Using an ontological representation of chemotherapy toxicities for guiding information extraction and integration from EHRs

Alice Rogier, Adrien Coulet, Bastien Rance

Équipe HeKa/Équipe22, Inserm, Inria, Centre de Recherche des Cordeliers, Université de Paris et Hopital Européen Georges Pompidou AP-HP

Context

To improve and adapt chemotherapy regimen, toxicity and treatment response must be monitored and analysed. Both toxicity and treatment response are complex events that are difficult to detect automatically from Electronic Health Records (EHRs). However their automatic detection would be of great help for the development of clinical decision support systems. This is the challenge of my PhD, under the supervision of AC and BR, and jointly funded by Inserm and Inria. Here we present our first work, the development of OntoTox, an ontology dedicated to chemotherapy toxicity extractions. [1]

Chemotherapy toxicities in EHRs

Chemotherapies against cancers are often interrupted due to severe drug toxicities, reducing treatment opportunities. For this reason, the detection of toxicities and their severity from EHRs is of importance for many downstream applications.
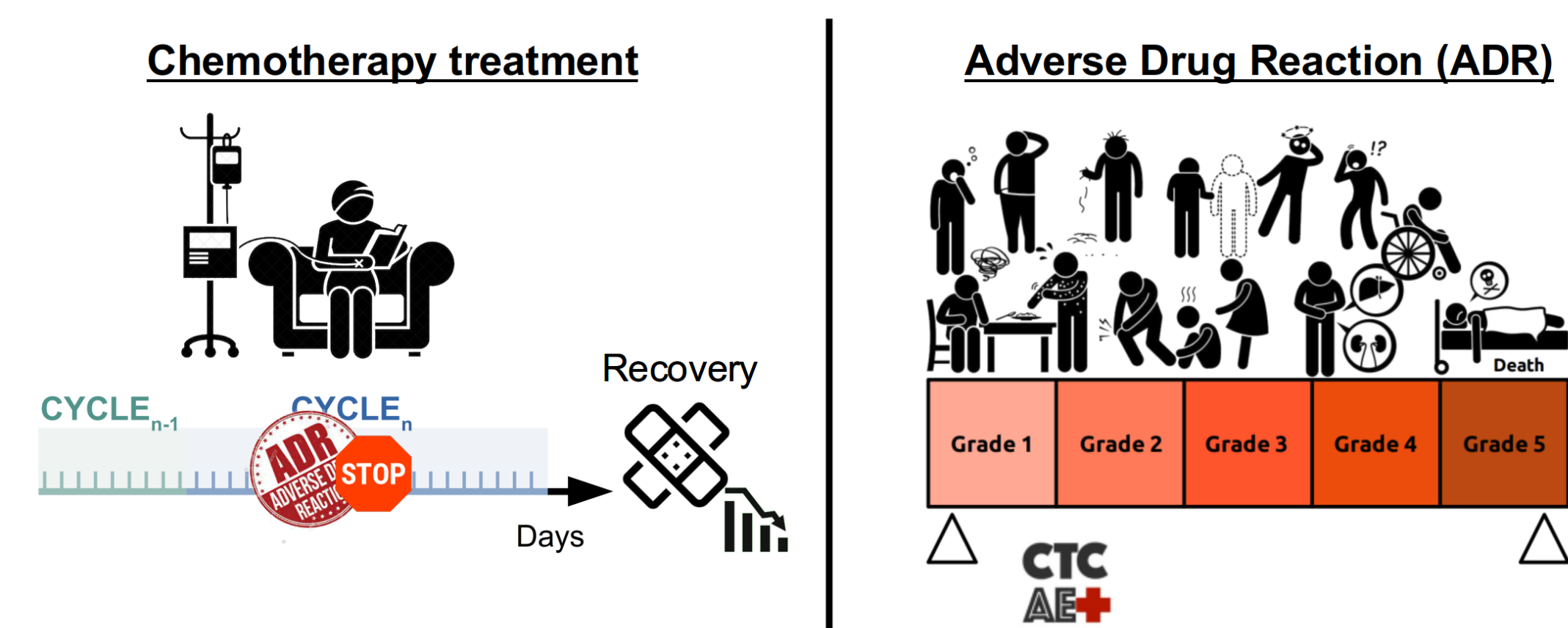

Fig. 1: Grading chemotherapy toxicity severity

However toxicity information is dispersed in various sources in the EHRs from Clinical DataWarehouses (CDW), making its extraction challenging.
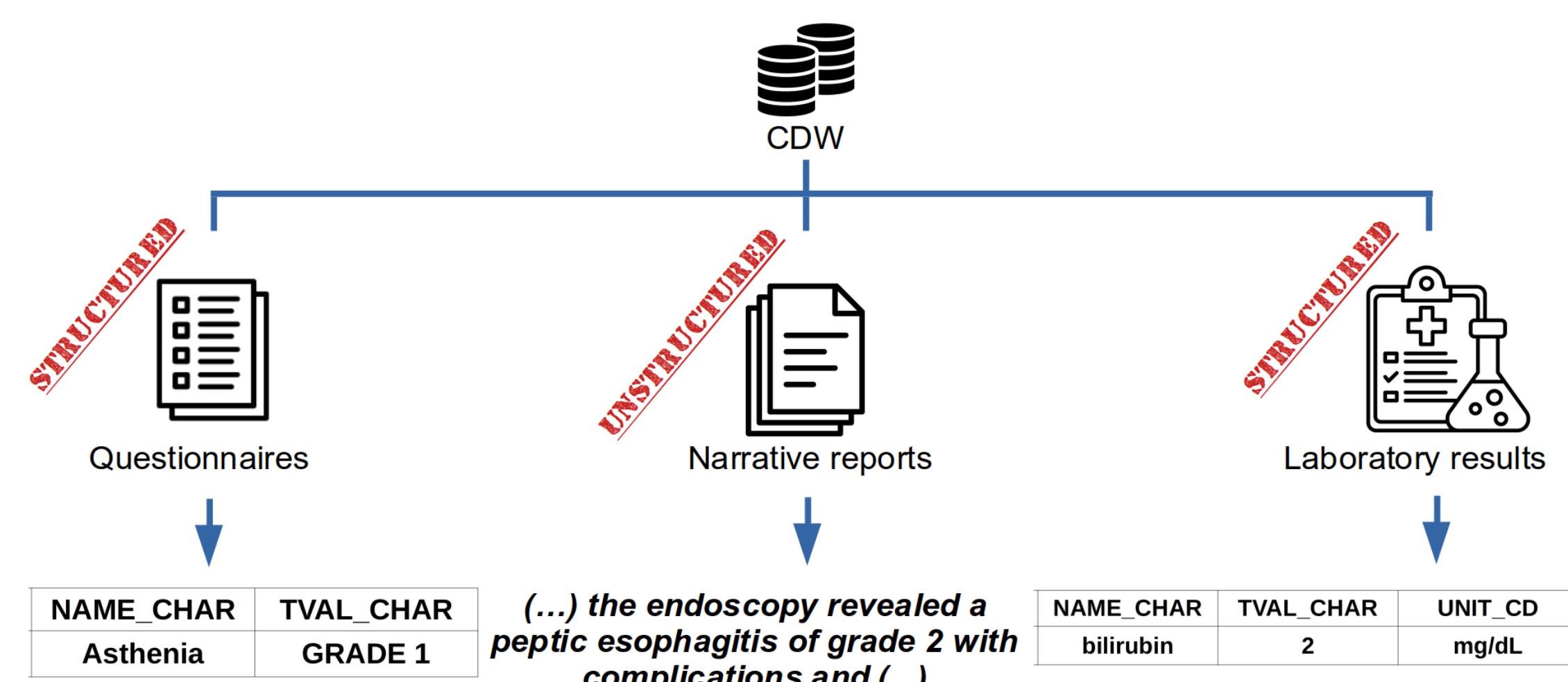

Fig. 2: Chemotherapy toxicity events in EHRs

alice.rogier@inserm.fr

Methods

We introduce OntoTox, an ontology designed to represent chemotherapy toxicities, its attributes and provenance.
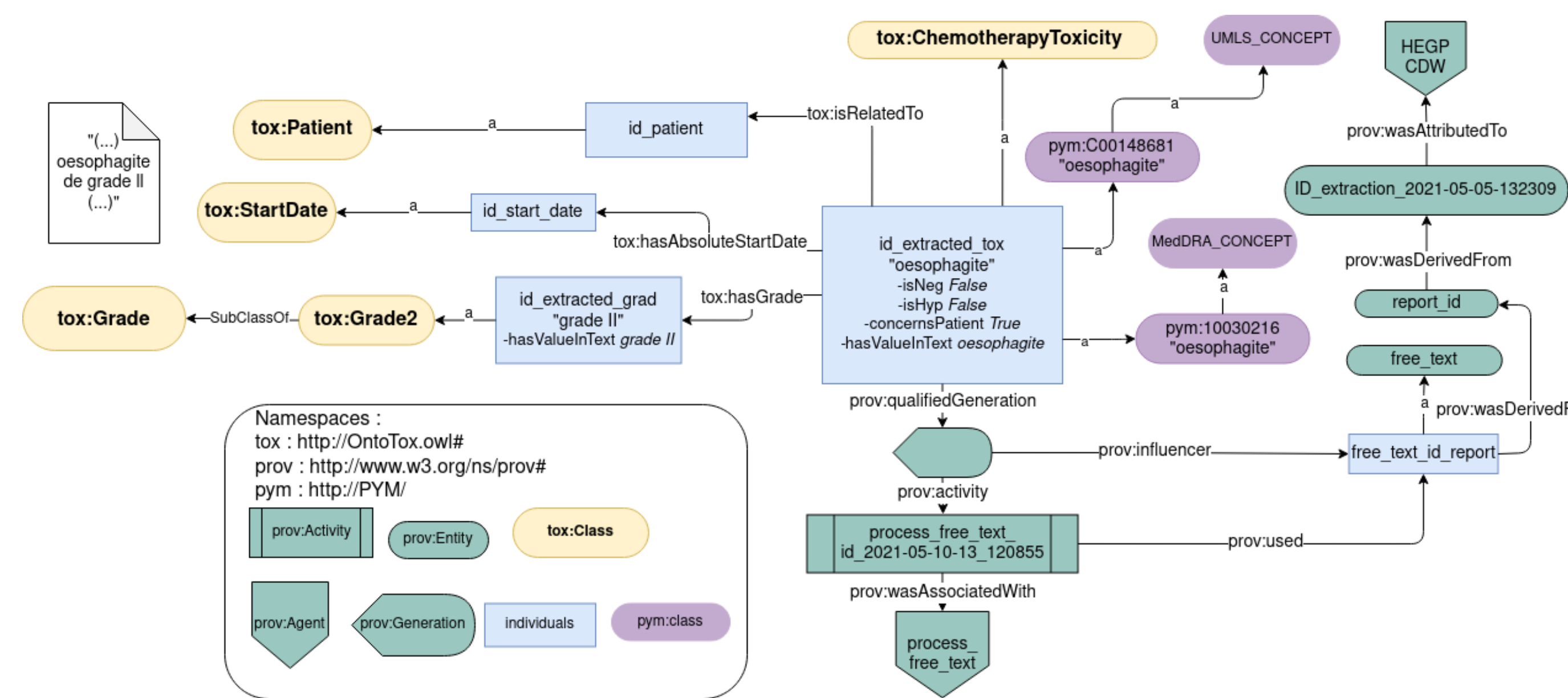

Fig. 3: OntoTox structure

We illustrated the interest of OntoTox by integrating toxicities and grading information extracted from three heterogeneous sources: EHR questionnaires, semi-structured tables, and free-text.
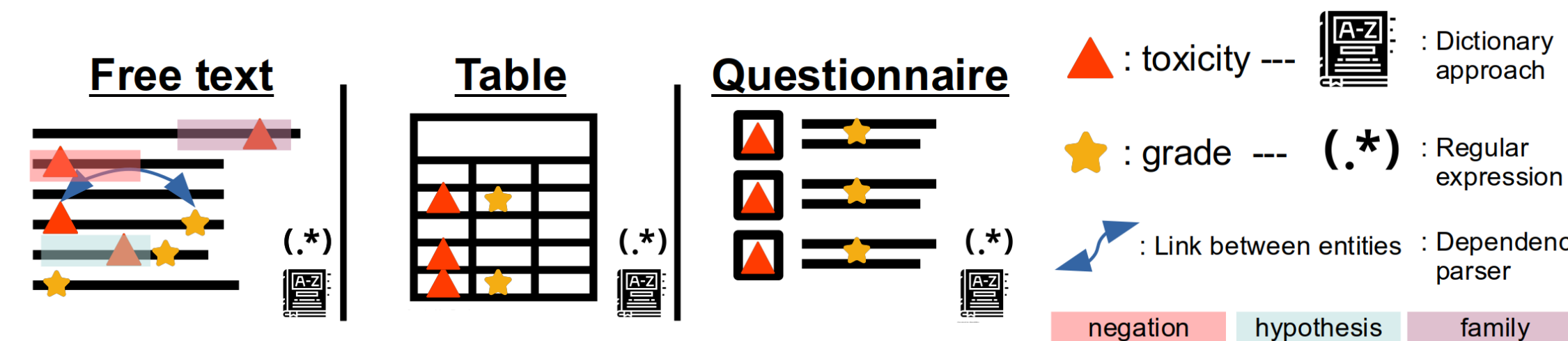

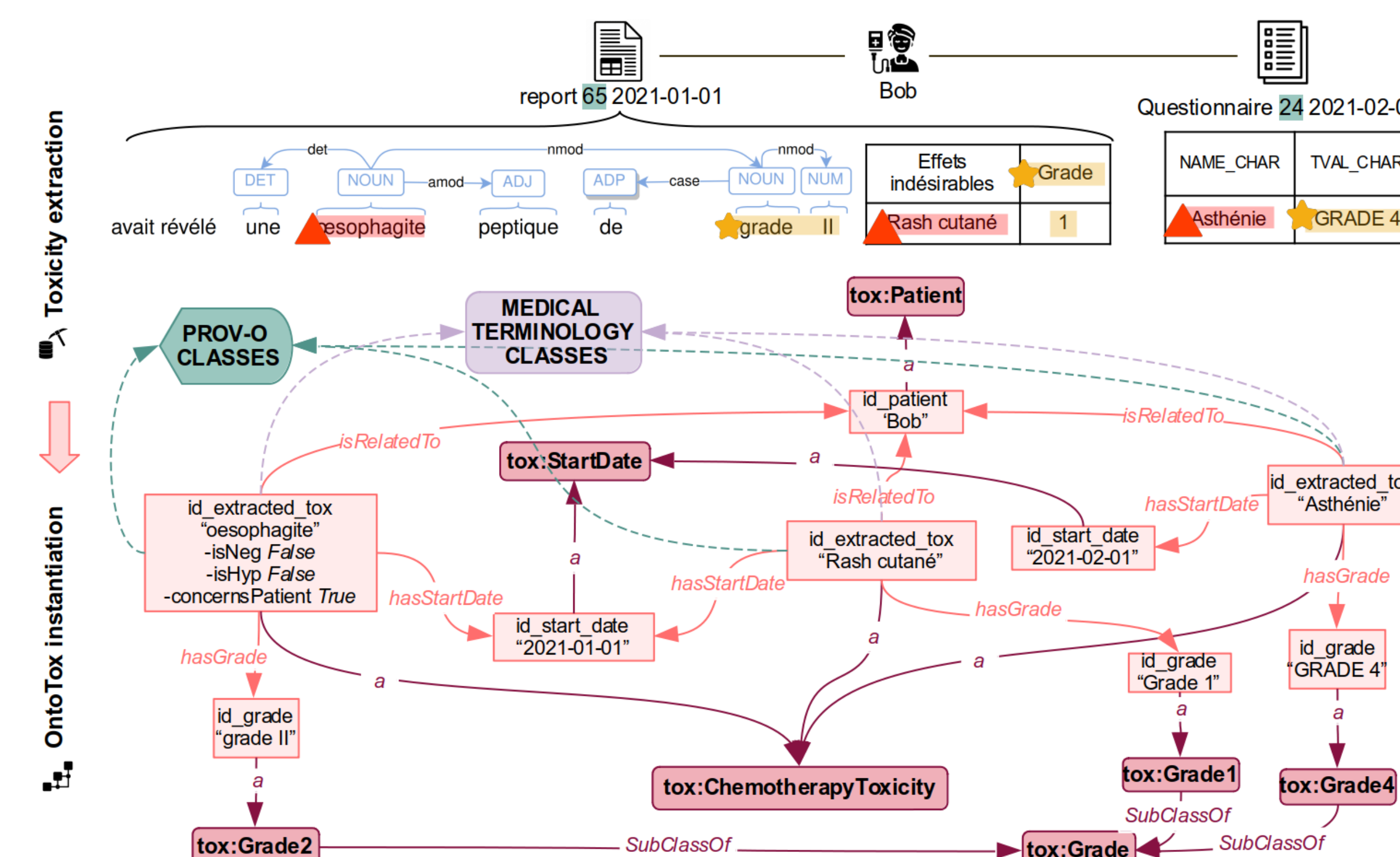Fig. 4: Toxicity information extraction processes

Fig. 5: Toxicity extraction to OntoTox intantiation

Results

We demonstrated the interest of OntoTox in a clinical use-case to gather and compare information regarding chemotherapy toxicities found in three types of heterogeneous sources: structured clinical questionnaires, tables and free-text from narrative reports.
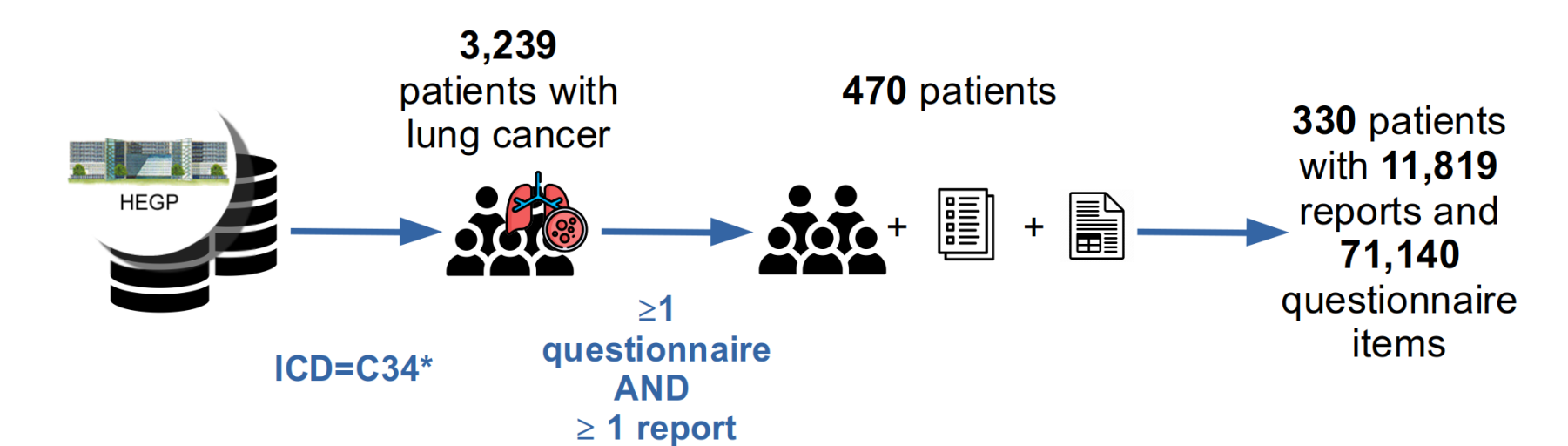

Fig. 6: Use-case

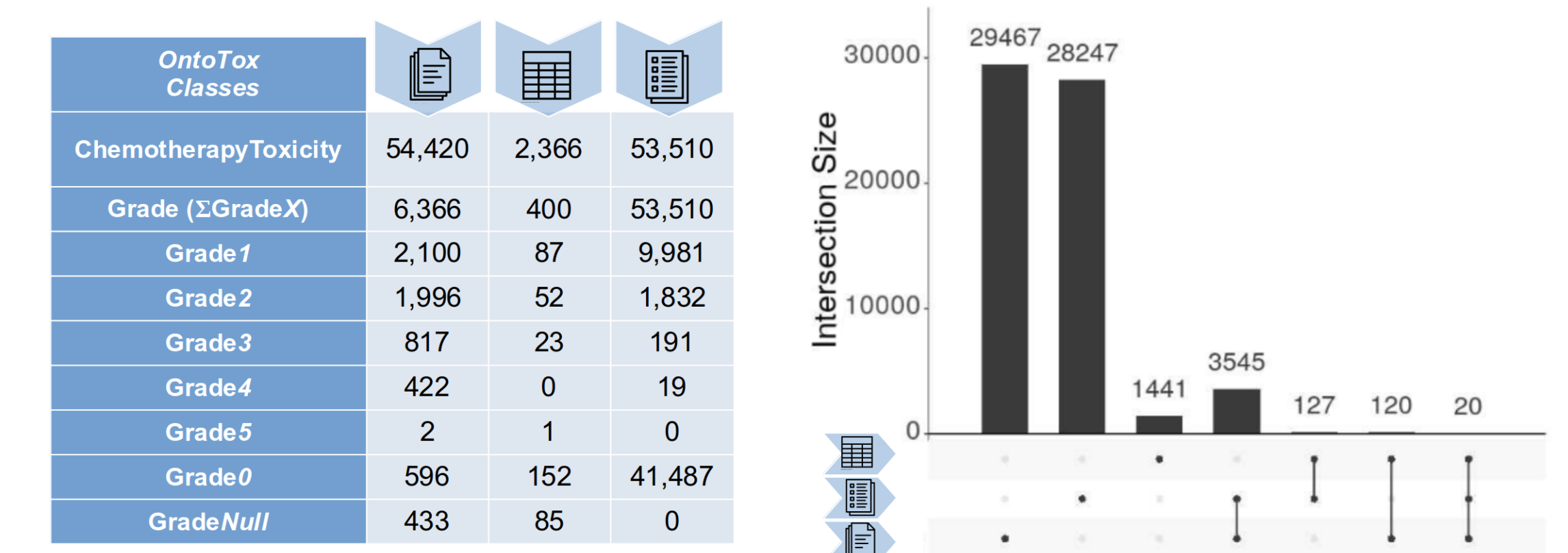| OntoTox Classes	📄	📊	📑
ChemotherapyToxicity	54,420	2,366	53,510
Grade (ΣGradeX)	6,366	400	53,510
Grade1	2,100	87	9,981
Grade2	1,996	52	1,832
Grade3	817	23	191
Grade4	422	0	19
Grade5	2	1	0
Grade0	596	152	41,487
GradeNull	433	85	0

Fig. 7: Number of OntoTox class instantiations (left) and Normalized toxicity per-patient and per-month intersection sets between the three sources (right)

Discussion

We illustrated with this preliminary study the potential of OntoTox to guide the integration of multiple sources, and identified that the three sources are only moderately overlapping, stressing the need for a common representation.
We identified some reasons for the information differences between the three sources, such as the fact that the start date is a metadata. Thus, OntoTox will be enriched with a better time representation.
We chose to create and instantiate an ontology rather than another data model to represent the field of chemotherapy toxicity. One reason for this choice is that we could easily link our ontology to other knowledge models, as PROV-O MedDRA and the UMLS. Furthermore, this enables the further use of a reasoner.

Conclusion

We showed that OntoTox can guide the integration of information from various data sources. OntoTox is rather small, but aim at being enriched to enable integrating a maximum of information qualifying chemotherapy toxicities and response that can be found in EHRs. OntoTox constitutes the seed of a valuable resource for oncology research and will further serve as a brick of a clinical decision support software.

[1] Alice Rogier, Adrien Coulet, and Bastien Rance. "Using an ontological representation of chemotherapy toxicities for guiding the information extraction and integration from EHRs". In: *MedInfo* (2021).