



Wikidata use in KDE's travel apps

Wikidata Data Reuse Days 2022

Volker Krause

vkrause@kde.org
[@VolkerKrause](https://twitter.com/VolkerKrause)





KTrip – Public Transport Journey Planner

The screenshot displays the KTrip application interface, divided into two main sections: 'Connections' and 'Details'.

Connections Section:

- ICE 706 (S42 S9):** 4:57 departure from München Hbf, 13:51 arrival at München Hbf, 18:48 arrival at Berlin Ostbahnhof (S). 2 changes.
- ICE 706 S 7:** 4:58 departure from München Hbf, 13:51 arrival at München Hbf, 18:49 arrival at Berlin Ostbahnhof (S). 1 change.
- ICE 504 (S42 S9):** 4:57 departure from München Hbf, 14:51 arrival at München Hbf, 19:48 arrival at Berlin Ostbahnhof (S). 2 changes.
- ICE 504 S 5:** 5:00 departure from München Hbf, 14:51 arrival at München Hbf, 19:51 arrival at Berlin Ostbahnhof (S). 1 change.
- ICE 704 S 3:** 4:56 departure from München Hbf, 15:50 arrival at München Hbf, 20:46 arrival at Berlin Ostbahnhof (S). 1 change.
- ICE 502 (S42 S9):** 4:57 departure from München Hbf, 16:51 arrival at München Hbf, 21:48 arrival at Berlin Ostbahnhof (S). 2 changes.
- ICE 502 S 7:** 4:58 departure from München Hbf, 16:51 arrival at München Hbf, 21:49 arrival at Berlin Ostbahnhof (S). 1 change.

Details Section:

ICE 706

- 13:51 München Hbf (24 connections)
- 18:19 Berlin Südkreuz (6 connections)
- Komfort Check-in possible (visit bahn.de/kci for more information)
- Please wear an FFP2 mask. You are legally required to do so
- Bordrestaurant
- From Bamberg as ICE 706 heading towards Hamburg-Altona

Walking (7 minutes)

S42 S 42

- 18:30 Berlin Südkreuz (12 connections)
- 18:43 Berlin Treptower Park (2 connections)
- Please wear an FFP2 mask. You are legally required to do so
- Number of bicycles conveyed limited
- 2nd class only
- Erfahrungsgemäß hat dieser Zug eine sehr hohe Auslastung zwischen Berlin-Neukölln und Berlin Treptower Park. Wir empfehlen Ihnen nach Möglichkeit eine andere Verbindung zu nehmen.

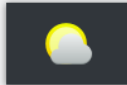
S9 S 9



KDE Itinerary – Digital Travel Assistant

My Itinerary

Friday, 10 Aug 2018



Temperature: 20.1 °C / 26.3 °C
Precipitation: 0 mm

✈ OS 232 TXL → VIE

17:05:00

Departure from Berlin - Tegel: 17:05:00

GERMANY



Terminal: - Gate: - Seat: -

Arrival at Wien: 10 Aug 2018 18:25:00

AUSTRIA



Temperature: 20.6 °C / 31.9 °C
Precipitation: 0 mm

🏠 a&o Wien Hauptbahnhof

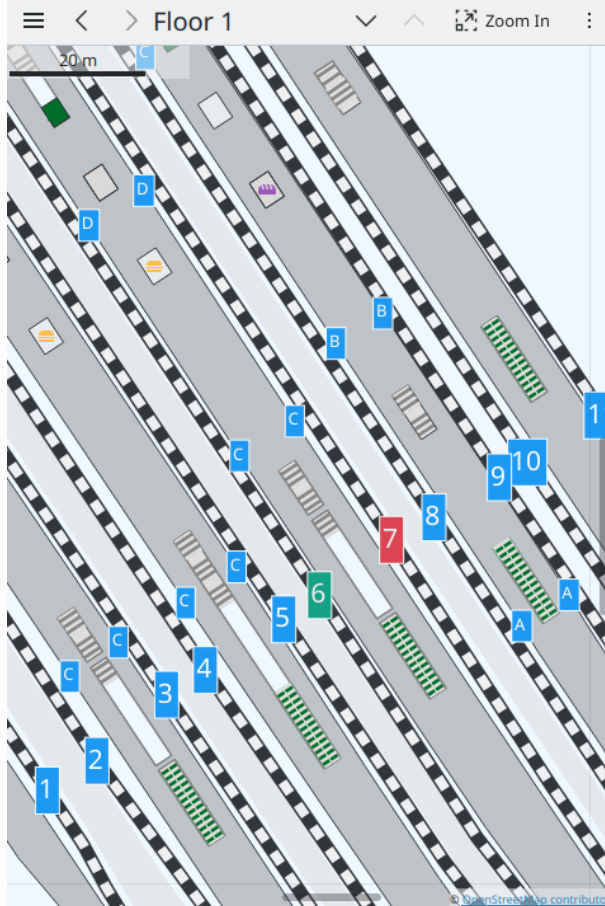
Sonnwendgasse 11
1100 Wien

AUSTRIA



Check-in time: 15:00:00

Check-out time: 18 Aug 2018 10:00:00



Vehicle Layout

Berlin Hbf (tief) - ICE 848 - 1/31/20 1:49 PM

Platform: 4

Coach: 34 Seat: 58

27 First class

31 Second class

32 Second class

33 Second class

34 Second class

35 Second class

D

C

B



Wikidata Use

- No online access for data
 - Public transport line logos are loaded on demand and then locally cached
- Data is pre-processed and shipped with the apps
 - Primary motivation: privacy
 - Side-effect: extra sanity checks/validation of the data
 - Allows expensive merging with OSM data

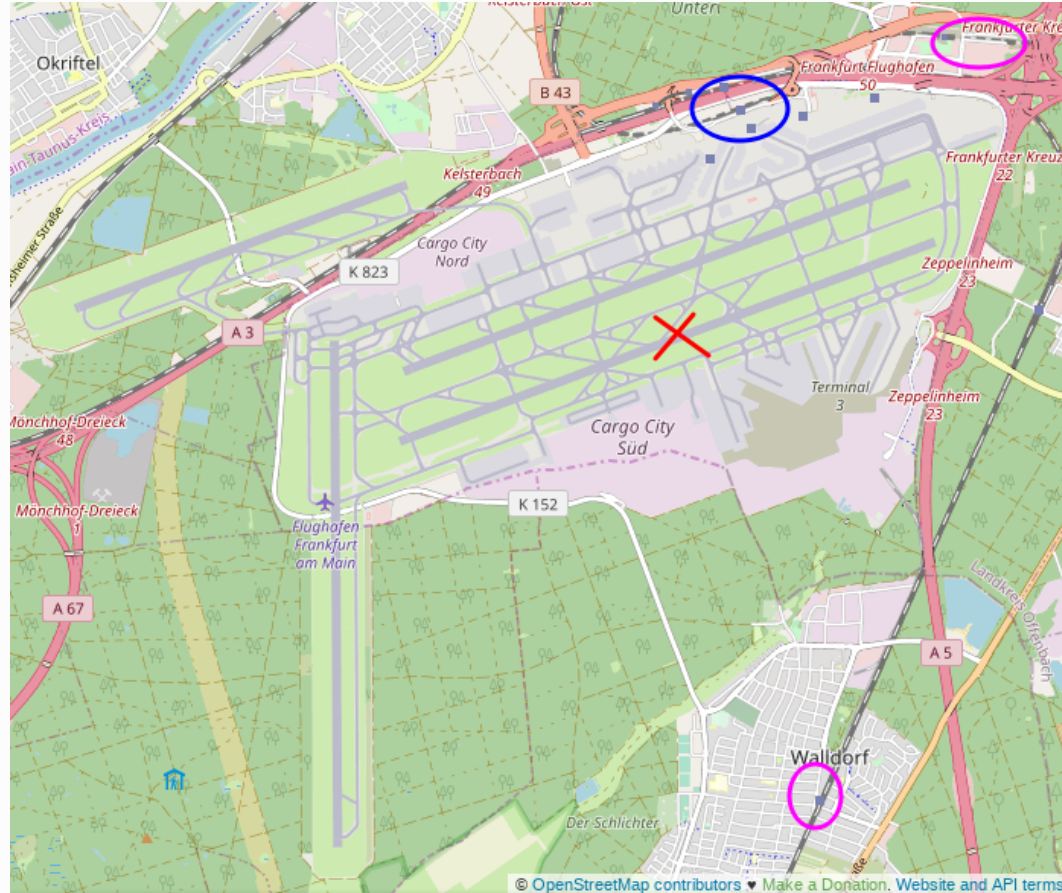


Airport Identification

- IATA airport code (P238)
 - Change over time and can be re-allocated
 - Exceptions to uniqueness...
 - Import errors in northern America and Argentina
 - Use for train stations
- Name
 - Can require complex disambiguation logic



Airport Location (P625)





Civil vs Military Airports

- Only airports with passenger services are relevant for us
- Irrelevant airports can impede name-based identification
- Civil, military and civil/military hybrid uses can change over time
 - Complex modeling
 - Name-based heuristic turned out to be easier



Train Stations

- Numerous identifier systems
 - Often operator- or country-specific
 - Specializations of P296
 - Newly added: P8181 (SNCF), P8448 (Benerail)
 - Proposed:
https://www.wikidata.org/wiki/Wikidata:Property_proposal/Via_Rail_station_code
- Name-based identification not reliable due to many (globally) ambiguous names and/or missing context



Train Stations - Open Issues

- IATA airport code (P238)
 - Separate property for train stations or constraint violation?
 - https://www.wikidata.org/wiki/Property_talk:P238#IATA_codes_for_train_stations
- Modeling multi-part / “virtual” stations
 - Technically separate stations considered as one by routing/ticketing
 - “Virtual” stations have their own identifiers



Country Information

- Country codes
 - ISO 3166-1 mappings moved to iso-codes data set
 - UIC country code
- Used power plugs
- Driving side
- The usual fun with countries...

📅 Donnerstag, 15.06.17

📍 Entering United Kingdom

People are driving on the wrong side.
No compatible power sockets: Type G

✈️ 4U 8462 TXL → LHR 12:55

Departure from Berlin-Tegel: 12:55

📅 Sonntag, 10.09.17

📍 Entering Switzerland

Some incompatible power sockets: Type J
Some incompatible power plugs: Schuko

✈️ LX 963 TXL → ZRH 06:15

Departure from Berlin Tegel: 06:45



Timezones

- Highly important for our use cases
- Specified in different formats in Wikidata (P421)
 - IANA timezone ids
 - UTC offsets for normal time/DST
- UTC offsets lack information about DST transitions
- Using a coordinate-based index rather than Wikidata now



Public Transport Lines

- Line logo and color (P154, P465)



- Mode or product logo





Public Transport Lines - Identification

- Names are highly ambiguous
- Use OSM to obtain a bounding box for each line
- Use name + coordinate to identify a line
- Relies on Wikidata <-> OSM mapping (P402)
 - Added/extended for a number of cities, in both directions



Public Transport Lines - Challenges

- Complex modeling
 - Distinction between modes of transport is very blurry
 - Unexpected things like historical lines or amusement rides
 - Level of detail varies
- Correct attribution for CC-BY licensed image assets

4:57 ICE 706 S42 S9
13:51 München Hbf 2 changes
18:48 Berlin Ostbahnhof (S)

4:58 ICE 706 S 7

The screenshot shows a public transport schedule. The first entry is '4:57 ICE 706 S42 S9', where 'S42' and 'S9' are circled in red. Below it are '13:51 München Hbf' with '2 changes' to the right, and '18:48 Berlin Ostbahnhof (S)'. A horizontal line separates this from the next entry, '4:58 ICE 706 S 7'. There is an upward-pointing arrow at the top right of the schedule area.



Thanks for you attention!

Questions?



References

- KDE Itinerary: <https://apps.kde.org/itinerary>
- KTrip: <https://apps.kde.org/ktrip>
- Travel document extraction: <https://invent.kde.org/pim/kitinerary>
- Airport locations: <https://volkerkrause.eu/2020/05/02/kde-itinerary-airport-locations.html>
- Public transport line meta data:
 - <https://volkerkrause.eu/2020/04/18/kpublictransport-line-metadata.html>
 - <https://invent.kde.org/libraries/kpublictransport>
- Timezone lookup:
<https://volkerkrause.eu/2020/05/23/kde-itinerary-coordinate-based-timezone-lookup.html>