



UNIKLINIK
KÖLN



HiGmed
Medical Informatics



Perspective of the Medical Informatics Initiative: today and future of Data Reuse with MeDIC

Prof. Dr. Andreas Beyer, University Hospital Cologne, CECAD Cologne & Prof. Dr. Oya Beyan, University Hospital Cologne, Director of the Institute for Biomedical Informatics at Köln (BI-K)

14. TMF-Jahreskongress, Köln, 18.04.2023

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Mitarbeiter des MeDIC Köln

MeDIC Leitung



Prof. Dr. Andreas Beyer
HiGHmed Standortsprecher



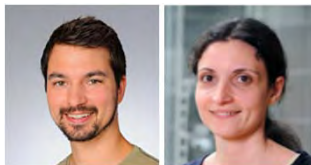
Prof. Dr. Oya Deniz Beyan
Professor for Med. Informatik



Rene Tielsch-Nebel
Operative MeDIC - Leitung

Projektkoordination

HiGHmed Use Cases

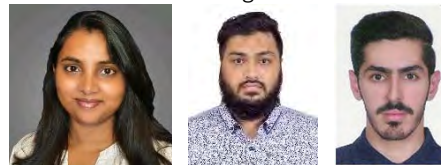


Dr. Simon
Schumacher

Dr. Ana
Grönke

Medizininformatiker*innen

Infrastruktur Daten-
management Entwickler



Usha
Pandit

Md. Mostafa
Kamal

Mehrshad
Jaberansary

Data Steward

medizinische
Daten klinische
Koordination



Dr. med. Samer
Alkaroukly

Dr. med. Annic
Weyersberg

IT-Administration

MeDIC UKK-IT



Muhammad
Adnan

Anna
Hannig



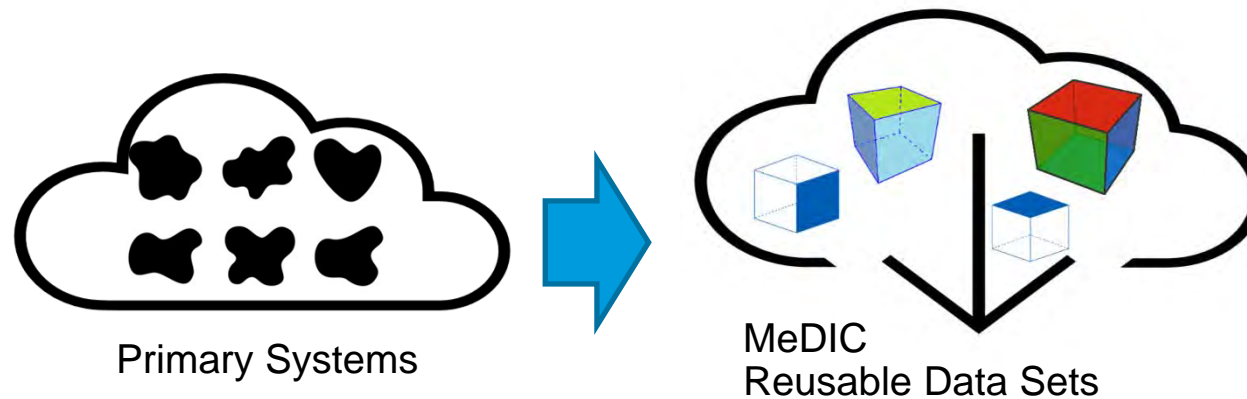
1. Was wir vorhaben
2. Wo wir stehen
3. Was als nächstes kommt



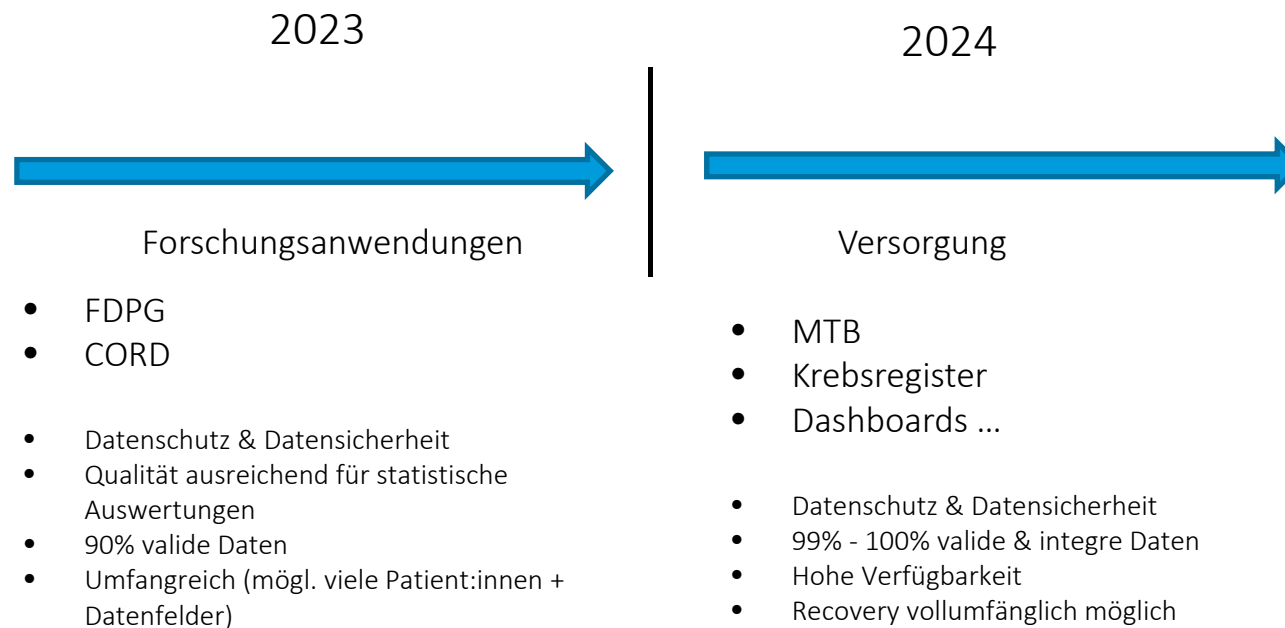
Was wir vorhaben

Cologne Medical Data Integration Center (MeDIC)

- Extracts and Integrates real patient data from operational systems
- Ensures the privacy and consent of the patient
- Follows domain standards for data exchange
- Shares data with approved projects
- Executes distributed analytics algorithms on data and shares results



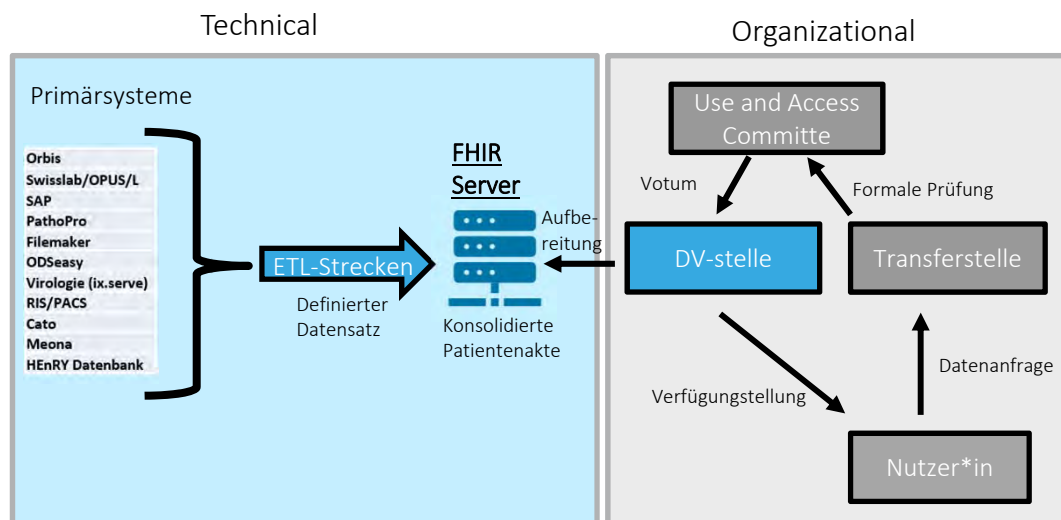
Perspektive MII | Förderphasen, Zeitplan, Status, Plan zur schrittweisen Produktivsetzung im MeDIC Köln



Data Sharing and Reuse

Supports reuse of patient data for research

Both for UKK researchers and other research networks



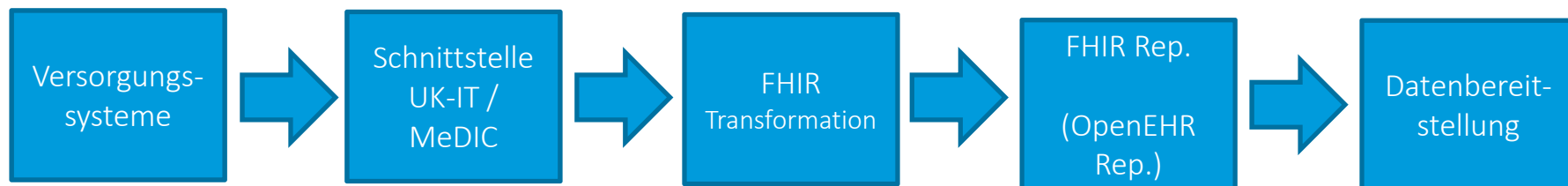


Wo wir stehen

ETL Strecken

- §21 Daten ab 2013 – FALL*, FAB, ICD, OPS (CSV)
- HL7 ADT*
- Labordaten OPUS/L (ORU)
- ORBIS (CSV)
- MEONA (XML)
- VIROLOGIE (CSV)

*erzeugen von Patient Ressourcen



Überblick Netzwerkprojekte und Usecases am Standort Köln

- MII, - SMITH, HiGHmed, Miracum, DIFUTURE
 - NUM-DIZ (Netzwerk Universitätsmedizin Datenintegrationszentren)
 - NUM-RDP (Routinedatenplattform)
- MeDIC Cologne HiGHmed – Usecases within MII Network
 - FDPG (ForschungsDatenPlattform Gesundheit)
 - SU-Term-Serv (Terminology Service)
 - FOrMe (Nephrology Dashboard)
 - CORD (Collaboration on Rare Diseases)
 - SnaP (Suche nach ähnlichen Patienten)
 - PM4ONCO (Oncology)
 - RISK Principe (Infection Control)
 - PEDREF (pädiatrische Labordatenanalyse)
 - PrivateAIM (Privacy-preserving AI for Future Medicine)
 - HIV CARE
 - Cover Child (COVID-19 Forschungsplattform für Kinder und Jugendliche)

Completed Usecases:

- NUM-RDP/Codex, Codex+
- B-FAST/SmICS (Smart Infection Control)
- NT-proBNP (Cardiology for SMITH)

CIO^C

Networks and Usecases

- NCT, CCCE (Nationales Centrum für Tumorerkrankungen)
- CIO^{ABCD} (Centrum Integrated Oncology Aachen, Bonn, Köln Düsseldorf)
- nNGM, DigiNet (nat. Netzwerk Genomische Medizin)
- DNPM (ZPM) (Deutsches Netzwerk Personalisierte Medizin)
- DKTK / C4 (Deutsches Konsortium für Translationale Krebsforschung)
- DK-FBREK (Konsortium (familiärer Brust- und Eierstockkrebs))
- genomDE
- GBA (German Biobank Alliance)
- UC: OncoConnect
- UC: radCIO (Radiologie und CIO Köln)

Completed:

- UC: iCUP (Cancer Unknown Primary)

UKK

Network

- Westdeutscher Radiologieverbund, TKmed

Nationale Forschungsdateninfrastruktur

- NFDI4Health
- NFDI4DataScience and AI
- GAIA-X (FAIR Data Spaces)
- FAIR4Rare (Evaluation Aufbauprozesses offenes Nationales Register)



Erreichte Ziele

Technische Erfolge

- Laden der Daten in das zentrale FHIR Repository
- MIRTH-Server, Mapping Engine - FHIR Transformationen
- MeDIC Webseite
- Orbis Schattenkopie
- Clinical Trialsform Consent Update
- Übermittlung Testdaten NUM-RDP / Codex via Treuhandstelle und Transferstelle
- Anschluss an FDPG mit DSF
- Aggregation der Daten für CORD
- Lieferung Forschungsdaten

Organisatorische Erfolge

- UAC mit Freigaben seit 18.1.2022
- Teilnahmeverträge NUM RDP, FDPG
- Sammlung Consents (BC-BioMaSota, CIO Consents)
- Terminologieserver für MII und NUM, Weiterführung mit SUTermserv2b
- Orbis/SAP Testpatienten im SAP Produktivsystem
- ISB&DSB Freigaben von MeDIC V2, NUMV1, NUMV2, VITU, FDPG, CORD, ARX Anonymisation





Was als nächstes kommt

ETL-Routen - Testen

Overview on Testing Concept:

Basic Tests for ETL Routes

- Populating error containing test data
- Populating data in the wrong format to the file and/or Pollaroid watch folder
- Manual content validation of imported data (querying the VONK)
- Integration Test on VONK (e.g. adding data to pre-existing patient record, Duplication Test)

Functionality Tests for ETL Routes

- Scalability Test
- Regression Test (e.g. after each new deployment to determine if changes have regressed other attributes of data and reports)
- Quality test (e.g. number check, date / time check, precision check for the values, empty values,...)
- Incremental ETL Testing (e.g. verifying that old data and updates (new data) are merged properly)

Performance Test for ETL Routes

- firely Server Performance
- Mapping Engine Performance

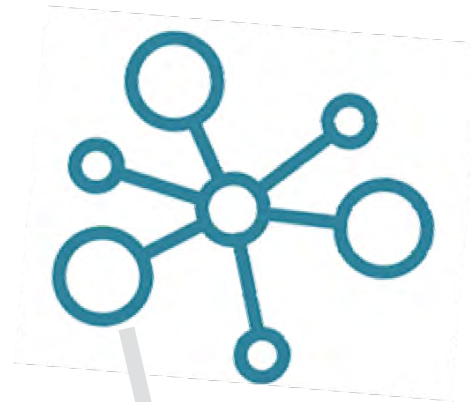
Rollout des MI-I Broad Consent

- Köln verwendet eigenen Broad Consent (BioMaSOTA) seit Jahren
 - Biobank, nur einige Kliniken
 - wichtige Erfahrungen für Prozesse
- Jetzt: MI-I Broad Consent
 - Ambulante und stationäre Patient:innen
 - Zentrale Aufnahme und in den Kliniken
 - Pilotstudie zweite Hälfte 2023



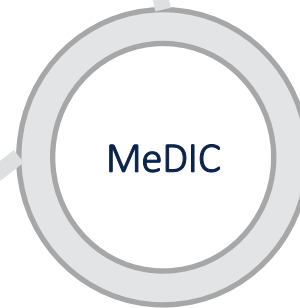
Moving Forward: What is next for UKK- MeDIC

Identifying Goals and Key Partners



More data reuse:
Developing
Partnership with
Data
Infrastructures

Strengthening research
capacity at home:
Clinicians and Data
Science Teams



Fostering innovation:
Precise AI and
Machine Learning

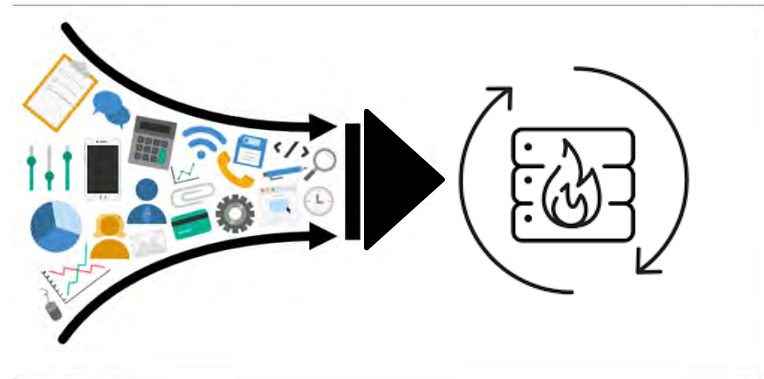
Moving Forward: Strengthening research capacity at home



Strengthening research capacity at home: Supporting our clinicians at UKK

- Be their DATA partners:

- Timely access to data: %70 of a data science project is to prepare data!
- Optimize the ethical & legal procedures:
special data access rights for their patients data
- Easy interfaces : to query and access their data
- Support for data enrichment:
RedCAP integration to FHIR



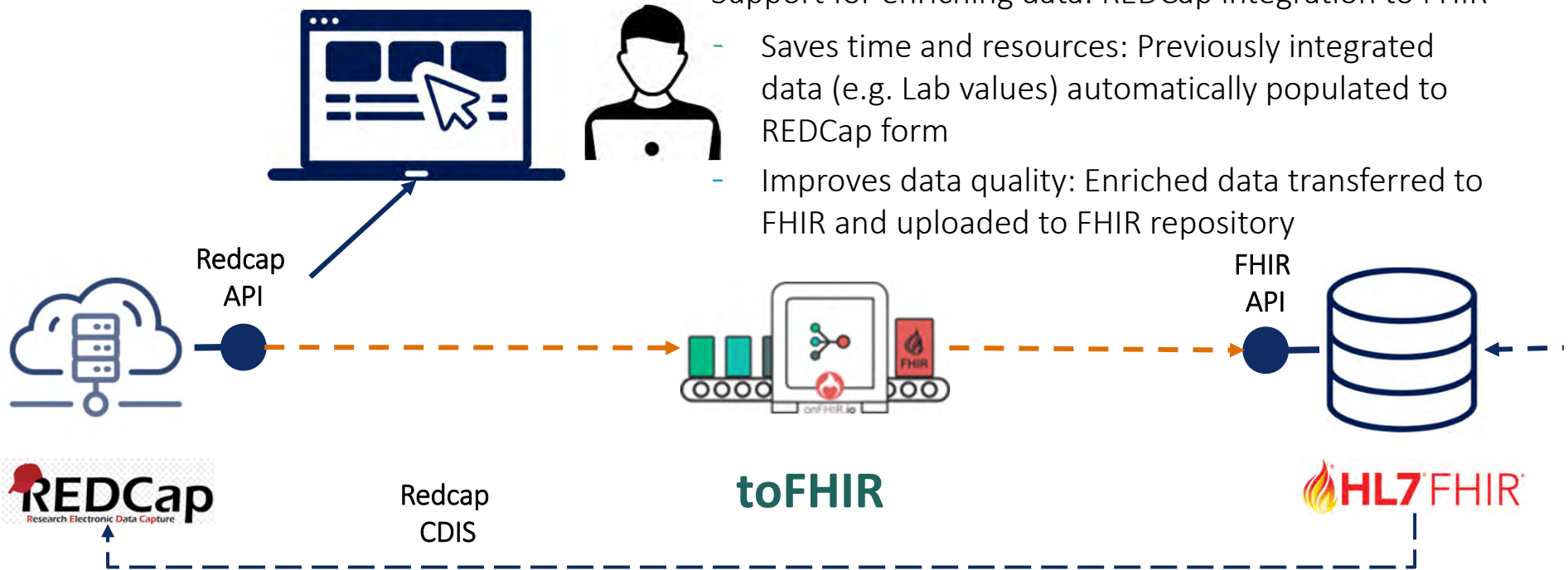
Integrated, unified, ready to use data

Moving Forward: Strengthening research capacity at home



Support for enriching data: REDCap integration to FHIR

- Saves time and resources: Previously integrated data (e.g. Lab values) automatically populated to REDCap form
- Improves data quality: Enriched data transferred to FHIR and uploaded to FHIR repository



Moving Forward: Strengthening research capacity at home



- Support for data analytics
 - Provide data analytics expertise together with Biomedical Informatics, Bioinformatics, medical statistics...
 - Visual and analytic tools analyse their data (dashboards, e.g. Smics)

Work together in research

- CIO: Unknow primary tumors CUP
- Radiology: Lung cancer
- Nephrology: Chronic kidney diseases- ADPKD, CKDu
- Pathology: Image analysis...



Moving Forward: An Active Partner of Data Infrastructures



Share It Generously

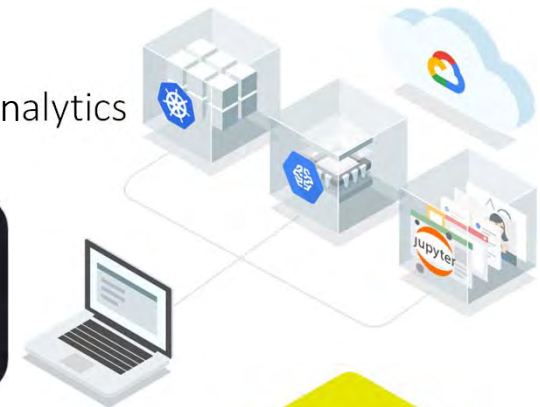
- Get experience in new data types > Image, omics..
- Integrate multi modal data > complete data sets

MII- new projects

- **OMI** : enable the discovery and use of remote AI services for Images
- **PM4Onco**: integrating and sharing data from clinical and biomedical research to improve personalized medicine in oncology

Guard It Jealously

- Secure Sharing
- PADME- Federated Analytics



- MII- new projects

PrivateAIM : Privacy-Preserving Analytics in Medicine



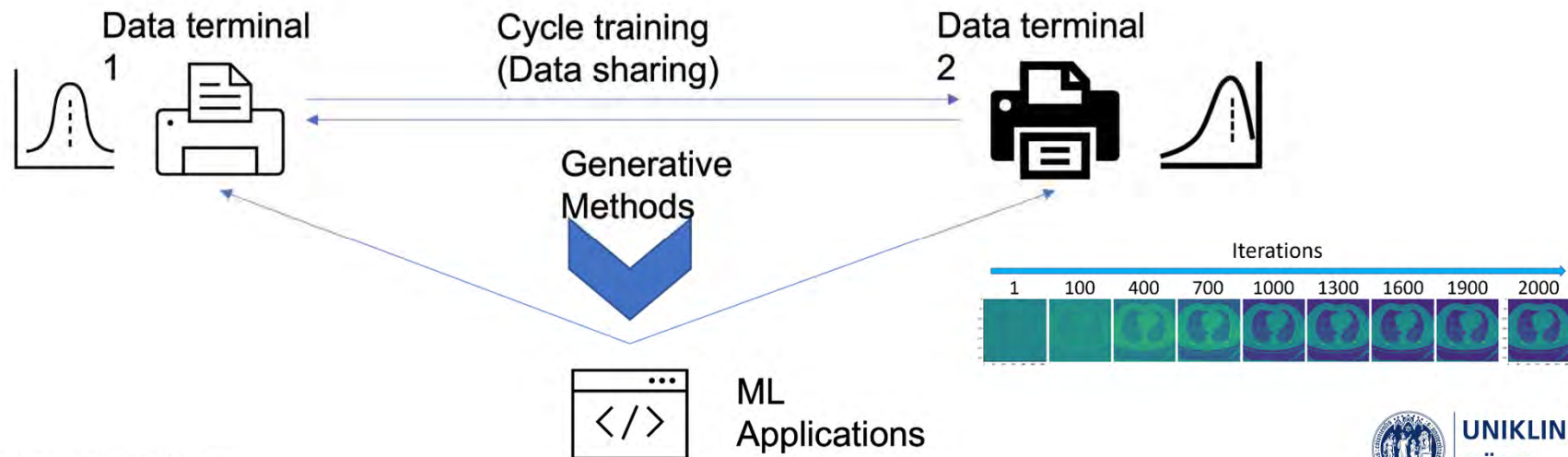


Moving Forward: An Active Partner of Data Infrastructures



A demonstrator of Federated Analytics with PADME:

- together with Radiology department: trained the generative model CycleGAN for style transfer and synthetic data generation to mitigate the distribution shifts in Lung Cancer CT data



Moving Forward: An Active Partner of Data Infrastructures



- ✓ AI robotic assistants in the ICU & ORs to cope with the shortage of medical staff in hospitals
- ✓ Develop interfaces for devices and integrate them
- ✓ Introduce an AI application using a collaborative robot (cobot) to conduct tasks originally performed by medical staff



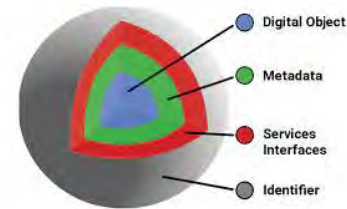
Photo Source: www.redeng.us



Moving Forward: An Active Partner of Data Infrastructures

And integrate data for more ...

- ✓ Forschungsdatenportal für Gesundheit (FDPG)
- ✓ European Health Data Space (EHDS)
- ✓ GAIA-X....



FAIR Data for

- Research
- Industry
- Clinical Trials
- Data Science
- App developers
- Public Health
- Smart cities
- ...

Source systems

- In silos
- In many formats
- Inconsistent vocabularies
- With access restrictions
-



Moving Forward: Precise AI and Machine Learning



Fostering innovation: Precise AI and Machine Learning

Bias in AI/ML : The main blocker to transfer AI/ML into practice

- Reproducibility / repeatability of AI results
- Non interpretable evidence leads to Opacity
- Accidentally fit confounders rather than true signals
- Systematically and unfairly discrimination of certain individuals/cases
- Data is not static artefact independent from the process that produced it
- Bias can be anywhere in the ML/AI pipeline



“We should really stop including drunk teenagers in the training data.”

Moving Forward: Precise AI and Machine Learning



Fostering innovation: Precise AI and Machine Learning

Bias in AI/ML : The main blocker to transfer AI/ML into practice

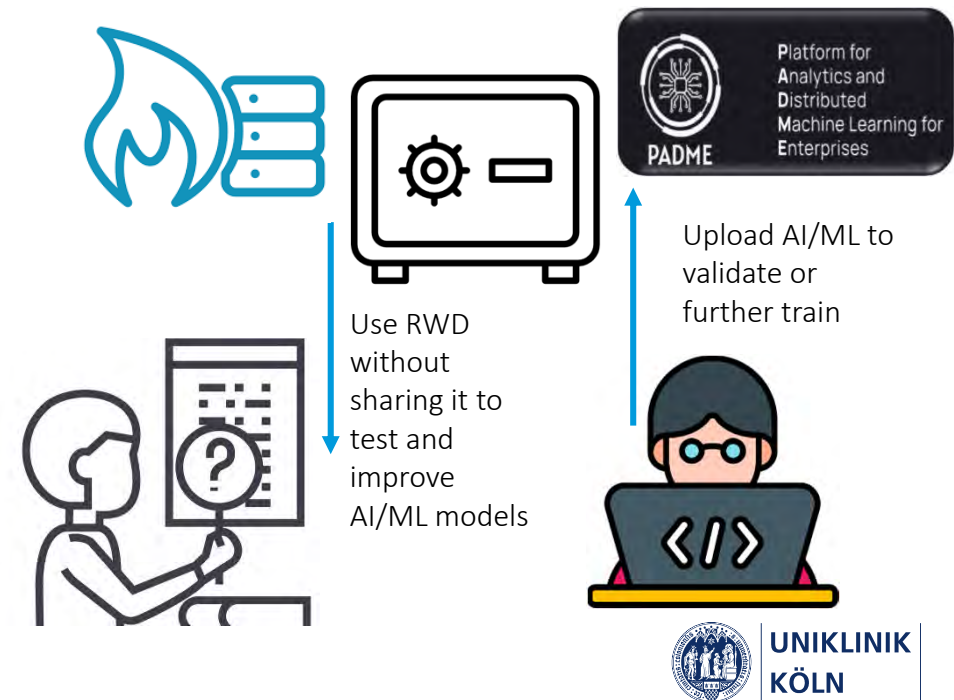
How to integrate AI/ML safely to the clinical practice?

More data to train + Trusted parties to validate

- Modelling and validating with more data
- Keeping the data provenance and training data

Federated Analytics:

- AI safe box to validate models,
- Identify algorithmic bias caused by the non-identical distribution of the data



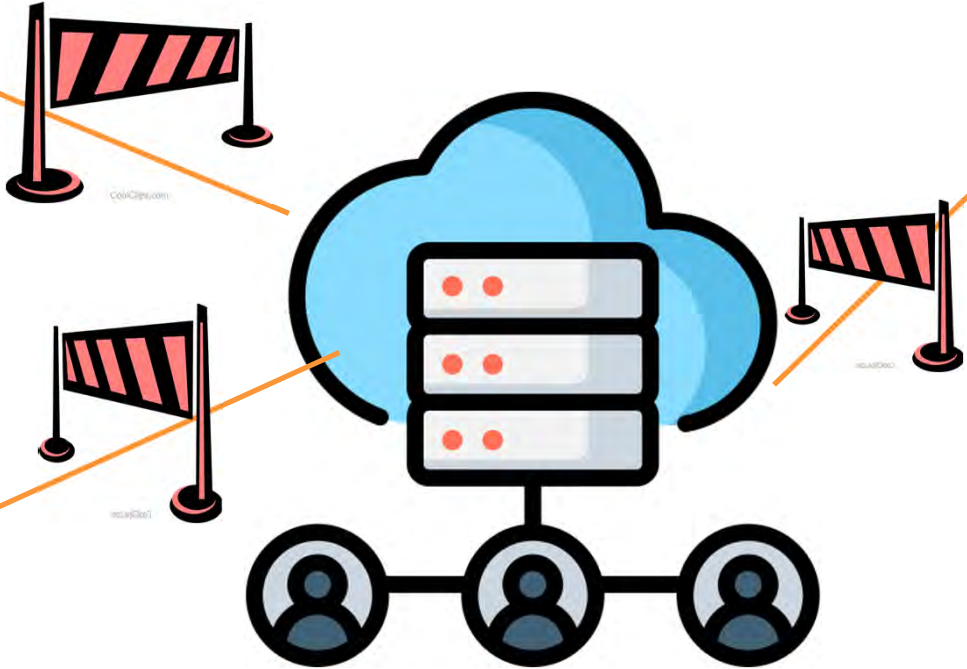


**What Prevents us
from reusing real-
world data in
research &
innovation ?**

What Prevents us from reusing real-world data in research & innovation

Insecurities, Uncertainties:
balancing benefits and harms of data reuse

Ineffective Planning :
mismatch between traditional study planning and RWD projects



No one size fits all:
everyone needs a bit of a different data

What Prevents us from reusing RWD ? : Insecurities, Uncertainties

Tension between to protect and get benefit: balancing Benefits and Harms

Sensitive data
Compromised privacies
Stigmatization of patients
Security nightmares
Data leaks
Information disclose
Legal compliance
Ethical concerns
....



Multi modal data analysis
Machine Learning / AI
Improved

- Clinical efficiency
- Patient safety
- Performant clinical decision support systems
- Healthcare sustainability...

with lower time and monetary costs.

Affected role

Patient

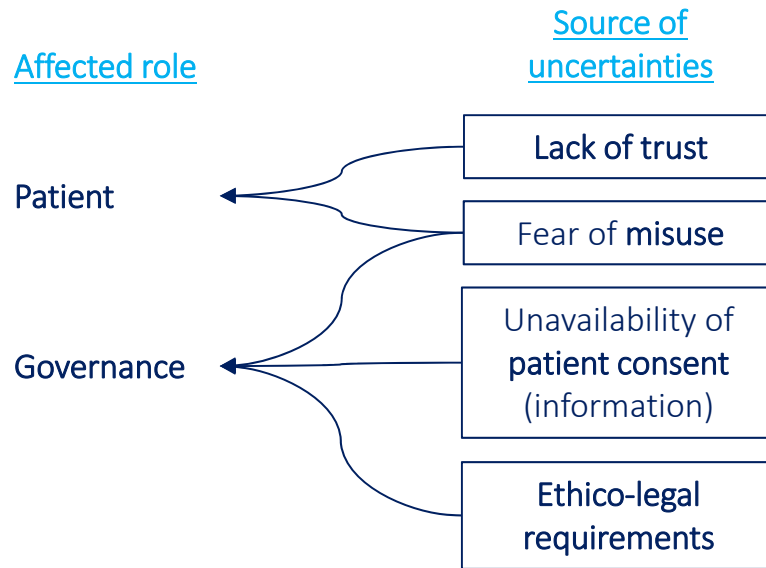
Source of uncertainties

Lack of trust

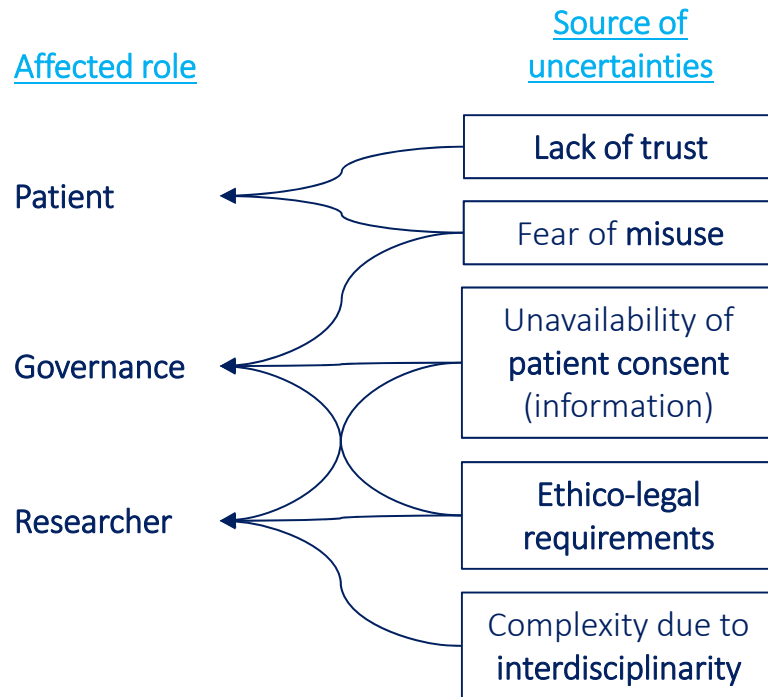
Fear of misuse

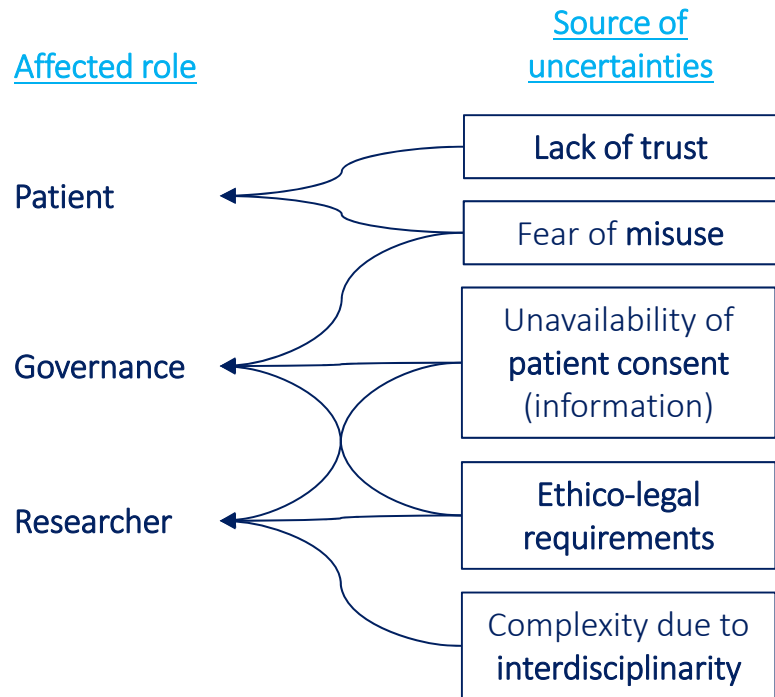


How long my data will be stored?
Who will use it ? For what purpose?
What if there is a data leakage?
I want to support common good , but...
I do not want to be transparent



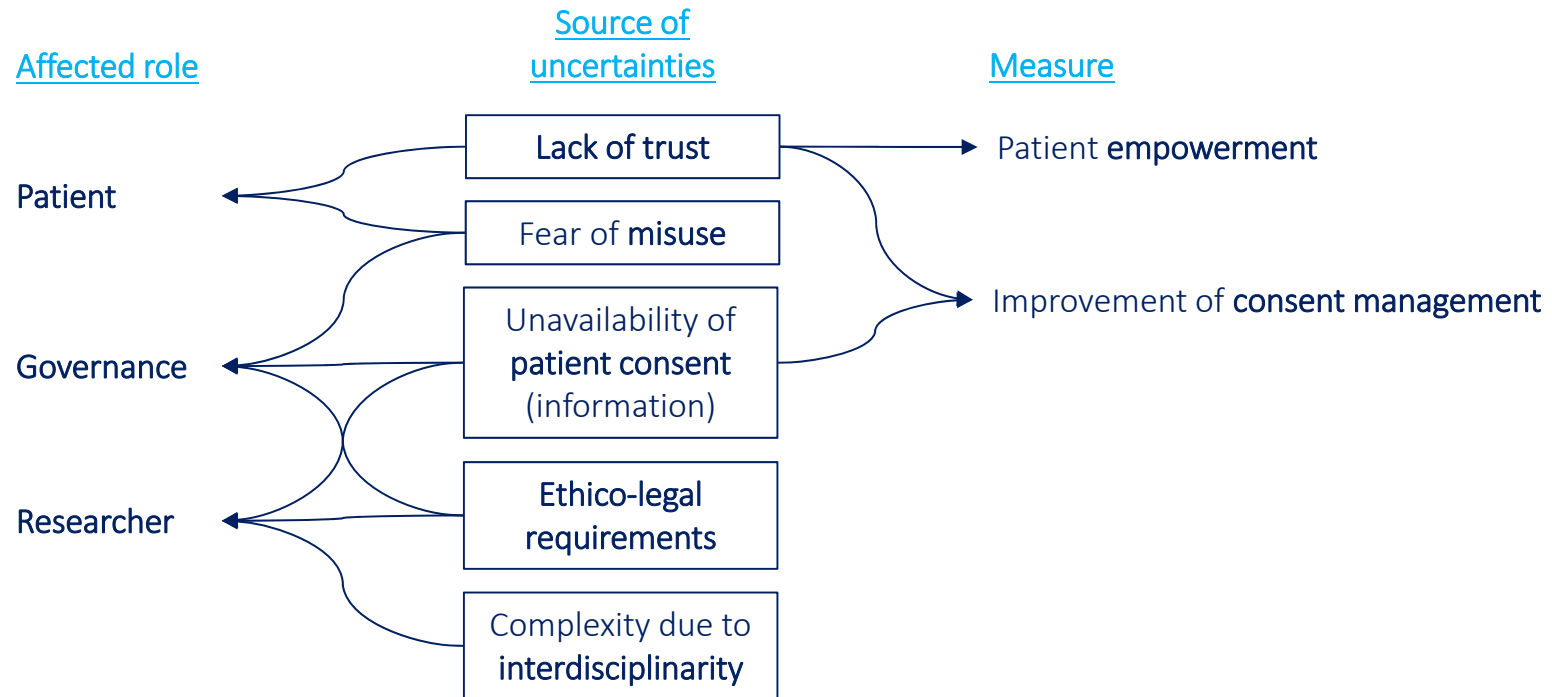
Complexity of ethical and legal requirements, variety of intervening laws, makes it difficult to identify correct legal basis
 Principle of data minimization, limitations of consent ...
 The project idea sounds good but,...I am not certain what can be allowed





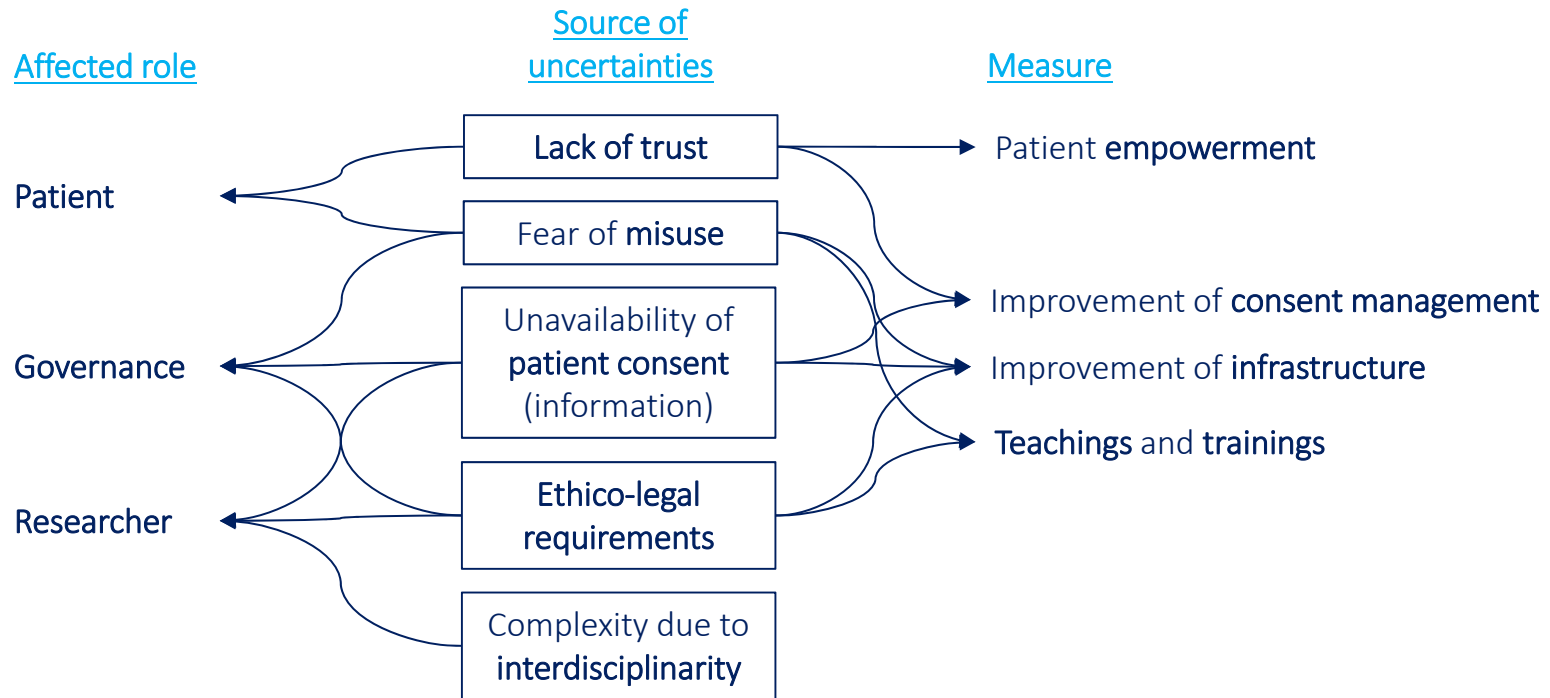
Consequences of uncertainties





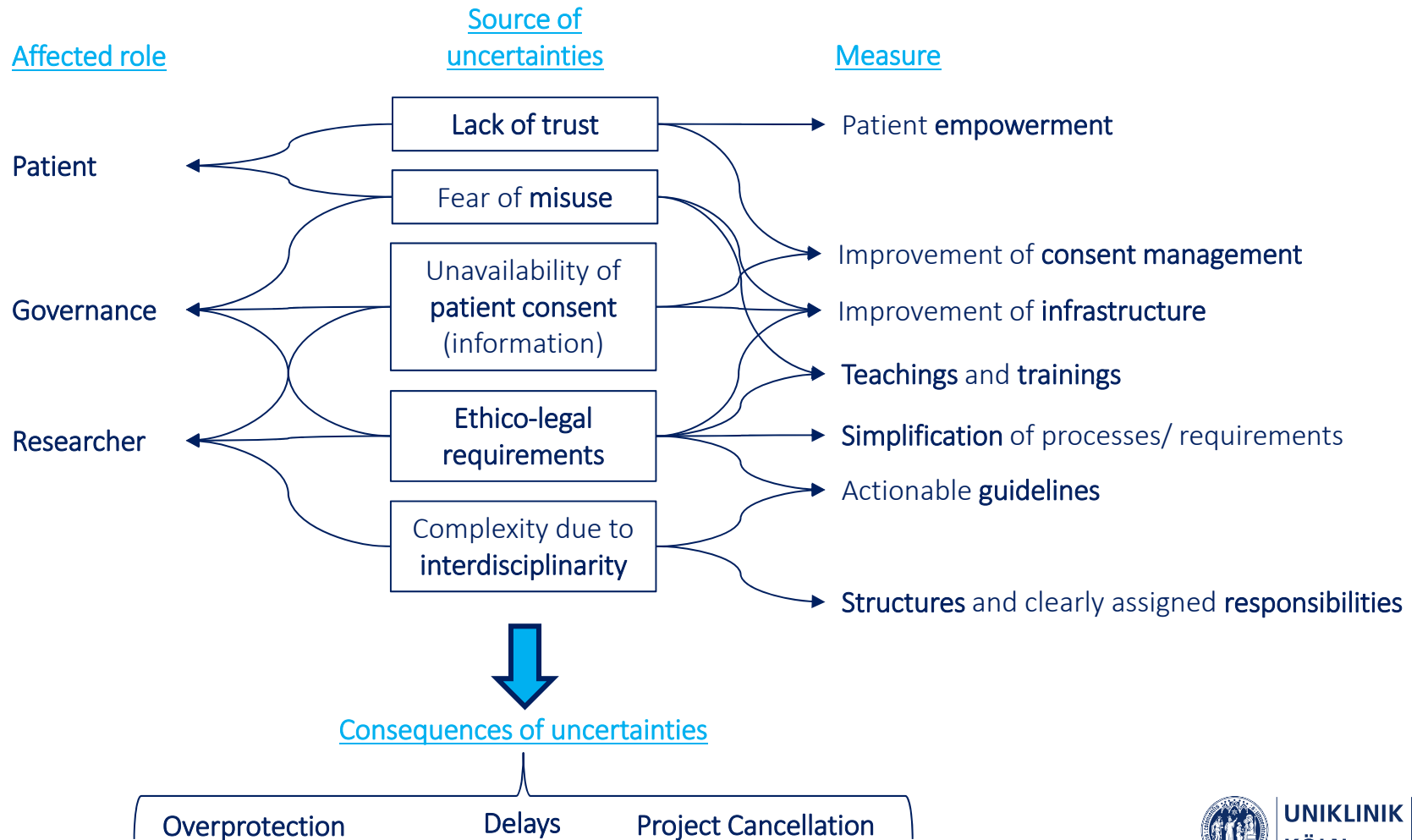
Consequences of uncertainties

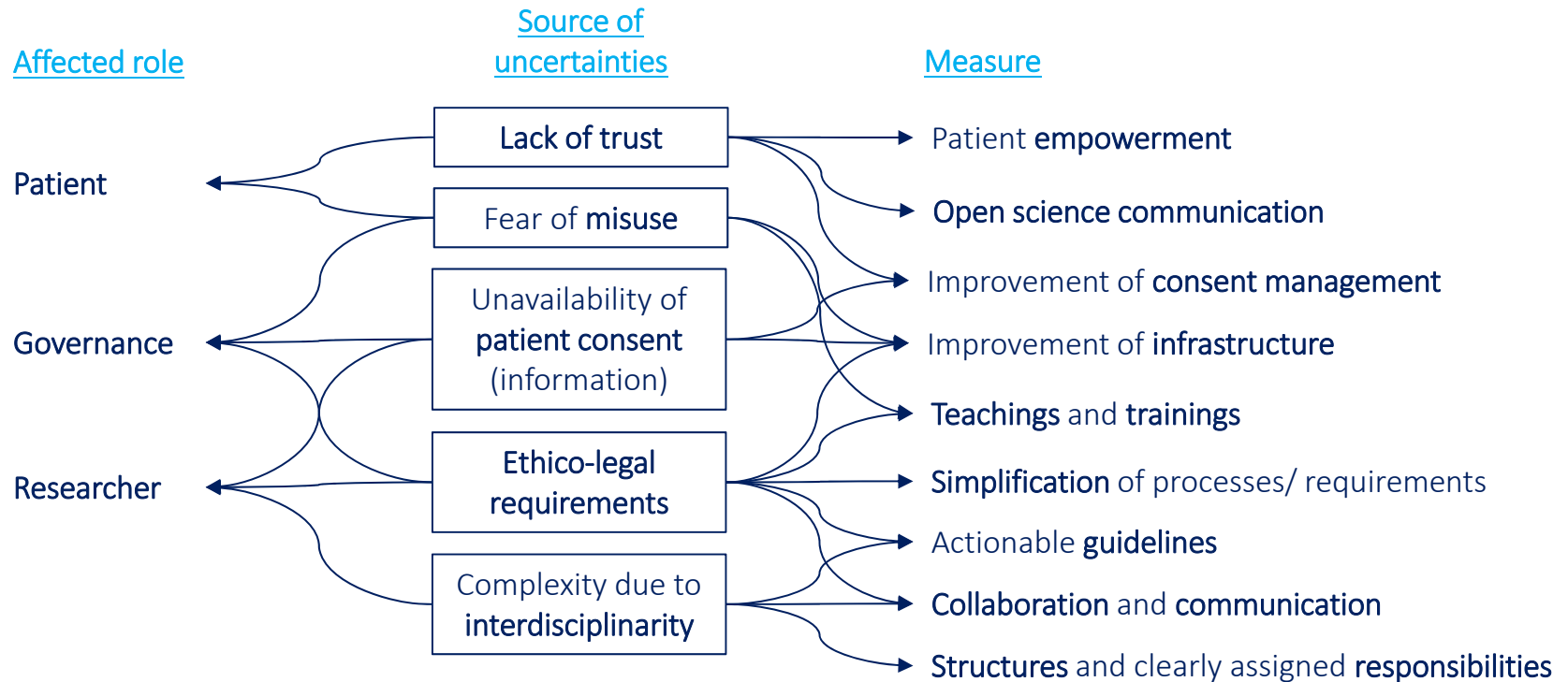




Consequences of uncertainties







Consequences of uncertainties



Ref: Gehrman, J., Herczog, E., Decker, D., and Beyan, O., 2023. What prevents us from reusing medical real-world-data in research. Scientific Data (Accepted).



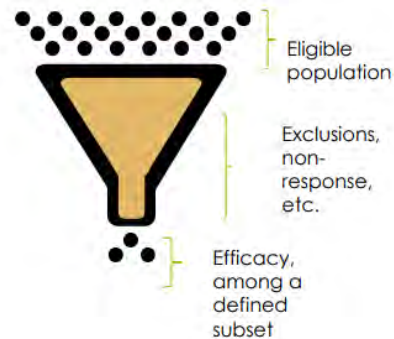
**UNIKLINIK
KÖLN**

What Prevents us from reusing RWD ? : Ineffective Planning

RWD studies are different than traditional clinical studies

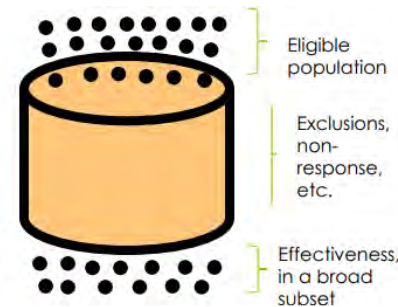
Clinical Trials:

- (-) Expensive: Selection bias
- (-) Prospective: limited time frame
- (-) Single-arm interventional trial has limitations
- (+) Designed to control variability
- (+) High data quality

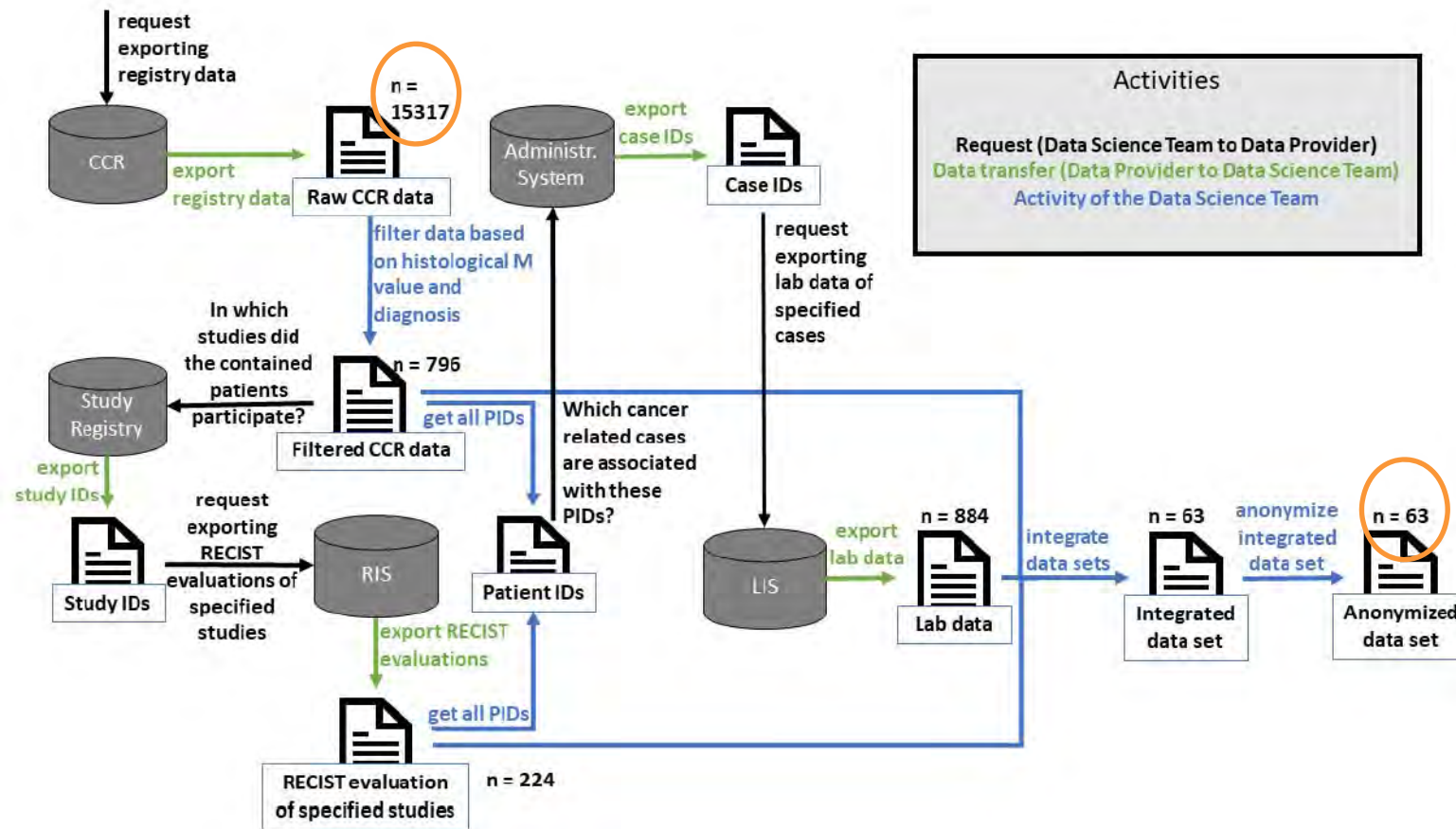


RWD:

- (+) Longitudinal patient-level EHR data
- (+) Large data sets from retrospective response
- (+) High dimensionality
- (-) Reliability: Low data quality
- (-) Sparse: high percentage of missing data.



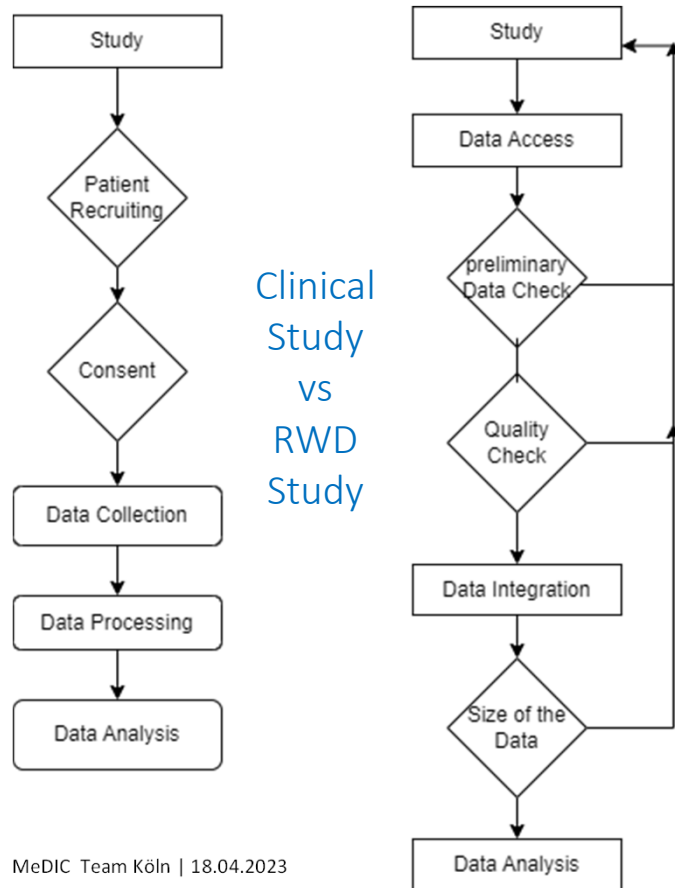
What Prevents us from reusing RWD ? : Ineffective Planning



Data Flow for the data access and integration processes in the CUP use case of CIO Cologne.

Gehrmann, Julia. (2022). Designing a data access and integration workflow for medical data science: a use case of compiling a reusable data set for primary tumor discovery at MeDIC Cologne. <https://doi.org/10.18154/RWTH-2023-03583>

What Prevents us from reusing RWD ? : Ineffective Planning



Mismatch between traditional study planning and RWD data science project:

- Aim of the AI/ML methods is not to test hypotheses but model the data: large data sets
 - Data is collected from wild, not generated in lab: No control on completeness, quality...
 - Can not assess the availability and usefulness of the data until the later stages
 - Study design and research questions might be tuned
- Updated Data Collection Plan - > new approval processes
-> leads to huge delay

New approach for RWD project planning and approval is needed with adaptable (flexible) approval processes

What Prevents us from reusing RWD ? : No one size fits all

Extract: Retrieves and verifies data from various sources

- Variety of the source systems
- Access difficulties, limited interfaces
- Hospital IT systems are living organisms: new systems, updates...

Transform: Processes and organizes data so it is usable

Challenges of

- mapping of terminologies
- complex data formats: letters, genomic...
- data integration, consent checking, anonymization..

Load: Moves transformed data to a sharable repository

- Each data request is different
- Additional requirements
- Variety of requested formats, limitation of existing profiles
- Reusability restrictions



What Prevents us from reusing RWD ? : No one size fits all

Data Integration Centers are not one-time development projects, but **Living Data Infrastructures!**

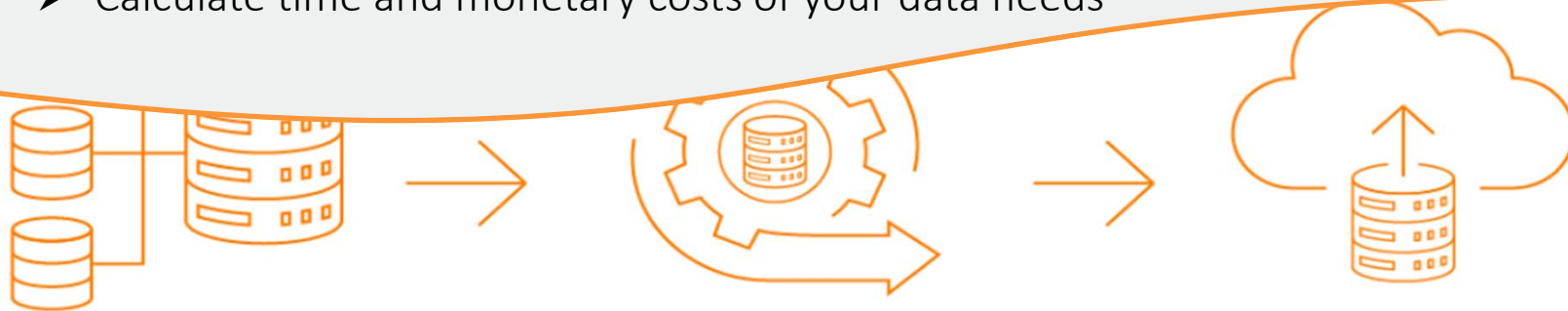
- Continuous development and testing with new ETLs are required

Each data request is a new project: always something additional to do

New data element, additional approvals, new standard to share

Involve MeDIC from project preparation stage!

- Coordinate the availability of data
- Calculate time and monetary costs of your data needs



Data Integration Centers are the Key for FAIR & Medical Data Science

Findable:

Hospital IT systems are fragmented.
Data can be findable only with intense communication with clinics or with previous experience of data stewards with clinical routine.

Interoperable:

Complex nature of patient data requires experience and investment in patient services: patient mapping, pseudo-anonymization, consent check, anonymization...

Terminology mapping is an art! Requires experts.

Accessible:

Long term storage and accessibility of data and metadata should be ensure: not in someones PC.

Easy accessibility processes and interfaces (fulfilling ethico-legal conditions) are need.

Reusable:

True data reuse beyond specific project can be only achieved with anonymized data or broad consent. MeDICs (1) publish machine readable consent of data, and (2) set up processes and tools for anonymization.



Data Integration Centers are the Key for FAIR & Medical Data Science

Findable:

Hospital IT systems are fragmented.
Data can be findable only with
communication with clinical
experience of data steward
routine.

Interoperable:

Complex nature of patient data requires experience
at services: patient
information, consent check,
! Requires experts.

Accessible:

Long term storage and
metadata should be easily
PC.

Easy accessibility processes and interfaces
(fulfilling ethico-legal conditions) are need.

Specific project can be only
data or broad consent.
the readable consent of
data, and (2) set up processes and tools for
anonymization.

MII and Data Integration Centers
leads the cultural and technical
transformation in RWD reuse for
research & innovation

<https://www.uk-koeln.de/forschung/medical-data-integration-center-medic/>

Kontakt

Für allgemeine Fragen bezüglich des MeDICs wenden Sie sich bitte an

E-Mail medic-info@uni-koeln.de

Bei Fragen bezüglich technischer Themen wenden Sie sich bitte an

E-Mail support-medic@uni-koeln.de

Für Fragen bezüglich Nutzungsanträge wenden Sie sich bitte an die Transferstelle des MeDICs Köln

E-Mail transferstelle-medic@uni-koeln.de



UNIKLINIK
KÖLN