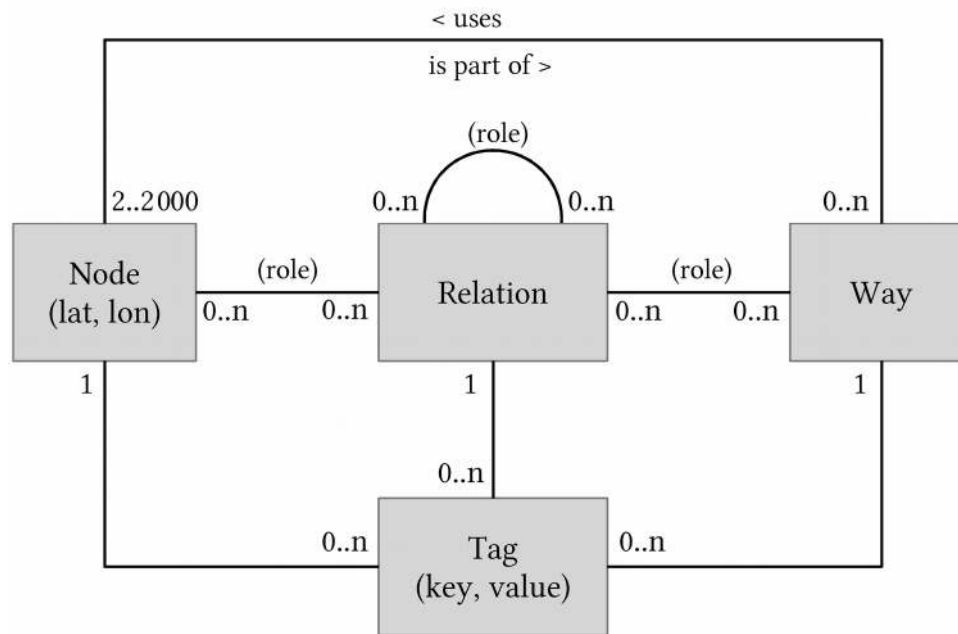
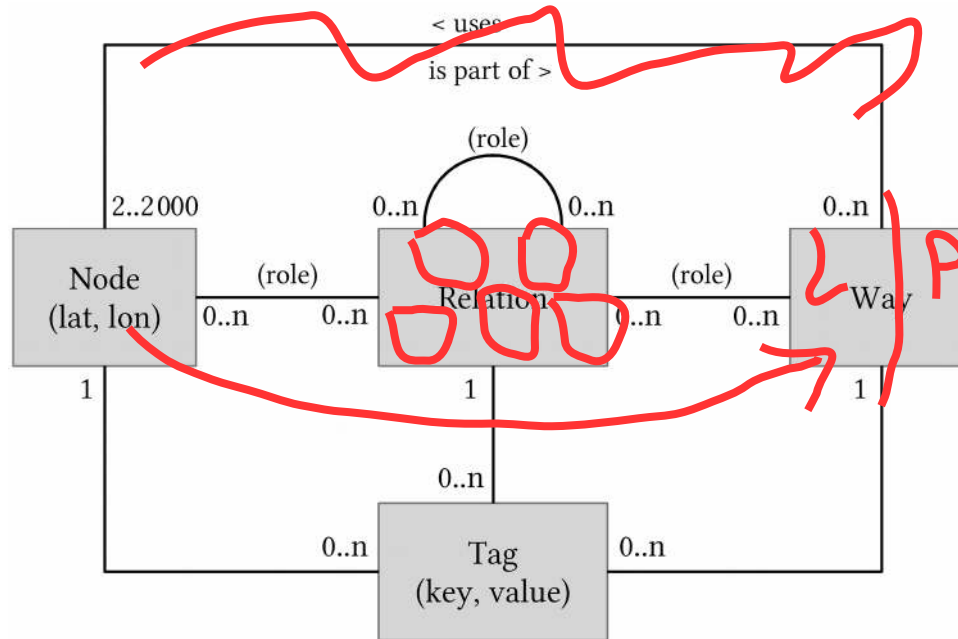


Modding the OSM Data Model



Jochen Topf

Modding the OSM Data Model



Jochen Topf

What we will talk about...

1. Background

Objects

Nodes

Ways

Relations

Tags

Object Identity and
Object Relationships

Locality

2. Problems

Way nodes

Way geometry

Relations

Areas

The way forward...

Background



ALBICCHE
Kilo-
2.00
3.00

SUSINE
Kilo-
3.00

PRUGINE
Kilo-
2.40
3.00

PESCHE
Kilo-
2.50
3.00

NATTEGNA
Kilo-
2.40
3.00

PESCHE-BIANCA
Kilo-
2.00

UVA
Kilo
3.50

Kilo-
4.50

Objects

Objects

Type (Node, Way, Relation)

ID

Tags

Version

Changeset

Timestamp

Uid/User

Nodes

Only objects that have a location!

Double duty as:

- Provide locations (no tags needed)
- Real Object (with tags, POI)
- Can be both (`highway=traffic_signals`)

Ways

Reference up to 2000 nodes

LineString or Polygon (or both)?

Relations

Bind other objects together.

Relations

Any number of members

Members have:

type

id

role

Members are mostly nodes or ways,
but can be other relations

Tags



Tags

Unlimited number of tags per object

Format: Key=Value

Key/Value each have up to 255 characters

Unique Keys

Tags have no types

Keys and Values are always strings

No structure beyond Key=Value

Tags have types

name = Москва  Any text

bridge = yes  Boolean

oneway = yes, no, -1  Bool? Enum?

highway = motorway, trunk, primary,  Enum

width = 125  Number

ref = I 40;US 270  List?

Keys and values have structure

Hierarchy with colon:

addr:street

name:ja

source:addr

Tag Types

maxspeed = 60

maxspeed = 55mph

maxspeed = walk

maxspeed = R0:rural

maxspeed = unknown

quirky, inconsistent, sometimes hard to use
but flexible!

→ need more „best practice“

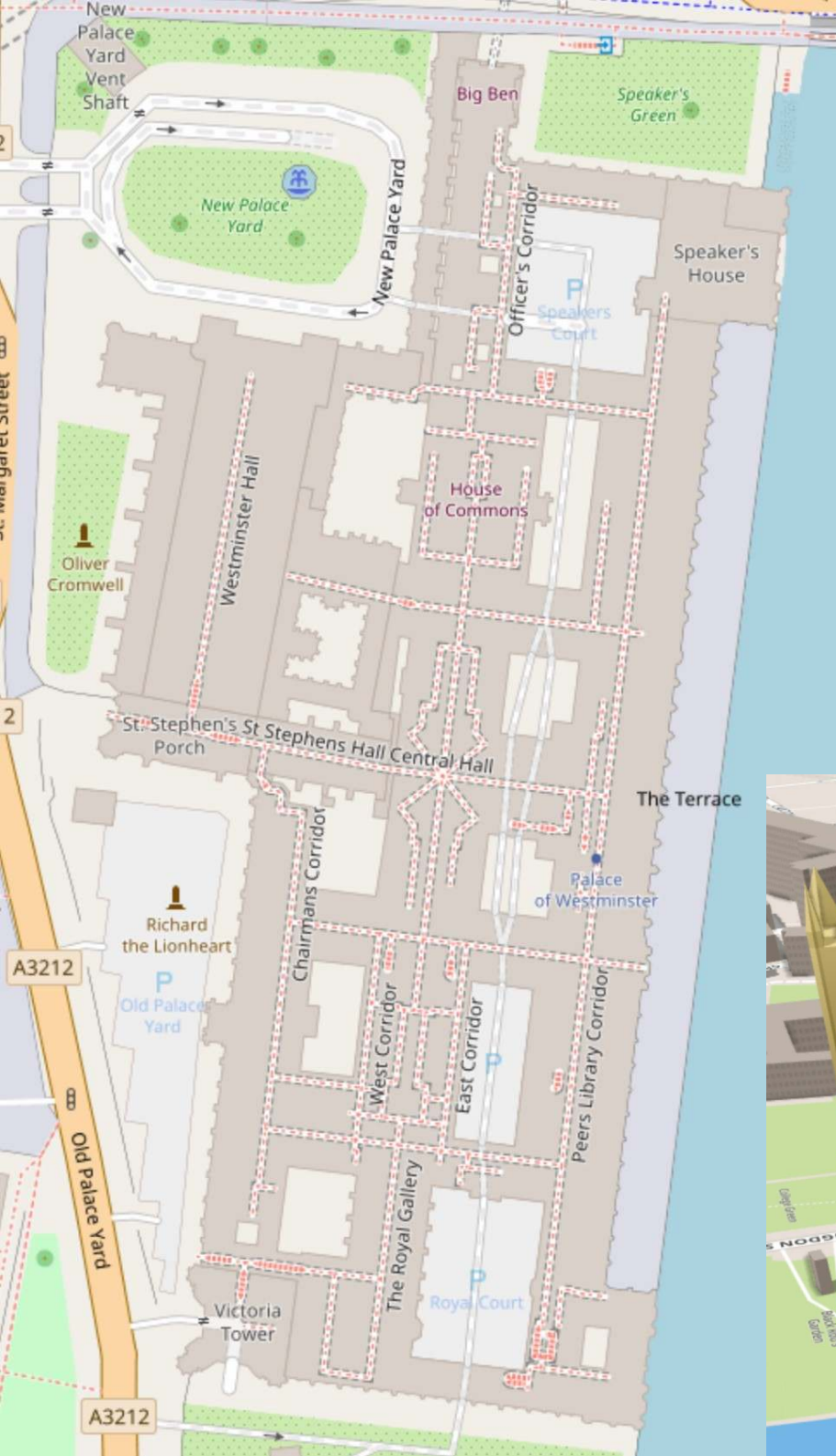
Object Identity



Palace of Westminster

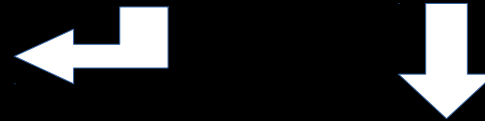


or... Houses of Parliament
or...



OSM

Lots of objects...
whatever is needed
to make
this and this



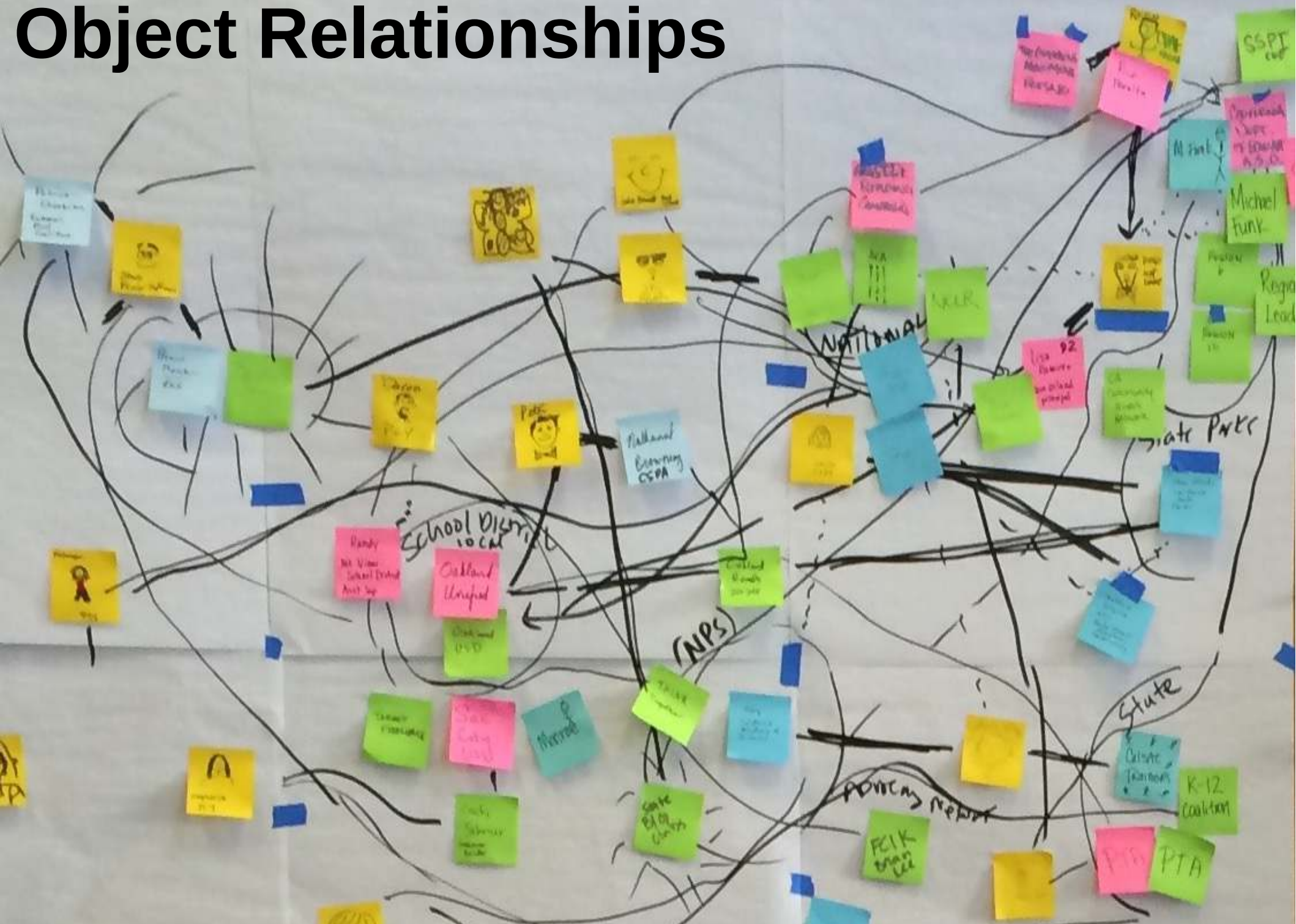
Object Identity in OSM

not about real-world objects

ID used to allow...

1. editing of OSM objects
2. relating one OSM object to another

Object Relationships



Object Relationships

Explicit (using ID):

Way → Node

Relation → Member Node, Way, Relation

Implicit:

through tags

through geography

Object Relationships: Explicit

No ambiguity, no duplication.

Often expensive to handle.

Explicit relationships break often.
(For instance when a way is split.)

Object Relationships: Implicit

Make objects independent.

Allows working without DB.

Can lead to inconsistent data.
(Which we can often detect and fix.)

Object Relationships

All motorways in Germany

Relation (Collection)?

All ways tagged highway=motorway
inside the Germany relation?

Object Relationships

Relation type=associatedStreet

Relation type=collection

is_in=...

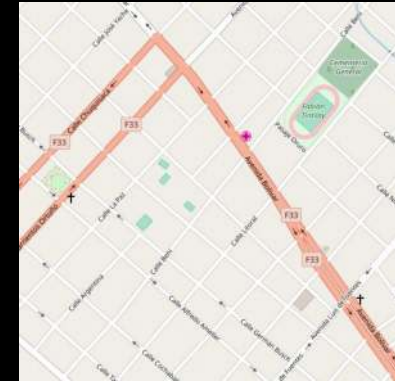
Object Relationships and Changes

Relationships make changes

- hard to understand
- hard to process

(worst offender: geometry change of way
only changes nodes)

Locality



Locality is important for a Geodatabase

because it allows splitting up
and
divide and conquer approach

Explicit references break locality

Problems

Problem Size

4600 million nodes

510 million ways

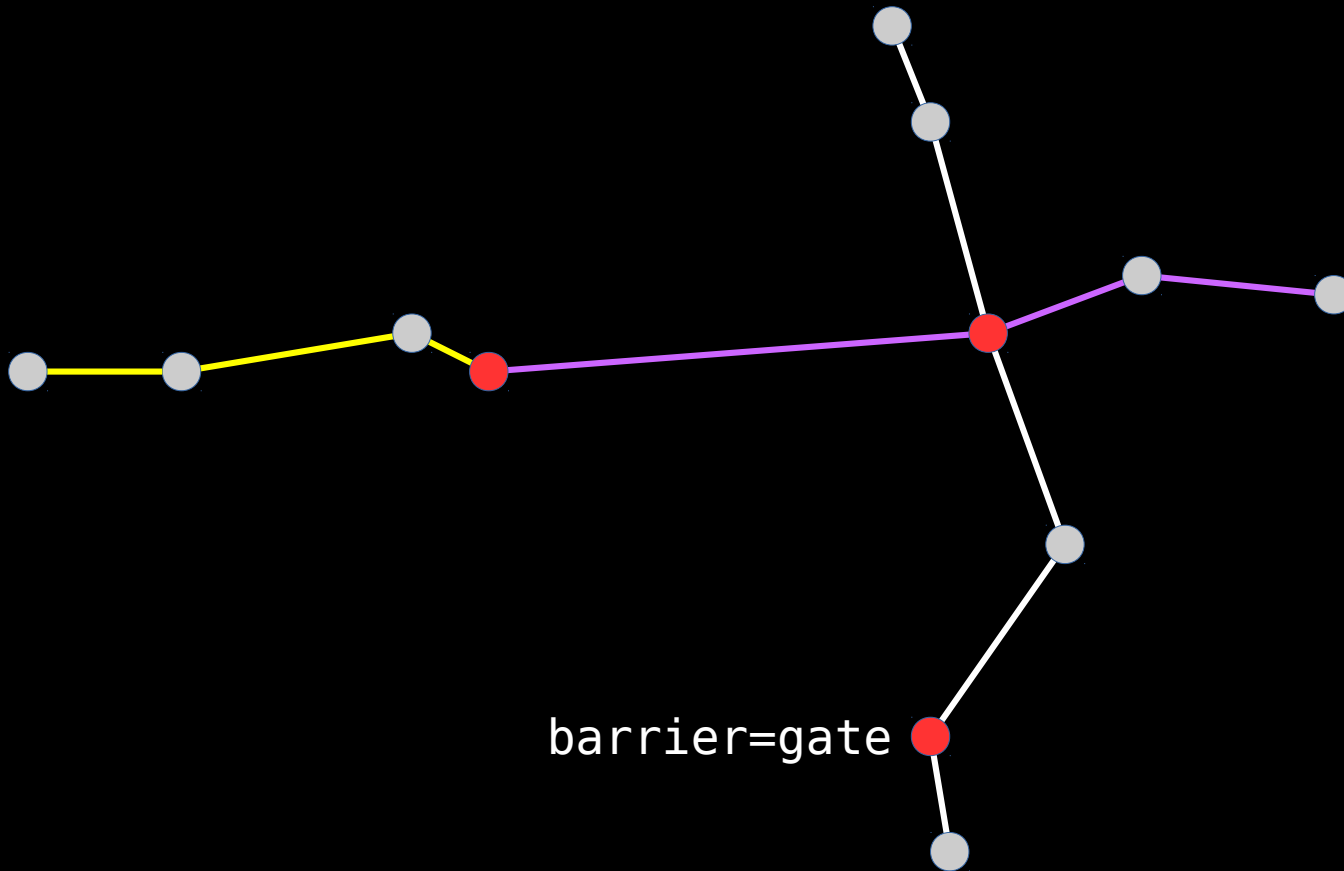
6 million relations

Way nodes

There are 4.6 billion nodes

98% of them only there
to provide locations to ways

Way nodes



Way nodes

- 43 GByte RAM needed for location index*
- Changes difficult to handle
- Prevent streaming operation
- Geometry checks in API expensive

*July 2018

Nodes and Ways

```
<osm>
  <node id="10" lat="1.1" lon="1.0"/>
  <node id="11" lat="1.2" lon="1.1"/>
  <node id="12" lat="1.3" lon="1.0">
    <tag k="barrier" v="gate"/>
  </node>
  <node id="13" lat="1.4" lon="1.1"/>

  <way id="20">
    <nd ref="10"/>
    <nd ref="11"/>
    <nd ref="12"/>
    <nd ref="13"/>
    <tag k="highway" v="residential"/>
  </way>
</osm>
```


Nodes and Ways

```
<osm>
  <node id="10" lat="1.1" lon="1.0"/>
  <node id="11" lat="1.2" lon="1.1"/>
  <node id="12" lat="1.3" lon="1.0">
    <tag k="barrier" v="gate"/>
  </node>
  <node id="13" lat="1.4" lon="1.1"/>

  <way id="20">
    <nd ref="10" lat="1.1" lon="1.0"/>
    <nd ref="11" lat="1.2" lon="1.1"/>
    <nd ref="12"/>
    <nd ref="13" lat="1.4" lon="1.1"/>
    <tag k="highway" v="residential"/>
  </way>
</osm>
```

Nodes and Ways

```
<osm>
  <node id="10" lat="1.1" lon="1.0"/>
  <node id="11" lat="1.2" lon="1.1"/>
  <node id="12" lat="1.3" lon="1.0">
    <tag k="barrier" v="gate"/>
  </node>
  <node id="13" lat="1.4" lon="1.1"/>

  <way id="20">
    <nd ref="10" lat="1.1" lon="1.0"/>
    <nd ref="11" lat="1.2" lon="1.1"/>
    <nd ref="12" lat="1.3" lon="1.0"/>
    <nd ref="13" lat="1.4" lon="1.1"/>
    <tag k="highway" v="residential"/>
  </way>
</osm>
```

Nodes and Ways

```
<osm>
  <node id="10" lat="1.1" lon="1.0"/>
  <node id="11" lat="1.2" lon="1.1"/>
  <node id="12" lat="1.3" lon="1.0">
    <tag k="barrier" v="gate"/>
  </node>
  <node id="13" lat="1.4" lon="1.1"/>

  <way id="20">
    <nd ref="10" lat="1.1" lon="1.0"/>
    <nd ref="11" lat="1.2" lon="1.1"/>
    <nd ref="12" lat="1.3" lon="1.0"/>
    <nd ref="13" lat="1.4" lon="1.1"/>
    <tag k="highway" v="residential"/>
  </way>
</osm>
```



Editors can handle this!

Editors download a complete area anyway

Snapping to other objects still possible

Limited resolution of Coordinates

Testing this now

```
osmium add-locations-to-ways
```

```
https://osmcode.org/osmium-tool
```

To be determined...

Do we keep nodes where ways connect?

Do we only allow ways to connect at the ends?

What about tagged nodes in ways?

What about common lines in areas?

...



landuse = presidential

barrier = fence

Processing

Lots of software needs lists of area tags:

- osm2pgsql
- editors
- exporters to other formats
- ...

...they are all incomplete and
make use of niche tags more difficult

Closed Ways

closed:	355.295.258	100%
linestring:	1.986.253	0%
polygon:	342.417.123	96%
both:	520.410	0%
no tags:	6.287.769	1%
error:	4.016	0%
unknown:	4.079.687	1%

*using 187 filter rules

Solution: Add an area flag

```
<way id="20" area="no">  
  <nd ref="10"/>  
  <nd ref="11"/>  
  <tag k="highway" v="residential"/>  
</way>
```

```
<way id="20" area="yes">  
  <nd ref="10"/>  
  <nd ref="11"/>  
  <tag k="landuse" v="forest"/>  
</way>
```

Introducing the Area Flag

Software that doesn't understand it, can ignore it.

Set to "unknown" initially.

Set automatically where possible,
ask mappers to do the rest.

Problems with Relations

Relations are very flexible!

Great for experimenting!



E2 Europäischer Fernwanderweg E2

[Show on map](#)
[GPX](#)
[KML](#)

Stranraer (UK) - Birmingham (UK) - Calais (FR) - Zoersel (BE) / Eastroute: Stranraer (UK) - Harwich (UK) - Rotterdam (NL) - Zoersel (BE) / Both Routes: Zoersel (BE) - Maastrich (NL) - Luxembourg (LUX) - Lake Geneva (CH) - Nice (FR)

Erfasste Länge: 4686 km
Offizielle Länge: 4850 km
Betreiber: European Ramblers Association

[Website](#)
[Wikipedia](#)

[Höhenprofil](#)

[Abschnitte](#)

[OpenStreetMap-Tags](#)

description	Stranraer (UK) - Birmingham (UK) - Calais (FR) - Zoersel (BE) / Eastroute: Stranraer (UK) - Harwich (UK) - Rotterdam (NL) - Zoersel (BE) / Both Routes: Zoersel (BE) - Maastrich (NL) - Luxembourg (LUX) - Lake Geneva (CH) - Nice (FR)
distance	4850
name	European long distance path E2
name:be	Europese wandelroute E2
name:de	Europäischer Fernwanderweg E2
name:en	European long distance path E2
name:fr	Chemin de randonnée Européen E2
name:nl	Europese wandelroute E2
name:pl	Europejski długodystansowy szlak pieszy E2
network	iwn
operator	European Ramblers Association
operator:be	Grote Routepaden
operator:fr	Fédération Française de la Randonnée Pédestre
operator:lux	Département Du Tourisme
operator:nl	Stichting Wandelplatform
operator:uk	Long Distance Walkers Association
osmc:symbol	blue:blue::E2:yellow
ref	E2
route	hiking
type	superroute
website	http://www.era-ewv-ferp.com/index.php?topmenu_id=29&id=29&page_id=100&module=text

Problems with Relations

but:

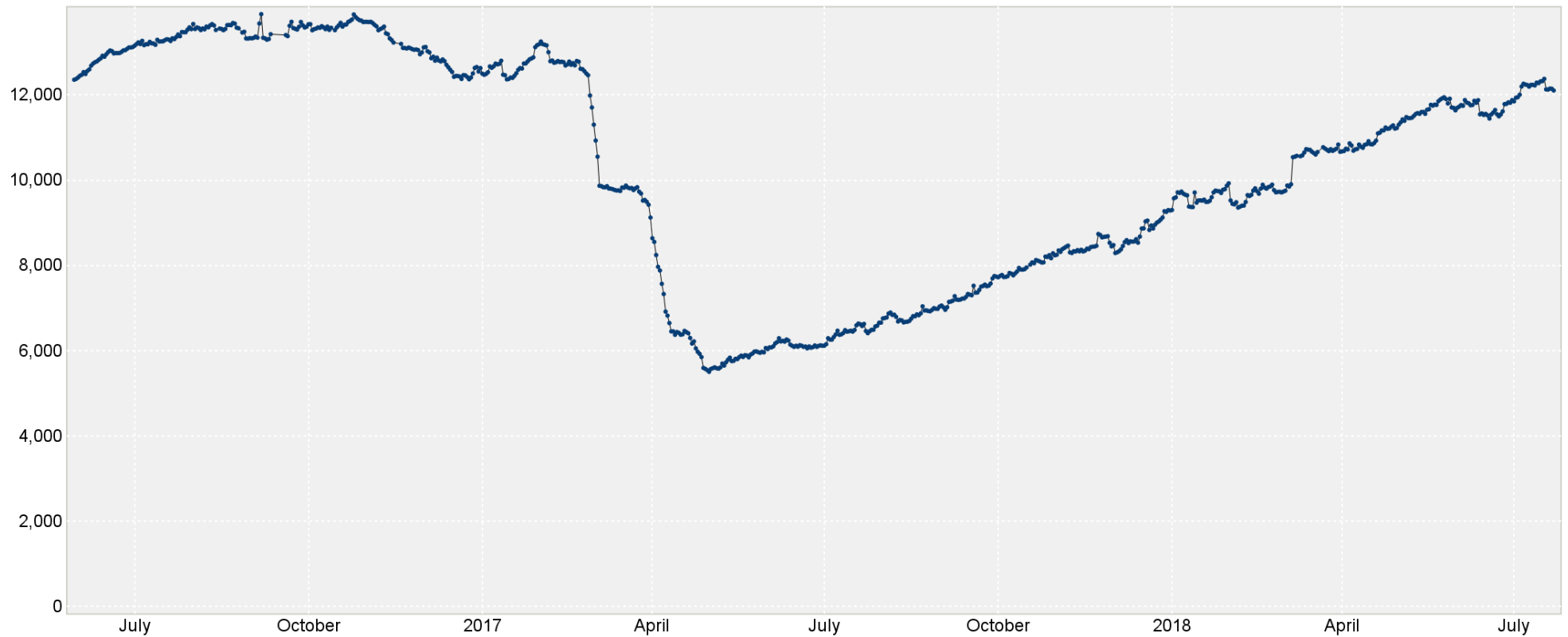
Huge (>18000 members)

Non-local (E2 4686 km)

Versionitis (version="3119")

Broken Relations

Errors: Open rings



Multipolygon relation where the outer or inner border is not closed. This is always an error and needs to be fixed.

Solutions?

Split up relations?
or allow changing only parts of them?

One size does not fit all!
Promote successful relations to their own types?
type=multipolygon, boundary, route, restriction

Do we need an area datatype?

Question has been around for a long time.

See also my talk at SOTM 2013:
“Towards an Area Datatype for OSM“

and

wiki.osm.org/wiki/Area/The_Future_of_Areas

See also...

Workshop:

Areas, Routing, and Diffs: Can we have
Something Better than Relations?

Roland Olbricht

14:00 in room S.1.5

The way forward...

Evolution, no revolution

Needs “rough consensus” in community

Needs buy in from developers

Perfomant central OSM database is key

The way forward...

Join us at

<https://github.com/osmlab/osm-data-model>
for docs and discussion

Thanks!



Thank you!

<https://github.com/osmlab/osm-data-model>

Jochen Topf

jochentopf.com

jochen@topf.org