



TECHNISCHE  
UNIVERSITÄT  
DRESDEN

Universitätsklinikum  
Carl Gustav Carus  
DIE DRESDNER.



# FHIR-to-OMOP ETL-Prozess

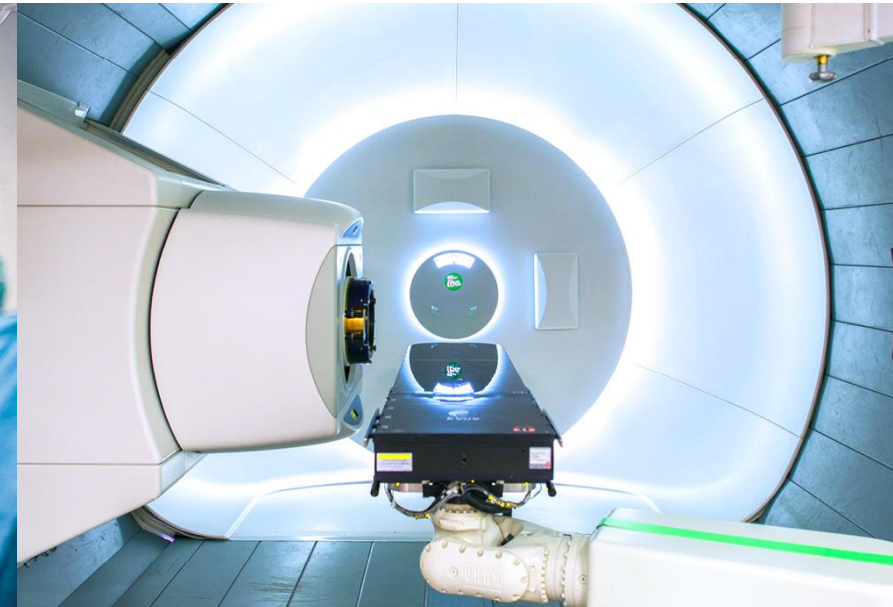
15.06.2023 HL7 Interoperabilitätsforum

**Elisa Henke und Yuan Peng**

Institut für Medizinische Informatik und Biometrie

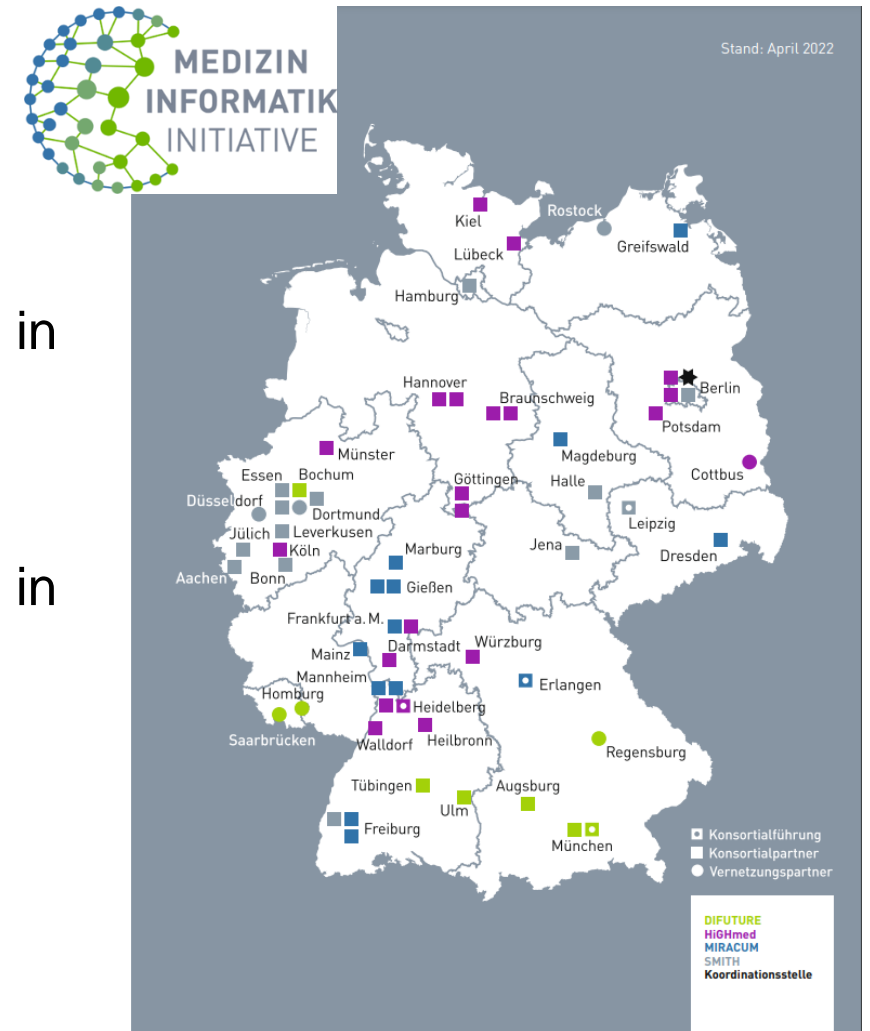
Medizinische Fakultät Carl Gustav Carus der TU Dresden

Universitätsklinikum Dresden



# Hintergrund – MII

- Zunehmende Bedeutung von Real World Data (RWD) in medizinischer Forschung
- Unterstützung der medizinischen Forschung mit RWD in Deutschland durch Medizininformatik Initiative (MII)
- Abgestimmter Kerndatensatz in FHIR durch MII



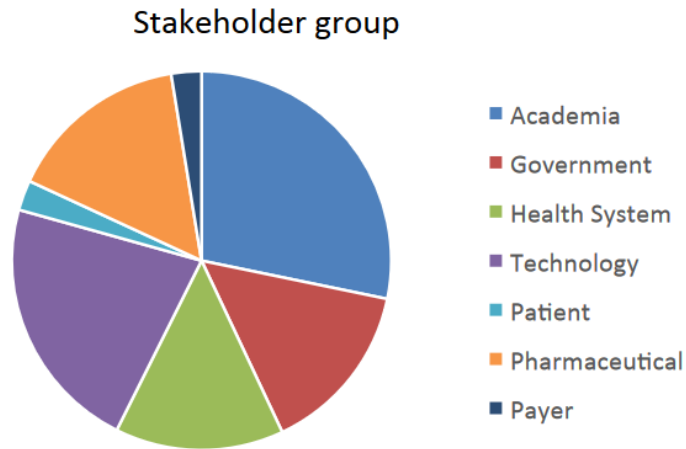
# Hintergrund – OHDSI



**OHDSI**  
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

## What OHDSI is:

- ✓ Open Source
- ✓ Community
- ✓ Data



### OHDSI Collaborators:

- 2,100 users
- 25 workgroups
- 20 open-source applications

### OHDSI Network:

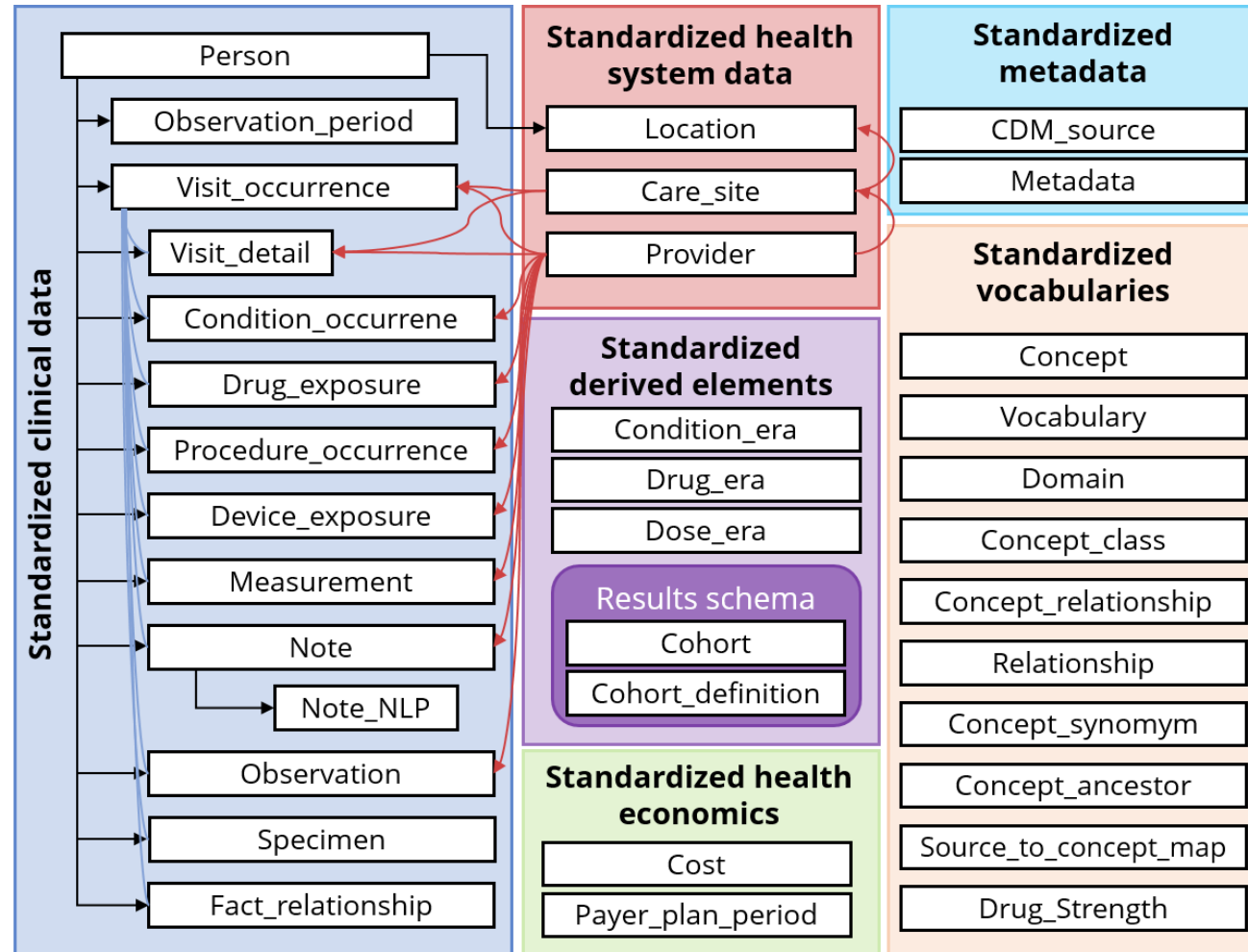
- >160+ databases
- 23 countries
- 578M distinct patients
- 2.7B de-identified patient records



# Hintergrund – OMOP CDM



EUROPEAN MEDICINES AGENCY  
SCIENCE MEDICINES HEALTH



# Hintergrund – Anforderungen



Transformation der MII KDS Basismodule nach OMOP CDM



Transformation großer Datenmengen



Flexibilität bei Änderungen der FHIR-Profile

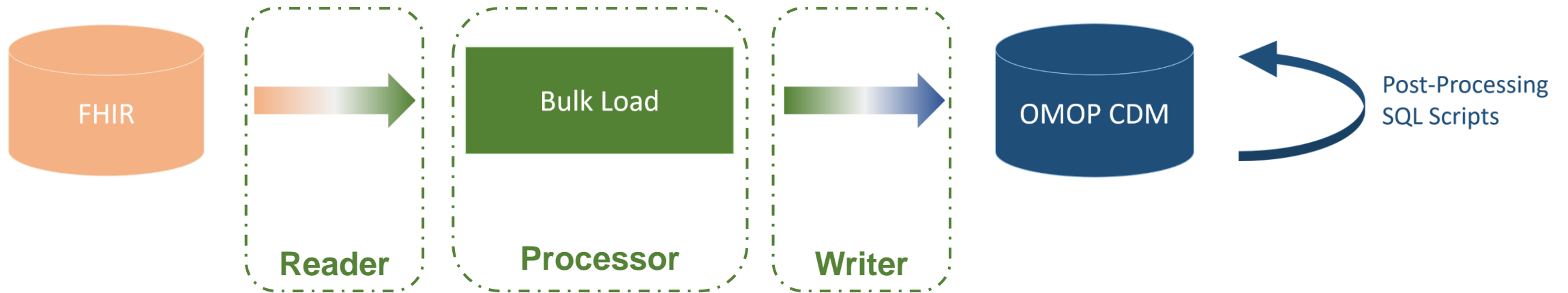
# Methode – Material

## ■ MII KDS Basismodule v1.0

Module	Ressourcentyp
Person	Patient
Fall	Encounter (Einrichtungskontakt)
	Encounter (Abteilungskontakt)
Diagnose	Condition
Medikation	Medication
	MedicationAdministration
	MedicationStatement
Prozedur	Procedure
Laborbefund	Observation

# Methode – ETL Design

- Java 17
- HAPI FHIR Library 6.1.3
- Spring-Batch Framework 2.6.7



# Methode – Post-Prozess

- Nutzung von SQL-Skripten, um Daten auf Datenbankebene zu transformieren
- ✓ Reduzierung der Komplexität des Mappings
- ✓ Reduzierung der RAM-Belastung
- ✓ Verbesserung der Performance
- ✓ Verhinderung von PostgreSQL-Datenbank Problemen während der Transformation, z.B. DeadLock Error



# Methode – Flexibilität

```
String singleStepName;  
  
Flow conditionalFlow{  
  on "Condition":  
    executeStepCondition(); //including Reader, Processor and Writer  
  on "Procedure":  
    executeStepProcedure(); //including Reader, Processor and Writer  
  on "Medication":  
    executeStepMedication(); //including Reader, Processor and Writer  
  on "MedicationStatement":  
    executeStepMedicationStatement(); //including Reader, Processor and Writer  
  on "MedicationAdministration":  
    executeStepMedicationAdministration(); //including Reader, Processor and Writer  
  on "Observation":  
    executeStepObservation(); //including Reader, Processor and Writer  
  on "DepartmentCase":  
    executeStepDepartmentCaseEncounter(); //including Reader, Processor and Writer  
}
```

# Ergebnis – OMOP CDM Tabellen

## OMOP CDM v5.3.1

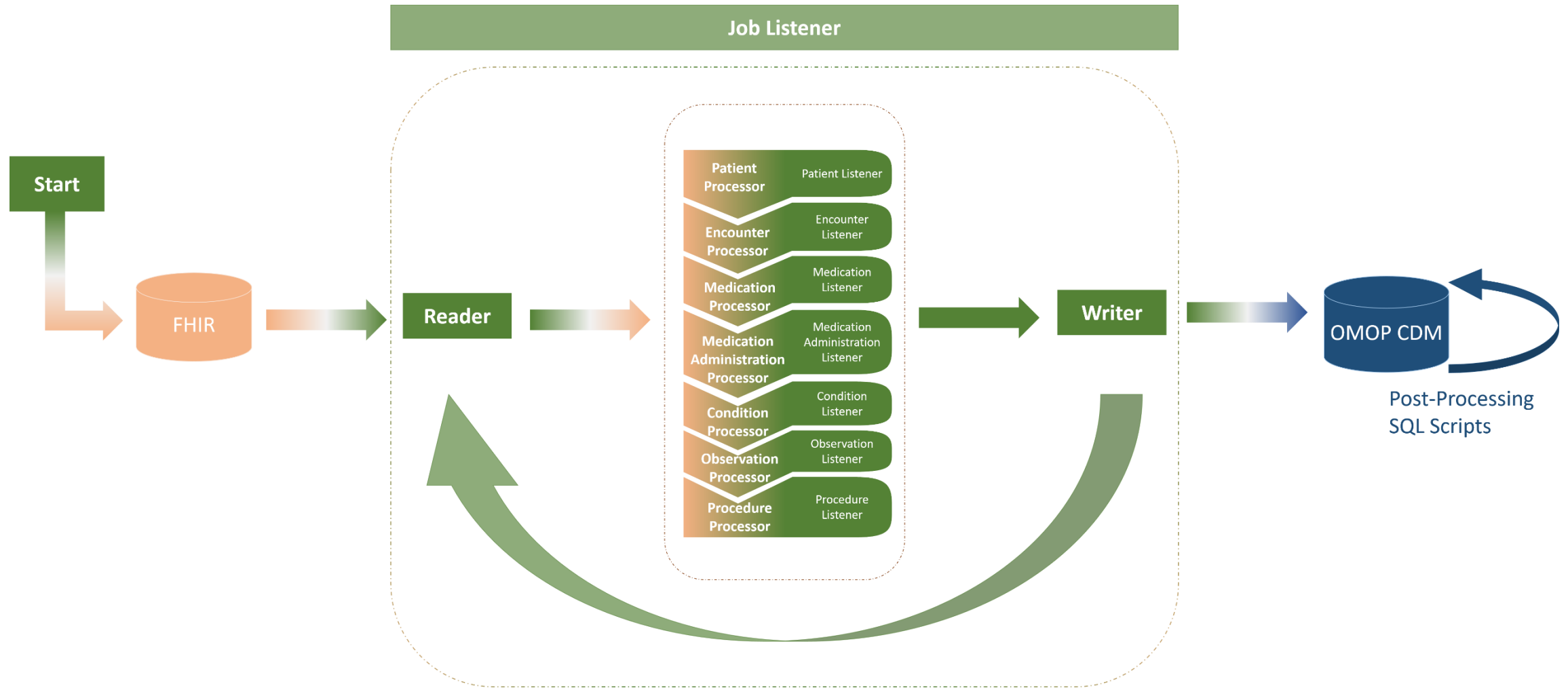
Schema	Tabellen-/Viewname	
cds_cdm	person	death
	visit_occurrence	visit_detail
	location	care_site
	observation	measurement
	procedure_occurrence	condition_occurrence
	drug_exposure	fact_relationship
	observation_period	cdm_source
	source_to_concept_map	
cds_etl_helper	medication_id_map	post_process_map
	icd_snomed_domain_lookup	

# Ergebnis – Datenqualität in OMOP CDM

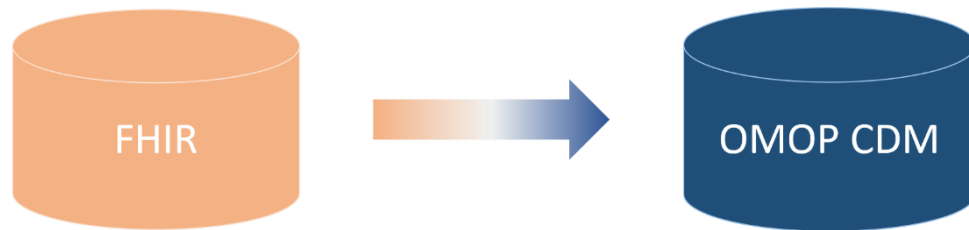
## Data Quality Dashboard (DQD)

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1991	<b>3</b>	1994	100%	281	<b>6</b>	287	98%	2272	<b>9</b>	2281	100%
Conformance	676	<b>5</b>	681	99%	104	0	104	100%	780	<b>5</b>	785	99%
Completeness	379	<b>7</b>	386	98%	12	<b>3</b>	15	80%	391	<b>10</b>	401	98%
<b>Total</b>	3046	<b>15</b>	3061	100%	397	<b>9</b>	406	98%	3443	<b>24</b>	3467	<b>99%</b>

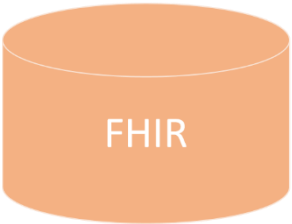
# Ergebnis – ETL-Architektur



# Anforderung – Inkrementelles Laden



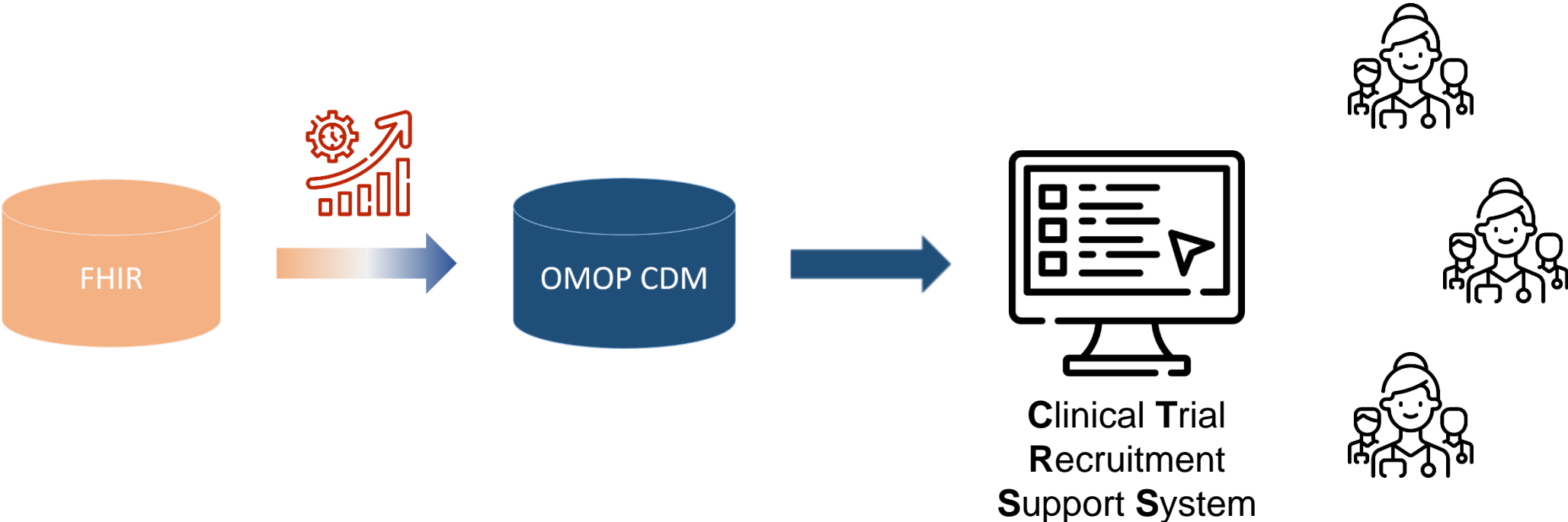
# Anforderung – Inkrementelles Laden



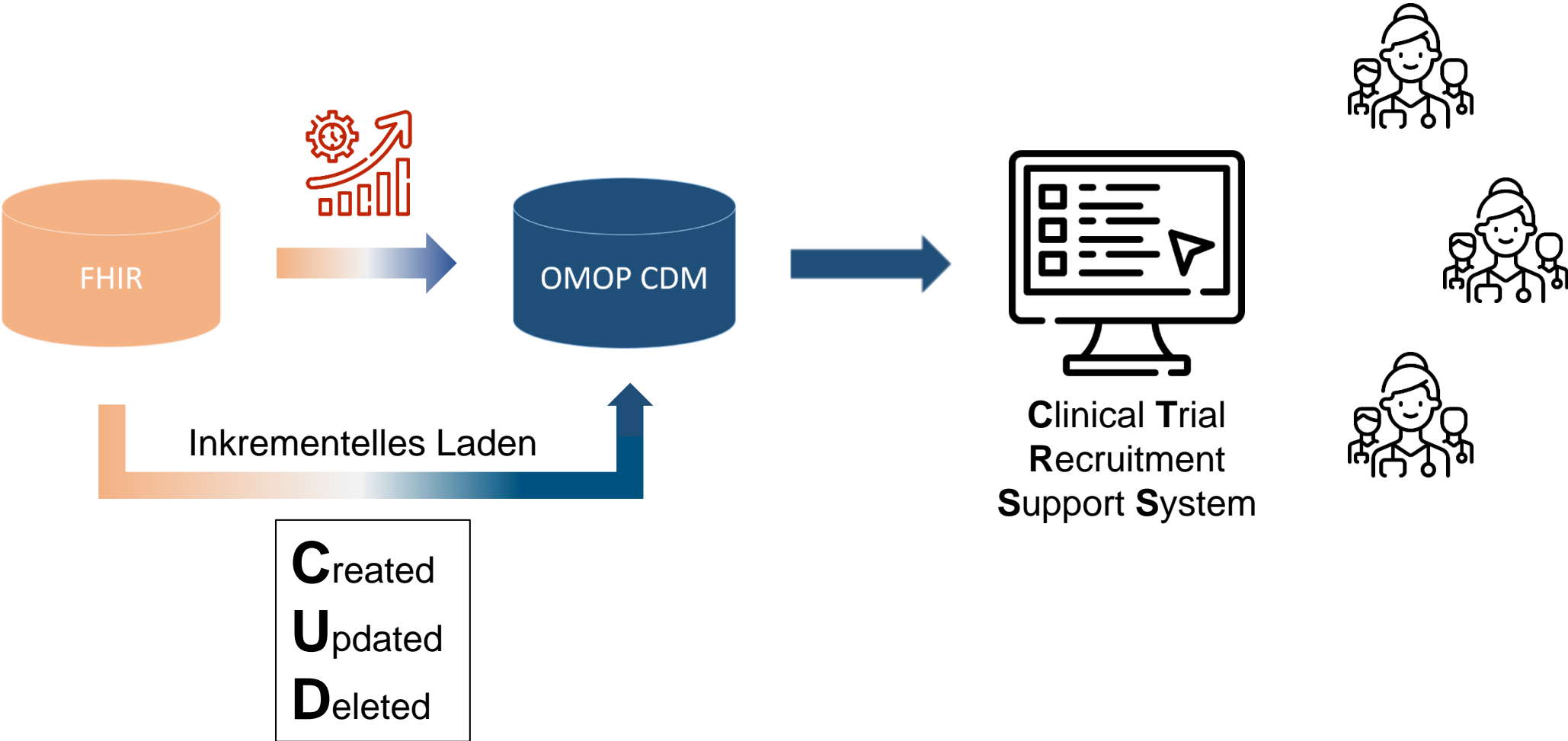
Clinical Trial  
Recruitment  
Support System



# Anforderung – Inkrementelles Laden



# Anforderung – Inkrementelles Laden





# Methode - Konfigurationsparameter

Parameter Merkmal	APP_BULKLOAD_ENABLED	DATA_BEGINDATE & DATA_ENDDATE
Fokus	Unterscheidung zwischen Bulk Laden und Inkrementellem Laden	Begrenzung des Einlesezeitraums von FHIR Ressourcen
Ansatz	Switch	Timestamp-based Change Data Capture (CDC)
Datentyp	Boolean	Date
Spezifikation	true: Bulk Laden false: Inkrementelles Laden	lastUpdated $\in$ [DATA_BEGINDATE, DATA_ENDDATE]

# Methode – Prozessieren von UD-FHIR Ressourcen

```
{
  "resourceType": "Condition",
  "id": "1234",
  "code": {
    "coding": [
      {
        "extension": [
          {
            "url": "http://fhir.de/StructureDefinition/icd-10-gm-diagnosesicherheit",
            "valueCoding": {
              "system": "https://fhir.kbv.de/CodeSystem/KBV_CS_SFHIR_ICD_DIAGNOSESICHERHEIT",
              "code": "G"
            }
          }
        ],
        "system": "http://fhir.de/CodeSystem/dimdi/icd-10-gm",
        "version": "2020",
        "code": "A54.4† M73.04*"
      }
    ],
    "subject": {
      "reference": "Patient/12345"
    },
    "onsetPeriod": {
      "start": "2019-09-26T12:45:00+01:00",
      "end": "2020-03-25T13:00:00+01:00"
    },
    "identifier": [
      {
        "value": "1234",
        "system": "https://diz.uniklinikum-dresden.de/fhir/core/NamingSystem/DiagnoseId"
      }
    ]
  },
  ...
}
```

# Methode – Prozessieren von UD-FHIR Ressourcen

## Extraction of id and identifier

FHIR id: 1234  
FHIR identifier: 1234



## Add prefix to id and identifier

FHIR id: con-1234  
FHIR identifier: con-1234



## Verification of Condition FHIR resource in OMOP CDM

condition_occurrence_id	person_id	condition_concept_id	condition_source_value	condition_status_source_value	condition_status_concept_id	fhir_logical_id	fhir_identifier	...
...	...	...	...	...	...	...	...	...
22	10	37116425	A54.4†	<null>	<null>	con-1234	con-1234	...
23	10	4211842	M73.04*	<null>	<null>	con-1234	con-1234	...
24	10	135466	M73.04*	<null>	<null>	con-1234	con-1234	...
25	5	78605	S50.0	V	4033240	con-9101	con-9101	...

# Methode – Prozessieren von UD-FHIR Ressourcen

Verification of Condition FHIR resource in OMOP CDM

condition_occurrence_id	person_id	condition_concept_id	condition_source_value	condition_status_source_value	condition_status_concept_id	fhir_logical_id	fhir_identifier	...
...	...	...	...	...	...	...	...	...
22	10	37116425	A54.4†	<null>	<null>	con-1234	con-1234	...
23	10	4211842	M73.04*	<null>	<null>	con-1234	con-1234	...
24	10	135466	M73.04*	<null>	<null>	con-1234	con-1234	...
25	5	78605	S50.0	V	4033240	con-9101	con-9101	...



Deletion of Condition FHIR resource in OMOP CDM

condition_occurrence_id	person_id	condition_concept_id	condition_source_value	condition_status_source_value	condition_status_concept_id	fhir_logical_id	fhir_identifier	...
...	...	...	...	...	...	...	...	...
25	5	78605	S50.0	V	4033240	con-9101	con-9101	...

# Methode – Prozessieren von UD-FHIR Ressourcen

Deletion of Condition FHIR resource in OMOP CDM

condition_occurrence_id	person_id	condition_concept_id	condition_source_value	condition_status_source_value	condition_status_concept_id	fhir_logical_id	fhir_identifier	...
...	...	...	...	...	...	...	...	...
25	5	78605	S50.0	V	4033240	con-9101	con-9101	...



Semantic mapping of Condition FHIR resource to OMOP CDM



# Methode – Prozessieren von UD-FHIR Ressourcen

Semantic mapping of Condition FHIR resource to OMOP CDM



Writing of Condition FHIR resource to OMOP CDM

condition_occurrence_id	person_id	condition_concept_id	condition_source_value	condition_status_source_value	condition_status_concept_id	fhir_logical_id	fhir_identifier	...
...	...	...	...	...	...	...	...	...
25	5	78605	S50.0	V	4033240	con-9101	con-9101	...
26	10	37116425	A54.4†	G	4230359	con-1234	con-1234	...
27	10	4211842	M73.04*	G	4230359	con-1234	con-1234	...
28	10	135466	M73.04*	G	4230359	con-1234	con-1234	...

new condition\_occurrence\_ids

updated data

# Methode – ETL-Test Design

## *Test focus*

## *Hypothesis*

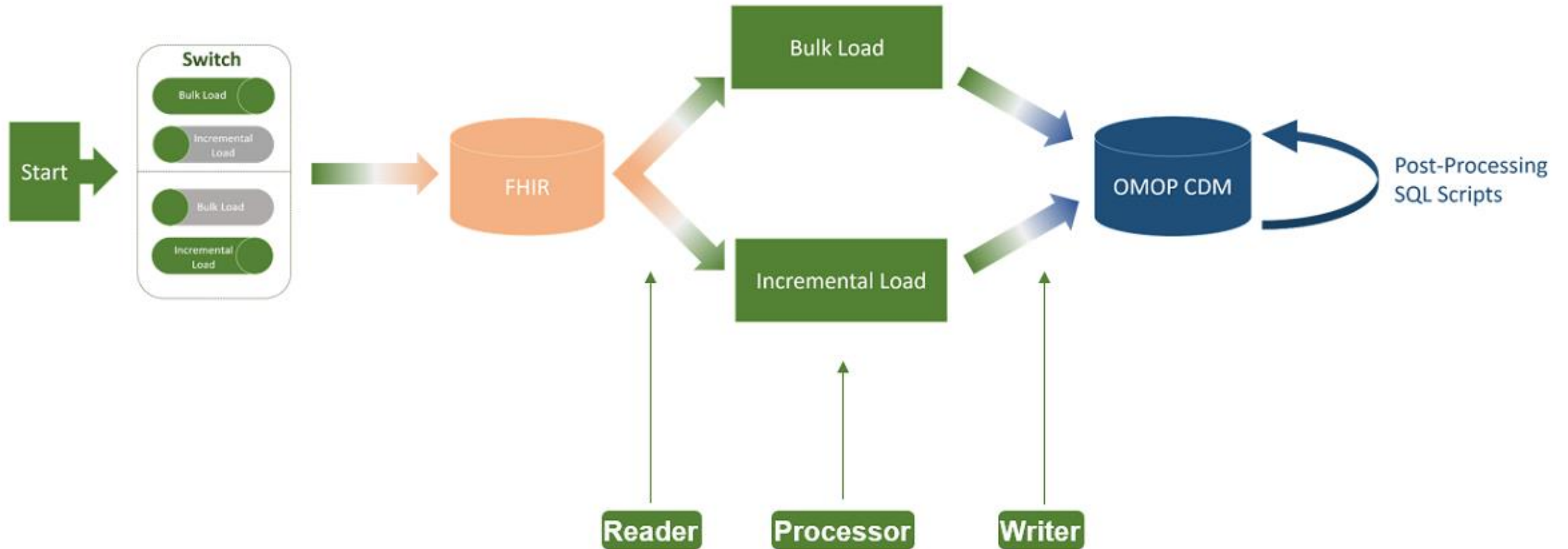
*Performance*

$t(\text{bulk loading (3 months)}) + t(\text{incremental loading (1 day)}) <$   
 $t(\text{bulk loading (3 months)}) + t(\text{bulk loading (3 months + 1 day)})$

*Data correctness*

$\#((\text{bulk loading (3 months)}) + (\text{incremental loading (1 day)})) =$   
 $\#(\text{bulk loading (3 months + 1 day)})$

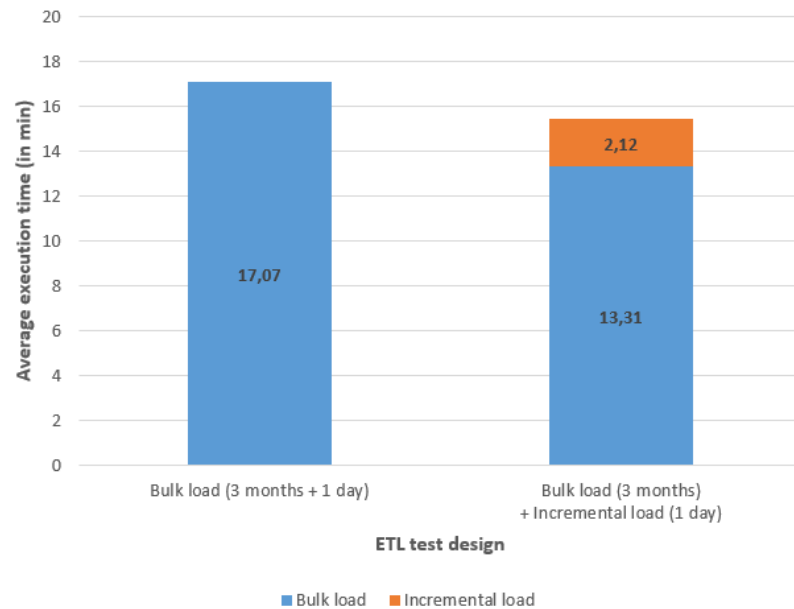
# Ergebnis – ETL-Architektur



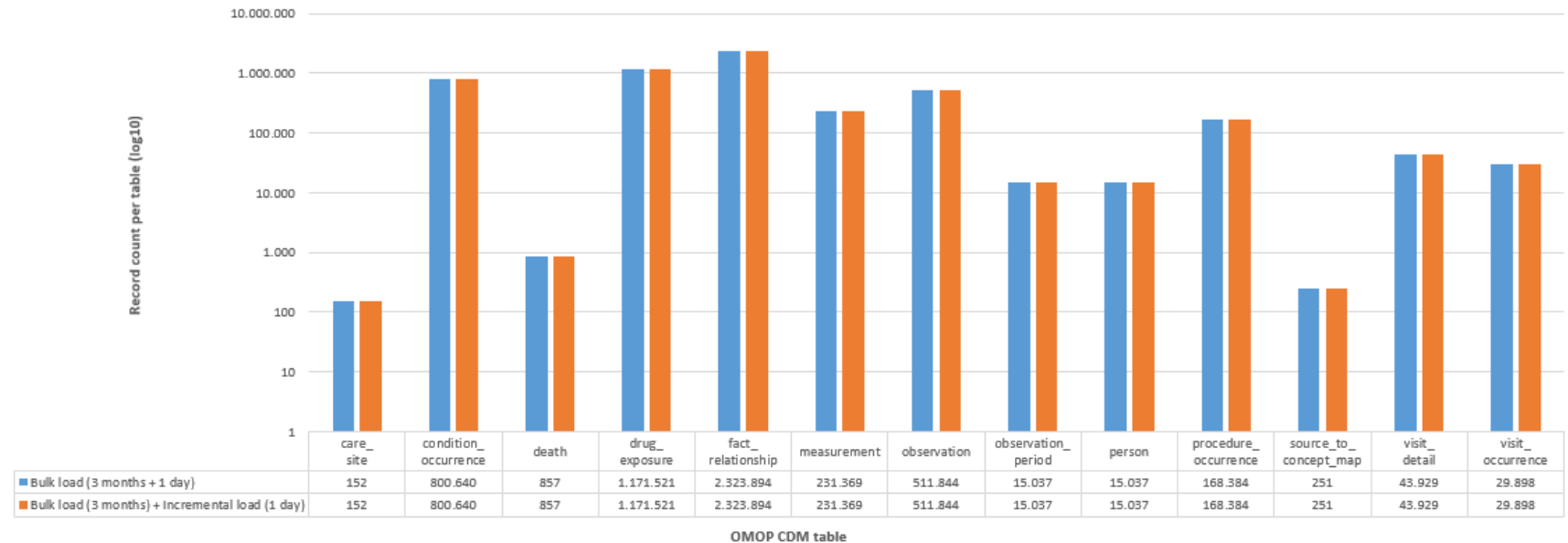


# Ergebnisse – ETL-Test

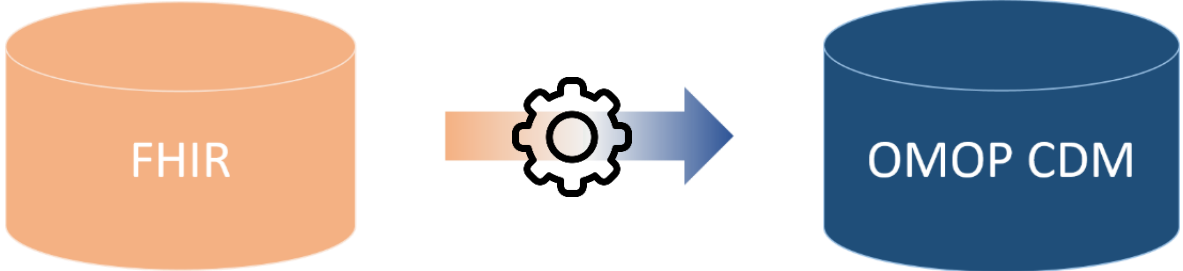
Comparison of execution time between bulk and incremental load



Comparison of data quantity in OMOP CDM between bulk and incremental load



# Zusammenfassung



MiHUBx



SATURN

Smartes Arztportal für unklare Erkrankungen



EUROPEAN MEDICINES AGENCY  
SCIENCE MEDICINES HEALTH



Universitätsklinikum  
Carl Gustav Carus  
DIE DRESDNER.



# Ausblick

- Verbesserung Error-Handling und Loggings
- weitere Verbesserung der Performance des ETL-Jobs
- Umsetzung neuer FHIR-Profile (KDS v2, Erweiterungsmodule)
  - Ansatz Metadaten getriebener ETL Prozess
- OMOP CDM v5.4
- ...

# Vielen Dank für Ihre Aufmerksamkeit.

## Kontakt

Medizinische Fakultät Carl Gustav Carus  
der Technischen Universität Dresden  
Fetscherstr. 74 (Haus M161)  
01307 Dresden  
E-Mail: [Elisa.Henke@ukdd.de](mailto:Elisa.Henke@ukdd.de)  
[Yuan.Peng@ukdd.de](mailto:Yuan.Peng@ukdd.de)



Weitere Informationen  
über FHIR-to-OMOP:



# FHIR-to-OMOP

