, during and following a study as a participant.

1. Introduction

Below you will find information on the aims and scope of the kidney (WP1-DIKI), liver (WP2-DILI), pancreas (WP3-DIPI), vascupar system (WP4-DIVI), nervous system (WP5-DINI) or body fluids (WP6-liquid Biopsies) TransBioLine study (referred to as "TransBioLine study") and the measures that are being taken to protect your privacy that will enable you to form your own opinion before making a decision.

We are working with a central biobank at Charité Hospital Berlin. This biobank stores human biological materials such as blood, urine or tissues linked to selected medical information.

The human biological materials and supporting data collected for this study will be made accessible for medical research in an effort to improve the prevention, diagnosis and treatment of human diseases. Only projects with a positive ethics vote will receive samples and data in a pseudonymous or anonymised form.

If any points regarding the collected samples and data remain unclear, please ask your attending physician or your study physician before giving your consent. If you have any further questions at a later stage, you may also contact your study doctor who can provide you a list of study contact details.

2. Aim of this study

The TransBioLine study will focus on the development of protein biomarkers of drug-induced injury to liver, kidney, pancreas, vasculature, and central nervous system (CNS) and the development of novel techniques based on microRNAs. The analysis of microRNAs in liquid biopsies is a minimal-invasive alternative to a surgical biopsy. In this case, it is a blood sample, which is specially treated to quantify very small RNA molecules.

When your body is affected by a particular disease, tissue injury occurs, which in turn causes certain molecules to be released into the blood. These molecules are called biomarkers.

The TransBioLine study aims to study these biomarkers and make the data from this study available to academic centers and private scientific-related institutions to aid further research.

The biomarkers are expected to improve safety of new drugs and to contribute to better diagnosis and management of acute and chronic diseases.

3. How is the study conducted?

This clinical study will be conducted at several sites across Europe, United Kingdom as well as in the United States of America.

In total, xxx patients are planned to be included.

Your doctor has identified you for the TransBioLine study, because your disease condition indicates Kidney Injury / Liver Injury / Pancreas Injury / Vascular Injury / Central Nervous System (CNS) Injury, or as a healthy volunteer.

If you agree to take part in this study, we will ask you questions about your general health, conduct a general examination, and measure your weight, height, blood pressure, heart rate and temperature. The aim of this screening is to confirm that all inclusion criteria are met and no exclusion criteria are applicable to your medical data and findings.

4. What type of biomaterials and data are collected?

(Note : Delete the case which is not applicable)

The biological materials which we would like to use for research are body fluids (blood, urine or CSF) that:

Case 1 have been collected for diagnostic/therapeutic purposes in the course of your present hospital stay, but which are no longer required for such purposes and would, therefore, be destroyed otherwise.

Case 2. will be drawn / collected by qualified health professionals.

Blood volumes collected will correspond to those used in clinical routine. Total volume of samples will not exceed **5** mL (**1** teaspoon).

The information collected for the TransBioLine study will include information about your person (e.g. age, ethnic origin, gender, body weight and height); past and present medical history, in addition to information about your health and test results from examinations and procedures done during this study. You doctor will ask you about your current medication too.

The biobank only stores a minimal dataset with the sample such as the patient ID (pseudonym), age, gender and sample information (date and time of sampling and freezing).

5. What personal benefit do I have from study participation?

You will not directly benefit by taking part in this study, because the most important health benefits will be realized many years from now. Rather, your participation will contribute to the advancement of scientific knowledge and help future patients by improving our understanding of factors that affect the health of the population.

6. Do any costs arise from my study participation? Do I receive any compensation or payments?

There is no charge to you for taking part in this study.

You will not receive any remuneration for donating your biomaterials and/or data for medical research purposes. Should such research result in commercially exploitable results, any profits will not be shared with you.

7. What are the risks associated with your donation?

a. Health risks:

(Note : Delete the case which is not applicable)

Case 1: *Only residual material will be used.*

As we intend to use only biomaterials that will be collected in the context of your diagnosis or treatment and that, otherwise, would be destroyed as residual material, the donation does not entail any additional health risk for you.

Case 2: We would like to draw xx ml of blood (corresponding to approximately xx teaspoons). Mild bruising around the area where the needle went into the vein is fairly common after a blood test. However, in some rare cases, transient inflammation may develop and the skin gets red and swollen. You may experience dizziness during or after a blood test; this is very common in people who have a fear of needles and injections. It is fairly common to have a haematoma but the bruising should heal independently over the course of time.

b. Further risks:

Any collection, storage and transfer of data related to your biomaterials in the context of (medical) research projects entails the risk of breaches of confidentiality (e.g. the possibility of identifying you). These risks cannot be completely excluded, Charité – Universitätsmedizin Berlin biobank will take all appropriate measures according to the current state of technology to protect your privacy and will transfer samples and/or data only to researchers/projects who can demonstrate appropriate data protection and confidentiality safeguards (*see* Item 8: “Who has access to your biomaterials and data?”).

8. Who has access to your biomaterials and/or data and how are they protected?

The object for the TransBioLine study is to make medical data available to academic centers and private scientific-related to aid further research, therefore your biomarker results will be used for a wide range of medical research. At present, it is not possible to describe all future medical research objectives. These can either refer to the above defined disease areas or to diseases that at present are still partially unknown (e.g. cancer, cardiovascular diseases, brain disorders). Thus it is possible that your biomaterials and data may also be used for research purposes which, at this stage, are unknown.

All personal data will only be gathered, processed and used in accordance with current applicable privacy protection acts. Personal data are information that can be used to determine your identity. Your study doctor and his study team will maintain a patient file. The patient file contains personal data such as age, gender, etc. as well as medical results of prior and/or current treatments and further medical information regarding your study participation.

The funding body (IMI) of this trial and its representatives as well as authorised persons from regulatory authorities or independent Ethics Committees must be allowed access to your patient file for inspection and supervision purposes. In case of premature withdrawal of your consent to study participation, your personal data including health data which has been stored up to that point, might be used for future evaluations of study treatment or compilation of application documentation.

Some study data will need further analysis. For this reason, the transfer of data to third parties will only be made in pseudonymised form.

We also expect to receive access requests from overseas researchers and international collaborators. These researchers must follow the same procedures as all other researchers. All access is subject to the strictest scientific and ethical scrutiny, as described above.

a. Any data that directly identifies you (name, date of birth, address, etc.) will not be collected and will be replaced by a code (pseudonymized, encoded). Following this, the encoded data set is re-coded again before it is stored. Based on current knowledge, this double encoding/pseudonymization procedure minimizes the possibility that you may be re-identified by unauthorized parties. The bio-materials and/or data will only be made available for research purposes in this form (i.e. double pseudonymized).

All biomaterial samples will be pseudonymised and cannot be linked to you in any way.

b. Data that directly identifies you (personal identifying data) remain at the hospital in which the biomaterials and data have been obtained and will be stored separately from the biomaterials and related clinical data. Access to personal identifying data is necessary only in case additional or missing medical data is needed from your medical records, or in case of a need to re-contact you personally. In no case will personal identifying data be transferred to scientists and/or other unauthorized third parties, such as insurance companies or employers.

A list that links the coded information with your identity will be kept secure by the study doctor, to allow for your re-identification in certain circumstances. Your unique code will enable us to link the information from different data sets to you, but at the same time, will enable us to keep your identity confidential when we give your data to other researchers to use.

c. Based on pre-defined criteria and following a request/application, the double-encoded biomaterials and medical data may be transferred to other universities, research institutes and research companies, including those in foreign countries, for medical research. Under certain circumstances these data may be linked to medical data from other databases, provided that all legal and regulatory requirements are met.

d. Biomaterials and/or data that are transferred to third parties may only be used for the research purpose indicated in the application and must not be passed on by the recipient for other purposes. Material that has not been utilized will be returned to the biobank or destroyed.

e. Research results for scientific publication will be anonymized, i.e. data will be published only in a form that does not allow re-identification.

While study information could be printed in journals or shared with other people at scientific meetings or for teaching purposes, it will not be possible to identify you. Your identity will be kept confidential. All data will be presented as group data, rather than individual data. Also, specific rules regulate access to your data and samples by researchers.

You can soon find out more about how we use your information at www.TransBioLine.com

9. What are the constraints and safeguards for the use of your biomaterials and data?

Your biomaterials that have been stored in the TransBioLine biobank (or Charite) would be used exclusively for the research aims of the TransBioLine. Only approved research institutions can gain access to your coded data and samples, in order to protect your privacy.

a. A mandatory pre-requisite for the acquisition and use of your biomaterials and related personal data for research purposes is your written consent. Your consent is voluntary and can be withdrawn at any time (see also Item 10 “What does your right of withdrawal include?”).

By giving TransBioLine permission to use your biomaterials and/or data, you also transfer ownership of the biomaterials to TransBioLine. You retain the right to correct any data that might have been incorrectly stored or processed at any time.

b. Your biomaterials will be stored in Central Biomaterial Bank (ZeBanC), Berlin under standardized quality and security conditions and are available for (medical) research purposes on request only. They are protected against unauthorized access according to the current state of technology. Your supporting health data will be stored at ITTM Luxembourg under standardized security conditions and are available for (medical) research purposes on request only. They are protected against unauthorized access according to the current state of technology.

c. *If applicable and required by your ethics:* A mandatory pre-requisite for the use of the biomaterials and data for a specific medical research project is a review by an ethics committee. The ethics committee assesses/evaluates the ethical and legal aspects of the respective research project.

The biomaterial samples are intended to be stored and made available for medical research for 10 years. The pseudonymised biomarker results and supporting data will be stored for at least 25 years after completion of Transbioline. Further use or deletion of data beyond 25 years will have to be decided upon at that time

10. What does your right of withdrawal include?

You are free to withdraw your consent for the use of your biomaterials and/or data at any time without giving a reason and without any fear of detriment. In case of withdrawal it is up to you to decide whether your biomaterials are to be destroyed and the corresponding data to be deleted, or whether they may be used in an anonymized form (that is, without any link to your person, *see* Item 8e) for further medical research projects. However, as soon as the link between the biomaterials and data and your person has been removed, previously donated biomaterials can no longer be destroyed. In addition, data cannot be removed from already completed studies/scientific analyses.

For withdrawal, please contact your local doctor.

11. Who can I contact for further information or in case of questions?

The present clinical study has received favourable opinion from the independent Ethics Committee.

If you have questions about this study, you should first discuss them with the clinic staff (contact details on the front page), study leaders or the ethics committee (details below):

For any questions about your rights as a study participant please contact:

Add details of responsible institution/contact person, e.g. complaints officer within the hospital or Patient Advice and Liaison Service (PALS)<address>

Telephone:

Email:

Or

[Research ethics committee name & contact details].

Additional Information on the New European Union General Data Protection Regulation (GDPR)

If you wish to raise a complaint on how we have handled your personal data, you can contact our Data Protection Officer who will investigate the matter. If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO).

Our Data Protection Officer is _____ and you can contact them at _____.

Consent Form

I have read this document/had its contents explained to me. I have been given the opportunity to ask questions about the research study procedures and the potential risks have been explained to me. I have been given time to discuss with others to decide whether to agree to take part in this study.

I, hereby, agree that *[provide name of institution (clinic)/location of record]*

- **collects and stores my personal identifying data**
- **collects/extracts additional information on my health from my health records,**
- **and makes the data together with my biomaterials available in pseudonymized (that is, encoded) form for medical research projects.**

My biomaterials and data may be used for medical research projects for up to 10 years (biomaterial) / at least 25 years (Data) after completion of Transbioline.

1. It has been explained to me that I am free to leave the study at any time, without any disadvantage to my future care. I do freely give my consent to join this study, as described to me in this document. I understand that I will receive a copy of this document as signed below.

2. By signing this consent form, I give permission for **my biomaterials and data to be transferred in a pseudonymized form to universities, research institutions and research companies for medical research purposes.**

This may include transfer of double-encoded biomaterials and/or data for research projects involving foreign countries with a lower level of data protection

3. At the end of the study unused blood samples will be stored and, if you agree, may be used in future investigations to contribute to better diagnosis and management of acute and chronic diseases.

You can still be part of this study, even if you choose for your unused samples to be destroyed after the study has ended. Approval from the relevant ethics committees will be sought before the stored blood samples are used for any such future investigations. Samples from those who do not give consent for use of their blood samples after the study has ended will be destroyed.

Please let us know what you would like us to do with any samples left over at the end of the study (tick one only):

☐ Destroy all left over blood/urine/csf samples

☐ Keep samples for use in future research to understanding and identifying biomarkers that provide insights into mechanisms of tissue injury

I have received one copy each of the patient information sheet and the Informed Consent Form. The original remains my study doctor.

_____

Patient's/Participant's name (in printed letters):

_____ _____

Date (to be completed by patient/participant) Signature of the patient/participant

I conducted the patient/study participant consultation and have obtained the patient's/ participant's consent.

_____

Physician name (in printed letters):

_____ _____

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.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: Biological sample-related Data dictionary

Annex_15_Biobank_Samples_Data_Dictionary.xlsx