

Zentrale Biobank der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster (ZBB-MFM)

Kerndatensatz der ZBB-MFM

für Daten-Management/-Migration/-Import sowie Reporting und Suche mittels
CentraXX

Autor: Dr. Lorenz Dobler

Leitung: Prof. Dr. Klaus Berger, Institut für Epidemiologie und Sozialmedizin an der
Westfälische Wilhelms-Universität Münster

Inhaltsverzeichnis

1. Hintergrund	3
2. ZBB-Kerndatensatz	3
3. Datenfluss	3
4. Attribut-Beschreibung	4

1. Hintergrund

Primärer Auftrag der Zentralen Biobank der Medizinischen Fakultät (ZBBMFM) ist die **qualitätsgesicherte Lagerung und Dokumentation von Biomaterialien bzw. Proben und -Aliquote sowie damit verknüpfter Daten zu Personen sowie zu beteiligten Organisationen**. Auf Basis möglichst einheitlicher interner Codierungen sollen neben dem Routinebetrieb (Registrierung von Personen und Proben, Aufbereitung, Aliquotierung, Herausgabe etc.) auch **einheitliche, d.h. organisations-übergreifende Personen- und Proben-Recherchen** sowie **Reportings und Abrechnungen** ermöglicht werden. Das zentrale IT-Tool der ZBB-MFM für die genannten Aufgabenbereiche ist die browser- und datenbank-gestützte Software CentraXX der Firma Kairos.

2. ZBB-Kerndatensatz

Für alle Aufgabenbereiche ist ein Kerndatensatz erforderlich, der sowohl im laufenden Routinebetrieb als auch im Zuge von Daten-Imports bzw. Daten-Migrationen aus Altsystemen nach CentraXX zu erbringen ist. Der **ZBB-Kerndatensatz** beschreibt also im Wesentlichen Attribute, die zum Zwecke des internen Daten-Managements für Personen und Proben mittels CentraXX absolut notwendig sind („Pflichtfelder“).

Die im lokalen CentraXX gespeicherten Daten sollen auch für den nationalen und internationalen elektronischen Austausch und die weitergehende digitale Nutzung medizinischer Daten geeignet sein (z.B. Im Rahmen des geplanten Deutschen Forschungsdatenportals für Gesundheit der MII¹). Dazu soll der ZBB-Kerndatensatz durch ausgewählte personen-bezogene Attribute, d.h. forschungsrelevante medizinische Annotationen erweitert werden und sich so weit wie möglich am Kerndatensatzmodul Biobank der MII² orientieren. Der so um forschungsrelevante „Such“-Attribute **erweiterte ZBB-Kerndatensatz** beinhaltet alle Pflichtfelder des ZBB-Kerndatensatzes und darüber hinaus weitere medizinische Annotationen sowie ggfs. weitere Attribute zur standardisierten Beschreibung von Proben und Organisationen. Genauer Umfang und Spezifikation zusätzlicher Attribute im erweiterten ZBB-Kerndatensatz sind im Moment noch nicht abschließend geklärt und müssen bei Bedarf darüber hinaus auch auf neue Anforderungen angepasst werden. Zu Beginn angestrebt sind hierbei zunächst Angaben zu Geschlecht und Alter der Personen sowie zu allgemeinen Diagnosen und Tumor-spezifischen Diagnosen.

3 Datenfluss

Im Routinebetrieb werden Daten direkt über lokale CentraXX-Workstations eingegeben. Aus dem UKM-KIS (ORBIS) werden im Routinebetrieb ausschließlich Personen (Stamm)Daten per HL7-Schnittstelle automatisch bei Vorliegen eines Broad Consents übertragen. Daten aus Studien außerhalb des UKM-Behandlungszusammenhanges werden bei Bedarf via CSV-/XML-Schnittstellen manuell in CentraXX importiert. IDat werden in CentraXX in einer extra zugriffsgeschützten Datenbank verwaltet und stehen mit andern Daten(banken) in CentraXX nur mit einer PersonenID (PID) in Verbindung. Consente- werden in CentraXX mit allen relevanten Attributen verknüpft, um eine unberechtigte Proben-Datennutzung zu

¹ <https://www.medizininformatik-initiative.de/de/use-cases-und-projekte/abidemi> (zuletzt besucht am 10.06.2022)

² Das Kerndatensatzmodul Biobank der Medizininformatik Initiative/MII berücksichtigt mehrere internationale Standards zum Austausch medizinischer Daten (inkl. Biobank-Attribute im Forschungskontext, u.a. MIABIS - Minimum Information About Biobank Data Sharing, SPREC - Standard Preanalytical Code; SNOMED CT - Systematized Nomenclature of Medicine Clinical Terms). Diese Standards bilden auch die Grundlage für die Erstellung der Suchdatensätze des Sample Locator des German Biobank Nodes/GBN sowie des BBMRI-ERIC-Directory. Siehe Deppenwiese, N. et al. (2022): Das Kerndatensatz-Modul „Biobank – Bioprobendaten“ der Medizininformatik-Initiative. – In: Altmann, H. et al. (Hrsg.): Tagungsband des 10. Nationalen Biobanken-Symposiums vom 01.-02. Juni in Berlin: 27-30.

verhindern. Für Datenrecherchen wird ausschließlich auf CentraXX-Teil-Datenbank zurückgegriffen, wobei im Vorfeld üblicherweise eine Datenaufbereitung (= ETL-/Extract-Load-Transform-Prozess) notwendig ist.

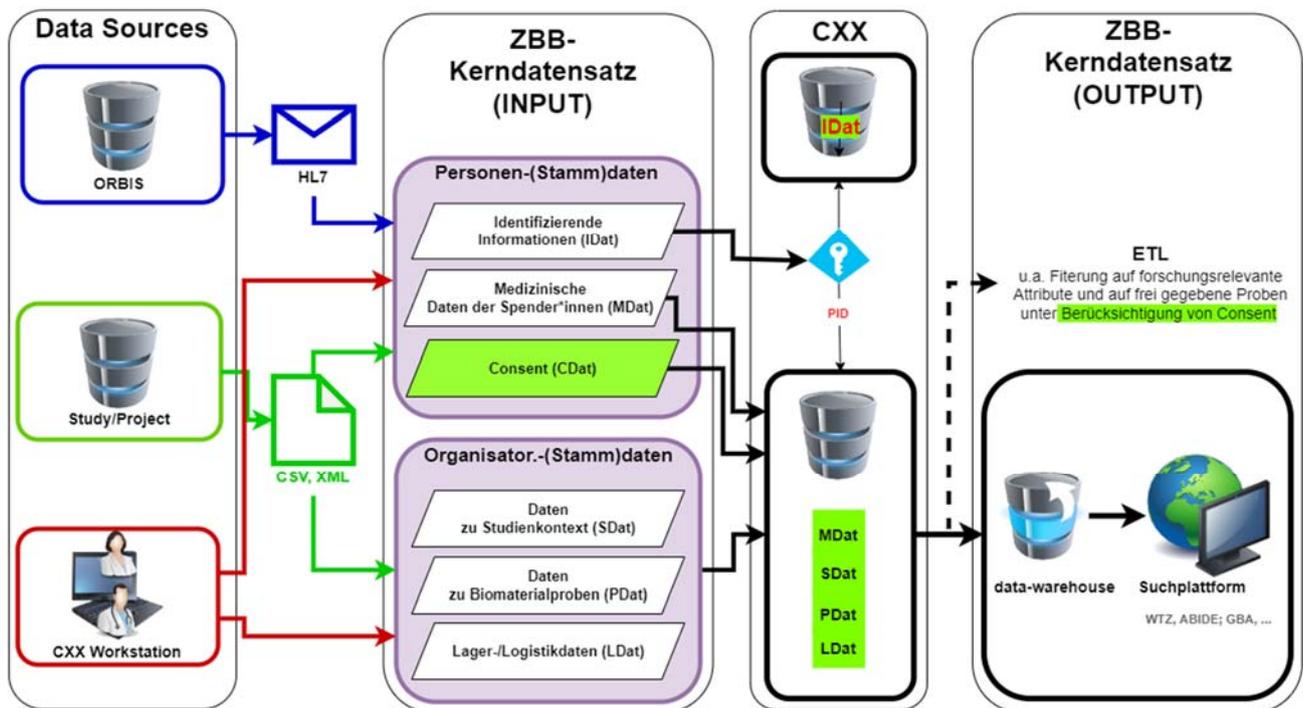


Abbildung 1 Schema Datenfluss ZBB-MFM.

4. Attribut-Beschreibung

Die Attribute des ZBB-Kerndatensatz sind inhaltlich in Gruppen und zwei Haupt-Kategorien gegliedert.

- **Personen-(Stamm)Daten (PersDat)**
 - **Identifizierende Informationen (IDat):** Eindeutige Kennzeichnung von Personen über ID, Name, Vorname, Geburtstag, ggfs. Eingangsdiagnose.
 - **Medizinische Daten der Spender*innen (MDat):** über die Eingangs-/Hauptdiagnose hinausgehende forschungsrelevante und personen-bezogene medizinische Annotationen
 - **Consent-Daten (CDat):** Art der Spender*innenzustimmung und ggfs. entsprechende Einschränkungen. Consent-Daten nehmen eine Sonderstellung ein, weil diese mit allen anderen consent-relevanten Attribute verknüpft werden müssen um eine unberechtigte Proben- Datennutzung zu verhindern.
- **Organisatorische (Stamm)Daten (OrgDat)**
 - **Daten zu Biomaterialproben (PDat):** Charakterisierung und eindeutige Identifizierung von Biomaterialien.
 - **Lagerdaten zu Biomaterialproben (LDat):** internen Organisation/Lagerung von von Biomaterialien.
 - **Daten über den Studienkontext (SDat):** Eindeutige Zuordnung jeder Probe zu einer Organisationseinheit (OE) zur Steuerung der Zugriffs- und Nutzerrechte.

In den folgenden Tabellen sind alle Attribute des ZBB-Kerndatensatzes nach Datengruppen getrennt aufgelistet und anhand von Beispielen erläutert. Die Attribute sind zusätzlich mit dreistelligen Sequenz-

Nummernbereichen versehen. In der Spalte „**Attribut-Format**“ beschreibt „**CXX**“ Feldausprägungen, die vor Import in CentraXX definiert sein müssen („Stammdaten“). Vor Import nach CXX müssen Datenfelder ggfs. entsprechend übersetzt werden.

Die Attribute in den Tabellen sind je nachdem ob es sich um ein Pflichtfeld oder ein optionales Feld sowie danach ob das Feld aus ORBIS stammt speziell formatiert.

Fett	<u>Pflicht-Attribut</u> <ul style="list-style-type: none"> • muss bei Daten-Import/-Migration geliefert werden • muss im laufenden Betrieb bei neuen Spender*innen und/oder Proben befüllt werden
<i>Kursiv</i>	<u>Zusätzliches/optionales (ORBIS-)Attribut</u> <ul style="list-style-type: none"> • wird für UKM-Patienten automatisch via HL7-Schnittstelle aus ORBIS befüllt und laufend aktualisiert • Im Moment nur für Patienten mit Broad Consent vorgesehen • Für Projekte/Studie bei Bedarf manuell nach CXX zu importieren und zu pflegen
Fett & Kursiv	<u>Pflicht-(ORBIS-)Attribut</u> <ul style="list-style-type: none"> • wird für UKM-Patienten automatisch via HL7-Schnittstelle aus ORBIS befüllt und laufend aktualisiert • Im Moment nur für Patienten mit Broad Consent vorgesehen • Für Projekte/Studie bei Bedarf manuell nach CXX zu importieren und zu pflegen
grau	<u>Nicht Such-relevante Attribute</u> <ul style="list-style-type: none"> • Rein intern/organisatorisch relevante Attribute • nicht relevant für Such-Maske und/oder Ausgabe der aggregierten Suchergebnisse

Tabelle 1 Beschreibung der Attribute und Formate des ZBB-Kerndatensatzes mit Beispielen

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
IDat (Patienten, Probanden, Spender)	101	Personen_ID (PID)	Code/ID für Person (SpenderIn/PatientIn/ProbandIn)	FREITEXT	6554k
	102	IDArt_Patient	Kategorie Personen_ID	CXX	ORBIS-PID, Projekt-PID
	103	Vorname	Vorname	FREITEXT	Max
	104	Nachname	Nachname	FREITEXT	Mustermann
	105	GeburtsDatum	GeburtsDatum	[DD.MM.JJJJ, XX.MM.JJJJ, XX.XX.JJJJ]	01.09. 2014, XX.09. 2014, XX.XX. 2014
	106	GeburtsDatum_Precision	Genauigkeit Geburtsdatum (nur wenn Datum nicht tagesgenau)	[DAY, MONTH, YEAR]	DAY, MONTH, YEAR

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
SDat (Studienkontext)	201	Episode_ID	Code/ID für Episode, Visite, FollowUp	FREITEXT	Visite1, Baseline, Followup1
	202	IDArt_Episode	Kategorie Episoden-/Visiten-/FollowUp-ID	CXX	ORBIS-Fallnr., Nexus-Fallnr., Baseline, FollowUp1
	203	EpisodeDatum	Datum Visite, Episode, FollowUp (i.d. Regel identisch mit ProbenDatum)	[DD.MM.JJJJ, XX.MM.JJJJ, XX.XX.JJJJ]	01.09. 2014, XX.09. 2014, XX.XX. 2014
	204	EpisodeDatum_Precision	ggfs. Genauigkeit Datum Visite/Episode/FollowUp (nur wenn Datum nicht tagesgenau)	[DAY, MONTH, YEAR]	DAY, MONTH, YEAR
	205	Organisationseinheit_ID	CentraXX-Organisationseinheit zur Steuerung von CXX-Zugriffsrechten	CXX	Pathologie, Neurologie

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
PDat (Probanddaten)	301	Proben_ID	Code/ID für Probe/Aliquot (Kennzeichnung Primärgebinde/Röhrchen)	FREITEXT	FGGJI123
	302	IDArt_Probe	Kategorie Proben-/Aliquot-ID	CXX	Etikett(-Beschriftung), Barcode
	303	ProbenOwner	Organisation und/oder Person mit Vetorecht auf die Verwendung von Proben und Daten	CXX	Neurologie, Pathologie
	304	ProbenFreigabe_SUCHE	Kennzeichnung Freigabe für allgemeine Suche (ZBB)	CXX	ZBB (= allg. Freigabe für globale Suche auch ausserhalb der Organisationseinheit)
	305	ProbenStatus	Probenstatus	CXX	[Draft, Disposed, Shipped, Released, ...]
	306	ProbenDatum	Datum Probenentnahme (i.d. Regel identisch mit EpisodeDatum)	[DD.MM.JJJJ, XX.MM.JJJJ, XX.XX.JJJJ]	01.09. 2014, XX.09. 2014, XX.XX. 2014
	307	ProbenDatum_Precision	ggfs. Genauigkeit von Datum Probenentnahme (nur wenn Datum nicht tagesgenau)	[DAY, MONTH, YEAR]	DAY, MONTH, YEAR
	308	ProbenTyp	Art der Probe	CXX	Plasma, (aus) EDTA(-Vollblut) & (mittels Monovette mit) Trenngel (gewonnen) ODER entsprechender CXX-Code (BF_PLA_ED-GEL)
	309	PrimGebindeTyp	Primäres Probenbehältnis (Röhrchen, Tube, Vial)	CXX	(Kryo-)Lager-Tube, 1 mL, Aussengew., Etikett o. Barcode ODER entsprechender CXX-Code (CTU_1000_3_04)
	310	ProbenMenge	Menge der Probe (Volumen, Gewicht, Anzahl, ...)	DEC	1,8
	311	ProbenMengeEinheit	Einheit für Menge der Proben (mL, mg, ...)	CXX	mL

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
LDat 2(Lager-/Logistikdaten)	401	SekGebinde_ID	ID Sekundäres Probenbehältnis (Box, Container)	FREITEXT	NC12746453263
	402	SekGebindeTyp	Typ Sekundäres Probenbehältnis (Box, Container)	CXX	SBS, 35 mm, 12x8, 1D Barcode integr. ODER entsprechender CXX-Code (SBS-PC_035_12x8_10)
	403	Lagerpfad	komplette Lagerhierarchie bis Sekundäres Gebinde (Box, Container)	CXX (bei Probenimport ggfs. aus vorhandenen CXX-Locations/Boxen ableiten)	Gebäude A -> Kühlgerät B -> Segment1 -> Gestell 3 -> Schublade4 -> Stellplatz 2
	404	LagerTemperatur	Aktuelle Betriebstemperatur des Gerätes in dem das Gebinde gelagert wird	INT	-80
	405	ProbePositionX	X-Position einer Probe in einer Box mit XY-Koordinaten	[A, B, C,...]	A
	406	ProbePositionY	Y-Position einer Probe in einer Box mit XY-Koordinaten	[1, 2, 3,...]	3

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
MDat (medizinische Annotation)	501	Geschlecht	Geschlecht	CXX	männlich
	502	Alter	Alter/Altersintervall bei Visite, FollowUp (zu berechnen aus EpisodeDatum - Geburtsdatum)	INT/?	38
	503	Diagnose_H1	Hauptdiagnose	ICD-10-Code	E11.01 (= Type 2 diabetes mellitus with hyperosmolarity with coma), ggfs. auch nur Haupt-Kategorie (Endocrine, nutritional and metabolic diseases E00-E89)
	504	Status_H1	Hauptdiagnose, Status	[aktuell, Z.n.]	aktuell
	505	Sicherheit_H1	Hauptdiagnose, Diagnosesicherheit	[Gesichert, Verdachtsdiagnose, Selbstauskunft, unbekannt]	gesichert
	506	Diagnose_T1	Haupt-Tumordiagnose	ICD-O-Code	
	507	Morphologie_T1	Haupt-Tumordiagnose, Morphologie	ICD-O-Code	
	508	Status_T1	Haupt-Tumordiagnose, Status	[aktuell, Z.n.]	Aktuell
	509	Sicherheit_T1	Haupt-Tumordiagnose, Diagnosesicherheit	[Gesichert, Verdachtsdiagnose, Selbstauskunft, unbekannt]	Verdachtsdiagnose
	510	Diagnose_N1/2/3	Nebendiagnose 1/2/3	ICD-10-Code	vgl. Hauptdiagnose
	511	Status_N1/2/3	Nebendiagnose 1/2/3, Status	[aktuell, Z.n.]	vgl. Hauptdiagnose
	512	Sicherheit__N1/2/3	Nebendiagnose 1/2/3, Diagnosesicherheit	[Gesichert, Verdachtsdiagnose, Selbstauskunft, unbekannt]	vgl. Hauptdiagnose

Attribut-Gruppe	Attribut-ID	Attribut-Name	Attribut Beschreibung	Attribut-Format	Beispiel
CDat (Consent)	601	<i>EinwilligungName</i>	<i>Name der Einwilligung</i>	CXX	<i>[Broad Consent, Projekt Consent]</i>
	602	<i>EinwilligungVollständigkeit</i>	<i>Vollständigkeit der Einwilligung</i>	[vollständig, teilweise, abgelehnt]	vollständig
	603	<i>EinwilligungBeschränkung</i>	<i>spezifische Beschränkungen einer Einwilligung</i>	CXX	keine Weitergabe von Proben und/oder Daten ins außereuropäische Ausland
	604	<i>EinwilligungGültigkeitVon</i>	<i>Gültigkeit der Einwilligung von Datum</i>	DD.MM.JJJJ	01.09. 2014
	605	<i>EinwilligungGültigkeitBis</i>	<i>Gültigkeit der Einwilligung bis Datum</i>	DD.MM.JJJJ	01.09. 2014
	211	<i>EinwilligungVersion</i>	<i>ggfs. Version der Einwilligung</i>	FREITEXT	Version2.0