

# The TOP Framework: A Next-generation Phenotype Repository

JRG “Terminology- and Ontology-based Phenotyping”

Christoph Beger<sup>a,1</sup>, Franz Matthies<sup>a</sup>, Ralph Schäfermeier<sup>a</sup>,  
Konrad Höffner<sup>a</sup>, Alexandr Uciteli<sup>a</sup>

MIRACUM-DIFUTURE Symposium 2024  
2024-10-11, Session 3 - Young Researchers

<sup>a</sup> Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University

<sup>1</sup> Corresponding Author: Christoph Beger, E-mail: [christoph.beger@imise.uni-leipzig.de](mailto:christoph.beger@imise.uni-leipzig.de), Mailing address: Härtelstraße 16-18, 04107 Leipzig, Germany

**Keywords:** Computable phenotypes, data-driven phenotyping, ontologies, patient stratification



**MIRACUM  
DIFUTURE**

Medical Informatics for Research and Care

# Overall Aims of the TOP JRG

1. Contribute to the **ontological foundation** of phenotyping
2. Develop a standardised phenotype **modelling language**
3. Combine phenotype models with data sources in the MII, enabling **search queries**
4. Integrate **external terminologies**
5. Enable **document search** in unstructured documents
6. Provide **user-friendly tools** for next-generation phenotyping



Next-generation phenotyping combines the **semantic and structured (standardised)** definition of phenotypes with the **creation of phenotype algorithms**. – Mo et al. [1].

# Ontological Foundation



# Phenotypes

- Phenotypes are observable (combinations of) characteristics of an organism [2]

**Two classes of phenotypes:**

## Basic phenotypes

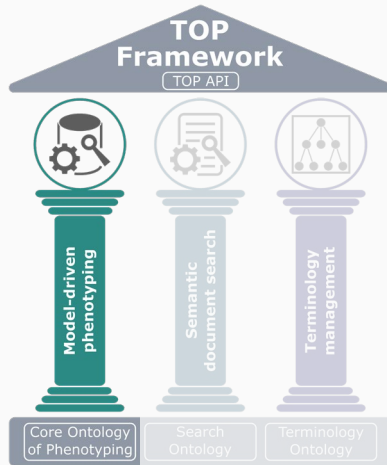
---

directly/indirectly

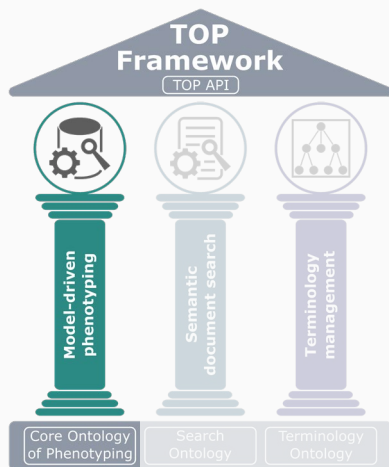
**observable or measurable**

---

eye color, blood pressure,  
insuline medication

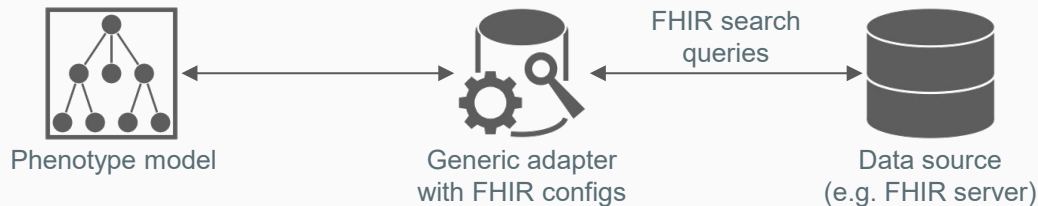


# Phenotype Models

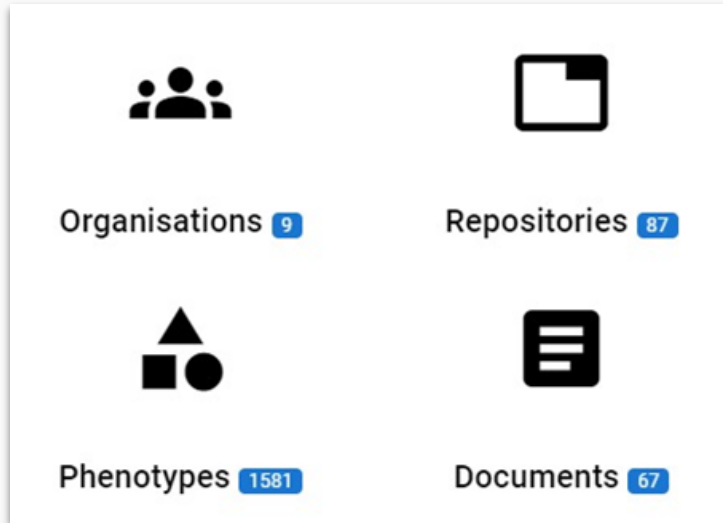


## Purposes:

- Detection of diseases
- Risk assessment
- Recruitment for clinical studies
- Individualised patient care and treatment
- Actual data is not stored in the framework!
- Models can be translated into query languages
- Data sources are accessible via adapters:
  - **generic adapters** with configurations (e.g., OMOP, POLAR, FHIR Search)
  - **specific adapters** (e.g., LIFE study)



# Model-driven Phenotyping With the TOP Framework



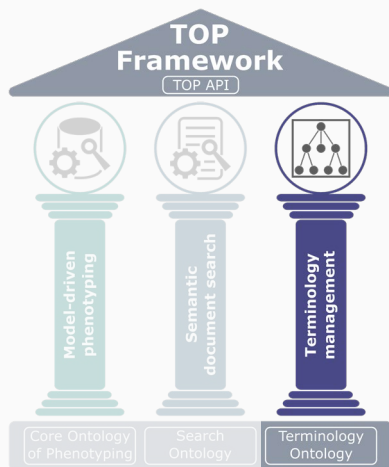
## Key functionalities:

- Collaborative web application for non-IT staff
- Development of computable phenotype models using standardised terminologies
- Efficient data exploration and analysis through a standardised query interface [3]
- Sharing and reusing phenotype models and their components [4]



**TOP Framework has the potential to enhance knowledge discovery, efficiency, and collaboration within clinical research.**

# Terminology Management

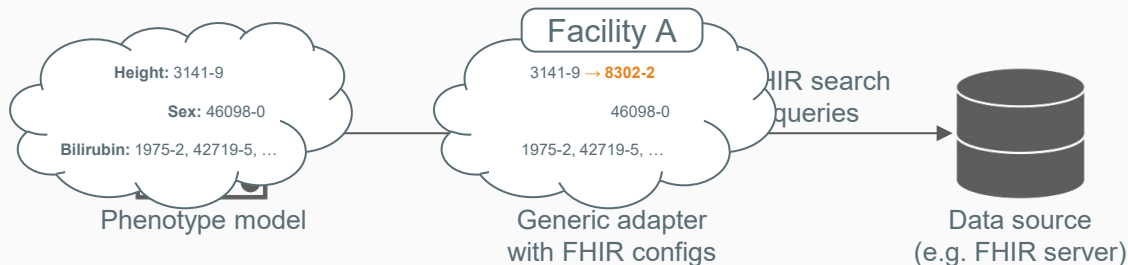


- Terminologies are stored in the **Ontology Lookup Service**



- Medication: ASK, ATC, PZN
- Diagnosis: ICD-10
- Observation: LOINC, SNOMED CT
- Procedure: OPS

- TOP Framework provides search widget for OLS codes
- Codes are used in phenotype models to enable interoperability with data sources



# Terminology Management

The screenshot shows a web-based interface for managing medical terminology. The main header is 'total bilirubin' with a close icon. Below it, it says 'total bilirubin Version: 3 (Jun 2019)'. There are three tabs: 'Synonyms', 'Descriptions', and 'Codes'. The 'Codes' tab is selected, showing a list of codes with their descriptions and a 'SNOMEDCT' label. The list includes:

- [104550005] Bilirubin renal clearance measurement
- [235043009] Bilirubin pigmentation of oral mucosa
- [707965004] Bilirubin crystal
- [104553007] Bilirubin, non-glucuronidated measurement
- [413054008] Bilirubin profile
- [47922009] Bilirubinuria
- [417567002] Bilirubin neonatal measurement

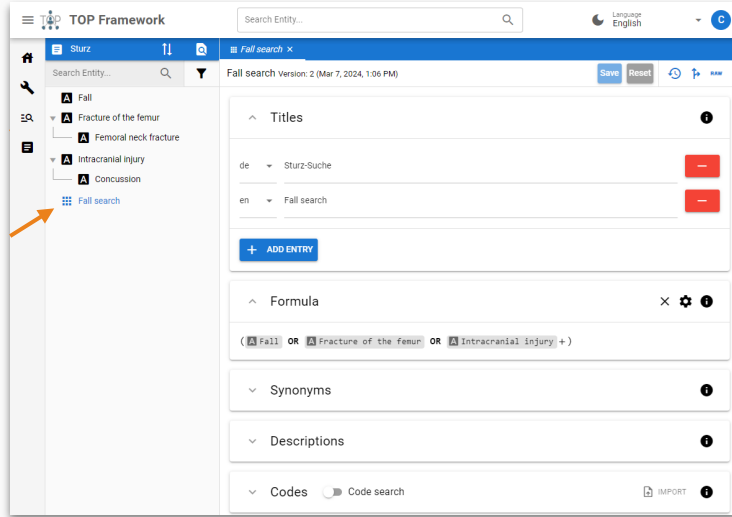
Below the list, there is a 'Code system' dropdown set to 'bilirubin' and an 'ADD CODE' button. At the bottom, there is a table of LOINC codes with their descriptions and a red minus button next to each:

| LOINC                                                            | Description | Action |
|------------------------------------------------------------------|-------------|--------|
| LOINC: Bilirubin.total [Mass/volume] in Blood [42719-5]          |             | —      |
| LOINC: Bilirubin [Mass/Vol] [1975-2]                             |             | —      |
| LOINC: Bilirubin [Moles/Vol] [14631-6]                           |             | —      |
| LOINC: Bilirubin Bld-sCnc [54363-7]                              |             | —      |
| LOINC: Bilirubin Fld Ql [43824-2]                                |             | —      |
| LOINC: Bilirubin Direct+Tot Pnl SerPl-mCnc [34543-9]             |             | —      |
| LOINC: Bilirubin.total [Mass/volume] in Arterial blood [59827-6] |             | —      |
| LOINC: Bilirubin [89872-6]                                       |             | —      |
| LOINC: Bilirubin.conjugated+indirect [Mass/Vol] [33898-8]        |             | —      |

## Future work:

- Handling of subtrees  
e.g. in ICD-10:
  - P05–P08: Disorders of newborn related to length of gestation and fetal growth  
⇒ 15 codes
- Comparison of terminology versions [5]
  - Integration into the TOP Framework
- Terminology updates may affect models!

# Semantic Document Search



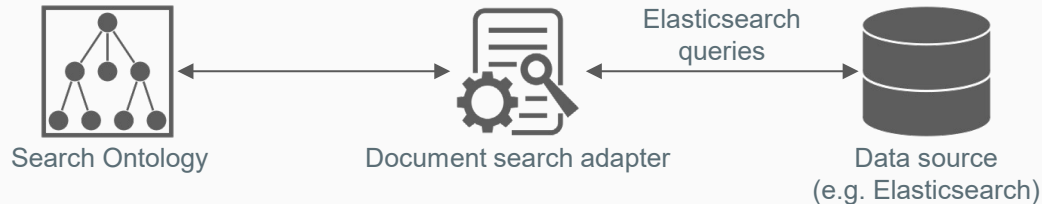
- Ontology-based modelling of text search queries with concept classes

Two classes of search concepts:

Single concepts

representing **sets of search terms**, organised **hierarchically**

Fall, Intracranial injury



# Semantic Document Search

TOP Framework

Pre-processing Cluster review

Document search

Data of query: Fall sea

Concepts

Selection mode: Union

Important nodes only: No

Records per page: 5 1-5 of 11

Popovic

Dear colleague,

We are reporting on our mutual patient Irina Popovic, born on 17.08.1958, who was treated on our trauma surgery ward 4 from 11.01.-14.01.2026.

Medical history

After a fall on the left side of the body in the dining room of the nursing home. Stat. admission by emergency doctor. On admission, severe pain on exertion and pain on movement as well as immobility of the left shoulder. She also complained of pain on the left side of her body (elbow, ribs, hip, thigh and lower leg).

Diagnoses:

- \* Known art. **hypertension**
- \* Known Parkinson's disease
- \* Known osteoporosis
- \* Currently proximal humeral head fracture on the left, abrasions and hematomas on the left side of the body, slight skull contusion

Physical examination

Patient with slightly reduced AZ and good EZ.

Vesicular breathing on both sides, **abdominal wall** compressible, DG present, no defensive tension, no DS, no KS renal bed. Liver and spleen palpable.

Pupils moderately dilated, positive light reaction, **neurological examination** without problems.

Local findings: Painful limitation of mobility in the affected joint. **Pressure** pain at the humeral head, incipient bruising in the armpit, laterally on the thoracic wall and on the medial side of the upper arm. DMS peripherally intact. Hematomas and minor abrasions on elbow, hip, knee joint and outer ankle. Bump over the left zygomatic bone.

X-ray in 2 planes:

Extra-articular, unifocal (proximal) fracture of the humeral head on the left.

Ultrasound examination showed no evidence of damage to the rotator cuff.

CT scan of the skull: 0.B.

- **Concept clusters** are the actually realized concept graphs in the TOP Framework (Graph DB in background)
  - Queries can be enhanced with graph algorithms
- Curation of concept clusters
  - Clusters are created **completely automatically**
  - Reasonable clusters are selected manually



**Search ontology and concept clusters can be combined or used individually to search for documents.**

# Conclusion

- We addressed challenges in standardising phenotype models for building a next-generation phenotype repository.
- The TOP Framework, built upon the COP and TOP API, empowers researchers with **standardised models** and facilitates **efficient data exploration and analysis**.
- It can be **integrated into the MII infrastructure** for streamlined data access.
- Real-world applications demonstrate the **potential to support clinical research** by enhancing efficiency, collaboration, and knowledge discovery.

# Publications

- Uciteli A, Beger C, Kirsten T, Meineke FA, Herre H. Ontological representation, classification and data-driven computing of phenotypes. J Biomed Semant **2020**;11:15. <https://doi.org/10.1186/s13326-020-00230-0>.
- Uciteli A, Beger C, Wagner J, Kirsten T, Meineke FA, Stäubert S, et al. Ontological modelling and FHIR Search based representation of basic eligibility criteria. GMS Medizinische Informatik **2021**:Biometrie und Epidemiologie; 17(2):Doc05. <https://doi.org/10.3205/MIBE000219>.
- Beger C, Matthies F, Schäfermeier R, Kirsten T, Herre H, Uciteli A. Towards an Ontology-Based Phenotypic Query Model. Applied Sciences **2022**;12:5214. <https://doi.org/10.3390/app12105214>.
- Matthies F, Beger C, Schäfermeier R, Uciteli A. Concept Graphs: A Novel Approach for Textual Analysis of Medical Documents. In: Röhrig R, Grabe N, Haag M, Hübner U, Sax U, Oliver Schmidt C, et al., editors. Studies in Health Technology and Informatics, IOS Press; **2023**. <https://doi.org/10.3233/SHTI230710>.
- Beger C, Matthies F, Schäfermeier R, Uciteli A. Model-driven execution of phenotype algorithms – introduction of the Terminology- and Ontology-based Phenotyping Framework. GMS Medizinische Informatik, Biometrie Und Epidemiologie **2023**;19. <https://doi.org/10.3205/MIBE000256>.
- Beger C, Boehmer AM, Mussawy B, Redeker L, Matthies F, Schäfermeier R, et al. Modelling Adverse Events with the TOP Phenotyping Framework. In: Röhrig R, Grabe N, Haag M, Hübner U, Sax U, Oliver Schmidt C, et al., editors. Studies in Health Technology and Informatics, IOS Press; **2023**. <https://doi.org/10.3233/SHTI230695>.
- Matthies F, Beger C, Schäfermeier R, Höffner K, Uciteli A. Extending the TOP Framework with an Ontology-Based Text Search Component. In: Röhrig R, Grabe N, Hübner UH, Jung K, Sax U, Schmidt CO, et al., editors. Studies in Health Technology and Informatics, IOS Press; **2024**. <https://doi.org/10.3233/SHTI240854>.
- Beger C, Eberl M, Dietz Y, Matthies F, Schäfermeier R, Höffner K, et al. Improving the LIFE Research Data Request Workflow with the TOP Phenotyping Framework. Gesundheit – Gemeinsam Kooperationstagung Der Deutschen Gesellschaft Für Medizinische Informatik, Biometrie Und Epidemiologie (GMDS), Deutschen Gesellschaft Für Sozialmedizin Und Prävention (DGSMP), Deutschen Gesellschaft Für Epidemiologie (DGEpi), Deutschen Gesellschaft Für Medizinische Soziologie (DGMS) Und Der Deutschen Gesellschaft Für Public Health (DGPH) **2024**. <https://doi.org/10.3205/24GMDS114>.
- Schäfermeier R, Beger C, Matthies F, Hoeffner K, Uciteli A. Tracking Changes for Inter-Version Interoperability in Heterogeneous Evolving Medical Terminologies. In: Röhrig R, Grabe N, Hübner UH, Jung K, Sax U, Schmidt CO, et al., editors. Studies in Health Technology and Informatics, IOS Press; **2024**. <https://doi.org/10.3233/SHTI240855>.

# Christoph Beger

[christoph.beger@imise.uni-leipzig.de](mailto:christoph.beger@imise.uni-leipzig.de)

<https://github.com/ChristophB>

Institute for Medical Informatics,  
Statistics and Epidemiology  
Leipzig University

Härtelstraße 16-18  
04107 Leipzig



UNIVERSITÄT  
LEIPZIG

Medizinische Fakultät

imise.



GEFÖRDET VOM

Bundesministerium  
für Bildung  
und Forschung



A. Uciteli



C. Beger



F. Matthies



R. Schäfermeier



K. Höffner

Photographer: Swen Reichhold

## Bibliography

1. Mo H, Thompson W K, Rasmussen L V, Pacheco J A, Jiang G, Kiefer R, et al. Desiderata for Computable Representations of Electronic Health Records-Driven Phenotype Algorithms. *Journal of the American Medical Informatics Association* 2015;22:1220–30. <https://doi.org/10.1093/jamia/ocv112>
2. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. *Summit Transl Bioinform* 2009;2009:116–20.
3. Beger C, Matthies F, Schäfermeier R, Kirsten T, Herre H, Uciteli A. Towards an Ontology-Based Phenotypic Query Model. *Applied Sciences* 2022;12:5214–5. <https://doi.org/10.3390/app12105214>
4. Beger C, Matthies F, Schäfermeier R, Uciteli A. Model-Driven Execution of Phenotype Algorithms – Introduction of the Terminology- and Ontology-based Phenotyping Framework. *GMS Medizinische Informatik, Biometrie Und Epidemiologie* 2023;19. <https://doi.org/10.3205/MIBE000256>
5. Schäfermeier R, Beger C, Matthies F, Hoeffner K, Uciteli A. Tracking Changes for Inter-Version Interoperability in Heterogeneous Evolving Medical Terminologies. In: Röhrig R, Grabe N, Hübner UH, Jung K, Sax U, Schmidt CO, et al., editors. *Studies in Health Technology and Informatics*, IOS Press; 2024. <https://doi.org/10.3233/SHTI240855>.