

OSCAL '19

---

# CATEGORIZING WIKIPEDIA



**USERNAME ALBINFO**

**LARS HAEFNER**

**ALL IMAGES BY LARS HAEFNER/ALBINFO**  
UNLESS OTHERWISE STATED — CC-BY-SA-4.0

A scenic landscape featuring a range of mountains with snow-capped peaks under a clear blue sky. In the foreground, a winding road curves through a valley with green vegetation. The overall scene is bright and clear.

ORGANIZING

---

WIKIPEDIA

# WIKIPEDIA IN NUMBERS

- ▶ English Wikipedia:  
5'853'889 articles
- ▶ German Wikipedia:  
2'301'735 articles
- ▶ Albanian Wikipedia:  
76'745 articles
- ▶ Wikimedia Commons:  
53'766'895 files



A scenic landscape featuring a road with a guardrail in the foreground, rolling hills in the middle ground, and snow-capped mountains in the background under a clear blue sky.

WIKIPEDIA

---

**CATEGORIES**

## WAYS TO FIND CONTENT

- ▶ Google
- ▶ Links
- ▶ Wikipedia search tool

and:

- ▶ Categories



# WHY CATEGORIES

Systematic structure, that allows ...

- ▶ Finding similar articles, files etc.
- ▶ Basis for analyzing statistically the content of the sites
- ▶ Tagging articles/files for internal purposes

# CATEGORIES? NEVER SEEN!



**WIKIPEDIA**  
The Free Encyclopedia

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)
- [Donate to Wikipedia](#)
- [Wikipedia store](#)

Interaction

- [Help](#)
- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact page](#)

Tools

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Permanent link](#)
- [Page information](#)

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article

[Talk](#)

Read

[View source](#)

[View history](#)

## Tirana

From Wikipedia, the free encyclopedia

**Tirana** (/tɪˈrɑːnə/ <sup>ⓘ</sup> <sup>ⓘ</sup> [listen](#)); Albanian pronunciation: [tiˈraːna]; **Gheg Albanian**: *Tirona*) is the **capital** and largest city of the **Republic of Albania**.

Tirana is located in the center of Albania and is enclosed by mountains and hills, with **Dajt** on the east and a slight valley on the northwest overlooking the **Adriatic Sea** in the distance. Due to its location within the **Plain of Tirana** and the close proximity to the **Mediterranean Sea**, the city is particularly influenced by a **Mediterranean** seasonal **climate**. It is among the wettest and sunniest cities in Europe, with 2,544 hours of sun per year.<sup>[3][4]</sup>

Tirana flourished as a city in 1614 but the region that today corresponds to the city's territory has been continuously inhabited since the **Iron Age**. The city's territory was inhabited by several **Illyrian tribes** but had no importance within **Illyria**. Indeed, it was annexed by **Rome** and became an integral part of the **Roman Empire** following the **Illyrian Wars**. The heritage of that period is still evident and represented by the **Mosaics of Tirana**. Later, in the 5th and 6th centuries, a **Paleochristian** basilica was built around this site.

After the Roman Empire split into East and West in the 4th century, its successor the **Byzantine Empire** took control over most of Albania, and built the **Petrelë Castle** in the reign of **Justinian I**. The city was fairly unimportant until the 20th century, when the **Congress of Lushnjë** proclaimed it as Albania's capital, after the **Albanian Declaration of Independence** in 1912.






# CATEGORIES? ONE OF WIKIPEDIA'S BEST KEPT SECRETS!

- Pojani, Dorina (2011). "From Carfree to Carfull: the Environmental and Health Impacts of Increasing Private Motorisation in Albania". *Journal of Environmental Planning and Management*. **54** (3): 319–335. doi:10.1080/09640568.2010.506076
- Pojani, Dorina (2011). "Urban and Suburban Retail Development in Albania's Capital After Socialism". *Land Use Policy*. **28** (4): 836–845. doi:10.1016/j.landusepol.2011.02.001

## External links

- [Municipality of Tirana](#) (in Albanian)
- [Visit Tirana Tourism Portal](#)
- ["Tirana"](#). *Encyclopædia Britannica* (11th ed.). 1911.

 Wikivoyage has a travel guide for [Tirana](#).

 Wikimedia Commons has media related to [Tirana](#).

## Topics related to Tirana [show]

### Authority control

BNF: [cb12460512r](#) (data) · GND: [4270905-2](#) · ISNI: [0000 0001 1431 9723](#) · LCCN: [n82213241](#) · MusicBrainz: [538e3b99-6e61-4fd0-b2ab-e3bbd507b18a](#) · NKC: [ge491566](#) · SUDOC: [067106854](#) · VIAF: [167355723](#) · WorldCat Identities (via VIAF): [167355723](#)

Categories: [Tirana](#) | [Capitals in Europe](#) | [Populated places established in 1614](#) | [Municipalities in Tirana County](#) | [Gegëri](#) | [Cities in Albania](#) | [1614 establishments in the Ottoman Empire](#)

This page was last edited on 19 April 2019, at 10:51 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

**CATEGORIES ARE HELPFUL**

**USE THEM!**

# BUT HOW?

- ▶ adding code:  
[[category:Tirana]]
- ▶ add HotCat
  - > Preferences > Gadgets > Editing
  - > HotCat
- ▶ Every Wikipedia page should belong to at least one category.






CATEGORIES

---

**CHOOSING THE RIGHT ONE**

# CHOOSING TOPICS – A RECENT EXAMPLE

Browser tabs: Comm..., File:Ne..., Catego..., File:Kis..., W Wikiped..., File:Ujë..., File:Ujë..., W Wikiped..., W Liste d..., W Benutz..., File:Tir..., W Eurasis..., W E...

	Version vom	Vorschaubild	Maße	Benutzer	Kommentar
aktuell	<a href="#">23:02, 7. Mai 2019</a>		4.608 × 3.456 (6,01 MB)	<a href="#">Rrozafa</a> ( <a href="#">Diskussion</a>   <a href="#">Beiträge</a> )	User created page with UploadWizard

- [Eine neue Version dieser Datei hochladen](#)

### Dateiverwendung

Die folgende Seite verwendet diese Datei:

- [User:OgreBot/Uploads by new users/2019 May 07 21:00](#)

### Metadaten

Diese Datei enthält weitere Informationen, die in der Regel von der Digitalkamera oder dem verwendeten Scanner stammen. Durch nachträgliche Bearbeitung der Originaldatei können einige Details verändert werden.

Hersteller	<a href="#">NIKON</a>
Modell	<a href="#">COOLPIX S8200 (en)</a>
Belichtungsdauer	1/60 Sekunden (0,016666666666667)
Blende	f/4,5
Film- oder Sensorempfindlichkeit (ISO)	280
Erfassungszeitpunkt	07:56, 16. Aug. 2014
Brennweite	14,1 mm

[Erweiterte Details einblenden](#)

**IMAGE: RROZAFa CC-BY-SA-4.0**  
COPIED FROM COMMONS:FILE:KISHA\_NË\_THETH.JPG – WIKIMEDIA COMMONS

**CATEGORIES: TOURISM | NATURE | VILLAGES |  
MOUNTAINS | TRADITIONS**

**TAGS!**

**Kategorien:** [Tourism](#) | [Nature](#) | [Villages](#) | [Mountains](#) | [Traditions](#)  
**Versteckte Kategorien:** [Natural heritage sites in Albania with known IDs](#) | [CC-BY-SA-4.0](#) | [Self-published work](#) | [Images from Wiki Loves Earth 2019](#) | [Images from Wiki Loves Earth 2019 in Albania](#) | [Landscape](#)  
[Uploaded via Campaign:wle-al](#)

# NO TAGGING!

- ▶ Wikipedia is 18 years old!
- ▶ (Hash)tags didn't exist at that time  
– they only became popular after the introduction in Delicious in 2003, Flickr soon afterwards and Twitter in 2007/09
- ▶ Tags are not part of Wikipedia, so:
- ▶ You must not collect various describing terms




# FIND THE PERFECT CATEGORY

- ▶ Each page/article should be placed in the *most specific* category to which it logically belongs
- ▶ Category names are usually in English („OSCAL 2018 - group photos“)
- ▶ Category names are usually in plural („2018 events in Albania“) unless specific („OSCAL 2019“)



# CHOOSING TOPICS – DOING IT THE RIGHT WAY

Browser tabs: Comm..., File:Ne..., Catego..., File:Kis..., W Wikip..., File:Ujë..., File:Ujë..., W Wikip..., W Liste d..., W Benutz..., File:Tir..., W Eurasis..., W E

	Version vom	Vorschaubild	Maße	Benutzer	Kommentar
aktuell	<b>23:02, 7. Mai 2019</b>		4.608 × 3.456 (6,01 MB)	<a href="#">Rrozafa</a> ( <a href="#">Diskussion</a>   <a href="#">Beiträge</a> )	User created page with UploadWizard

[Eine neue Version dieser Datei hochladen](#)

### Dateiverwendung

Die folgende Seite verwendet diese Datei:

- [User:OgreBot/Uploads by new users/2019 May 07 21:00](#)

### Metadaten


Diese Datei enthält weitere Informationen, die in der Regel von der Digitalkamera oder dem verwendeten

Hersteller	<a href="#">NIKON</a>
Modell	<a href="#">COOLPIX S8200 (en)</a>
Belichtungsdauer	1/60 Sekunden (0,016666666666667)
Blende	f/4,5
Film- oder Sensorempfindlichkeit (ISO)	280
Erfassungszeitpunkt	07:56, 16. Aug. 2014
Brennweite	14,1 mm

[Erweiterte Details einblenden](#)

**IMAGE: RROZAF** CC-BY-SA-4.0  
COPIED FROM COMMONS:FILE:KISHA\_NË\_THETH.JPG – WIKIMEDIA COMMONS

**CATEGORIES: ~~TOURISM~~ | ~~NATURE~~ | ~~VILLAGES~~ | ~~MOUNTAINS~~ | ~~TRADITIONS~~**

**CATEGORIES: CHURCH OF THETH** 

**Kategorien:** ~~Tourism~~ | ~~Nature~~ | ~~Villages~~ | ~~Mountains~~ | ~~Traditions~~

**Versteckte Kategorien:** [Natural heritage sites in Albania with known IDs](#) | [CC-BY-SA-4.0](#) | [Self-published work](#) | [Images from Wiki Loves Earth 2019](#) | [Images from Wiki Loves Earth 2019 in Albania](#) | [Landscape](#) | [Uploaded via Campaign:wle-al](#)

## FIND THE PERFECT CATEGORY: EXAMPLE 2, STEP ONE

- ▶ Finding generic topics
- ▶ like tagging:

#STREET

#SHKODËR

#CASTLE

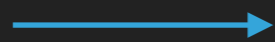
#MERCEDES



## FIND THE PERFECT CATEGORY: EXAMPLE 2, STEP TWO

- ▶ Translate it to specific categories:

**#STREET**



Streets in Shkodër

**#SHKODËR**



Streets in Shkodër

**#CASTLE**



Rozafa

**#MERCEDES**



Mercedes-Benz vehicles in  
Albania

2003 in Albania

Cycling in Albania

Street trees (in Albania)



# THE SYSTEM OF CATEGORIES

- ▶ Hierarchy - top down getting more and more specific
- ▶ Get as low as possible in the hierarchy
- ▶ Categories are usually combinations of locations and topics (e.g. Mercedes-Benz vehicles in Albania)
- ▶ Or: Category for a specific object



# HIERARCHY OF „CATEGORY:BUSHTRICA BRIDGE“

## RAILWAY BRIDGES IN ALBANIA

- A. BRIDGES IN ALBANIA
  - 1. BRIDGES BY COUNTRY
  - 2. TRANSPORT INFRASTRUCTURE IN AL
- B. RAIL TRANSPORT INFRASTRUCTURE IN ALBANIA
  - 1. RAIL TRANSPORT IN AL
  - 2. RAIL TRANSPORT INFRASTRUCTURE BY CTRY
  - 3. TRANSPORT INFRASTRUCTURE IN AL
- C. RAILWAY BRIDGES BY COUNTRY
  - 1. BRIDGES BY COUNTRY BY FUNCTION
  - 2. CATEGORIES BY COUNTRY
  - 3. RAIL TRANSPORT INFRASTRUCTURE BY CTRY
  - 4. RAILWAY BRIDGES
    - ▶ BRIDGES BY FUNCTION
    - ▶ RAIL TRANSPORT INFRASTRUCTURE

## RROGOZHINË— POGRADEC RAILWAY

- A. RAILWAY LINES IN ALBANIA
  - 1. RAIL TRANSPORT INFRASTRUCTURE IN AL
    - ▶ RAIL TRANSPORT IN AL
    - ▶ RAIL TRANSPORT INFRASTRUCTURE BY CTRY
  - 2. RAILWAY LINES BY COUNTRY
    - ▶ RAILWAY LINES
    - ▶ RAIL CATEGORIES BY COUNTRY
    - ▶ RAIL TRANSPORT INFRASTRUCTURE BY CTRY

## LIBRAZHD DISTRICT

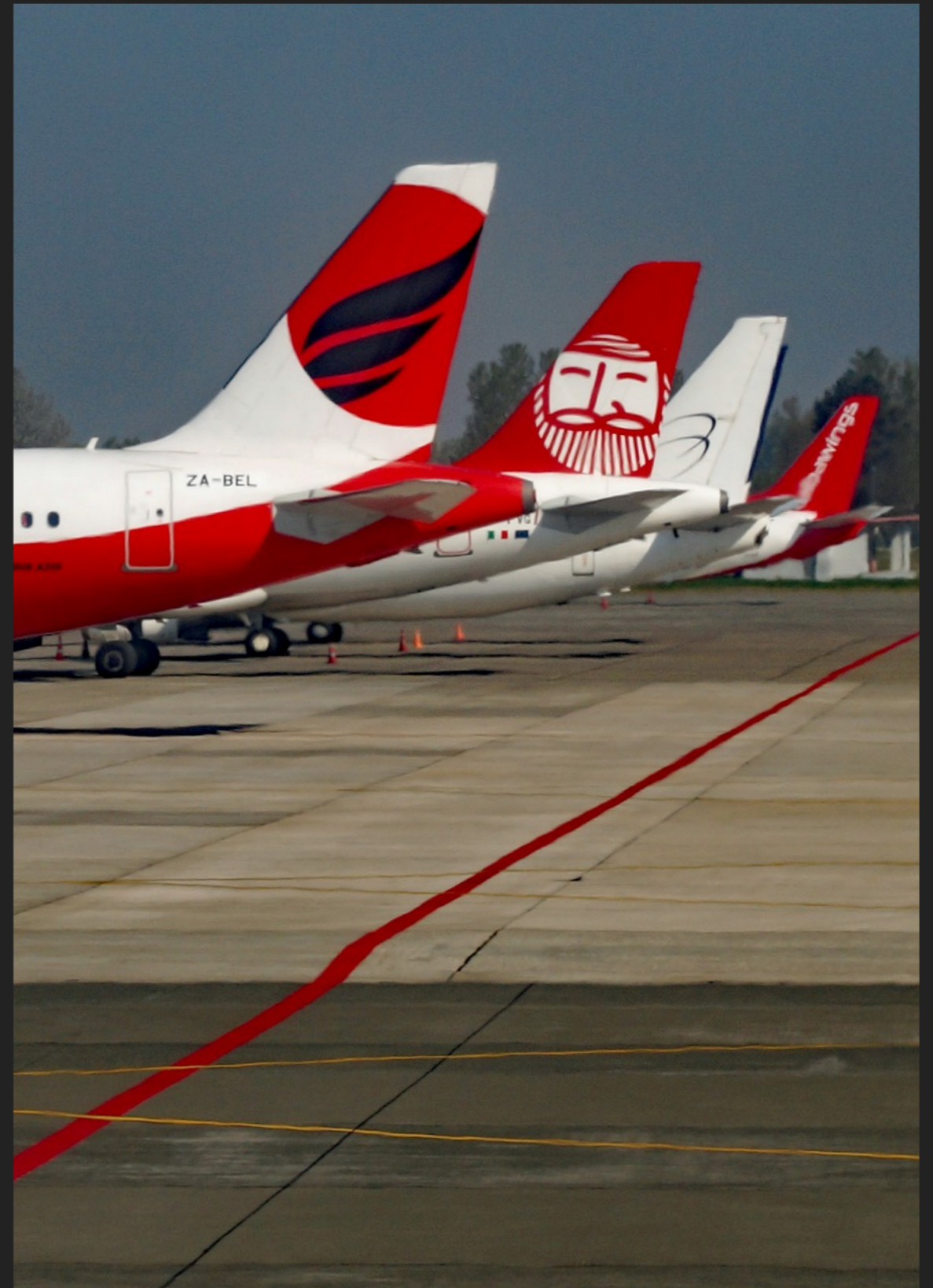
- A. ELBSAN COUNTY
  - 1. COUNTIES OF AL
    - ▶ SUBDIVISIONS OF AL
    - ▶ COUNTIES BY CTRY
- B. DISTRICTS OF ALBANIA
  - 1. SUBDIVISIONS OF ALBANIA
    - ▶ POLITICAL GEOGRAPHY OF ALBANIA
      - GEOGRAPHY OF AL
      - POL. GEOG. BY CTRY
      - POLITICS OF ALBANIA
    - ▶ COUNTRY SUBDIVISIONS BY COUNTRY
      - CTRY SUBDIVISIONS
      - POL. GEOG. BY CTRY
    - ▶ COUNTRY SUBDIVISIONS IN EUROPE
      - TERRITORIAL ENTITIES IN EUROPE
      - CTRY SUBDIVISIONS BY CONTINENT

## BRIDGES COMPLETED IN 1973

- A. 1970S BRIDGES
  - 1. 1970S ARCHITECTURE
  - 2. TRANSPORT IN THE 1970S
  - 3. 20TH-CENTURY BRIDGES
- B. 1973 IN TRANSPORT
  - 1. 1973 BY TOPIC
  - 2. TRANSPORT IN THE 1970S
  - 3. TRANSPORT BY YEAR
- C. BUILT IN 1973
  - 1. 1970S ARCHITECTURE
  - 2. 1973 IN CONSTRUCTION
  - 3. 1973 WORKS
- D. BRIDGES BY YEAR OF COMPLETION
  - 1. BRIDGES BY YEAR
    - ▶ BRIDGES BY DATE
  - 2. BUILDINGS BY YEAR OF COMPLETION
    - ▶ BUILDINGS BY DATE
    - ▶ CATEGORIES BY YEAR
    - ▶ ART BY YEAR
    - ▶ ESTABLISHMENTS BY YR

# MORE THAN ONE CAT

- ▶ Where?
- ▶ What?
- ▶ When?
- ▶ Who?
- ▶ Which status? (e.g. ruins, protection)
- ▶ Technical (small) cats, often added automatically:  
license, status, author, source, issues  
etc.



# CREATING CATEGORIES

- ▶ is allowed
- ▶ 5-10 articles/files
- ▶ copy structure of other categories (e.g. other countries)
- ▶ add to Wikidata and Wikipedia articles



**DON'T USE TAGS —  
FIND THE RIGHT BOX (OR TWO  
OR THREE)**





CATEGORIES

---

**STRUCTURED DATA**

# CHALLENGES ON COMMONS

- ▶ Categories have disadvantages
- ▶ Files are hard to search
- ▶ Descriptions are usually not translated
- ▶ Metadata is not structured
- ▶ Most information is not readable for machines
- ▶ Hard to manage 53+ millions of files without structured data



# THE FUTURE

- ▶ **Structured data** for media files on Wikimedia Commons
  - > search, > translations,
  - > related information, > import,
  - > smart suggestions
- ▶ easier for users
  - > analysis > machine-readable
- ▶ 3rd party institutions
  - > re-use, archive



# STRUCTURED DATA

- ▶ Introduction step by step:
- ▶ file information (January 2019)
- ▶ depicts (23 April 2019)
- ▶ to be extended with more features



# FILE INFORMATION

- ▶ Multilingual file captions
- ▶ Short description about a file
- ▶ Captions are simple and short. While descriptions can be very expansive, captions are limited to 255 characters.
- ▶ Part of structured data: In technical terms, a description is plain wikitext wrapped in a language template, wrapped again in an Information template
- ▶ see c:Commons:File captions

WIKIMEDIA COMMONS

Main page  
Welcome  
Community portal  
Village pump  
Help center

Sprachauswahl  
English

Participate  
Upload file  
Recent changes  
Latest files  
Random file  
Contact us

Print/export  
Download as PDF

Tools  
What links here  
Related changes  
Special pages  
Permanent link  
Page information  
Cite this page  
Nominate for deletion

File Discussion

File:Rruga Kombit (WPWTR17).jpg

From Wikimedia Commons, the free media repository

File File history File us

Size of this preview: 800 × 532 pixels. Other resolutions: 320 × 213 pixels | 640 × 425 pixels | 1,024 × 680 p  
Original file (4,912 × 3,264 pixels, file size: 1.83 MB, MIME type: image/jpeg); ZoomViewer: flash/no flash

Open in Media Viewer

File information Structured data

Captions Edit

English Highway "Rruga e Kombit" in Northern Albania

See 2 more languages

Summary [ edit ]

Description	English: Rruga e Kombit in Mirdita
Date	December 2017
Source	Own work
Author	Albinfo

The making of this doc  
For all the files cor

العربية | বাংলা | Čeština | Deutsch | English | Esperanto | Espa

# DEPICTS

- ▶ Kind of tagging
- ▶ Connecting files with structured data
- ▶ Searchable for machines
- ▶ see `c:Commons:Depicts`

The screenshot shows the Wikimedia Commons interface for the file `File:Rruga Kombit (WPWTR17).jpg`. The page includes a navigation sidebar on the left with links like 'Main page', 'Welcome', and 'Community portal'. The main content area features the file name, a description 'From Wikimedia Commons, the free media repository', and a large image of a road winding through a mountainous landscape. Below the image, there is a 'File information' section with tabs for 'File information' and 'Structured data'. A red arrow points to the 'Structured data' tab. Under the 'Structured data' tab, there is a section titled 'Items portrayed in this file' with a search box and a list of items, including 'A1 (Albania)' with the Wikidata ID 'Q751952'.

## HOW TO CHOOSE DEPICTS

- ▶ Again: NO GENERIC TAGGING!
- ▶ No
  - Tourism
  - Nature
  - Villages
  - Mountains
  - Traditions



IMAGE: RROZAFI CC-BY-4.0

COPIED FROM COMMONS:FILE:NEGLECT 2.JPG – WIKIMEDIA COMMONS

# HOW TO CHOOSE DEPICTS

## ▶ NO „FUN“ TAGGING!

File information	Structured data
<b>Items portrayed in this file</b> <i>depicts</i>	
<input type="text" value="uncle tom's cabin"/>	
Uncle Tom's Cabin <i>anti-slavery novel by Harriet Beecher Stowe</i>	
Uncle Tom's Cabin <i>1918 film by J. Searle Dawley</i>	
Uncle Tom's Cabin <i>1965 film by Géza von Radványi</i>	
Uncle Tom's Cabin <i>1927 film by Harry A. Pollard</i>	
Uncle Tom's Cabin <i>1914 film by William Robert Daly</i>	
Uncle Tom's Cabin <i>1910 film by Barry O'Neil</i>	
Uncle Tom's Cabin <i>1991 single by Warrant</i>	



IMAGE: RROZAFa CC-BY-4.0

COPIED FROM COMMONS:FILE:NEGLECT 2.JPG – WIKIMEDIA COMMONS



# DEPICTS

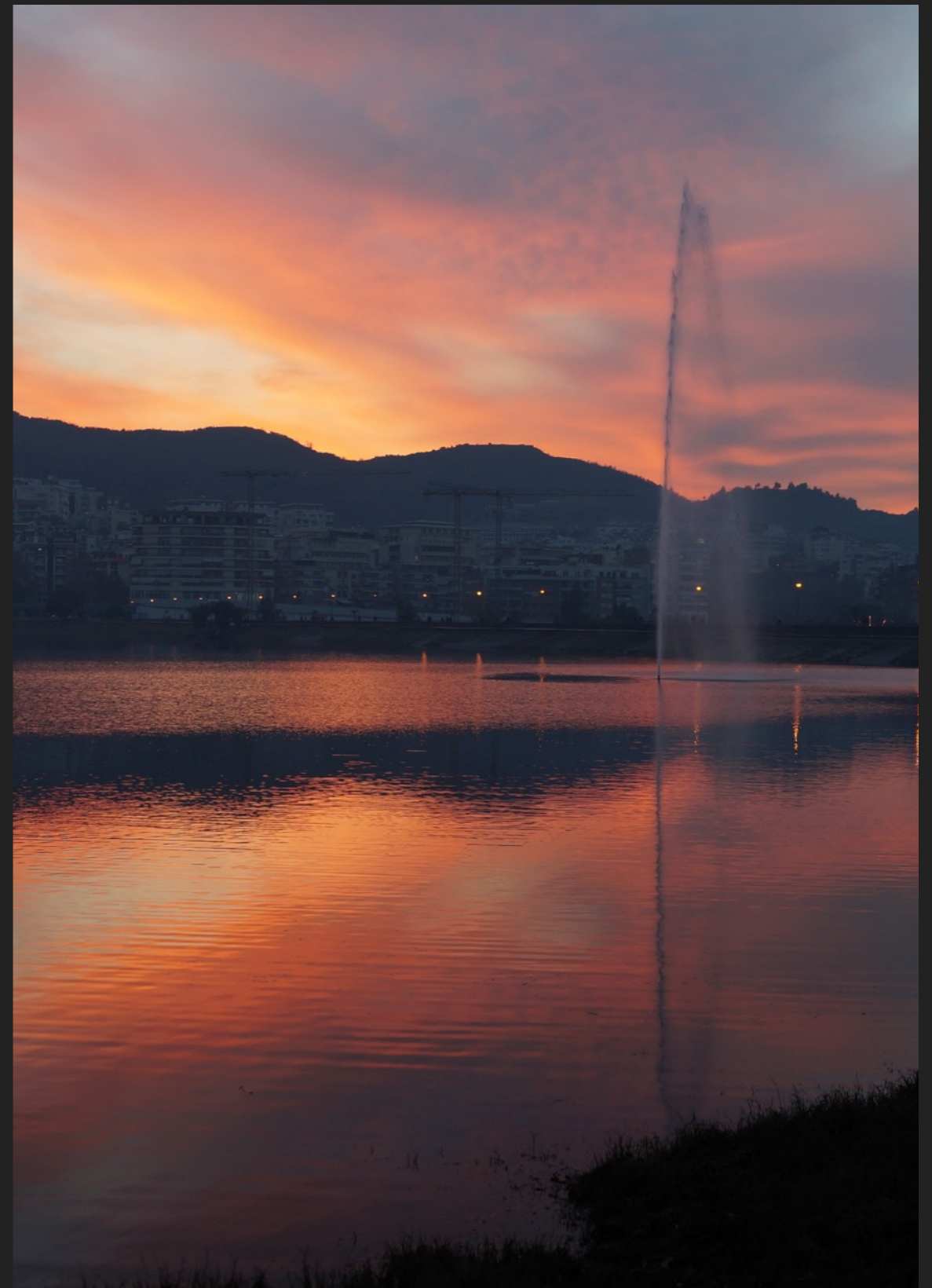
- ▶ Be as specific as you can with the primary tag.
- ▶ Every tag is connected with a Wikidata item.
- ▶ Generic tags should not currently be added.

The screenshot shows the Wikimedia Commons interface for the file 'File:Rruga Kombit (WPWTR17).jpg'. The page includes a navigation menu on the left with links like 'Main page', 'Welcome', and 'Upload file'. The main content area features a large image of a road in a mountainous region. Below the image, there is a section titled 'Items portrayed in this file' which contains a search bar and a list of Wikidata items. A red arrow points to the item 'A1 (Albania)' in this list.

**PLEASE ONLY TAG CONSERVATIVELY  
AT FIRST, WHILE THE COMMUNITY  
LEARNS HOW TO USE THIS TOOL  
BEST.**

## TRY IT

- ▶ We all need to learn
- ▶ Guidelines under progress
- ▶ Ask or check the help pages  
[commons.wikimedia.org/wiki/  
Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)
- ▶ 50+ millions of items need to be enriched with structured data



An aerial photograph of a coastal town nestled in a valley, with a clear blue bay and mountains in the background. The sky is a deep, clear blue. The foreground shows a grassy hillside with some rocks.

SUPPORT NEEDED?

---

**ASK!**