



# When is my flight?

Semantic data extraction in KMail and Nextcloud Mail



**FOSDEM 2024** 

Volker Krause

- Hard to find relevant information in ad overloaded HTML email
- Tedious to manually enter all booking information into calendar/travel app/etc.
- Had to figure out my departure gate in an unreadable email, looked at source code, things got out of hand...



# schema.org annotations

```
"@context": "http://schema.org",
"@type": "FlightReservation",
"reservationNumber": "RXJ34P",
"underName": { "@type": "Person", "name": "Eva Green" },
"reservationFor": {
  "@tvpe": "Flight",
  "flightNumber": "110",
  "airline": { "@type": "Airline", "name": "United", "iataCode": "UA" },
  "departureAirport": { "@type": "Airport", "name": "San Francisco Airport",
                        "iataCode": "SFO" },
  "departureTime": "2027-03-04T20:15:00-08:00",
  "arrivalAirport": { "@type": "Airport", "iataCode": "JFK",
                      "name": "John F. Kennedy International Airport" },
  "arrivalTime": "2027-03-05T06:30:00-05:00"
```

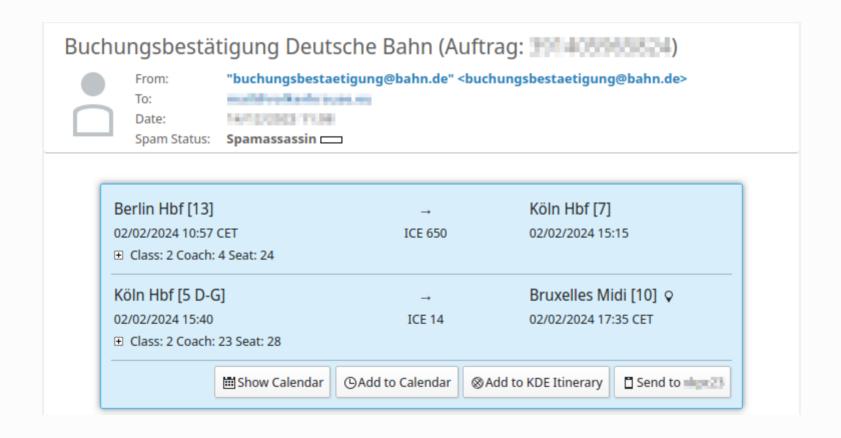
- JSON, Microdata or "creative" JSON
- Incomplete or incorrect data
  - Wrong timezones, mismatching address and coordinates, etc
- Limited use
  - Some airlines, few hotels, practically no train operators
  - Less common in Europe than in the US, very rare in Asia (?)

- Structured
  - Barcodes, in particular for international flights/trains
- Semi-structured
  - Apple Wallet passes
- Unstructured
  - Generic extractors for e.g. boarding passes
  - 130+ provides specific scripts

- Normalization
- Apply context and domain knowledge to resolve ambiguities
- Augment from other data sources
  - Airport/train station code lookup in Wikidata/OSM
  - Timezone from country/location
- Validation and plausibility checks

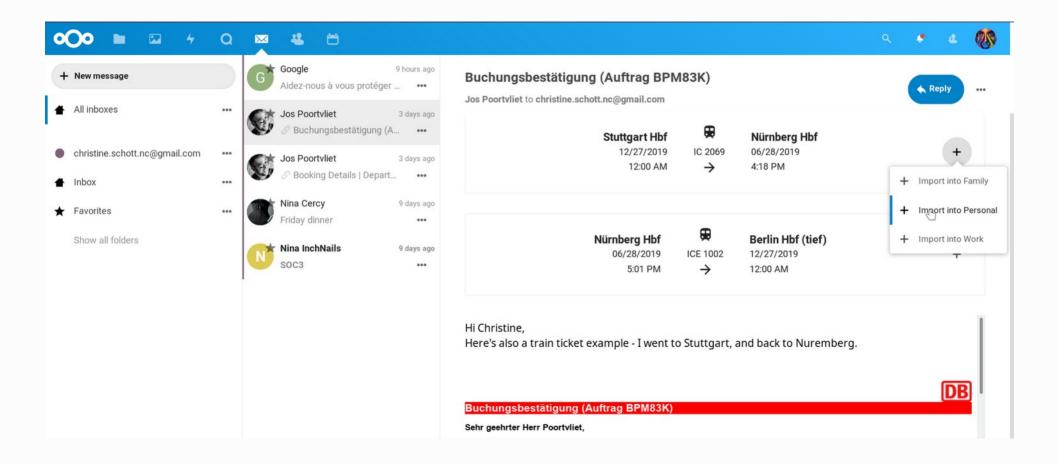


# Integration in KMail



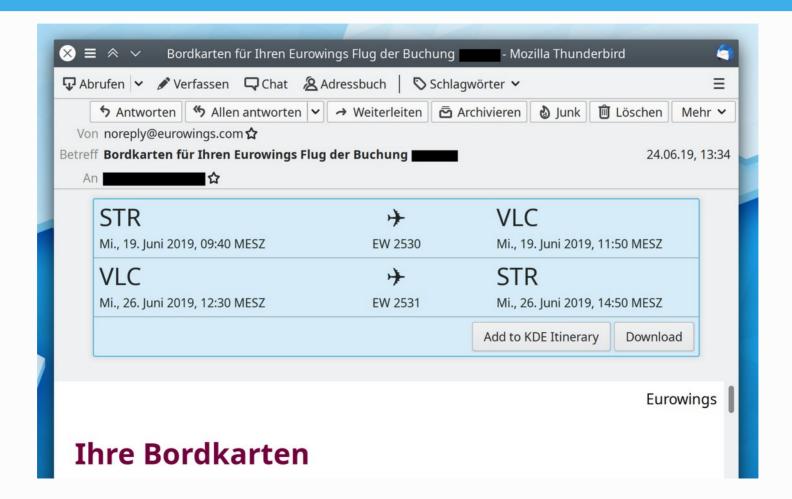


#### Integration in Nextcloud Mail



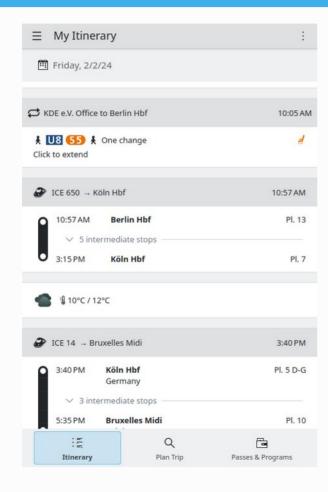


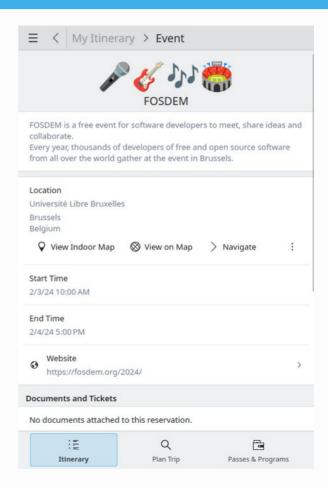
#### Integration in Thunderbird





# Integration in KDE Itinerary





- Semantic annotations in emails enable productivity features
- Incorrect data is worse than no data
- Please add semantic annotations to emails you produce!
- Donations of sample documents are very welcome:)



# Questions?

Meet KDE: KDE stand, building H, level 1

**Contact:** 

Email: vkrause@kde.org

Matrix: @vkrause:kde.org

Mastodon: @VolkerKrause@floss.social

- Semantic data extraction engine: https://invent.kde.org/pim/kitinerary
- KMail plug-in: https://invent.kde.org/pim/kdepim-addons
- Nextcloud Mail integration: https://github.com/nextcloud/mail
- Thunderbird plug-in: https://invent.kde.org/broulik/itinerary-travelbird
- KDE Itinerary: https://kde.org/for/travelers and https://apps.kde.org/itinerary