

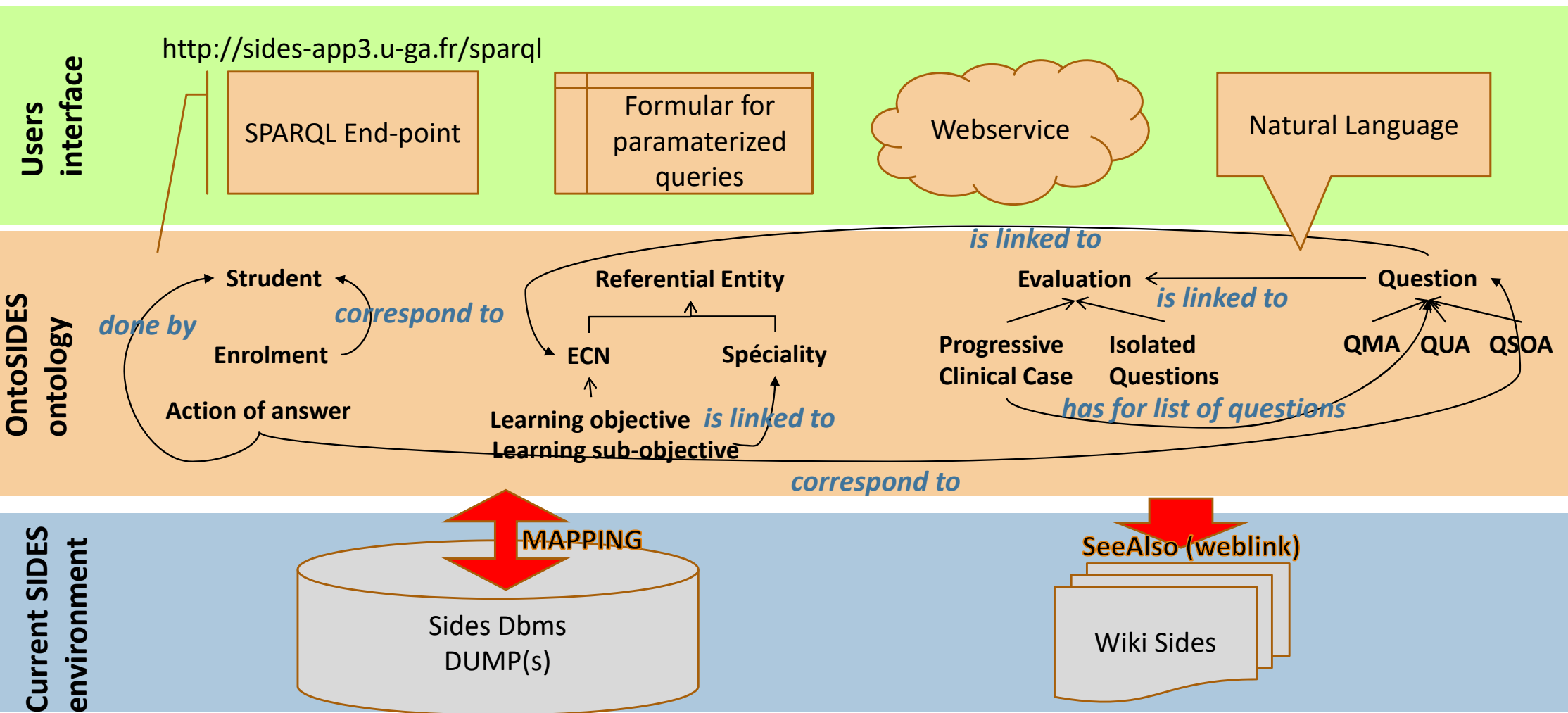


# Ontology Mediated Query Rewriting challenged by aggregated queries in OntoSIDES

---

Fabrice Jouanot, Olivier Palombi, M-Christine Rousset, Adam Sanchez

# OntoSIDES: the OBDA layer of SIDES 3.0



# OntoSIDES ontology

67 classes, 42 properties, 1419 instances (medical specialities, official items of the ECN programme) + 8 rules

The screenshot displays the OntoSIDES ontology editor interface. The left pane shows a tree view of classes, with 'sides:speciality (31)' selected. The central pane shows the 'Class Form' for 'sides:speciality', including annotations for 'Speciality' and 'Spécialité', and a class axiom 'sides:referential\_entity'. The right pane lists various properties such as 'sides:has\_for\_proposal\_of\_answer' and 'sides:correspond\_to'. The bottom pane shows a table of instances for the 'sides:speciality' class.

Instance	Property	Label
[Resource]	rdf:type	rdfs:label
◆ sides:addictology	sides:speciality	Addictologie
◆ sides:anesthesiology_resuscitati...	sides:speciality	Anesthesiologie - Reanimation - ...
◆ sides:cancerology_radiotherapy	sides:speciality	Cancerologie - Radiotherapie
◆ sides:cardiovascular	sides:speciality	Cardio-vasculaire
◆ sides:dermatology	sides:speciality	Dermatologie

# Ontop: a framework for implementing OBDA

---

An open source system for querying relational data sources through an ontology using SPARQL

- a virtual approach based on query rewriting using mappings
- support SPARQL 1.0

D. Calvanese, B. Cogrel, E. G. Kalayci, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, and G. Xiao.

OBDA with the ontop framework.

*In 23rd Italian Symposium on Advanced Database Systems, SEBD 2015, Gaeta, Italy, June 14-17, 2015., pages 296–303, 2015*

# Ontop mappings 1/2

The screenshot displays the Ontop Mapping Assistant interface. On the left, a class hierarchy for 'Action to answer' is shown, including subclasses like 'Action', 'City', 'Content', 'Educational content', 'Evaluation content', 'Evaluation type', 'Proposal of answer', 'Question', 'Test', 'Media', 'Degree', and 'Enrolment'. Below this is an 'Object property hierarchy' listing various properties such as 'has for author', 'has for learning objective', 'has for list of questions', etc.

The main area is the 'Mapping editor', which contains a 'Mapping manager' with buttons for 'Create', 'Remove', and 'Copy'. It lists several mappings with their corresponding SQL queries:

- urn:test**  
sides:test{id} a sides:test .  
`select a.id as id, a.title, a.startdate, a.enddate from public.assessment a`
- urn:test\_has\_for\_title**  
sides:test{id} sides:has\_for\_title {title}^^xsd:string .  
`select a.id as id, a.title as title, a.startdate, a.enddate from public.assessment a`
- urn:test\_start\_and\_end\_dates**  
sides:test{id} sides:starting\_date\_of\_test {startdate}^^xsd:dateTime ; sides:ending\_date\_of\_test {enddate}^^xsd:dateTime .  
`select a.id as id, a.title, a.startdate as startdate, a.enddate as enddate from public.assessment a`
- urn:evaluation\_has\_for\_textual\_content**  
sides:eval{id} sides:has\_for\_textual\_content {eval\_introduction}^^xsd:string .  
`SELECT di.id as id, string_agg(di.introduction, '') as eval_introduction from ontosides.docimocontent_introduction di inner join ontosides.evaluation ev on ev.pool_id = di.id group by di.id`
- urn:relation\_evaluation\_type\_question**  
sides:eval{pool\_id} sides:has\_for\_question sides:q{question\_id} .  
`select pq.pool_id as pool_id, pq.question_id as question_id, pq.position from pool_question pq inner join ontosides.evaluation ev on pq.pool_id = ev.pool_id`
- urn:relation\_test\_evaluation\_type**  
sides:test{assessment\_id} sides:is\_made\_of sides:eval{docimocontent\_id} .  
`select dca.assessment_id as assessment_id, dca.docimocontent_id as docimocontent_id from docimocontentassessment dca inner join ontosides.evaluation ev on ev.pool_id = dca.docimocontent_id`
- urn:relation\_test\_student**  
sides:test{assessment\_id} sides:has\_for\_registrant sides:stu{participant\_id} .  
`select p.assessment_id as assessment_id, p.participant_id as participant_id from public.participant p`

At the bottom, the 'Mapping count' is 189, and the search filter is 'pred:action\_to\_answer'. There is also a checkbox for 'Enable filter' and a note at the bottom right: 'To use the reasoner click Reasoner > Start reasoner' with a checked 'Show Inferences' checkbox.

# Ontop mappings 2/2

Mapping editor: Mapping manager Mapping Assistant - BETA

Mapping manager

+ Create - Remove Copy Select all Select none

**urn:test**  
sides:test{id} a sides:test .  
select a.id as id, a.title, a.startdate, a...

**urn:test\_has\_for\_title**  
sides:test{id} sides:has\_for\_title {title}^^xsd:string .  
select a.id as id, a.title as title, a.startdate as startdate

**urn:test\_start\_and\_end\_dates**  
sides:test{id} sides:starting\_date\_of\_test {startdate}^^xsd:date .  
select a.id as id, a.title, a.startdate as startdate

**urn:evaluation\_has\_for\_textual\_content**  
sides:eval{id} sides:has\_for\_textual\_content {eval}^^xsd:string .  
SELECT di.id as id, string\_agg(di.introduction, ' ')  
inner join ontosides.evaluation ev  
group by di.id

**urn:relation\_evaluation\_type\_question**  
sides:eval{pool\_id} sides:has\_for\_question sides:evaluation\_type {eval}^^xsd:string .  
select pq.pool\_id as pool\_id, pq.question\_id as question\_id,  
inner join ontosides.evaluation ev

**urn:relation\_test\_evaluation\_type**  
sides:test{assessment\_id} sides:is\_made\_of sides:evaluation\_type {test}^^xsd:string .  
select dca.assessment\_id as assessment\_id, dca.evaluation\_type as evaluation\_type,  
inner join ontosides.evaluation ev

**urn:relation\_test\_student**  
sides:test{assessment\_id} sides:has\_for\_registrant {test}^^xsd:string .  
select p.assessment\_id as assessment\_id, p.participant\_id as participant\_id from public.participant p

**Edit Mapping**

Mapping ID: urn:test\_has\_for\_title

Target (Triples Template): sides:test{id} sides:has\_for\_title {title}^^xsd:string .

Source (SQL Query): select a.id as id, a.title as title from public.assessment a limit 10

id	title
1	CONF-2015-2016-HEMATO n°1 - M2-Cpe F-DM
13894	PHARMA 3A EVAL UE 2014-2015 - 3A S05 - 1er déce...
11	C6-ORL-02/09/2015-DP1
15567	5DP
16	CONFBDI/CHIRVISC/DFASM3/01
17	CONFBDI/MAL INF/DFASM3/01

Test SQL Query

Update Cancel

Mapping count: 189 Search: pred:action\_to\_answer Enable filter

To use the reasoner click Reasoner > Start reasoner Show Inferences

# Limitations of Ontop

---

- Mappings do not enable to capture RDF containers
  - needed for relating a test of type « progressive\_clinical\_case » to its ordered list of questions
    - ⇒ we extended the class ModellOManager to generate such mappings
- Query rewriting does not support SPARQL 1.1 queries
  - Aggregated queries are needed
    - ⇒ we used mappings to materialize data
    - ⇒ accessible using a sparql endpoint: <http://virtuoso.ontosides.network/sparql>

# Actual status: ~ 1,5 billions of triplets

---

▪ **64 957 students** `select (count (distinct ?student)) where {?student a sides:student}`

▪ **947 students from Grenoble**

`select (count (distinct ?student)) where {?student a sides:student. ?e sides:correspond_to_student ?student. ?e sides:has_for_registration_place sides:grenoble_alpes_university }`

▪ **224 919 141 action of answer** (68 126 for Grenoble Strudents)  
for **457 873 questions** (**66 531 questions with unique answer** and **390 323 questions with multiple answers**)

`select (count (?a)) where {?a sides:done_by ?student}`

`select (count (distinct ?q)) where { ?a sides:correspond_to_question ?q }`

`select (count (distinct ?q)) where { ?a sides:correspond_to_question ?q. ?q a sides:QUA }`

`select (count (distinct ?q)) where { ?a sides:correspond_to_question ?q. ?q a sides:QMA }`



# Number of students by University

```
select (count (distinct ?student)) as ?NbEtu ?univ where {?student a sides:student.  
?e sides:correspond_to_student ?student. ?e sides:has_for_registration_place ?univ } group by ?univ
```

Sample of answers:

NbEtu	univ
456	<a href="http://www.side-sante.fr/sides#paris_xi_university">http://www.side-sante.fr/sides#paris_xi_university</a>
824	<a href="http://www.side-sante.fr/sides#caen_university">http://www.side-sante.fr/sides#caen_university</a>
1419	<a href="http://www.side-sante.fr/sides#bordeaux_university">http://www.side-sante.fr/sides#bordeaux_university</a>
537	<a href="http://www.side-sante.fr/sides#paris-ile_de_france_ouest_university">http://www.side-sante.fr/sides#paris-ile_de_france_ouest_university</a>
1127	<a href="http://www.side-sante.fr/sides#lille_university">http://www.side-sante.fr/sides#lille_university</a>
39	<a href="http://www.side-sante.fr/sides#paris_xii_university">http://www.side-sante.fr/sides#paris_xii_university</a>
451	<a href="http://www.side-sante.fr/sides#paris_xiii_university">http://www.side-sante.fr/sides#paris_xiii_university</a>
3842	<a href="http://www.side-sante.fr/sides#lyon_est_university">http://www.side-sante.fr/sides#lyon_est_university</a>
83	<a href="http://www.side-sante.fr/sides#la_reunion_university">http://www.side-sante.fr/sides#la_reunion_university</a>
757	<a href="http://www.side-sante.fr/sides#lyon_sud_university">http://www.side-sante.fr/sides#lyon_sud_university</a>
732	<a href="http://www.side-sante.fr/sides#tours_university">http://www.side-sante.fr/sides#tours_university</a>
947	<a href="http://www.side-sante.fr/sides#grenoble_alpes_university">http://www.side-sante.fr/sides#grenoble_alpes_university</a>
743	<a href="http://www.side-sante.fr/sides#nantes_university">http://www.side-sante.fr/sides#nantes_university</a>
554	<a href="http://www.side-sante.fr/sides#brest_university">http://www.side-sante.fr/sides#brest_university</a>

# Some statistics about questions

---

## ■ Questions with multiple answers (QMA): 467 498

- ✓ with 77 175 without action of answer

```
select (count (distinct ?q)) where { ?q a sides:QM
      FILTER NOT EXISTS {?a sides:correspond_to_question ?q} }
```

- ✓ with 50 550 related to a medical speciality, and 54 497 to an item of the ECN referential

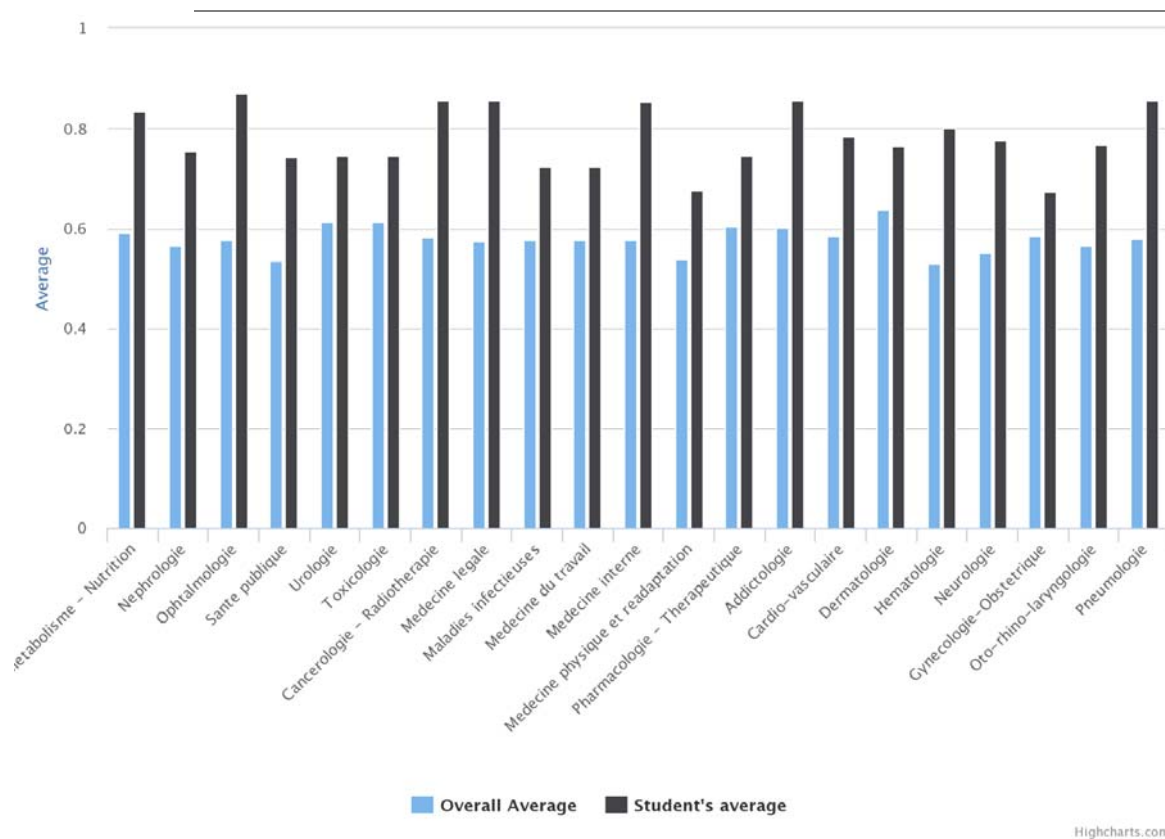
```
select (count (distinct ?q)) where { ?q a sides:QMA.
      ?q sides:is_linked_to_the_medical_speciality ?s }
```

## ■ Questions with unique answers (QUA): 81 155

- ✓ with 14 624 without action of answer
- ✓ with 5181 related to a medical speciality, and 5912 to an item the ECN referential

# Focus on a student: stu12402

## average comparison of his results with average results of all students



```

SELECT ?specialite ?label ((?moyennespecialite) AS ?moy) ?moyenneetu
WHERE {GRAPH <http://ontosides.en> {
  { SELECT ?specialite ((AVG(?note1)) AS ?moyennespecialite)
    ((COUNT(DISTINCT ?q1)) AS ?nbquestionspecialite)
    WHERE { ?adr1 sides:has_for_result ?note1 .
      ?adr1 sides:done_by ?etu .
      ?adr1 sides:correspond_to_a_question ?q1 .
      ?q1 sides:is_linked_to_ENC_referential_entity ?item .
      ?item sides:is_linked_to_the_medical_speciality ?specialite . }
    GROUP BY ?specialite } .
  { SELECT ?specialite ((AVG(?note)) AS ?moyenneetu)
    ((COUNT(DISTINCT ?q)) AS ?nbquestionItemEtu)
    WHERE { ?adr sides:has_for_result ?note .
      ?adr sides:done_by ?etudiant .
      ?adr sides:correspond_to_a_question ?q .
      ?q sides:is_linked_to_ENC_referential_entity ?item .
      ?item sides:is_linked_to_the_medical_speciality ?specialite .
      FILTER(?etudiant = sides:etu12402) }
    GROUP BY ?specialite } .
  OPTIONAL { SELECT ?specialite str(?label) as ?label
    WHERE { ?specialite a sides:speciality.
      ?specialite rdfs:label ?label. }
    GROUP BY ?specialite } . }
  
```

# On-demand data extraction at different granularity level and driven by analysts

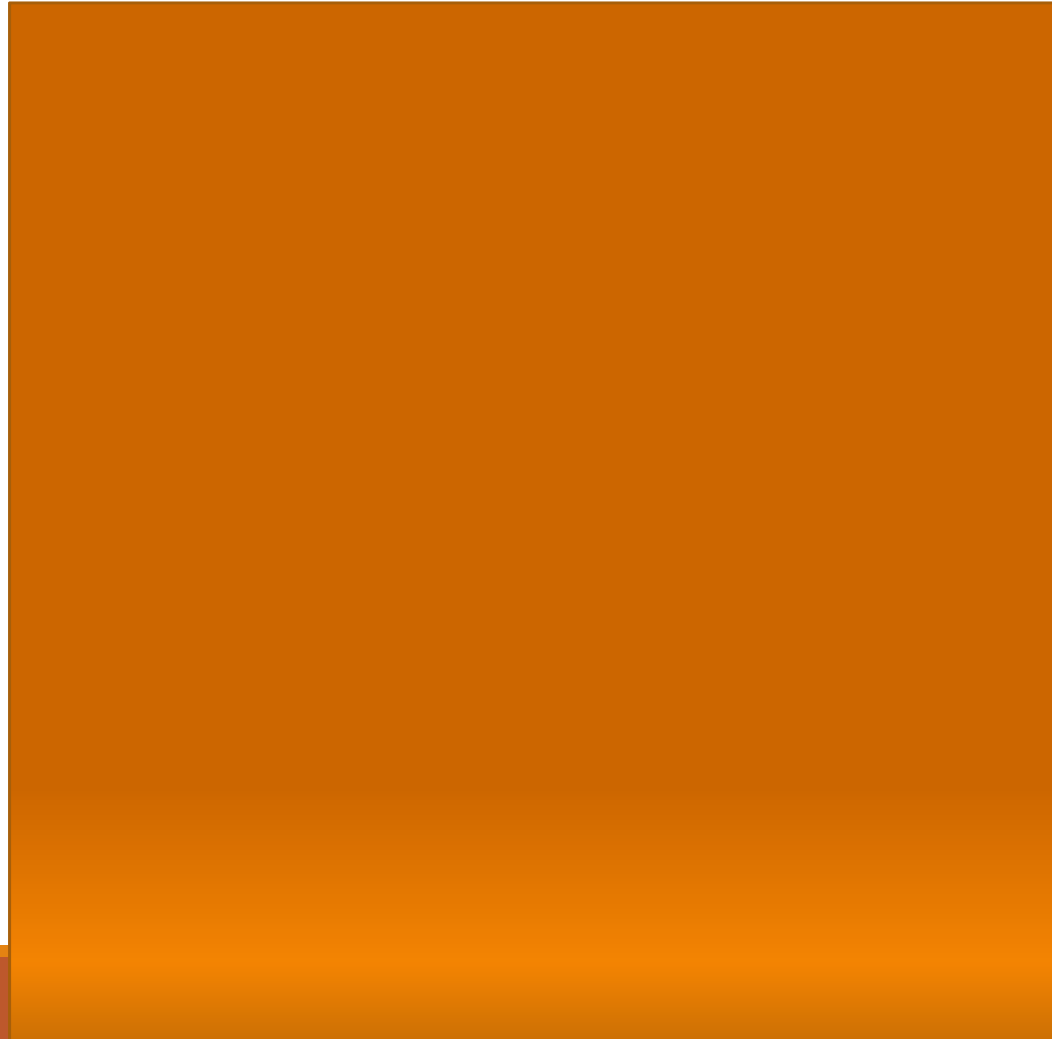
---

- **Use case 1:** Define explainable and computable criteria from data about student questions and their related answers
  - ✓ For computing difficulty level, discriminatory factor, and quality level of questions (inside a test or a pool of questions, ...) or distractors (inside a question content)
- **Use case 2:** Define and compute a student **trajectory** (or for a group of students) related to items from the referential or to medical specialities
  - ✓ Temporality is central !

For each QUA question linked to the speciality about infectious disease, with at least 100 action of answers and at least one correct reponse, give the level of difficulty (**number of correct answers/** number of action of answer)

```
SELECT ?question ?NbReponsesCorrectes ?NbReponses
      |(1-(xsd:decimal(?NbReponsesCorrectes)/xsd:decimal(?NbReponses)) AS ?NiveauDifficulte)
FROM <http://ontosides>
WHERE {
  {SELECT ?question (COUNT(?action) AS ?NbReponses)
  WHERE
  {?question a sides:QUA.
  ?question sides:is_linked_to_the_medical_speciality sides:speciality_infectious_diseases.
  ?action sides:correspond_to_question ?question.
  FILTER EXISTS {?actionCorrect sides:correspond_to_question ?question.
                  ?actionCorrect sides:has_rightly_ticked ?rep.}}
  GROUP BY ?question
  HAVING(COUNT (?action) > 100)
  }
  {SELECT ?question (COUNT(?actionCorrect) AS ?NbReponsesCorrectes)
  WHERE
  {?question a sides:QUA.
  ?question sides:is_linked_to_the_medical_speciality sides:speciality_infectious_diseases.
  ?actionCorrect sides:correspond_to_question ?question.
  ?actionCorrect sides:has_rightly_ticked ?rep.}
  GROUP BY ?question
  }
}
ORDER BY DESC(?NiveauDifficulte)
```

For each QUA question linked to the speciality about Infectious disease, with at least 100 action of answers and at least one correct reponse, give the level of difficulty



# Challenges

---

## ■ Technical Challenges

- ✓ Tracking student history across SIDES DB dumps
- ✓ Developing an efficient method for scaling forward reasoning on big data (triple store saturation)

## ■ Methodological Challenges

- ✓ Managing the whole workflow for an analyst use case: from data extraction to recommendation