

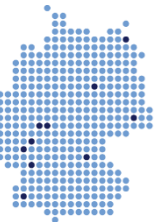
## Use Case 2: Gemeinsame statistische Analysen für verteilte Daten

**Daniela Zöller**

**15.07.2021**

**MIRACUM Symposium 2021**

# Recap: Ziele in Use Case 2

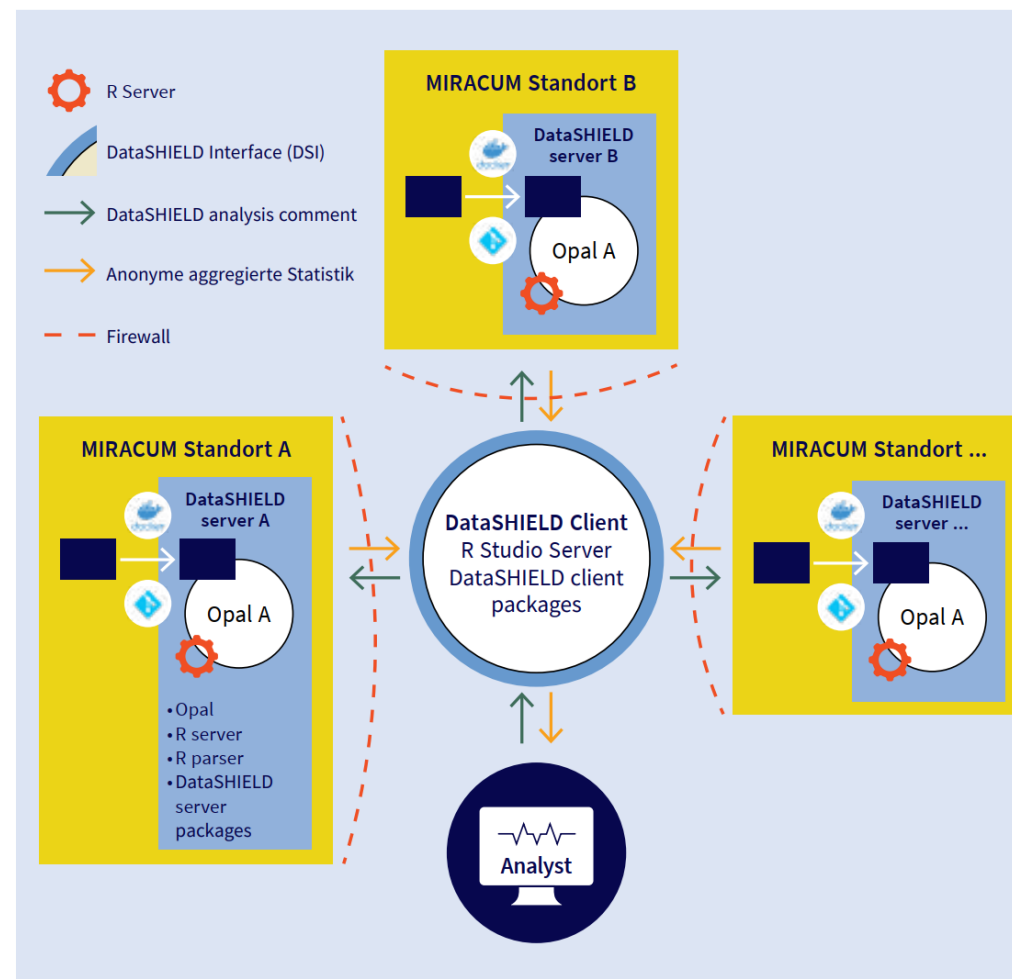
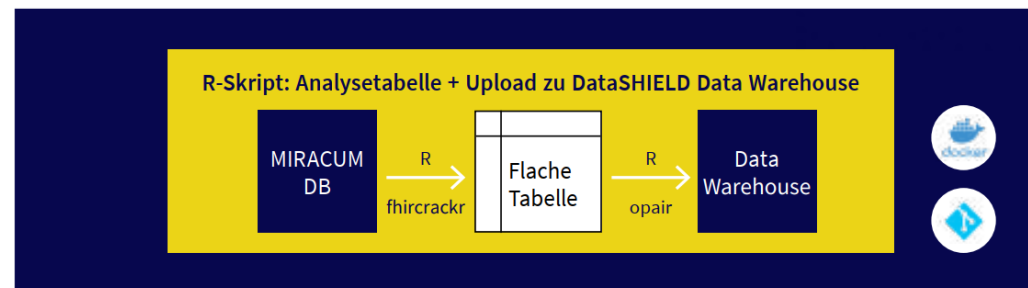
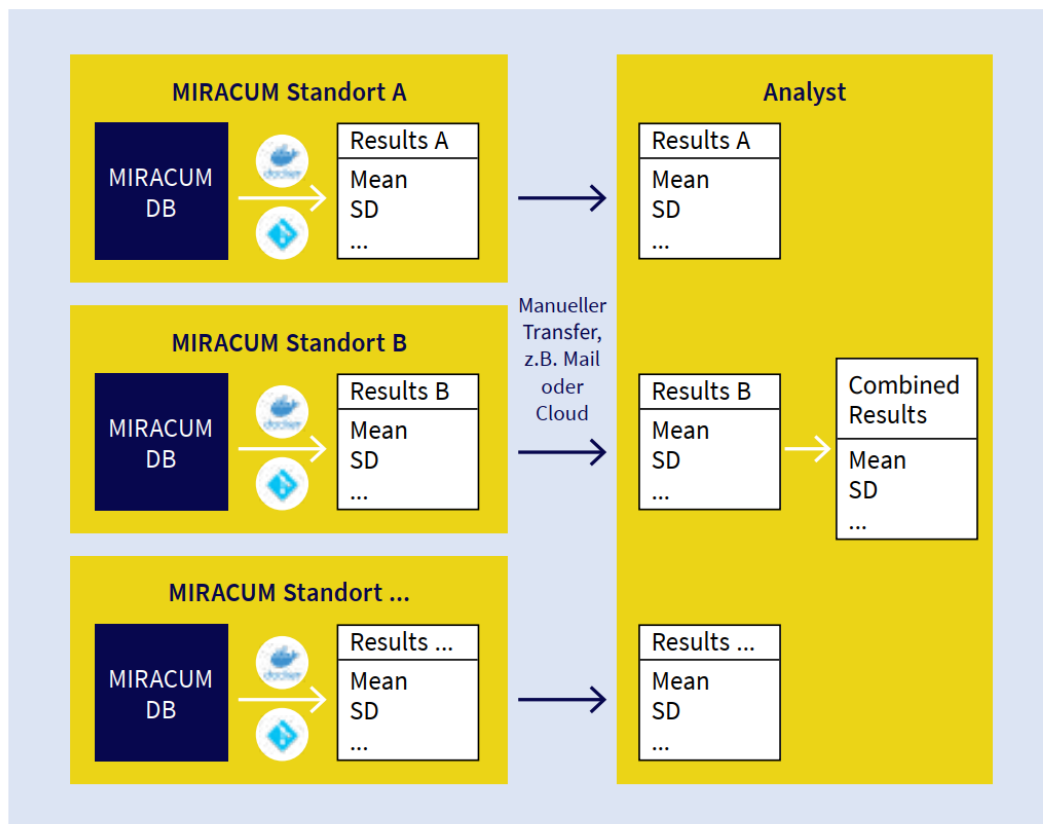
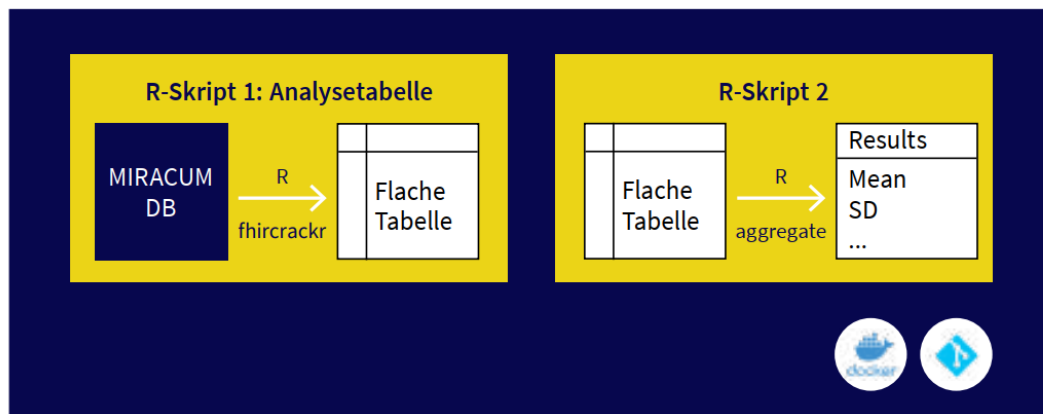


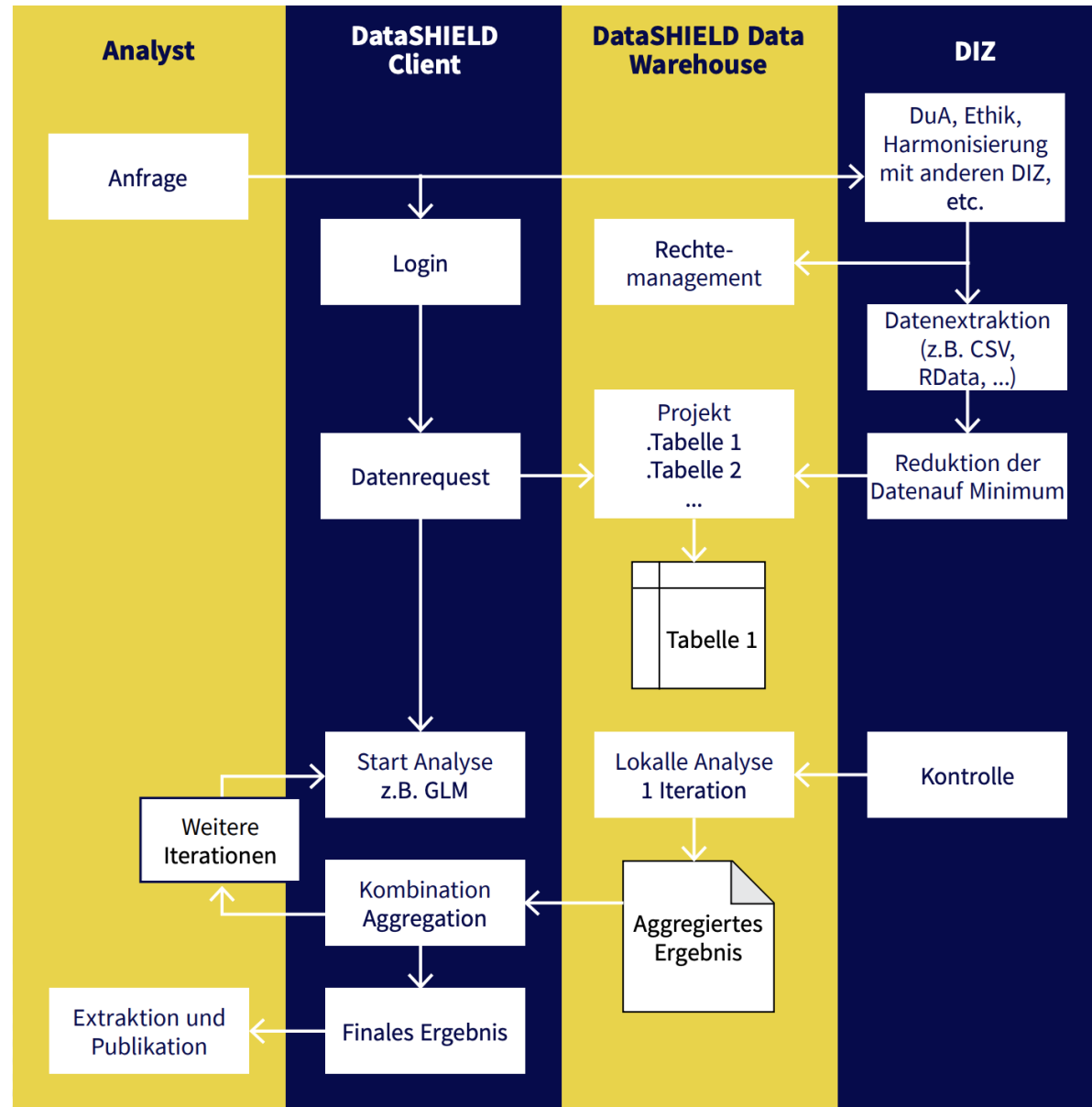
miracum

## UC 2

### FROM DATA TO KNOWLEDGE – KLINISCHE PROZESSE REVOLUTIONIEREN

Gigantische Datenmengen werden jeden Tag im deutschen Gesundheitswesen erzeugt – bislang zumeist weitgehend ungenutzt. Dabei sind unüberschaubare Mengen an Daten genau die richtige Voraussetzung, um damit valide Prädiktionsmodelle zu entwickeln – ganz so wie es in Use Case 2 angestrebt wird.







## DataSHIELD Rollen

### RACI Matrix

<b>D</b>	<b>Driver</b>	Assists those who are responsible for a task.
<b>R</b>	<b>Responsible</b>	Assigned to complete the task or deliverable.
<b>A</b>	<b>Accountable</b>	Has final decision-making authority and accountability for completion. Only 1 per task.
<b>S</b>	<b>Support</b>	Provides support during implementation.
<b>C</b>	<b>Consulted</b>	An adviser, stakeholder, or subject matter expert who is consulted before a decision or action.
<b>I</b>	<b>Informed</b>	Must be informed after a decision or action.

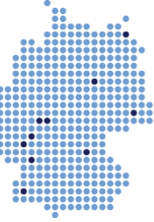
### Roles and Responsibilities

Responsible, Accountable, Consulted, Informed

ability for completion. Only 1 per task.

who is consulted before a decision or action.

	Prozesse	Status	ROLES								
			Netzwerkadmin	System Admin	DS Admin	KeyCloak Admin	Interer User	Externer User	Teamleiter	Studienleiter	Datenmanger
			Administratoren				Anwender		Projektspezifisch		
Installation											
	Netzwerkinfrastruktur		R	I							I
	Bereitstellen von Servern			R	I						
	DS Installation			I	R					A	I
	Freigabe von Nutzern				I	R	I	I	I	A	
	Benennung von Administratoren			I						A	I
Datenimport											
	Datenimport				S		I	I	C	A	R
Analyse											
	Ausführen von analysen						R	R		A	C



# Pub 1: Quality and utility of German routine care data

- COSYCONET: **C**OPD and **S**ystemic Consequences – **C**omorbidities **N**etwork
- Ziel: Exemplarisch die Verteilung und Aussagekraft der MIRACUM Daten untersuchen
  - COSYCONET Register: Goldstandard
  - Grafischer Vergleich der Verteilungen
  - Tabellarischer Vergleich von Lage- und Streuungsparametern
  - Vergleich von absoluten und relativen Statistiken

• Kohort

• Variabl

• P21:

• Lebe

• ProB

• Lung

	Ohne AATM	Mit AATM
Kein Emphysem	25815 (93.43%)	166 (84.69%)
Emphysem	1815 (6.57%)	30 (15.31%)

p-Wert Chi-Quadrat-Test mit Yates: <.001

Mar

abet

tte (

	Ohne AATM	Mit AATM
Keine Bronchiktase	27329 (98.91%)	184 (93.88%)
Bronchiktase	301 (1.09%)	12 (6.23%)

p-Wert Chi-Quadrat-Test mit Yates: <.001

# Pub 2: Differences in lipid metabolism of COPD patients correlated to Alpha-1-Antitrypsin-Deficiency



miracum

- Ziel: Unterschied im Fettstoffwechsel von COPD mit und ohne AATM
  - Erwartete Beobachtungen:  
Unterschiedliche Leberfunktion
- Kohorte: COPD mit und ohne Alpha-1-Antitripsin-Mangel
- Variablen:
  - P21: Bronchiektase, Lungenemphysem, Lipoprotein, Diabetes
  - Leberwerte (GOT, GPT, INR, Quick, Bilirubin) und Blutfette (LDL, HDL, Triglyzeride)
  - ProBNP und Troponin
  - Lungenfunktionsparameter

Vermutlich schwierig

-

Sehr wenige Messungen  
der nötigen Laborwerte

# Pub 3: Differences in germ profiles and infectious diseases of COPD patients correlated to AATD



miracum

- Ziel: Detektion von Hinweisen auf Unterschiede im Keim- und Infektionsprofil
  - Erwartete Beobachtungen:  
AATM Patienten sind stärker von Pseudomonas betroffen
- Kohorte: COPD mit und ohne Alpha-1-Antritripsin-Mangel
- Variablen:
  - P21: Bronchiektase, Lungenemphysem, Lipoprotein, Diabetes
  - Leberwerte (GOT, GPT, INR, Quick, Bilirubin) und Blutfette (LDL, HDL, Triglyzeride)
  - ProBNP und Troponin
  - Lungenfunktionsparameter

## Pneumokokkeninfektion

Mit AATM: 1.53%

Ohne AATM: 0.98%

p-Wert=0.6833

Adju. p-Wert=0,083

## Pseudomonasinfektionen

Mit AATM: 3.57%

Ohne AATM: 1.54%

p-Wert= 0.04502

Adju. p-Wert<0,001





# Pub 4/5: COVID-19 complications in COPD / Asthma patients

- Ziel: Vergleich COVID-19 Patienten mit und ohne COPD / Asthma
  - Erwartete Beobachtungen:  
Unterschiedliche Stärke von Komplikationen
- Kohorte: COVID-19 Patienten
- Variablen:
  - Lungenembolie, Gefäßschäden, Herzinfarkt, Schlaganfälle, Abszesse, Beatmung, Mortalität
  - Laborparameter
  - Lungenfunktionsparameter

Marburg hat mit  
Analysen gerade  
begonnen (Asthma)

-

Qualität Labor schwierig



# Pub 6: Implications of OxyHem in COPD patients

- Ziel: Untersuchung des Einfluss von OxyHem
  - $\text{OxyHem} = \text{Haemoglobin} * \text{fractional SaO}_2$
  - Erwartete Beobachtungen:  
OxyHem ist ein unabhängiger Prediktor für den Krankheitsverlauf
- Kohorte: COPD Patienten
- Variablen:
  - Lungenembolie, Gefäßschäden, Herzinfarkt, Schlaganfälle, Abszesse, Beatmung, Mortalität
  - Laborparameter
  - Lungenfunktionsparameter

Zeitaspekt in Arbeit  
—  
Aufnahme und  
Entlassung sehen bereits  
vielversprechend aus!

# Beispielhafte Ergebnisse aus Freiburg

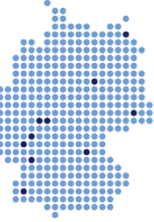
## adjustiert für Alter, Geschlecht, mehrere Aufenthalte



miracum

	LP	OR (>1 => Higher Odds for AATM)	
Flag Beatmung	2,366492	10,6599316	<0,001
COPD MAIN	0,511081	1,66709235	0,09042
BRONCHIEKTASE	1,72943	5,63743965	0,0181
Emphysem	1,75671	5,7933459	<0,001
Atelektase	0,903696	2,46871062	0,0263
Lung cancer	-2,8703415	0,05667957	0,005591
Metastasis	-0,66449	0,51453587	0,17339
Artrial Fibrillation	1,47663	4,37816637	0,000654
Valvular disease	0,62532	1,86884389	0,12986
Ischemic Stroke	1,41888	4,13248946	0,138
Arrhythmias	2,07133	7,93537014	<0,001
Renal Failure	1,857445	6,40734507	<0,001
Infect	0,890426	2,43616724	0,125
Bronchitis	0,848506	2,33615403	0,0138
Pneumokokken	1,16292	3,19926149	0,083381
Klebsielle	1,22551	3,40590265	0,0774
Pseudomonoas	2,33726	10,3528309	<0,001
Staph	1,65732	5,24523476	0,000695
Alcohol abuse	1,0723187	2,92214723	0,0791
Dementia	-17,65418	2,1522E-08	0,0556

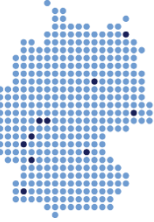
max	TLCbest	1,00142	<0,001
	FVCbestpred	-9,2515	0,0378
	Rvbest	0,08952	0,00249
	FEV1bestpred	-7,9201	0,071774
	SaO2	0,05674	0,01305
	TLCbestpred	12,34718	0,0128
	Rvbestpred	65,8302	0,000136
	GPT	131,2099	<0,001
	GOT	301,389	<0,001
	BiliGesamt	1,065	0,000101
	CRP	36,3884	<0,001
	Krea	0,1475	<0,001
	GammaGT	64,2304	0,00153
	LDH	333,046	<0,001
	Kalium	0,3558	<0,001
	Natrium	2,032	0,000944
	Harnstoff	31,59	<0,001
	Leuko	3,754	<0,001
	Thrombo	30,1574	0,003638
	Chlorid	5,82903	<0,001
	Calcium	-0,03721	0,114961
	KreaKinase	434,097	0,019
	Magnesium	0,01051	0,000352
	ProBNP	3598,55	0,00584
	AP	66,5615	<0,001
	Neutro	17,105	<0,001
	Mono	-0,08216	0,12043
	NA	9,09529	<0,001
	HbA1c	-0,41647	0,078492
	Ddimere	2,02157	0,03331



# Abstimmung Laborparameter für Publikationen (LOINC)

Laborparameter	Erlangen		Frankfurt		Freiburg		Gießen		Greifswald		Mainz		Mannheim	
		Einheiten		Einheiten		Einheiten		Einheiten		Einheiten		Einheiten		Einheiten
Albumin			1751-7	mg/L	1751-7 14957-5	g/dL mg/dL	1751-7	g/dL			1751-7	g/dL	1751-7	g/l
alkalische Phosphatase (AP)			6768-6	U/L	6768-6	U/L	6768-6	U/L			6768-6	U/L	6768-6	U/L
Alpha-Amylase			1798-8	U/L	1798-8	U/L	1798-8	U/L			1798-8	U/L	1798-8	U/L
Bilirubin					42719-5 1975-2 50551-1 5770-3	mg/dL mg/dL	nicht vorhanden							
Bilirubin Gesamt			1975-2	mg/dL	54363-7	umol/L	1975-2	mg/dL			1975-2	mg/dL		
Bilirubin direkt			1968-7	mg/dL	1968-7	mg/dL	1968-7	mg/dL			1968-7	mg/dL		
Bilirubin indirekt			1971-1	mg/dL	1971-1	mg/dL								
Blutglucose			2339-0 2345-7 41653-7	mg/dL mg/dL mg/dL	2339-0	mg/dL	2345-7	mg/dL			2345-7	mg/dL		
BNP (Natriuretic peptide B)					30934-4	pg/mL	30934-4	pg/mL						

# Lungenfunktionsparameter



miracum

Laborparameter	Erlangen		Frankfurt		Freiburg		Gießen		Greifswald		Mainz		Mannheim	
Lungenfunktionsparameter														
Diffusion capacity.carbon monoxide					19911-7	cm3/min/mm[Hg]							19911-7	mmol/min/kF
Diffusion capacity/Alveolar volume					19916-6	mL/min/mm[Hg]/L							19916-6	mmol/min/kF
Diffusion capacity/Alveolar volume pre bronchodilatation					82620-6	mL/min/mm[Hg]/L								
Diffusion capacity/Alveolar volume post bronchodilatation					82619-8	mL/min/mm[Hg]/L								
FEV1 best			20150-9	L	20150-9	L							20157-4	L
FEV1 best predicted			20149-1	L	20152-5	%								
FEV1 post bronchodilatation			20155-8	L	20155-8	L							20155-8	L
FEV1 pre bronchodilatation			20157-4	L	20157-4	L							20157-4	L
FVC best			19870-5	L	19868-9	L							19876-2	L
Forced Vital Capacity best predicted			19869-7	L	19872-1	%								
Forced Vital Capacity post bronchodilatation			19875-4 (unspez.) 19874-7 (Spiro)	L	19874-7	L							19874-7	L
Forced Vital Capacity pre			19877-0	L	19876-2	L							19876-2	L



# TMF Taskforce Verteilte Analyse

Konsortium	Nachname	Vorname	Titel	Institution
DIFUTURE	Fürstberger	Axel	Dr.	Universität Ulm
DIFUTURE	Kestler	Hans	Prof. Dr.	Universität Ulm
DIFUTURE	Ballhausen	Hendrik	PD Dr.	Universitätsklinikum München
DIFUTURE	Herr	Marius		Universität Tübingen
DIFUTURE	Hinske	LudwigChristia n	Prof. Dr.	Universitätsklinikum München
HiGHmed	Floca	Ralf Omar	Dr.	DKFZ Heidelberg
HiGHmed	Prasser	Fabian	Prof. Dr.	Charité Berlin
HiGHmed	Wolf	Andreas	Dr.	Universitätsklinikum Schleswig- Holstein,Campus Kiel
MIRACUM	Binder	Harald	Prof. Dr.	Universitätsklinikum Freiburg
MIRACUM	Boeker	Martin	Prof. Dr.	Universitätsklinikum Freiburg
MIRACUM	Tabatabaei	Amir	Dr.	Universität Gießen
MIRACUM	Zöller	Daniela	Dr.	Universitätsklinikum Freiburg
SMITH	Beyan	Oya Deniz	Dr.	RWTH Aachen
SMITH	Kirsten	Toralf	Dr.	Universität Leipzig
SMITH	Löffler	Markus	Prof. Dr.	Universität Leipzig
SMITH	Schuppert	Andreas	Prof. Dr.	RWTH Aachen
SMITH	Ucer Yediel	Yeliz		Fraunhofer FIT
SMITH	Zink	Corinna		Universitätsklinikum Aachen
SMITH	Pietzner	Diana	Dr.	Universitätsklinikum Halle
SMITH	Gentges	Christian		Universitätsklinikum Aachen
SMITH	Kusch	Nina		Universitätsklinikum Aachen
SMITH	Erdfelder	Felix	Dr.	Universitätsklinikum Bonn
POLAR	Kesselmeier	Miriam	Dr.	Universitätsklinikum Jena



# Erste Ziele der TF Verteilte Analysen

- **Ziel 1:**  
Erfolgreiche Umsetzung **eines Anwendungsfalls** des verteilten Rechnens in Phase 1 der MII
- **Ziel 2:**  
Erfolgreiche **Etablierung einer nationalen Struktur** für verteiltes Rechnen in Phase 2 der MII

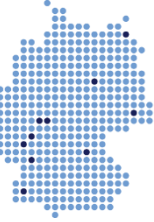


# Erste Ergebnisse der TF Verteilte Analysen

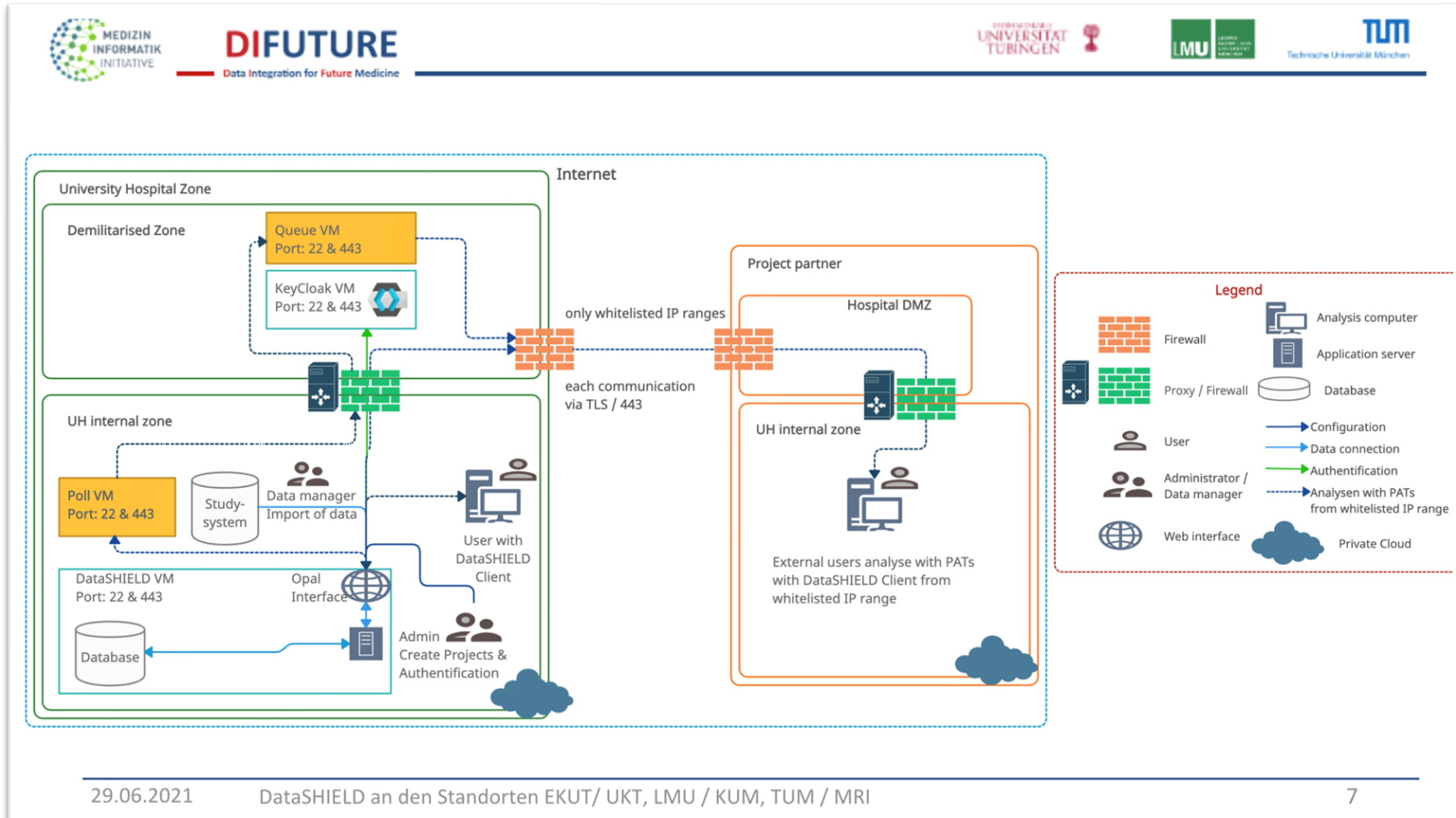
- **Aktuell: Keine** der bisherigen Lösungen erfüllt **Ziel 1 UND 2**
  - Betrachtete Lösungen haben sehr unterschiedliche Stärken und Schwächen
- **Ziel 2** hat deutlich **breiteres/anspruchsvolleres Anforderungsprofil**
  - Entwicklung Kriterienkatalog abgeschlossen
  - Bewertung der Lösungen ausstehend
  - Ziel: Zusammenführung der Lösung in synergetischer Systemarchitektur
- **Ziel 1** hat **kurzfristig höhere Bedeutung** und ist leichter erreichbar
  - Höchste Priorität für die Bewertung: Verwendung in existierenden Projekten („Installed Base“)

➔ **DataSHIELD** für Ziel 1 als die **geeignetste Lösung**

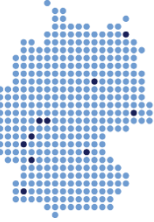
# Einsatz von DataSHIELD in anderen Konsortien



miracum



# Danke an das UC 2 Team!



miracum



@miracum\_de

- Verantwortliche:

Harald Binder, Till Acker und Harald Renz

- Team (vermutlich nicht vollständig):

Dennis Toddenroth, Julian Gründner, Stefanie Schild

Abishaa Vengadeswaran, Jannik Schaaf

Saskia Kiefer, Stefan Lenz, Kiana Farhadyar, Christian Haverkamp, Johannes Wolf, Denis Gebele

Patrick Fischer, Amir Tabatabaei, Jochen Zohner, Attila Nemeth, Hildegard Dohmen

Roland Buhl, Dativa Tibyampansa, Stefanie Korn,

Michael Neumaier, Jan Scheer, Alejandro Cosa Linan, Frederik Trinkmann

Fabian Simons, Christian Seidemann

Sebastian Schindler, Frederike Euchner, Christian Bruns, Anke Lux