



Modelling protected areas

A proposal

Ismael Olea

[User:Olea](#)

This session is recorded: Please mute your microphone and camera when you're not speaking.

The problem

- In Wikidata we are not having a pattern/good practice/standard for modeling [protected area \(Q473972\)](#) items.
- Some cases are using [instance of \(P31\)](#) and others [heritage designation \(P1435\)](#).
- We are mixing instances of natural geographic with protected areas in the same item but
- conceptually both are different:
 - A protected area is created by an administrative act associating an «arbitrary» geographic area with a [conservation designation \(Q5162954\)](#).
 - A [protected area \(Q473972\)](#) is related with one or more:
 - other protected areas,
 - geographic objects and
 - administrative areas.

This situation needs to be fixed and addressed, probably, at [Wikidata:WikiProject Protected areas](#).

Being protected is not easy

See [Caenlochan](#) (P31: valley ([Q39816](#)))

| | |
|---|--|
| located in protected area |  Caenlochan Special Area of Conservation  |
| | applies to part part of landform |
| | ▼ 0 references + add reference |
|  Caenlochan Special Protection Area  | |
| applies to part part of landform | |
| ▼ 0 references + add reference | |
|  Caenlochan SSSI  | |
| applies to part part of landform | |
| ▼ 0 references + add reference | |
| + add value | |

How (supposedly) to query protected areas

```
SPARQL
# list designated protected natural areas
# about 148352 items
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
SELECT DISTINCT ?place WHERE {
  VALUES ?prop { wdt:P757 wdt:P809 wdt:P1732 wdt:P1848
    wdt:P2516 wdt:P2520 wdt:P2525 wdt:P2584 wdt:P2621 wdt:P2763
    wdt:P3009 wdt:P3296 wdt:P3327 wdt:P3401 wdt:P3412 wdt:P3425
    wdt:P3498 wdt:P3515 wdt:P3516 wdt:P3609 wdt:P3613 wdt:P3810
    wdt:P3974 wdt:P4001 wdt:P4029 wdt:P4083 wdt:P4154 wdt:P4171
    wdt:P4170 wdt:P4172 wdt:P4182 wdt:P4190 wdt:P4762 wdt:P4800
    wdt:P4977 wdt:P5965 wdt:P6070 wdt:P6230 wdt:P6280 wdt:P6478
    wdt:P6560 wdt:P6602 wdt:P6659 wdt:P6700 wdt:P7129 wdt:P7308
    wdt:P8114 wdt:P9516 wdt:P9552 wdt:P9755 wdt:P10015 wdt:P10392
    wdt:P814 wdt:P10513 wdt:P10514 wdt:P10343 wdt:P2467 wdt:P677
  }
  ?place ?prop ?natural.
}
```

QLever query: <https://qlever.cs.uni-freiburg.de/wikidata/PbiY9H>

WDQS: <https://w.wiki/8MEJ> (from time to time it can't be run)

Properties used both for natural and cultural immovable heritage

- Heritage designation ([P1435](#))
 - technically fixable
- and UNESCO's World Heritage Site ID ([P757](#))
 - used by UNESCO for both types

The problem with «heritage designation» ([P1435](#))

- semantic drift:
 - [first description](#) (2014): the heritage status of a historical site
 - [current description](#) (2023): heritage designation of a cultural or natural site
- today: just a non-exhaustive [sample query of protected area items using it](#) (1824 items)



Some data

Protection figure and jurisdictions

subclass of (P279) of protected area (Q473972):

- non applies to jurisdiction (P1001): <https://w.wiki/8MEc> (221 items)
- applies to jurisdiction (P1001): <https://w.wiki/8MEe> (282 items)

Because, we are expecting a protected area should be related with a jurisdiction, right?

Instances of «conservation designation»

```
# Intances of «conservation designation» without jurisdiction nor country
# about 15
SELECT ?item ?itemLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
  ?item wdt:P31 wd:Q5162954.
  FILTER(NOT EXISTS { ?item wdt:P17 []. })
  FILTER(NOT EXISTS { ?item wdt:P1001 []. })
}
```

SPARQL

Query: <https://w.wiki/8JXs>

Instances of «conservation designation»

```
# Intances of «conservation designation» with jurisdiction or country
# about 48
SELECT DISTINCT ?item ?itemLabel ?P17Label ?P1001Label WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "
[AUTO_LANGUAGE],en". }
  {
    ?item wdt:P31 wd:Q5162954;
    wdt:P17 ?P17.
    OPTIONAL { ?item wdt:P1001 ?P1001. }
  }
  UNION
  {
    ?item wdt:P31 wd:Q5162954;
    wdt:P1001 ?P1001.
    OPTIONAL { ?item wdt:P17 ?P17. }
  }
}
```

SPARQL

Query: <https://w.wiki/8JY5>

A subclassing party

```
# Number of subclasses+ of «protected area»  
# >1000 !!!  
SELECT (COUNT(DISTINCT ?item) AS ?count) WHERE  
{ ?item wdt:P279|wdt:P279+ wd:Q473972. }
```

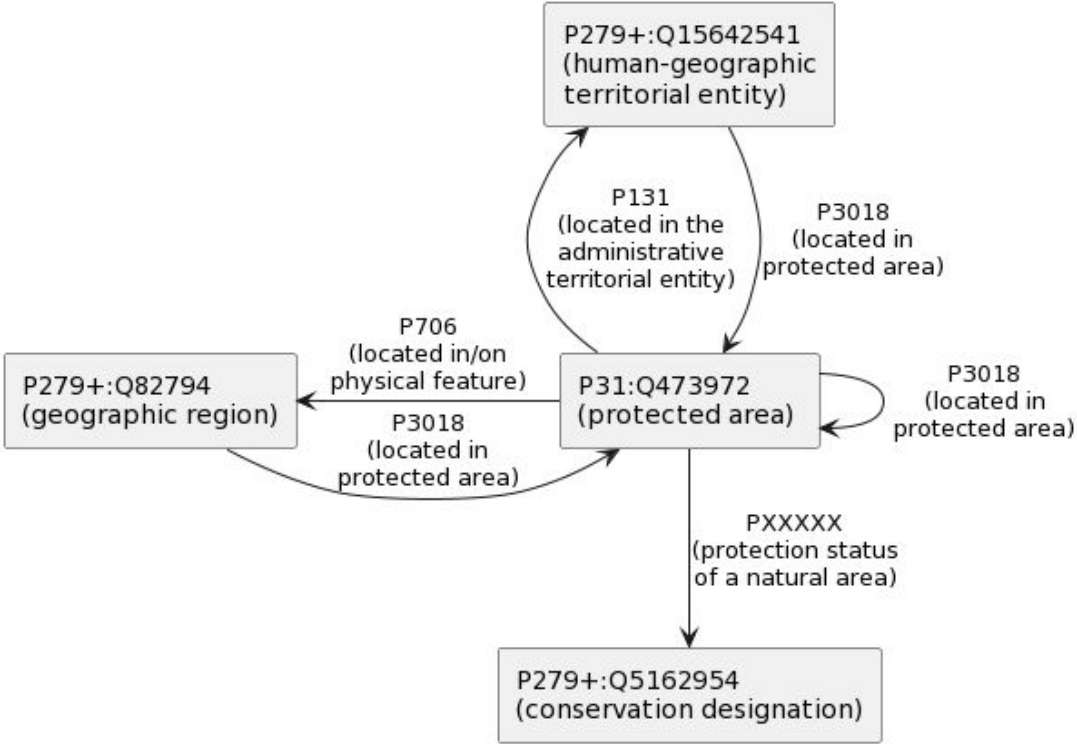
Query: <https://w.wiki/8Jj8>





Proposed solution

New property: «protection status of a natural area» (PXXXXX)



New property: «protection status of a natural area» (PXXXXX)

Setting a common best practice pattern:

- [heritage designation \(P1435\)](#)
- [intangible cultural heritage status \(P3259\)](#)
- **protection status of a natural area (PXXXXX)**

Rationale

As described at [Wikidata:WikiProject Protected areas/Taxonomy](https://www.wikidata.org/wiki/Wikidata:WikiProject_Protected_areas/Taxonomy):

- «Areas are administratively defined territories and thus are deep subclasses of territory» (Q1496967).
- The protected portion(s) of land (or sea) should be distinct entities, instances of a different subtree of geographic region (Q82794).»

Property PXXXXX restrictions

- Attribute value:
 - P279+ conservation designation (Q5162954)
 - conservation designation item should include:
 - [applies to jurisdiction \(P1001\)](#)
 - country (P17)
 - [maintained by \(P126\)](#)
 - or [operator \(P137\)](#) ?
 - [main regulatory text \(P92\)](#)
- P31 required
 - «protected area»
 - none any other
 - open question: protected area subclasses instances?
- P31 prohibited:
 - geographical items
 - cultural items
- P31 exception:
 - items using [«World Heritage Site ID \(P757\)»,](#) used by UNESCO natural and heritage sites;
 - open question: some other case?

Property PXXXXX restrictions

- should use properties:
 - [applies to jurisdiction \(P1001\)](#)
 - [inception \(P571\)](#)
 - country (P17)
 - located in the administrative territorial entity (P131)
 - [maintained by \(P126\)](#)
 - or [operator \(P137\)](#) ?
 - [area \(P2046\)](#)
- optionally can use:
 - [IUCN protected areas category \(P814\)](#)
 - coordinate location (P625)
 - [located in/on physical feature \(P706\)](#)

Justification

- designations are not abstract concepts, are human administrative acts (WD [notability criteria 2](#));
- also, a designation act specifies a set of relevant technical details as:
 - extension,
 - precise protection status,
 - applies to jurisdiction,
 - operators,
 - links to maps (in OSM, Commons...),
 - etc;
- simpler and faster queries by using:
 - an specific protection status property (PXXXXXX) and
 - a single instancing class (protected area (Q473972))
- easier data reuse (because an specified data model)

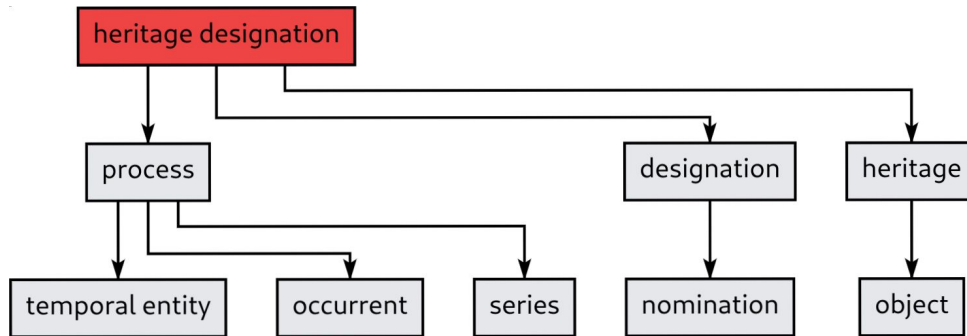
Justification

- disambiguates cases as [UNESCO World Heritage Site \(Q9259\)](#) sites;
- better way to keep a 1:1 mapping with important databases/inventories and existing external ID properties ([Common Database on Designated Areas ID \(P4762\)](#), [WDPA ID \(P809\)](#), [Natura 2000 site ID \(P3425\)](#) and [many others](#));
- this way we would not need to mix different designations data in the same item (example: [Parque nacional y natural de Sierra Nevada \(Q1969050\)](#));
- and any geographical object (natural or not) can be explicitly related with their related designations with [located in protected area \(P3018\)](#).

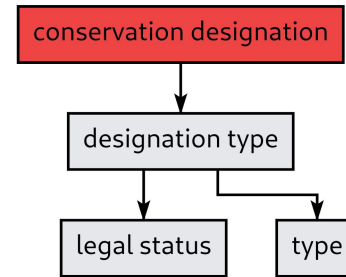


Related action plan

Analogous modeling of «conservation designation» and «heritage designation»



heritage designation ([Q30634609](#))
([wikidata-graph-builder](#)):



conservation designation ([Q5162954](#))
([wikidata-graph-builder](#)):

Open question: should we apply the heritage model practice to conservation?

Check «conservation designation» (Q5162954) instances

- Fix jurisdiction & country,
- distinguish items for legal based designations from their equivalent subclasses of protected areas,
- remove unrelated cultural heritage statements,
- related: remove unrelated natural area subclasses in cultural heritage items (as an example, seems to be the case of [Q6626844](#)),
- Open question: how to model international jurisdiction (Ramsar, UNESCO...)

Fix protected area subclasses

- from subclasses of «protected area»,
- to instances of «conservation designation»,
- open question: should be apply to ALL subclasses?
- remove unrelated cultural heritage statements,
- related: remove unrelated natural area subclasses in cultural heritage items.
- flatten/merge as much as possible the protected area subclasses:
 - just to get a meaningful conceptual abstract set,
 - to be used, for example, with XXwiki sitelinks.
 - Open question: do we want to apply this set to the conservation designations?
This can be troublesome if related legislation is not explicit.

Restrict «heritage designation» (P1435) to cultural heritage

- remove/exclude everything related with natural areas
- strict restriction to tangible cultural heritage

Massive modifications

- Identify and accordingly model «conservation designation» concepts, used as instances of:
 - geographical objects
 - protected areas
 - (some other?)
- Identify and create new protected area items from instances attributes of geographical objects

Move from P1435 to PXXXXX

For protected area items:

- remove all uses of P1435 with conservation designation attributes
- moving them to new PXXXXX statement (except UNESCO, maybe).
- For non protected area items:
- same but creating and linking the protected area required by the model.

Disambiguate XXwiki sitelinks

- An unclear amount of sitelinks of geographical items describe only/also protected areas.
- And viceversa.
- Also, sitelinks in different language for a same item can be any combination of both.



Create an EntitySchema

Inspired, maybe, by:

- [basic schema for national parks in Australia.](#)

but at the moment my personal experience here is zero.

How to manage different surface units

Different surface units, from different sources for the same protected area:

Example: [Q63378447](#)

Some related work:

- Vargas-Rojas, Felipe, et al. [“PhyQus: Automatic Unit Conversions for Wikidata Physical Quantities.”](#) 4rd Wikidata Workshop (co-located with ISWC2023). 2023.



| | | |
|------|---|---|
| area |  191.2585 square kilometre  |  |
| | 1 reference | |
| |  19,125,850 hectare  |  |
| | 1 reference | |
| | + add value | |

How to manage supposedly identical protected areas (1)

Different surface units, from different sources for the same protected area:

Example: [Q63378447](#)



The screenshot shows the Wikidata instance page for Q63378447. The left sidebar indicates it is an instance of a protected area. The main content area lists two types of protected areas:

- site of community importance** (with an edit icon):
 - start time: 2008
 - 1 reference: WDPA ID 555623600 (with an add reference icon)
- Natura 2000 protected area** (with an edit icon):
 - 1 reference: stated in Nationally designated areas inventory (with an add reference icon)
 - + add value (at the bottom of the section)

Maybe different items linked by [said to be the same as \(P460\)](#)?

How to manage supposedly identical protected areas (2)

There are **many** cases in which *Common Database on Designated Areas ID* ([P4762](#)) and *Natura 2000 site ID* ([P3425](#)) attributes are exactly same value:

- do we know they refer exactly to the same protected area+conservation designation?
- is this true in all the cases?

(happens with other properties too)

How to model UNESCO natural and heritage sites

items using [«World Heritage Site ID \(P757\)»](#), should use:

- [heritage designation \(P1435\)](#) ?
- **protection status of a natural area (PXXXXX)** ?
- both?

Closing

- This is just a draft proposal.
- Not yet being formalized in the Wikidata community.
- Suggestions and comments are welcome.
- Thanks to Susanna Ånäs ([User:Susannaanas](#)) and Camillo Pellizzari ([User:Epìdosis](#)) for their feedback.



Thanks for your attention!

Get in touch:

Ismael Olea

[User:Olea](#)

<https://olea.org>

Credits

Original content licensed CC-BY

