Observation Facilities in the VO
[Data Discovery Use Case]

B. Cecconi (1), L. Debisschop (1)
M. Louys, (2), E. Perret (2),

(1) Observatoire de Paris, Meudon, France;
(2) CDS, Strasbourg, France

The Europlanet-2024 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.

Why

- **ObsTAP** / **EPN-TAP** have “instrument_host_name” / “facility” keywords.

  Standard nomenclature needed for efficient data mining

- No official nomenclature/standard for “observation facilities” names.

- Several lists identified (some have >1000 items).

- Merging is complex (by hand or programmatically)

Goal for this study

- Use case: **Data discovery**
  - step 1: define what should be stored (observatory/ telescope/space mission/spacecraft…), and if relations are needed (e.g., telescope to observatory)
  - step 2: match lists and catalogues, build a lookup table with alternate names
  - step 3: define maintenance procedure
  - step 4: build a name resolver for data discovery clients, or to help provider to select a name

Connecting with WikiData

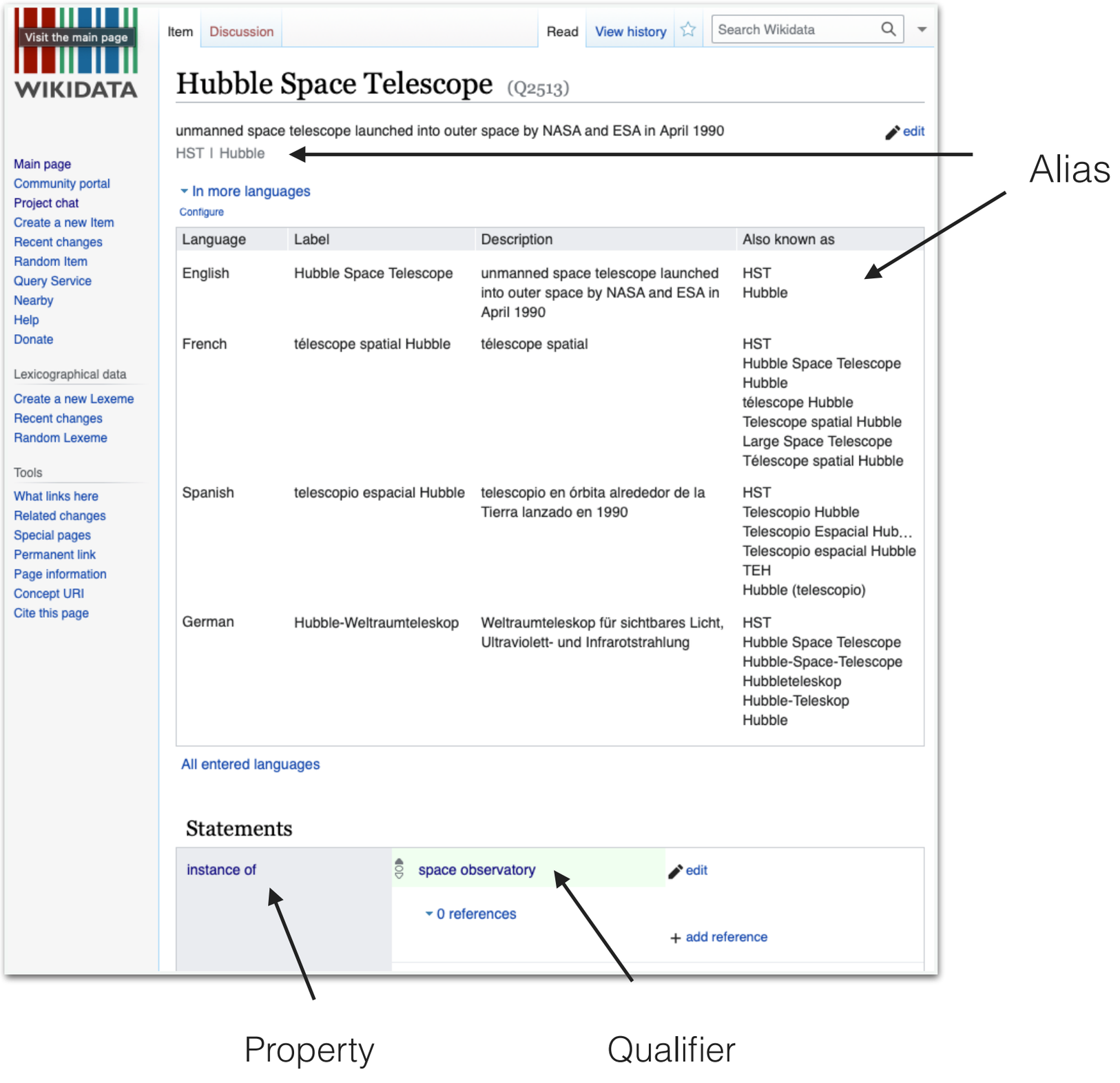
- Wikidata = free and open knowledge base
  => structured data (export in RDF, JSON)
  (each item has properties + values)
  => connected with wikipedia, Wiktionary

- Wikidata can be **queried** and **edited** manually, or with API, SPARQL and other tools (e.g., QuickStatements, OpenRefine…)
  https://meta.wikimedia.org/wiki/QuickStatements
  https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine

- For *Observation Facilities*:
  - many identifiers are already connected
  - model/properties are fuzzy but adequate
  - curation / extension is feasible
  - semantics quality is variable

WIKIDATA

Wikidata record examples



Wikidata record examples

WIKIDATA

Item | Discussion | Read | View history

Mars Orbiter Mission (Q2156739)

Indian Mars orbiter, launched in 2013

MOM | Mangalyaan

edit

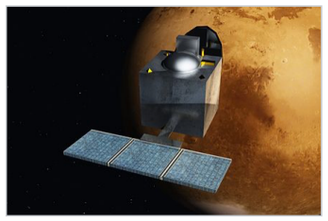
In more languages

Configure

Language	Label	Description	Also known as
English	Mars Orbiter Mission	Indian Mars orbiter, launched in 2013	MOM Mangalyaan
French	Mangalyaan	No description defined	
Spanish	Mars Orbiter Mission	sonda espacial cuyo lanzamiento fue el 5 de noviembre de 2013	
German	Mars Orbiter Mission	indische Marsmission	

All entered languages

Statements

instance of	space probe 1 reference
image	

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

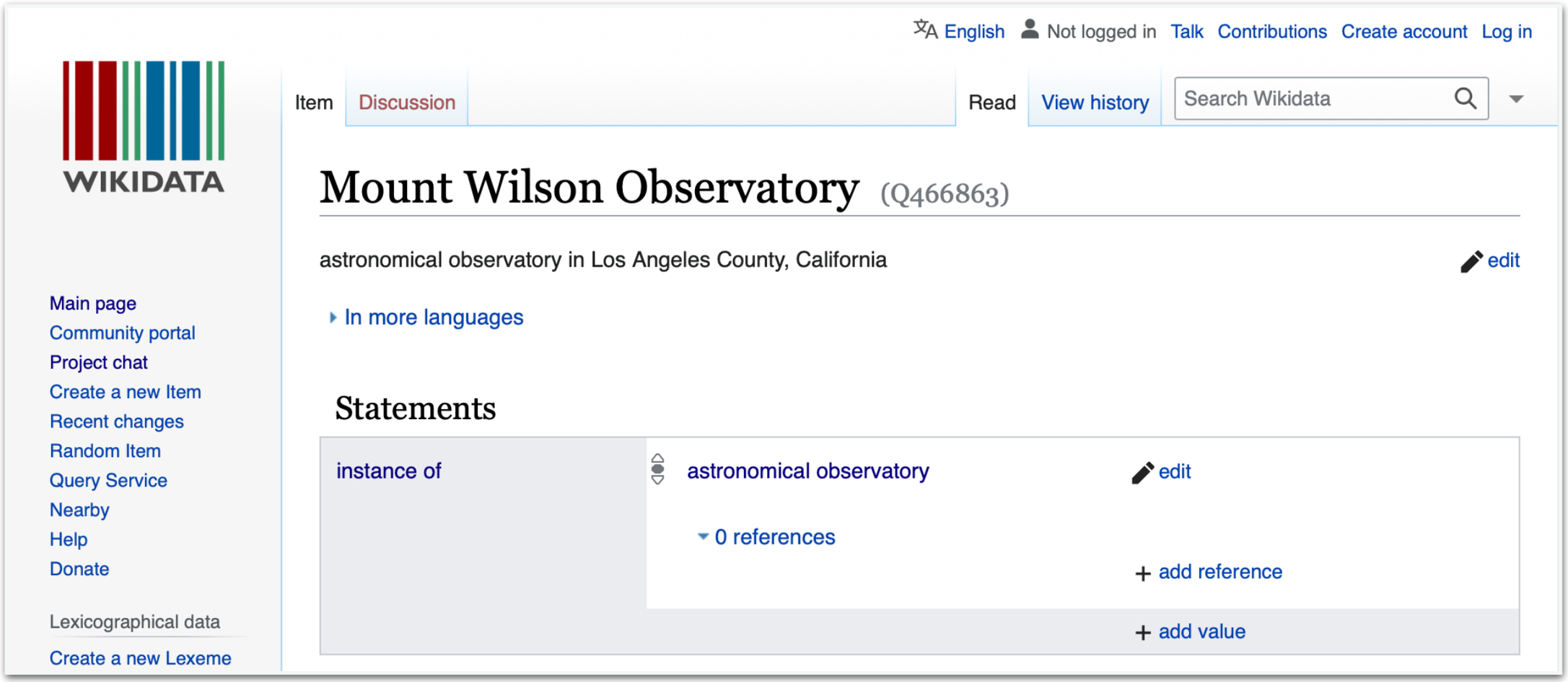
Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

Identifiers

COSPAR ID	2013-060A 2 references	edit + add value
Encyclopædia Britannica Online ID	topic/Mars-Orbiter-Mission 0 references	edit + add reference + add value
Freebase ID	/m/0knt6hh 1 reference	edit + add value
NAIF ID	-3 1 reference	edit + add value

Wikidata record examples



Identifiers

Minor Planet Center observatory code	672 0 references	edit add reference
		add value

SparQL query example

Wikidata Query Service | Exemples | Assistant de requêtes | Aide | Davantage d'outils | français

```
1  PREFIX schema: <http://schema.org/>
2  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3  PREFIX wikibase: <http://wikiba.se/ontology#>
4  PREFIX bd: <http://www.bigdata.com/rdf#>
5  SELECT
6    ?item
7    ?itemLabel
8    (GROUP_CONCAT(DISTINCT ?alias; SEPARATOR="|") AS ?aliases)
9  WHERE
10 {
11   ?item p:P31 ?stat .
12   #item instance of
13   {?stat ps:P31 wd:Q148578 .}  # space observatory
14   OPTIONAL {?item skos:altLabel ?alias .}
15   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
16 }
17 GROUP BY ?item ?itemLabel
```

147 résultats en 1109ms | Code | Télécharger | Lien

item	itemLabel	aliases
wd:Q19764	Astron	
wd:Q48633	Corot	CoRoT|Convection Rotation and Planetary Transits|COnvection ROtation and planetary Transits|Convection, Rotation and planetary Transits|CoRoT|コロー衛星|COROT space telescope|COnvection ROtation and planetary Transits
wd:Q49694	EXOSAT	Sat|European X-ray Observatory Satellite

Name resolver prototype

- Implemented on ElasticSearch + Kibana.

- Export from WikiData:
  - identifier (Qxxxxxx), label, aliases, known external identifiers (IAU, NSSDC, UAT…)

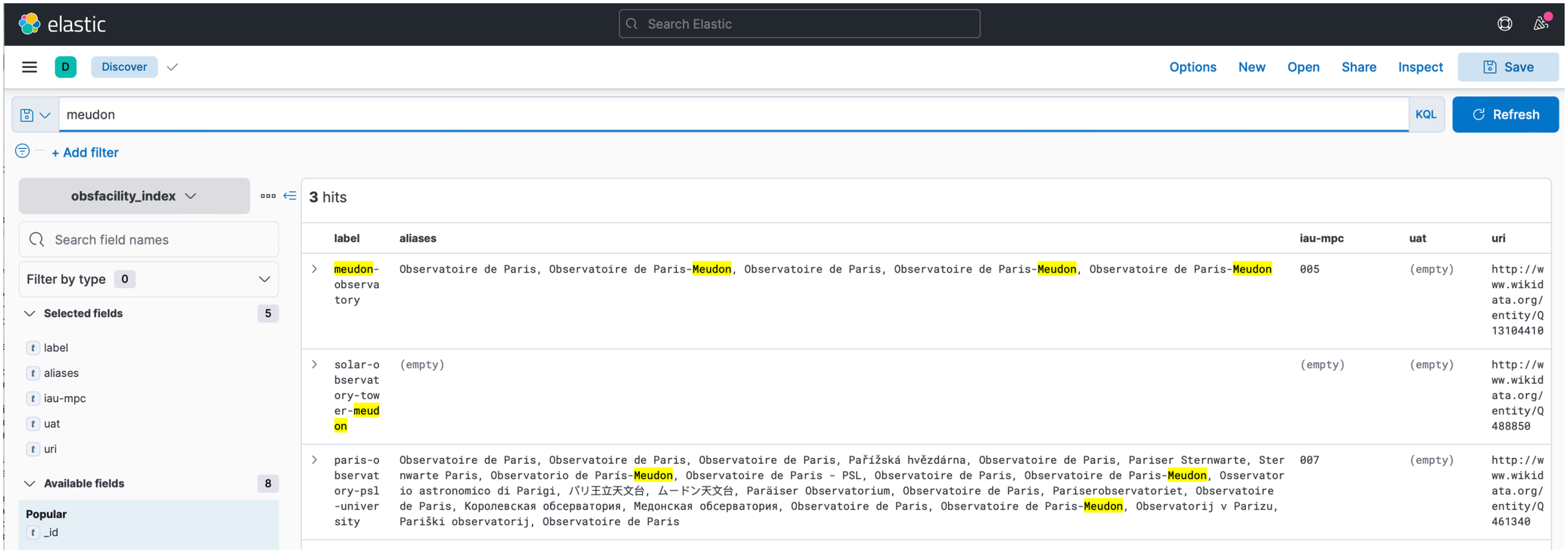
- Import into « obsfacility_index »

- Use Kibana for exploring dataset

- Currently: 14200 items ingested
  Curation still needed.

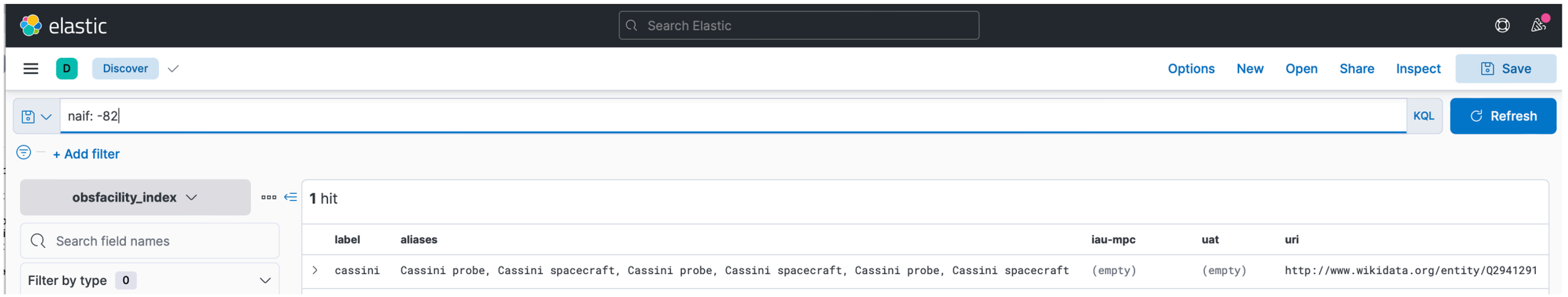
Examples

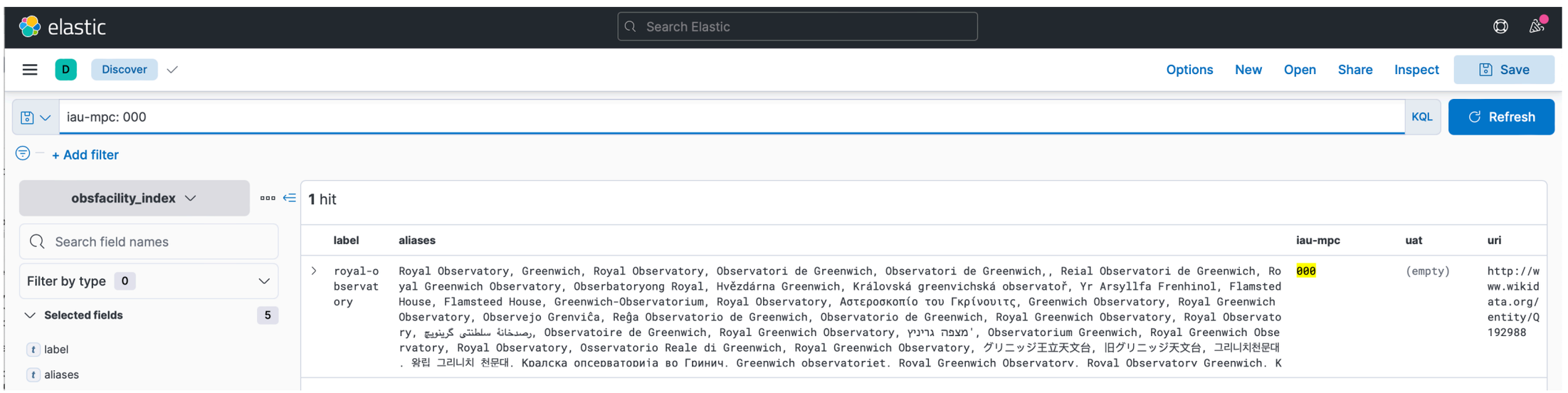
free-text search



Examples

known id search





Operation & Maintenance

- Very preliminary… *doc in preparation, currently in French*
- Scripts to export from WikiData: SPARQL => json
- Comparison with known lists (fuzzy matching)
  => tagged results:
  - ‘not_found’: check for wikidata records, add record if necessary.
  - ‘very_good_match’: Ingest into Elastic Search database
- TBD:
  - SPARQL query retrieving only updated or new records.
  - procedure to ingest results with existing content in Elastic Search
  - translate doc in english…

Wikidata
as a knowledge base

- **Proposal**: use Wikidata as our curated knowledge base

- **Data Discovery** use case: *looks good.*
  => regular extraction into a name resolver database
  (using SPARQL queries)
  => curation of metadata directly in Wikidata
  => name-resolver (e.g., with elastic-search API)

- Two documents in preparation (IVOA Notes):
  - https://github.com/BaptisteCecconi/ObsFacilityWikidata
  - update and maintenance document (not online yet)