

24. Informationstagung Tumordokumentation und Krebsregistrierung

Qualitativ hochwertige Tumordokumentation - Was können wir voneinander lernen?

Integration von Biobank-Daten in die Tumordokumentation

Anja Franziska Schmidt / Moritz Uthe

CCC Hannover (Claudia von Schilling-Zentrum) - Medizinische Hochschule Hannover (MHH)

Hintergrund / Zielsetzung

Bioproben

- Untersuchung molekularer Marker
- Analyse der Eigenschaften des Tumors
- Schaffung personalisierter Therapiekonzepte

Biobank

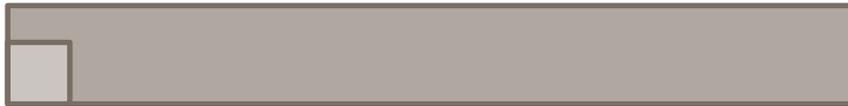
- Sammlung von Bioproben (Gewebe- oder Flüssigproben)
- Bereitstellen der Proben für Forschungsprojekte

Tumordokumentation

- Bereitstellen klinischer Daten für Forschungsprojekte
- Lieferung von Informationen zur Verfügbarkeit von Bioproben
- Verbesserung Service, Nutzen und Attraktivität der Tumordokumentation

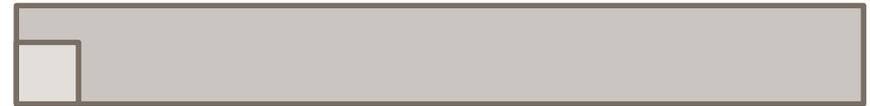
Ausgangslage

Tumordokumentation



- Onkostar
- Diagnosen
- Therapien
- Verläufe

Biobank (HUB) / Pathologie



- Centraxx / PathoPro
- Gewebeproben
- Flüssigproben
- Molekulare Untersuchungen

Die Daten befinden sich abgeschlossen in jedem System für sich!

Lösungskonzept

Entscheidung Datenübernahme

- Positionierung innerhalb Onkostar u.a. durch hohe Flexibilität des Systems
- Keine 1:1 Übernahme, sondern als Verweis gedacht

Vorarbeiten

- Definition Minimaldatensatz
- Formularerstellung in Onkostar

Script-Programmierung

- Programmiersprache Python
- Exportdatei aus Biobank über das Data Warehouse der MHH
- Importdatei für Onkostar

Ablauf Datenabgleich

Export

- Monatliche Datenlieferung (csv-Format)
- Daten aus Biobank und Pathologie
- Enthält u.a. Proben-ID, Probenart, Aufbewahrungsart, Verfügbarkeit

Python

- Einlesen des Exports
- Plausibilitätsprüfungen
- Unterscheidung zwischen bekannten und neuen Bioproben

Import

- SQL-Script zur Aktualisierung bekannter Bioproben
- Importdatei (csv-Format) für neue Bioproben

Aufbau des Scripts

Python

- Version 3.8.5
- Pakete Pandas, Numpy und Datetime

Dataframes

- Import und Umwandlung der Exporte (Biobank und Onkostar) in Panda Dataframes
- Anlegen eines leeren Dataframes mit Struktur der zukünftigen Importdatei für Onkostar

Anpassungen

- Befüllen des Dataframes aus den Daten der Biobank
- Umwandlungen durch Mapping-Tabellen

Abgleich

- Bereits bekannte Proben -> Generierung SQL-Script zur Aktualisierung (Datum, Status)
- Neue Proben -> Erzeugen einer Import-Datei inkl. Plausibilitätsprüfungen (Codes, Datum)

IZahl	ProbenID	ProbenTyp	EntnahmeZeit	ProbenArtCode	PrimaerContainerCode	ExportZeitpunkt	Biobank	PathoProbenart	Histologie	Lokalisation
		LIQUID	10.05.2021	PL1	PED	27.05.2021	Biobank MHH			
			13.08.2021			20.08.2021	Pathologie MHH	PE Kolon	C18.9	8140/3



```
In [2]: import pandas as pd
import numpy as np
import datetime
from datetime import datetime

In [5]: # Festlegen der Spaltennamen des Dataframes
columns = ['patientenId', 'tumorid', 'dokumentierendeeinrichtung', 'dokumentierendefachabteilung',
'BiobankProbendatum', 'BiobankStand', 'BiobankID', 'BiobankProbenID', 'BiobankLokalisation',
'BiobankICD03', 'BiobankSampleType', 'BiobankProbenTyp', 'BiobankFixationType',
'BiobankProbenStatus']

In [25]: # Kombinieren der Dataframes von Onkostar und der Biobank über die Proben ID
merged_bio = pd.merge(df_bio_onk, df_bio_exp, on=['BiobankProbenID'], how='inner')

In [29]: # Prüfung, ob die Biobank Export IDs im Onkostar Export vorhanden sind - Es werden alle Proben gelistet,
# die nicht in Onkostar vorkommen
df_bio_unbekannt = df_bio_exp[~df_bio_exp.BiobankProbenID.isin(df_bio_onk.BiobankProbenID)]
df_bio_unbekannt.BiobankProbenstatus = 'V'

In [55]: # Generierung des SQL-Scripts für die bekannten Biobank-Daten
for i in range(len(df_bio_sql.BiobankStand)):

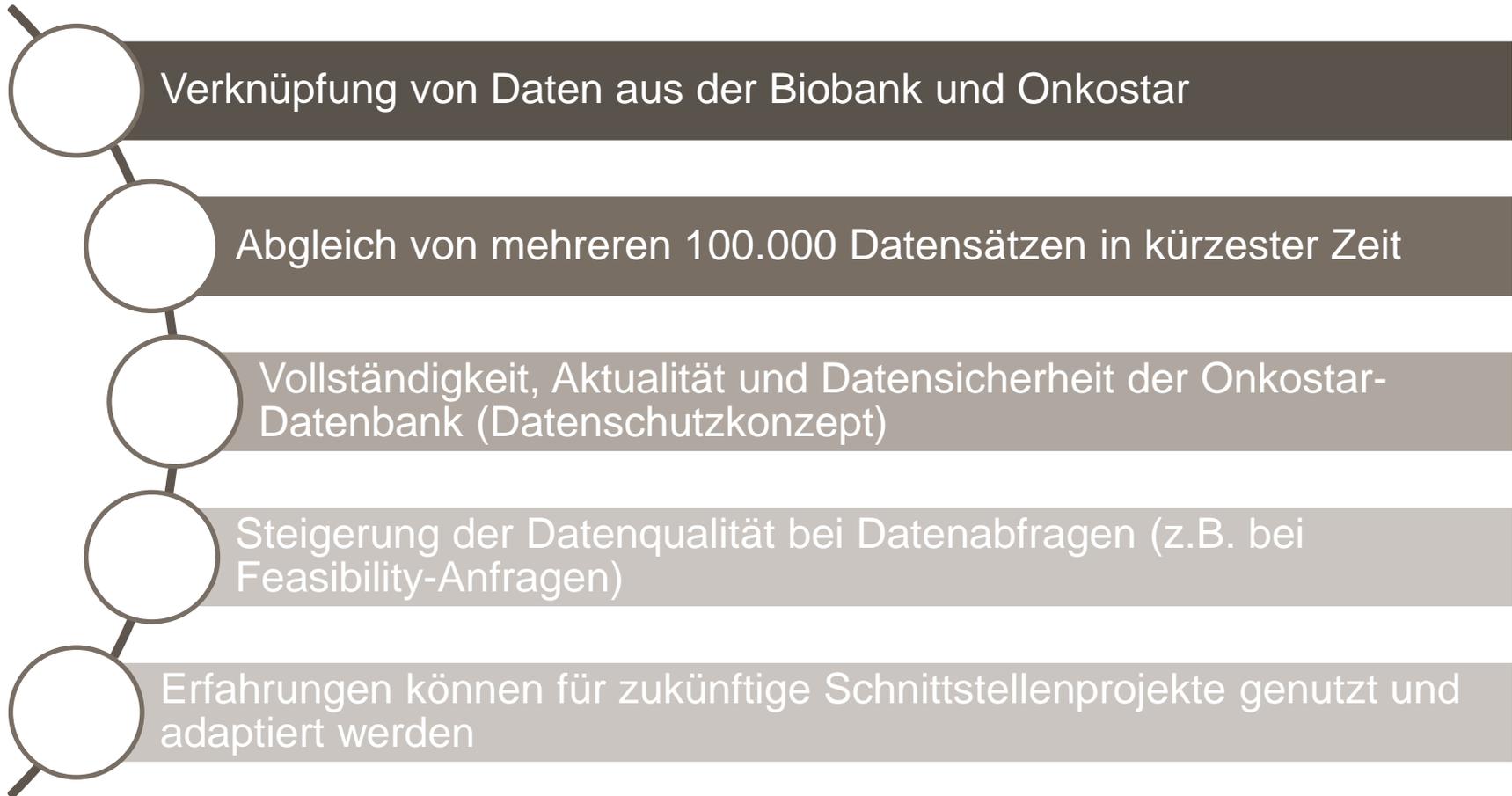
# CSV Exporte
df_comb_unbekannt4.to_csv("Output/Import_Daten_Biobank_Proben_"+ str(Biobank_Stand[0:10]) + ".csv",
AM + "" + ", biobank_proben:"
```



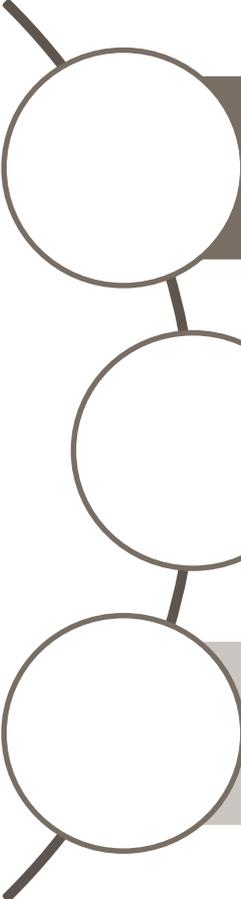
Biobank Probe	
Entnahmedatum und -Uhrzeit:	10.05.2021 11:19
Stand der Information:	29.08.2021 00:00
Biobank:	Biobank MHH
Proben-ID:	
Probentyp [SPREC]:	Plasma, einmal zentrifugiert [PL1]
Probentyp:	Vollblut
Probenfixierung (flüssig) [SPREC]:	Kalium EDTA (PED)
Probenstatus:	Probe vorhanden

Biobank Probe	
Entnahmedatum und -Uhrzeit:	13.08.2021 00:00
Stand der Information:	29.08.2021 00:00
Biobank:	Pathologie MHH
Proben-ID:	
Lokalisation:	PE Kolon
ICD-O-3:	C18.9
Probentyp:	Unbekannt
Probenstatus:	Probe vorhanden

Ergebnisse / Vorteile



Limitationen / Ausblick



MHH-spezifische Lösung

Script ist nur bedingt auf andere Einrichtungen übertragbar. Prozesslösung allerdings für vergleichbare Register relevant

Manuelles Auslösen des Scripts nötig. Weiterentwicklung einer vollautomatisierten Ausführung wird angestrebt

Kontakt Daten

Anja Franziska Schmidt
Fachliche Bereichsleiterin Tumordokumentation
CCC Hannover (Claudia von Schilling-Zentrum)
Medizinische Hochschule Hannover (MHH)
Schmidt.AnjaFranziska@mh-hannover.de

Moritz Uthe
IT Koordination
CCC Hannover (Claudia von Schilling-Zentrum)
Medizinische Hochschule Hannover (MHH)
Uthe.Moritz@mh-hannover.de

Vielen Dank für Ihre Aufmerksamkeit!