MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

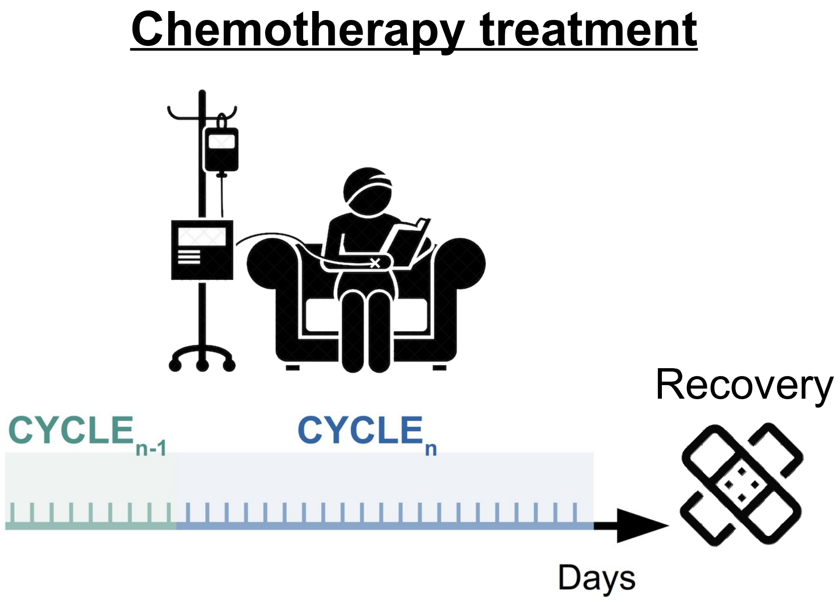
Using an ontological representation of chemotherapy toxicities for guiding information extraction and integration from Electronics Health Records (EHRs)

Alice ROGIER, Adrien COULET, Bastien RANCE

Inria – INVENTEURS DU MONDE NUMÉRIQUE
Inserm – La science pour la santé – From science to health
ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS
HÔPITAUX UNIVERSITAIRES PARIS OUEST – Hôpital européen Georges-Pompidou
Université de Paris

Context (1/2)

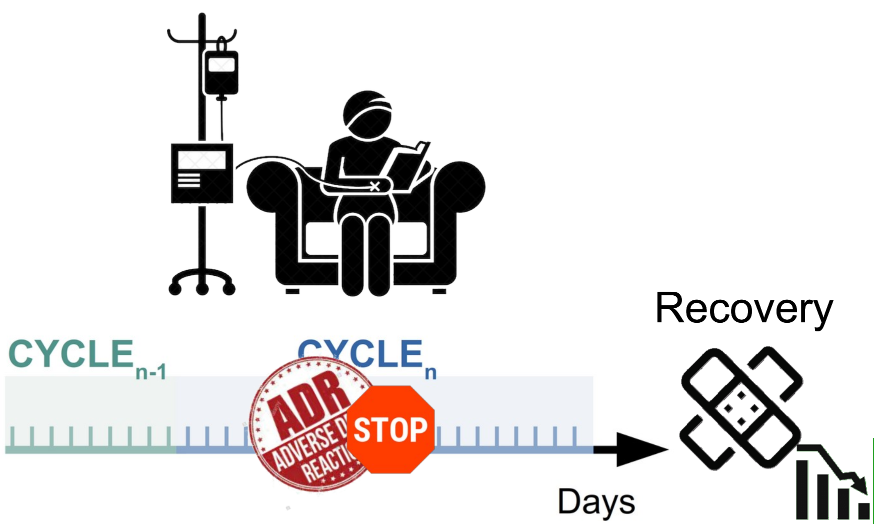
Adverse Drug Reaction (ADR) due to chemotherapy treatment



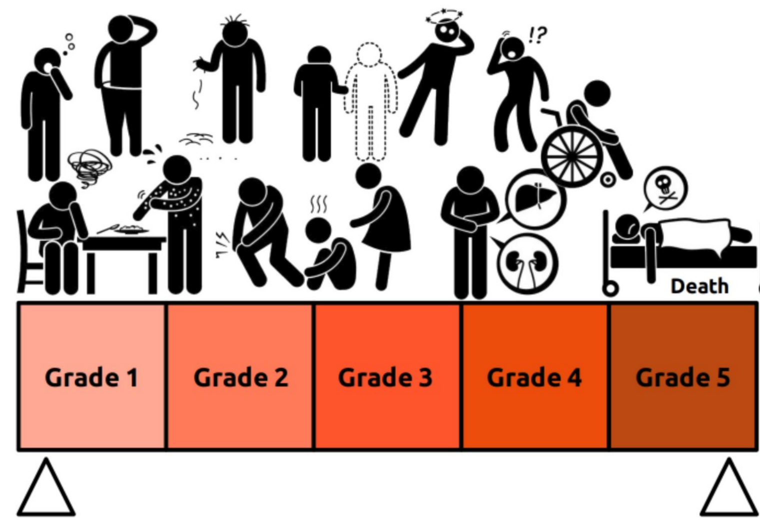
Context (1/2)

Adverse Drug Reaction (ADR) due to chemotherapy treatment

Chemotherapy treatment



Adverse Drug Reaction (ADR)

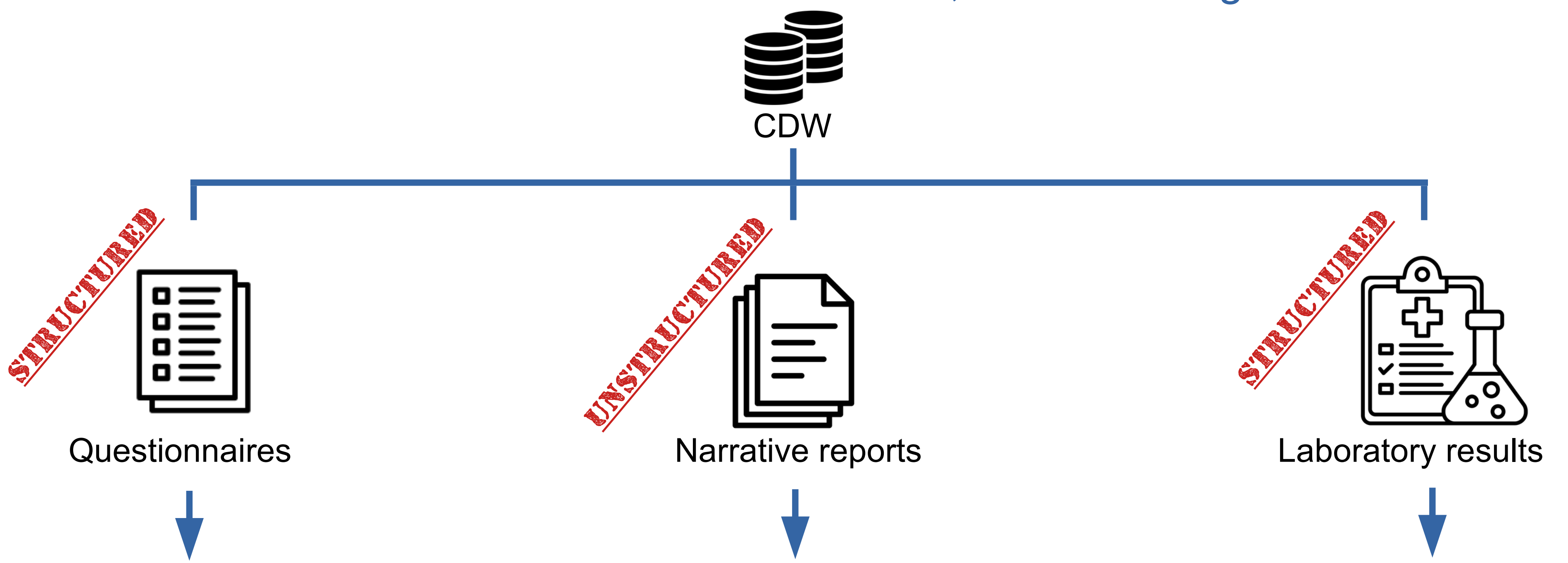


Gastrointestinal disorders					
CTCAE Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
Esophageal stenosis	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; altered GI function	Severely altered GI function; tube feeding or hospitalization indicated; elective operative intervention indicated	Life-threatening consequences; urgent operative intervention indicated	Death

Context (2/2)

Toxicity sources in Clinical Data Warehouses (CDW)

Information about toxicities is available in CDW, but in heterogeneous forms



NAME_CHAR	TVAL_CHAR
Asthenia	GRADE 1

(…) the endoscopy revealed a peptic esophagitis of grade 2 with complications and (…)

NAME_CHAR	TVAL_CHAR	UNIT_CD
bilirubin	2	mg/dL

Objective

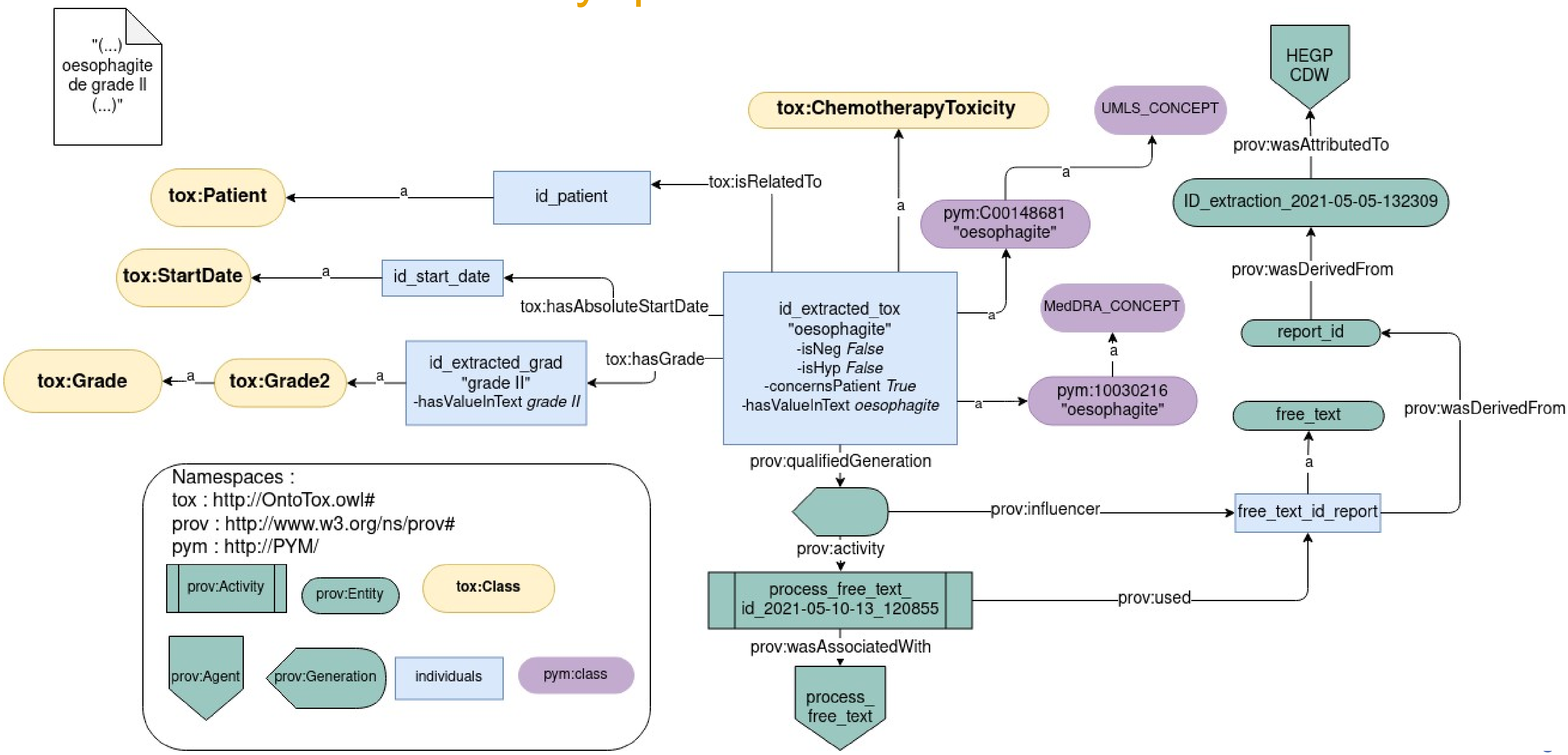
Integrating chemotherapy toxicities information in a **common data model**

Our contributions :

- Creation of OntoTox, an **ontology** for chemotherapy toxicities guided by data
- Toxicity **extractions** from different data sources
- Demonstration of OntoTox interest with a **clinical use case**

OntoTox structure (1/3)

OntoTox classes: toxicity qualifications

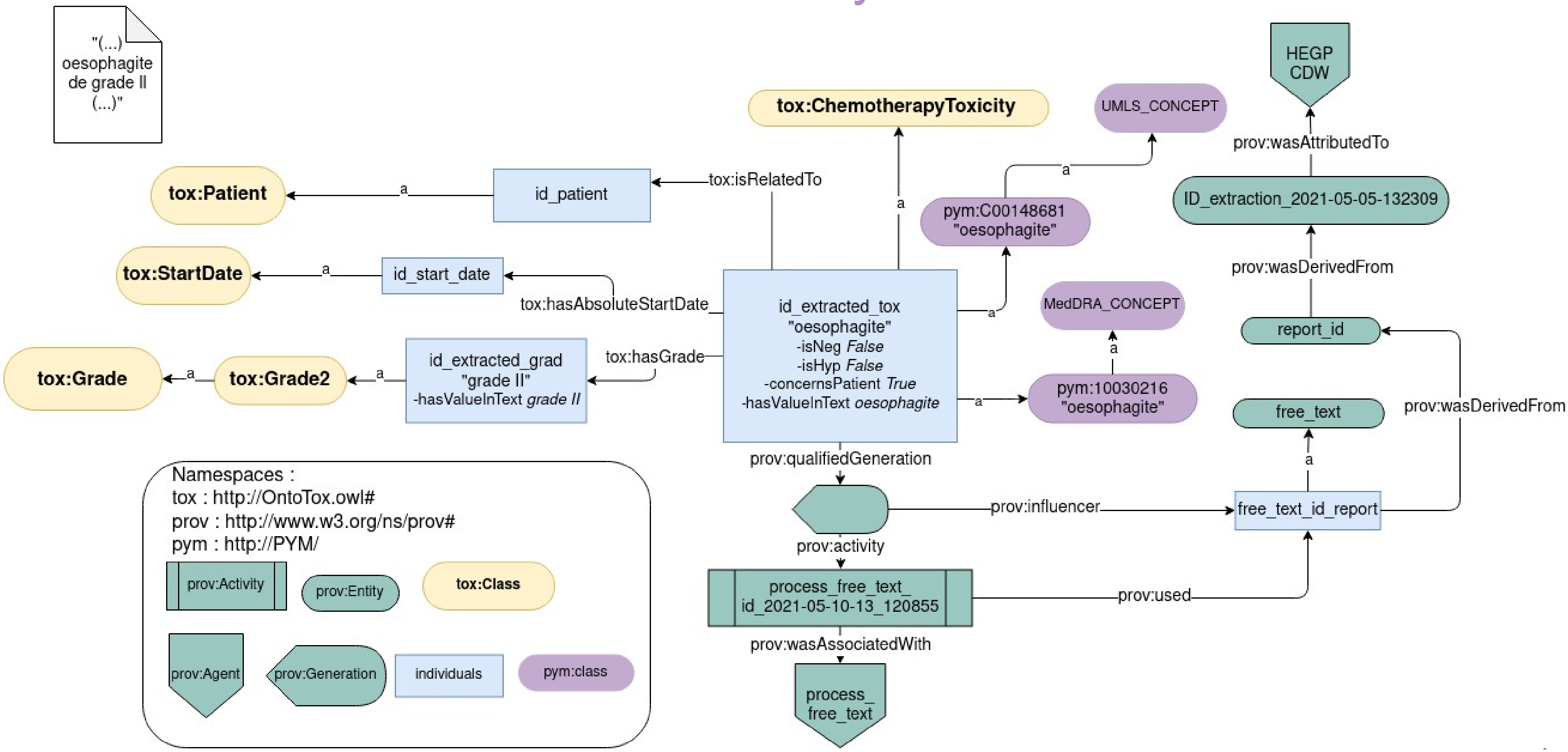


MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

IMIA

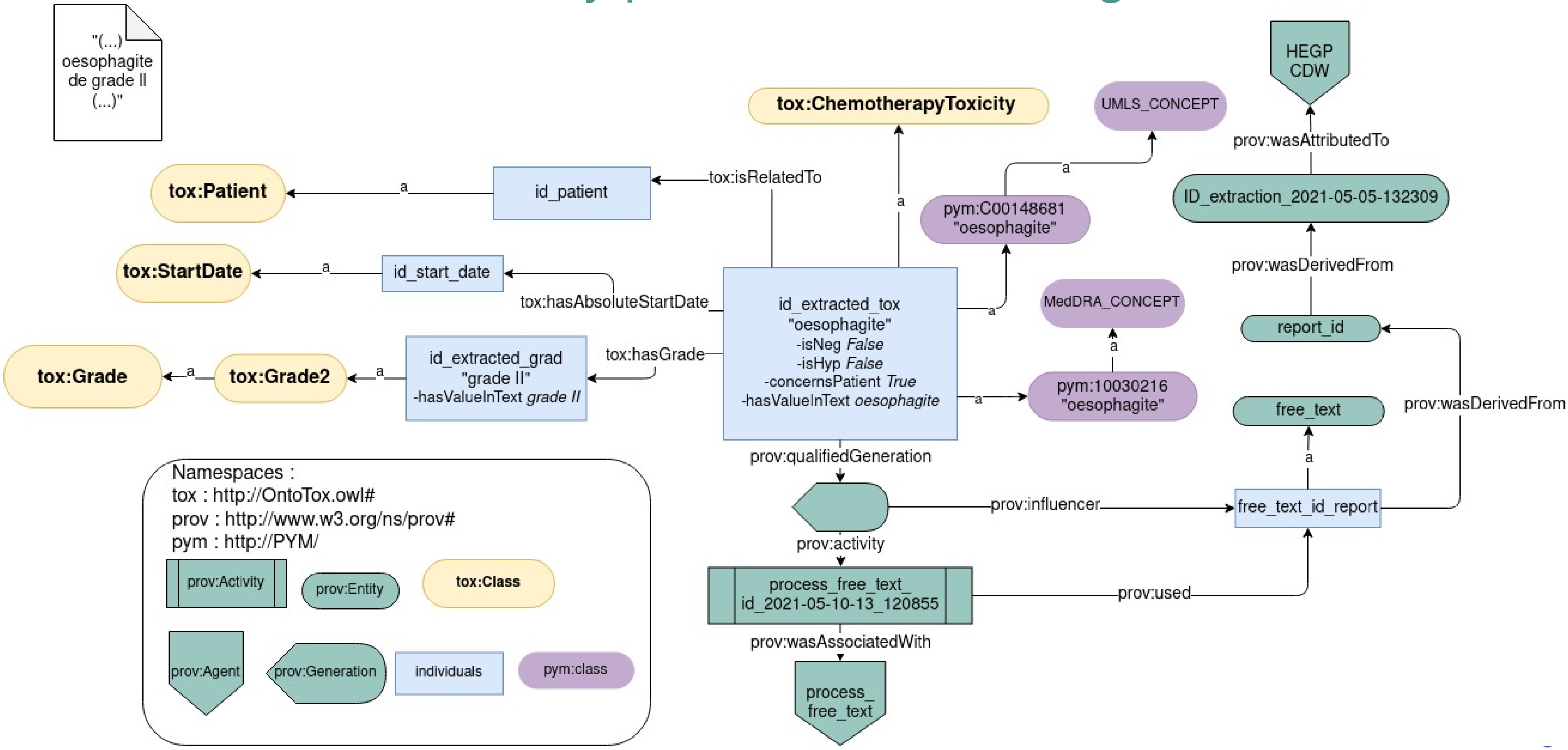
OntoTox structure (2/3)

MedDRA and UMLS classes: toxicity normalization



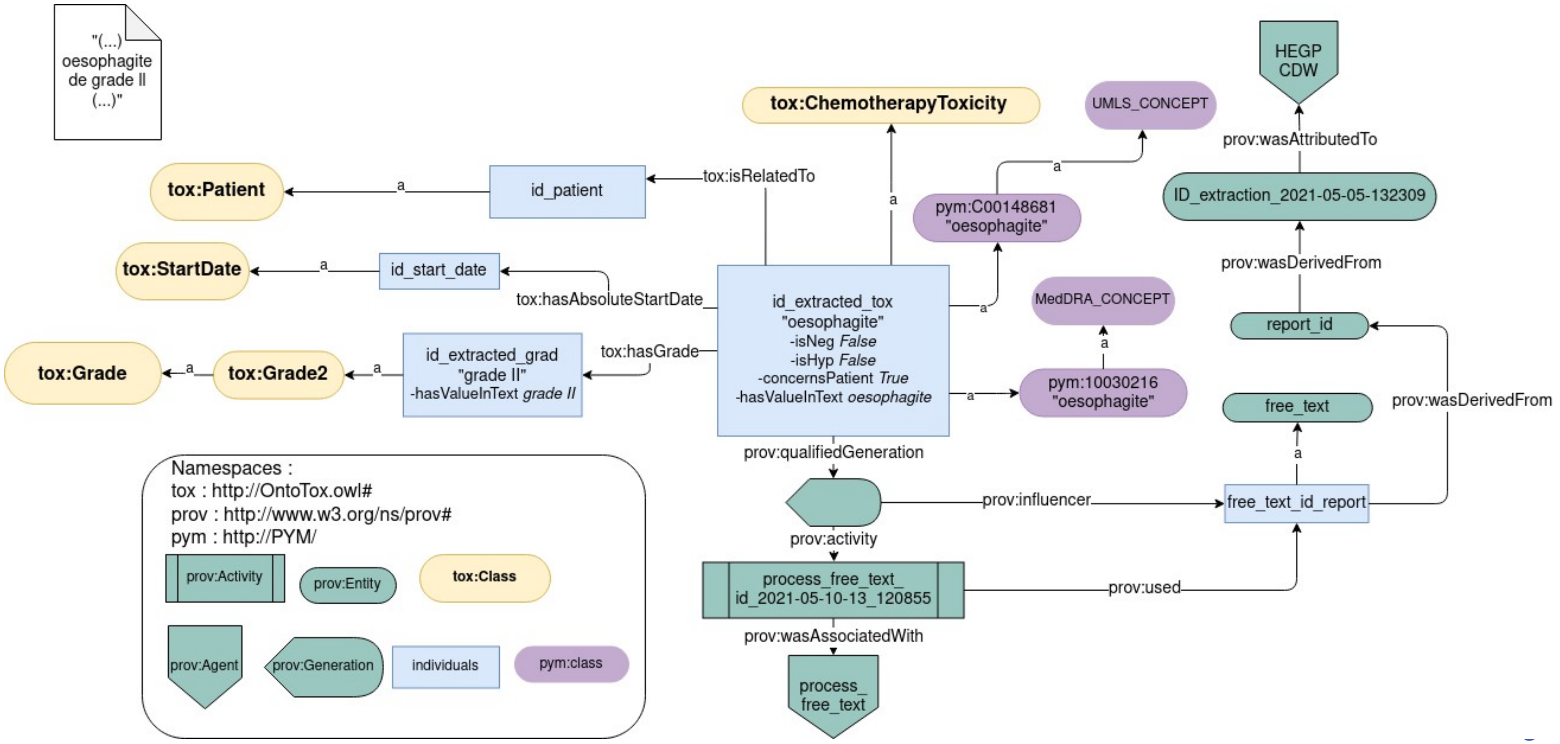
OntoTox structure (3/3)

PROV-O classes: toxicity provenance encoding

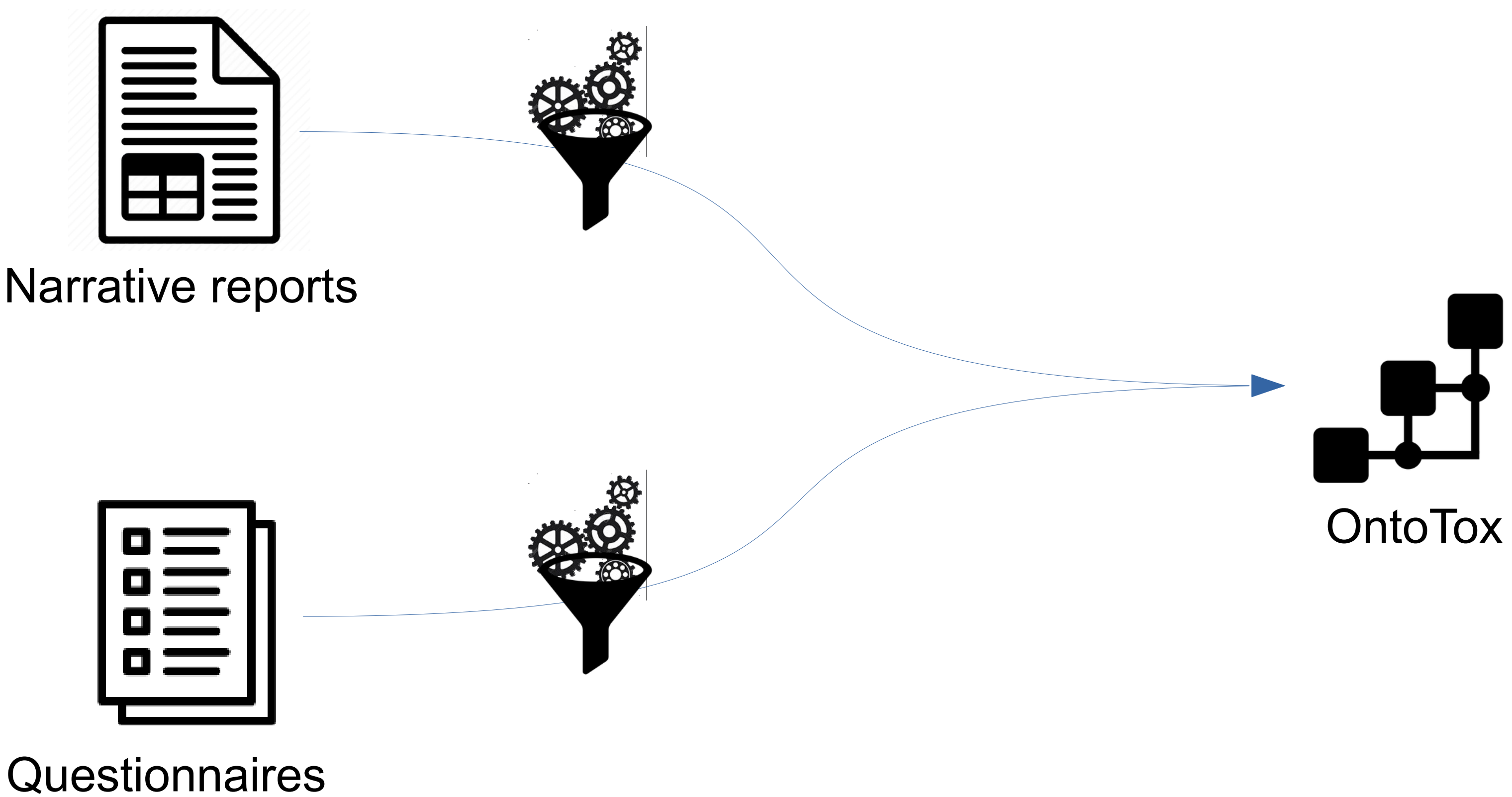


OntoTox implementation

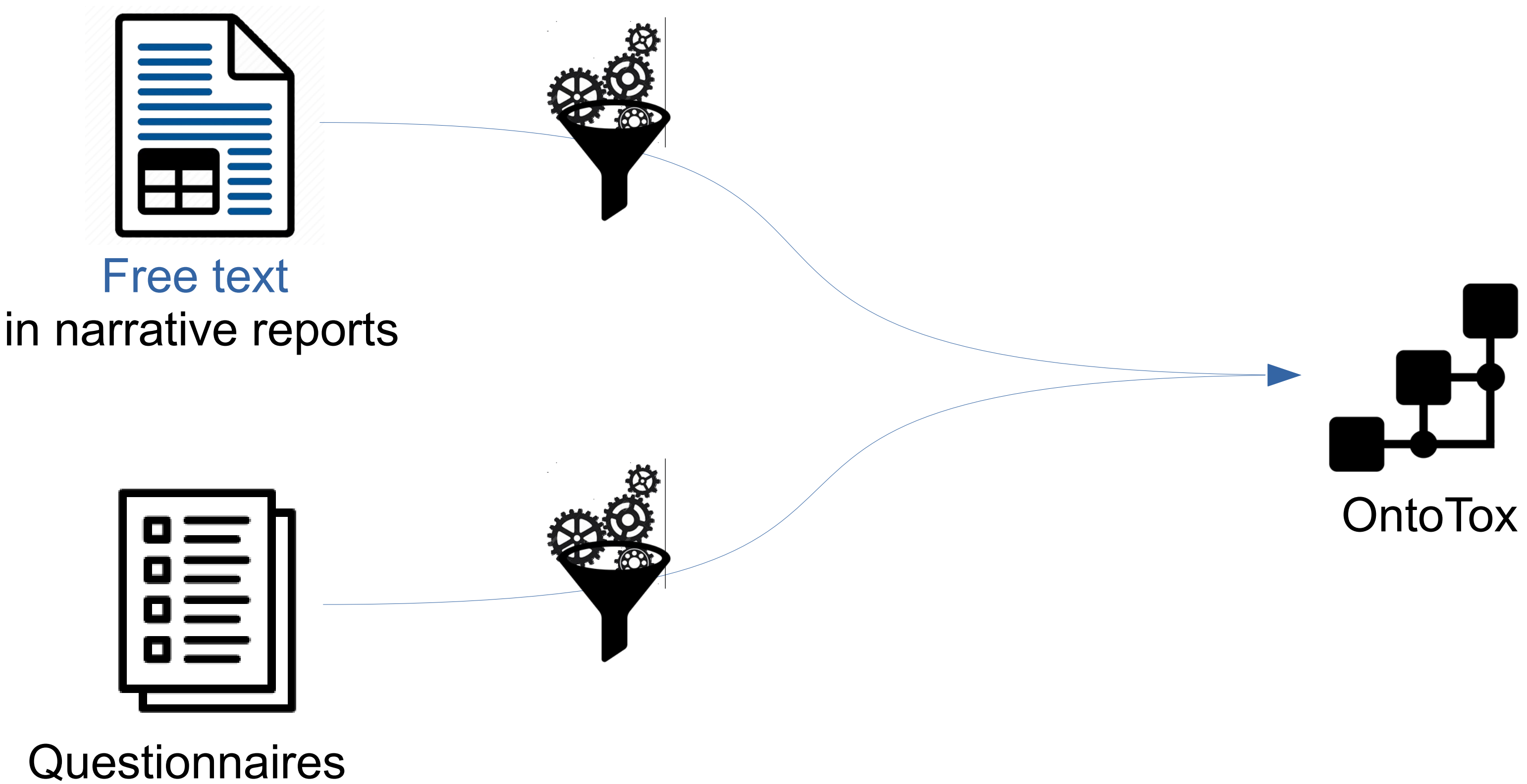
- Data driven
- Owlready, PyMedTermino
- Protégé



Extract and integrate chemotherapy toxicities in OntoTox



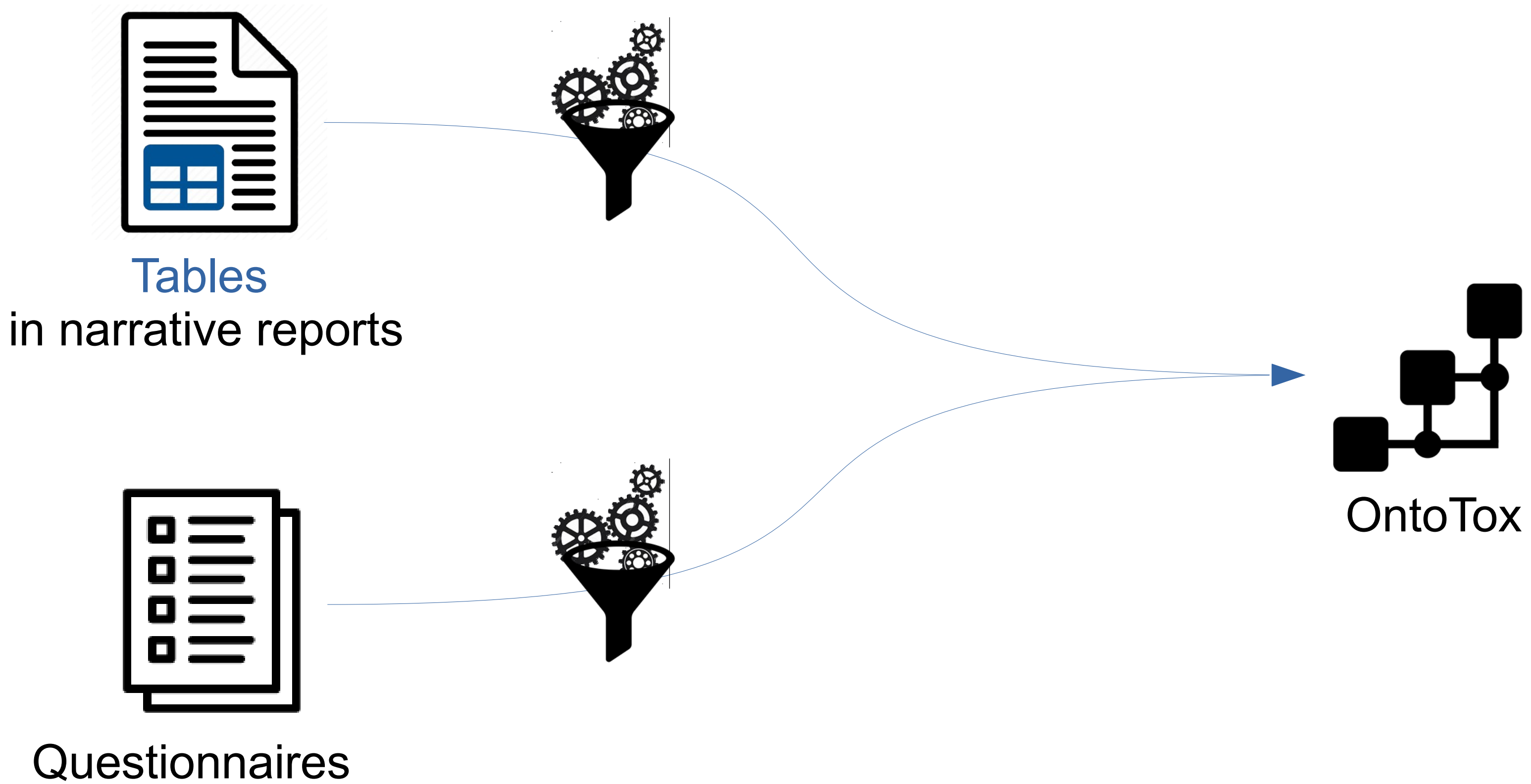
Extract and gather chemotherapy toxicities in OntoTox



MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

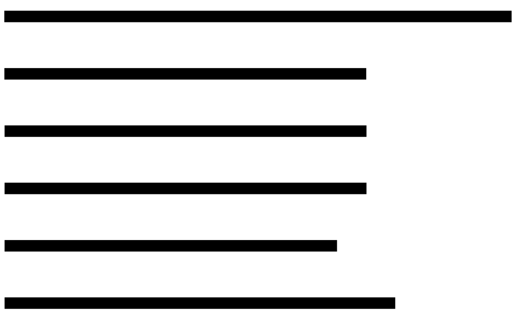
IMIA

Extract and gather chemotherapy toxicities in OntoTox

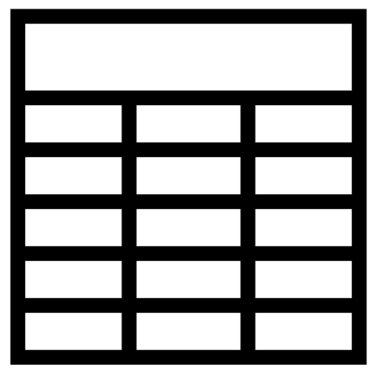


Toxicity extraction processes

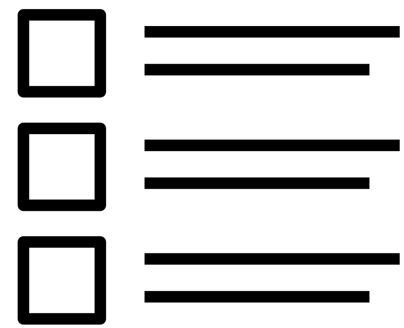
Free text



Table

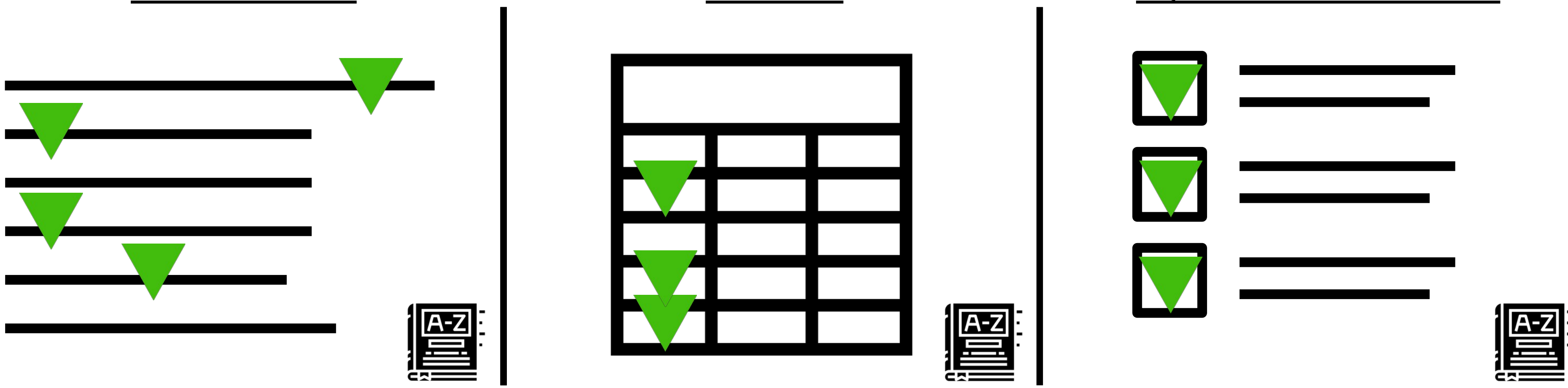


Questionnaire



Toxicity extraction processes

Free text | **Table** | **Questionnaire**

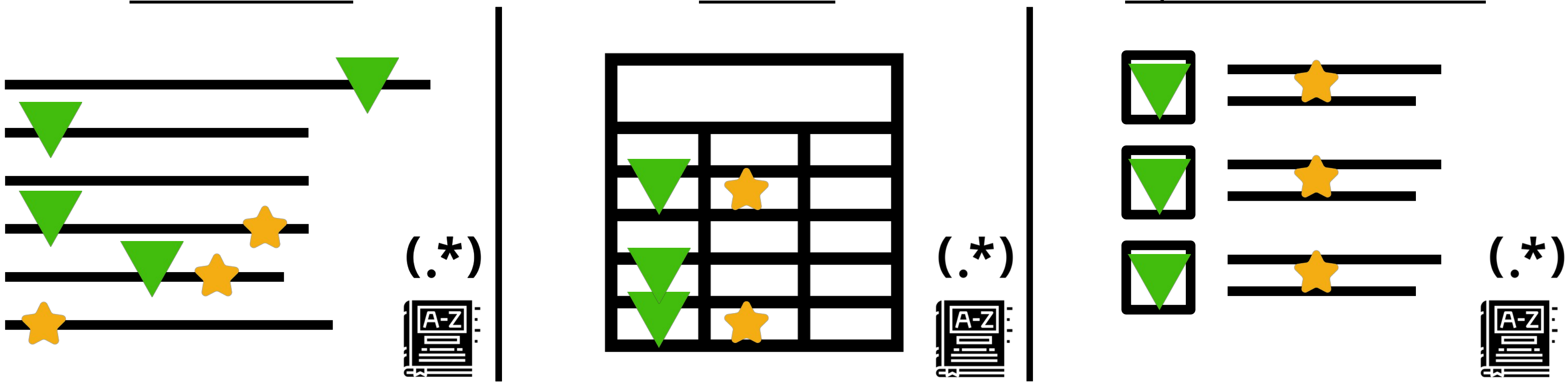


Entitiy recognition :

: toxicity ------- : 4,038 toxicity terms dictionary

Toxicity extraction processes

Free text | **Table** | **Questionnaire**



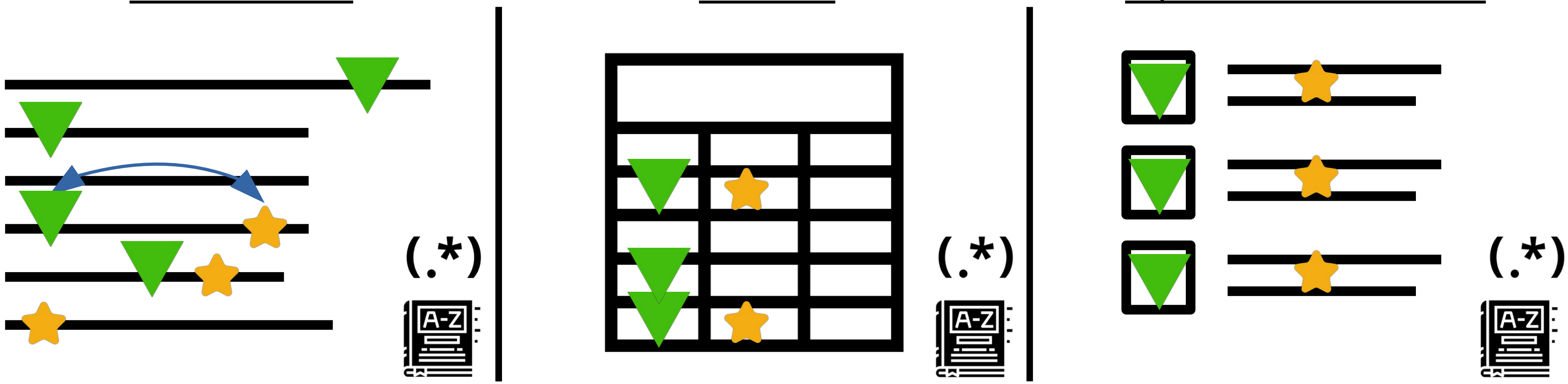
Entitiy recognition :

: toxicity ------- A-Z : 4,038 toxicity terms dictionary

: grade ------- (.*) : regular expression

Toxicity extraction processes

Free text | **Table** | **Questionnaire**



Entitiy recognition :

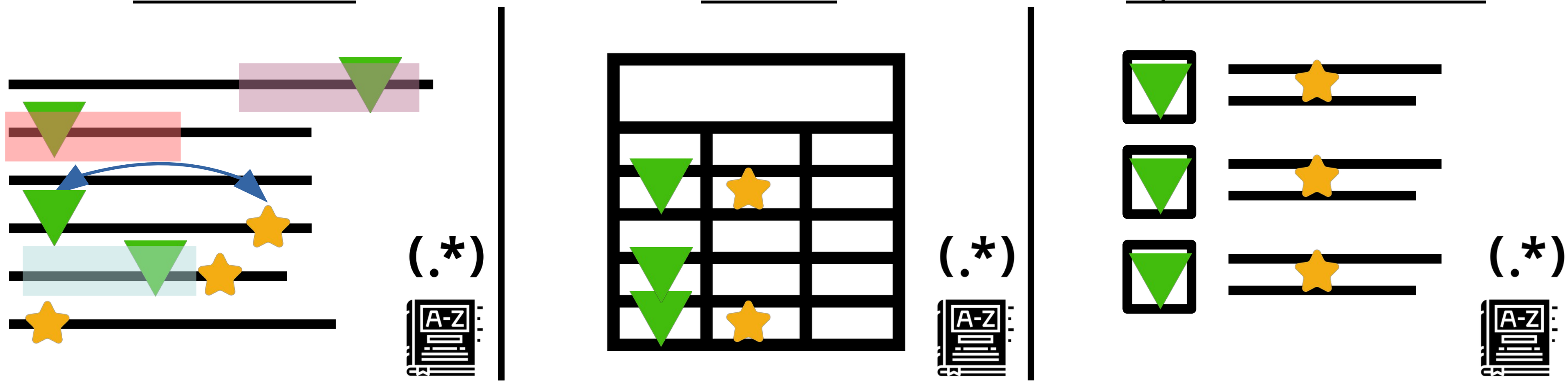
: toxicity ------- A-Z : 4,038 toxicity terms dictionary

: grade ------- (.*) : regular expression

: Link between entities with dependency parsing

Toxicity extraction processes

Free text | Table | Questionnaire



Entitiy recognition :

: toxicity ------- A-Z : 4,038 toxicity terms dictionary

: grade ------- (.*) : regular expression

Context detection:

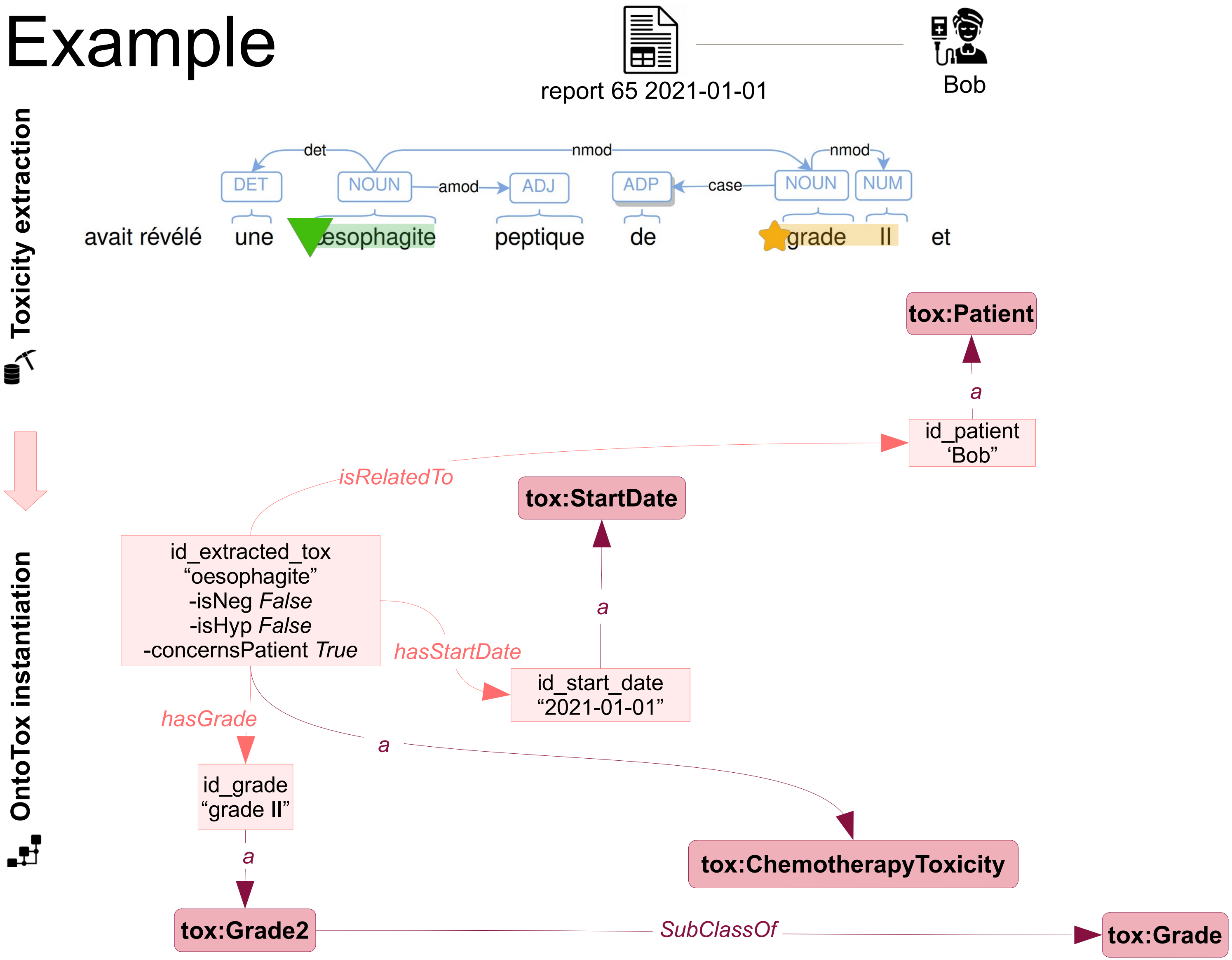
negation

hypothesis

family

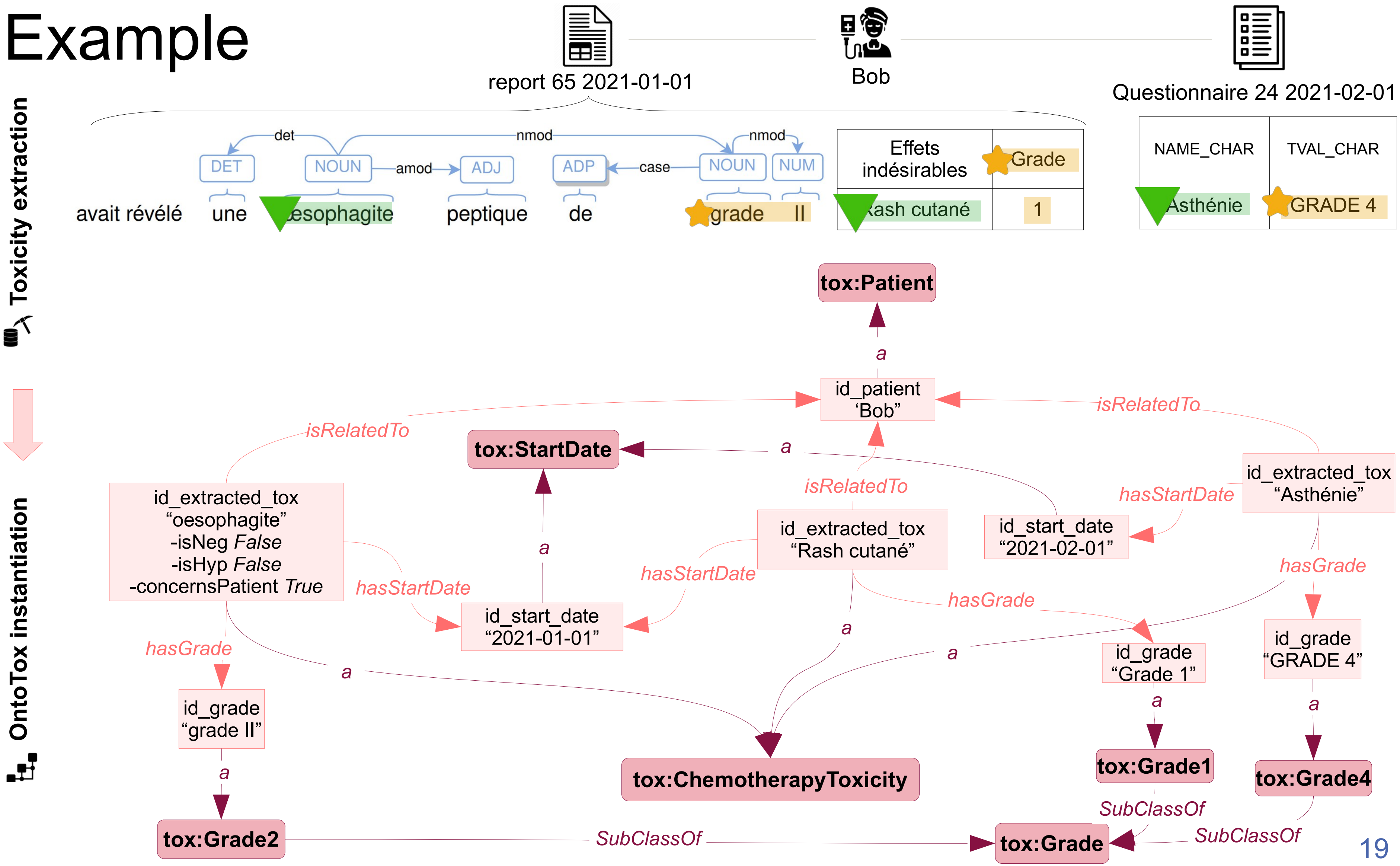
: Link between entities with dependency parsing

Example



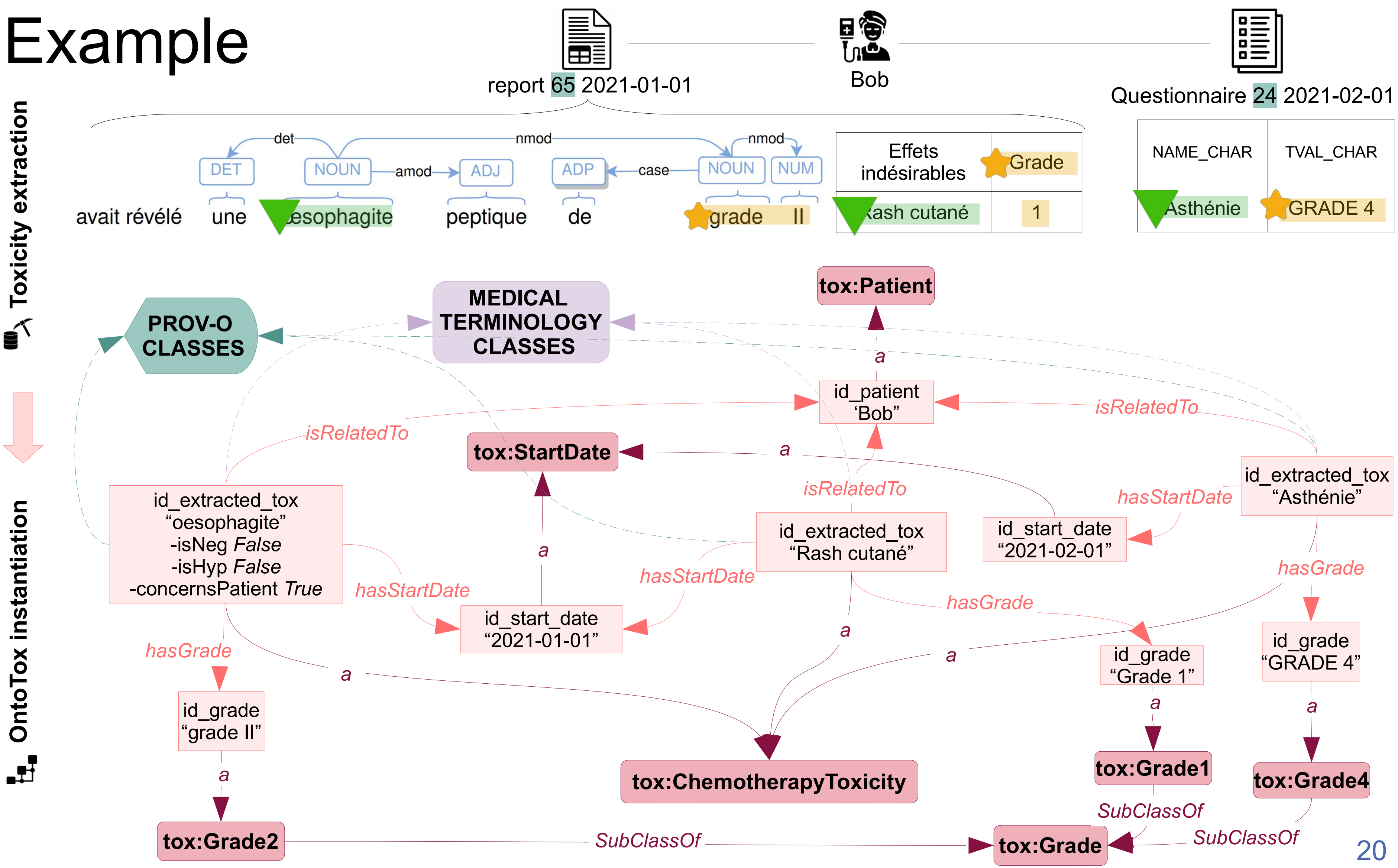
Toxicity extraction
OntoTox instantiation
report 65 2021-01-01
Bob
det
nmod
nmod
DET
NOUN
amod
ADJ
ADP
case
NOUN
NUM
avait révélé
une
oesophagite
peptique
de
grade
II
et
tox:Patient
a
id_patient
'Bob"
isRelatedTo
tox:StartDate
id_extracted_tox
"oesophagite"
-isNeg False
-isHyp False
-concernsPatient True
a
hasStartDate
id_start_date
"2021-01-01"
hasGrade
a
id_grade
"grade II"
a
tox:ChemotherapyToxicity
tox:Grade2
SubClassOf
tox:Grade

Example



report 65 2021-01-01
Bob
Questionnaire 24 2021-02-01
Toxicity extraction
OntoTox instantiation
det
nmod
nmod
DET
NOUN
amod
ADJ
ADP
case
NOUN
NUM
avait révélé
une
oesophagite
peptique
de
grade
II
Effets indésirables
Grade
Rash cutané
1
NAME_CHAR
TVAL_CHAR
Asthénie
GRADE 4
tox:Patient
a
id_patient
'Bob"
isRelatedTo
isRelatedTo
tox:StartDate
a
isRelatedTo
hasStartDate
id_extracted_tox
"Asthénie"
id_extracted_tox
"oesophagite"
-isNeg *False*
-isHyp *False*
-concernsPatient *True*
a
id_extracted_tox
"Rash cutané"
id_start_date
"2021-02-01"
hasGrade
hasStartDate
hasStartDate
hasGrade
id_start_date
"2021-01-01"
a
a
hasGrade
id_grade
"Grade 1"
id_grade
"GRADE 4"
a
id_grade
"grade II"
a
a
a
tox:ChemotherapyToxicity
tox:Grade1
tox:Grade4
SubClassOf
SubClassOf
tox:Grade2
SubClassOf
tox:Grade

Example



report 65 2021-01-01
Bob
Questionnaire 24 2021-02-01
Toxicity extraction
det
nmod
nmod
DET
NOUN
amod
ADJ
ADP
case
NOUN
NUM
avait révélé
une
oesophagite
peptique
de
grade
II
Effets indésirables
Grade
Rash cutané
1
NAME_CHAR
TVAL_CHAR
Asthénie
GRADE 4
OntoTox instantiation
PROV-O CLASSES
MEDICAL TERMINOLOGY CLASSES
tox:Patient
a
id_patient ‘Bob”
isRelatedTo
isRelatedTo
tox:StartDate
a
isRelatedTo
id_extracted_tox “Asthénie”
hasStartDate
id_extracted_tox “oesophagite” -isNeg False -isHyp False -concernsPatient True
a
id_start_date “2021-02-01”
id_extracted_tox “Rash cutané”
hasGrade
hasStartDate
hasStartDate
hasGrade
id_start_date “2021-01-01”
a
a
hasGrade
a
id_grade “Grade 1”
id_grade “GRADE 4”
id_grade “grade II”
a
a
a
tox:ChemotherapyToxicity
tox:Grade1
tox:Grade4
SubClassOf
SubClassOf
tox:Grade2
SubClassOf
tox:Grade

MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

Use case

HEGP

ICD=C34*

3,239 patients with lung cancer

≥1 questionnaire AND ≥ 1 report

+ +

470 patients

Random sample

330 patients with **11,819** reports and **71,140** questionnaire items

Results (1/2): What is extracted?

OntoTox Classes			
ChemotherapyToxicity	54,420	2,366	53,510
Grade (ΣGrade*X*)	6,366	400	53,510
Grade*1*	2,100	87	9,981
Grade*2*	1,996	52	1,832
Grade*3*	817	23	191
Grade*4*	422	0	19
Grade*5*	2	1	0
Grade*0*	596	152	41,487
Grade*Null*	433	85	0

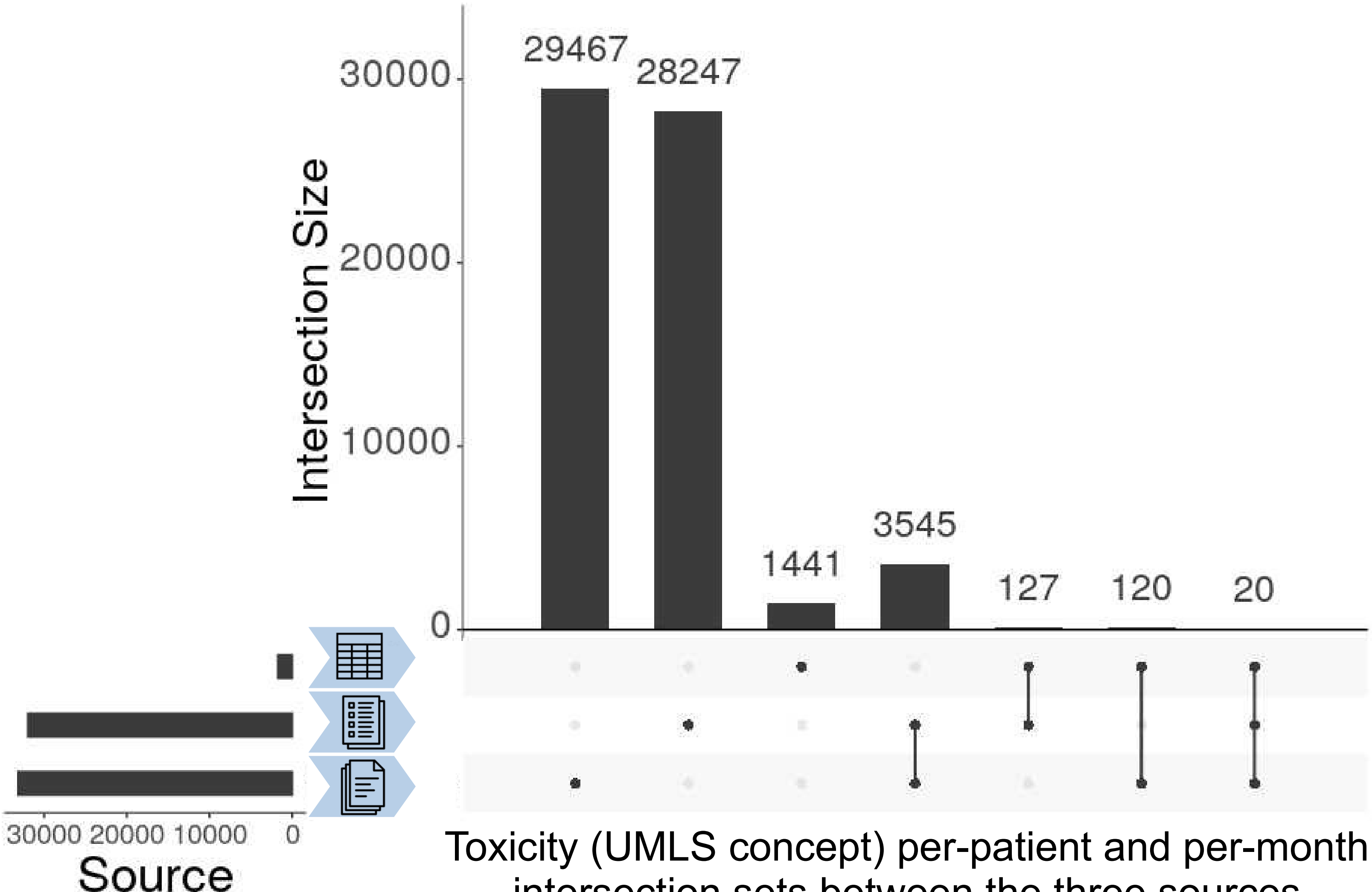
Results (1/2): What is extracted?

OntoTox Classes			
ChemotherapyToxicity	**54,420** >>	2,366 <<	**53,510**
Grade (ΣGrade*X*)	6,366	400	53,510
Grade*1*	2,100	87	9,981
Grade*2*	1,996	52	1,832
Grade*3*	817	23	191
Grade*4*	422	0	19
Grade*5*	2	1	0
Grade*0*	596	152	41,487
Grade*Null*	433	85	0

Results (1/2): What is extracted?

OntoTox Classes	UNSTRUCTURED	SEMI STRUCRURED	STRUCTURED
ChemotherapyToxicity	**54,420**	**2,366**	**53,510**
Grade (ΣGrade*X*)	**6,366**	**400**	**53,510**
Grade*1*	2,100	87	9,981
Grade*2*	1,996	52	1,832
Grade*3*	817	23	191
Grade*4*	422	0	19
Grade*5*	2	1	0
Grade*0*	596	152	41,487
Grade*Null*	**433**	**85**	**0**

Results (2/2): How sources agree?

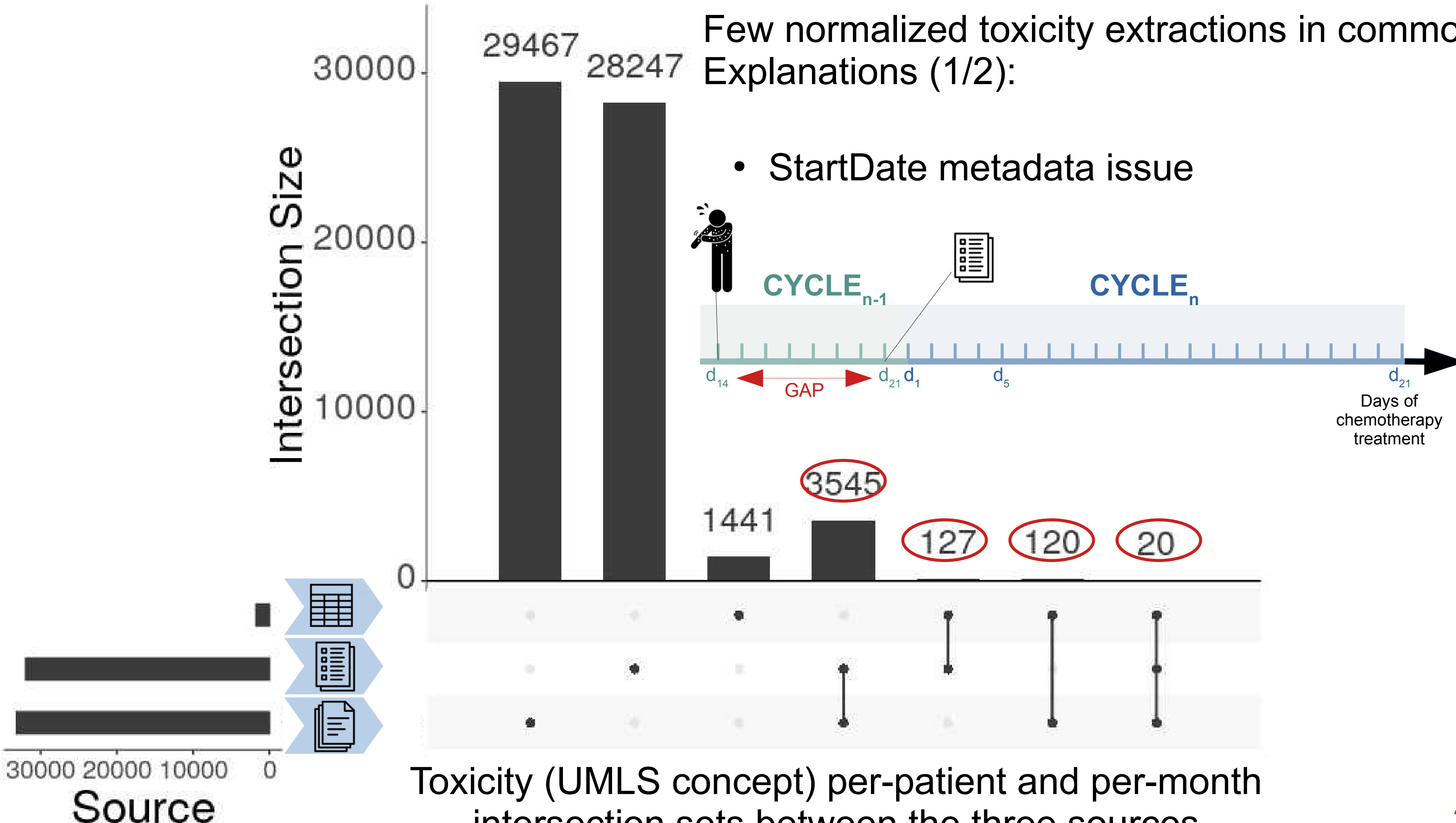


Toxicity (UMLS concept) per-patient and per-month intersection sets between the three sources

Results (2/2): How sources agree?

Few normalized toxicity extractions in common.
Explanations (1/2):

- StartDate metadata issue



Toxicity (UMLS concept) per-patient and per-month intersection sets between the three sources

Results (2/2): How sources agree?

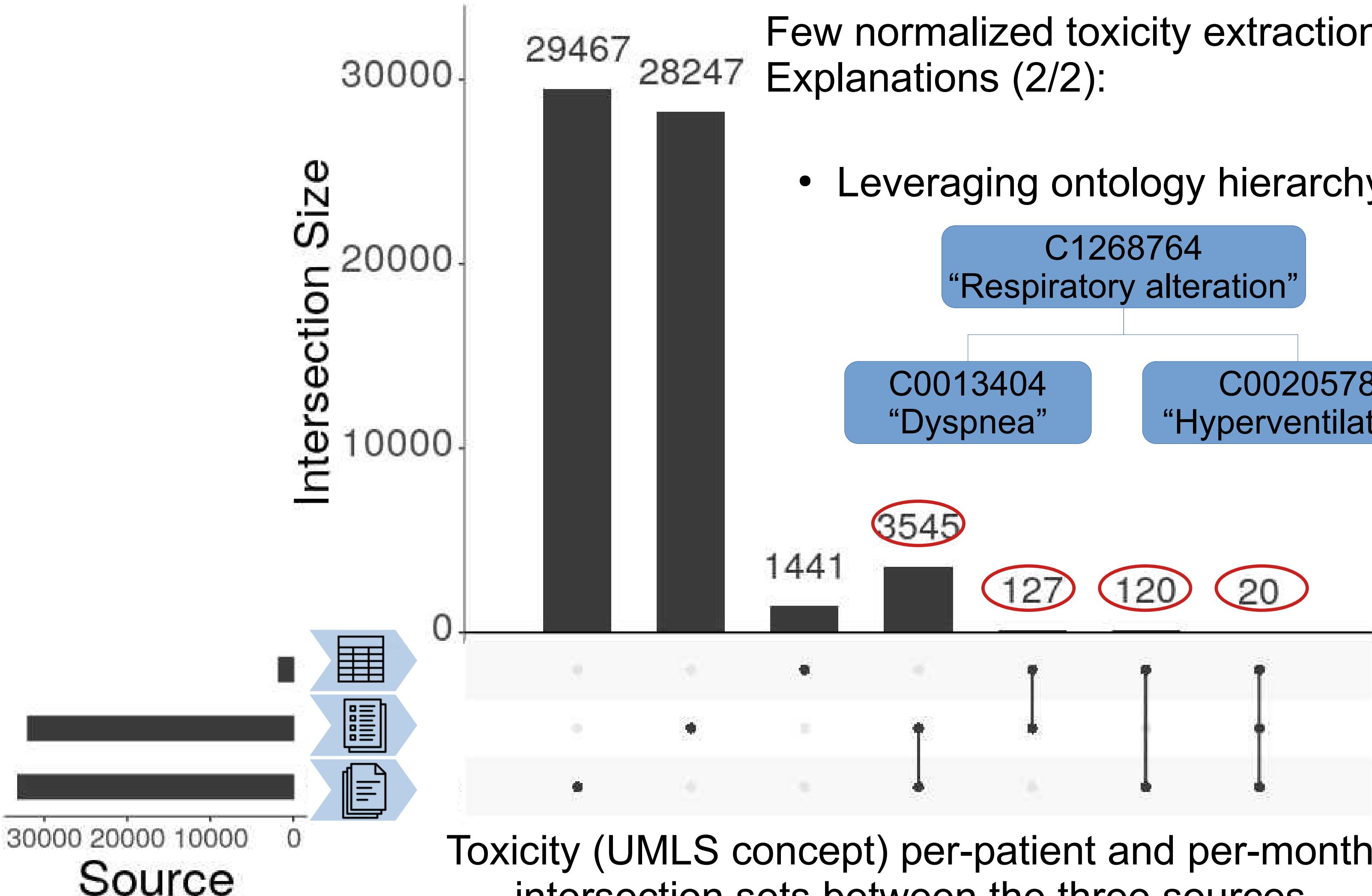
Few normalized toxicity extractions in common.
Explanations (2/2):

- Leveraging ontology hierarchy

C1268764
"Respiratory alteration"

C0013404
"Dyspnea"

C0020578
"Hyperventilation"



Toxicity (UMLS concept) per-patient and per-month intersection sets between the three sources

MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

IMIA

Discussion

OntoTox an ontology

- link to other knowledge models (PROV-O, MedDRA, UMLS)
- further use of reasoner

Room for improvement

- treatment representation
- time representation
- leveraging ontology hierarchy for granularity
- Evaluate and improve extractions algorithms

MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

Conclusion

OntoTox...

- very first ontology for chemotherapy toxicities
- can guide the data integration from various sources
- will be be enriched
- will further serve as a brick for clinical decision support systems

MEDINFO'21
ONE WORLD, ONE HEALTH
GLOBAL PARTNERSHIP FOR DIGITAL INNOVATION

IMIA

Thank you !



Alice ROGIER, Adrien COULET, Bastien RANCE

alice.rogier@inserm.fr

OntoTox GitHub:
https://github.com/TeamHeka/OntoTox

OntoTox BioPortal:
https://bioportal.bioontology.org/ontologies/ONTOTOX

Icons from Noun Project:

Fengquan Li
Eucalyp
Creative Stall
Adrien Coquet
Chanut is Indusries
Prettycons
H Alberto Gongora
Smalllike
visual world
Pham Thi Dieu Linh
icons
alberto galindo
Justin Blake
Hrbon
SA Family
Ifki riantos
Fajar
ProSymbols
Magicon
Danil Polshin
Gregor Cresnar
Eynav Raphael
angelina
Becris
23 icons
Hasanudin

Bibliography (1/2)

Basch, Ethan, Bryce B. Reeve, Sandra A. Mitchell, Steven B. Clauser, Lori M. Minasian, Amylou C. Dueck, Tito R. Mendoza, et al. « Development of the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) ». *JNCI: Journal of the National Cancer Institute* 106, nº dju244 (1 septembre 2014). https://doi.org/10.1093/jnci/dju244.

Brown, Elliot G., Louise Wood, et Sue Wood. « The Medical Dictionary for Regulatory Activities (MedDRA) ». *Drug Safety* 20, nº 2 (1 février 1999): 109-17. https://doi.org/10.2165/00002018-199920020-00002.

Coulet, Adrien, Nigam H. Shah, Yael Garten, Mark Musen, et Russ B. Altman. « Using Text to Build Semantic Networks for Pharmacogenomics ». *Journal of Biomedical Informatics* 43, nº 6 (1 décembre 2010): 1009-19. https://doi.org/10.1016/j.jbi.2010.08.005.

Datta, Surabhi, Elmer V. Bernstam, et Kirk Roberts. « A Frame Semantic Overview of NLP-Based Information Extraction for Cancer-Related EHR Notes ». *Journal of Biomedical Informatics* 100 (1 décembre 2019): 103301. https://doi.org/10.1016/j.jbi.2019.103301.

Lamy, Jean-Baptiste. « Owlready: Ontology-Oriented Programming in Python with Automatic Classification and High Level Constructs for Biomedical Ontologies ». *Artificial Intelligence in Medicine* 80 (1 juillet 2017): 11-28. https://doi.org/10.1016/j.artmed.2017.07.002.

Lamy, Jean-Baptiste, Alain Venot, et Catherine Duclos. « PyMedTermino: An Open-Source Generic API for Advanced Terminology Services », s. d., 5.

Lebo, Timothy, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, et Jun Zhao. « PROV-O: The PROV Ontology », 30 avril 2013. https://www.research.manchester.ac.uk/portal/en/publications/provo-the-prov-ontology(733f89c6-5e48-44f9-aabc-ae1c276a5602).html.

Murphy, Shawn N, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, et Isaac Kohane. « Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2) ». *Journal of the American Medical Informatics Association* 17, nº 2 (1 mars 2010): 124-30. https://doi.org/10.1136/jamia.2009.000893.

Neuraz, Antoine, Ivan Lerner, William Digan, Nicolas Paris, Rosy Tsopra, Alice Rogier, David Baudoin, et al. « Natural Language Processing for Rapid Response to Emergent Diseases: Case Study of Calcium Channel Blockers and Hypertension in the COVID-19 Pandemic ». *Journal of Medical Internet Research* 22, nº 8 (14 août 2020): e20773. https://doi.org/10.2196/20773.

Bibliography (2/2)

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajicˇ, Christopher D Manning, Ryan McDonald, et al. « Universal Dependencies v1: A Multilingual Treebank Collection », s. d., 8.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, et Christopher D. Manning. « Stanza: A Python Natural Language Processing Toolkit for Many Human Languages ». *arXiv:2003.07082 [cs]*, 23 avril 2020. http://arxiv.org/abs/2003.07082.

Soldaini, Luca, et Nazli Goharian. « QuickUMLS: A Fast, Unsupervised Approach for Medical Concept Extraction », s. d., 4.

Stenetorp, Pontus, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, et Jun'ichi Tsujii. « Brat: A Web-Based Tool for NLP-Assisted Text Annotation », s. d., 6.

« The protégé project: a look back and a look forward: AI Matters: Vol 1, No 4 ». Consulté le 16 mai 2021. https://dl.acm.org/doi/abs/10.1145/2757001.2757003.

Warner, Jeremy L., Andrew J. Cowan, Aric C. Hall, et Peter C. Yang. « HemOnc.org: A Collaborative Online Knowledge Platform for Oncology Professionals ». *Journal of Oncology Practice* 11, nº 3 (10 mars 2015): e336-50. https://doi.org/10.1200/JOP.2014.001511.

Warner, Jeremy L., Dmitry Dymshyts, Christian G. Reich, Michael J. Gurley, Harry Hochheiser, Zachary H. Moldwin, Rimma Belenkaya, Andrew E. Williams, et Peter C. Yang. « HemOnc: A New Standard Vocabulary for Chemotherapy Regimen Representation in the OMOP Common Data Model ». *Journal of Biomedical Informatics* 96 (1 août 2019): 103239. https://doi.org/10.1016/j.jbi.2019.103239.