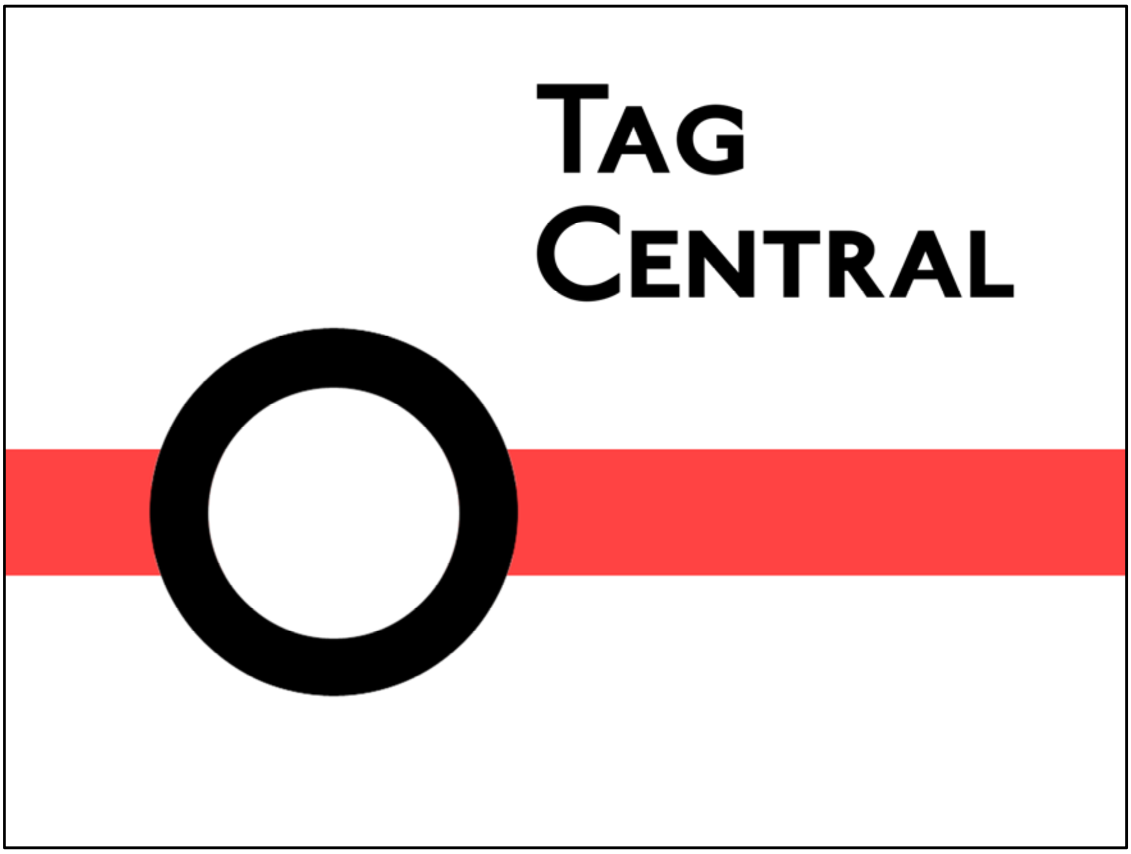


# Tag Central

Tag Central – a schema for OpenStreetMap

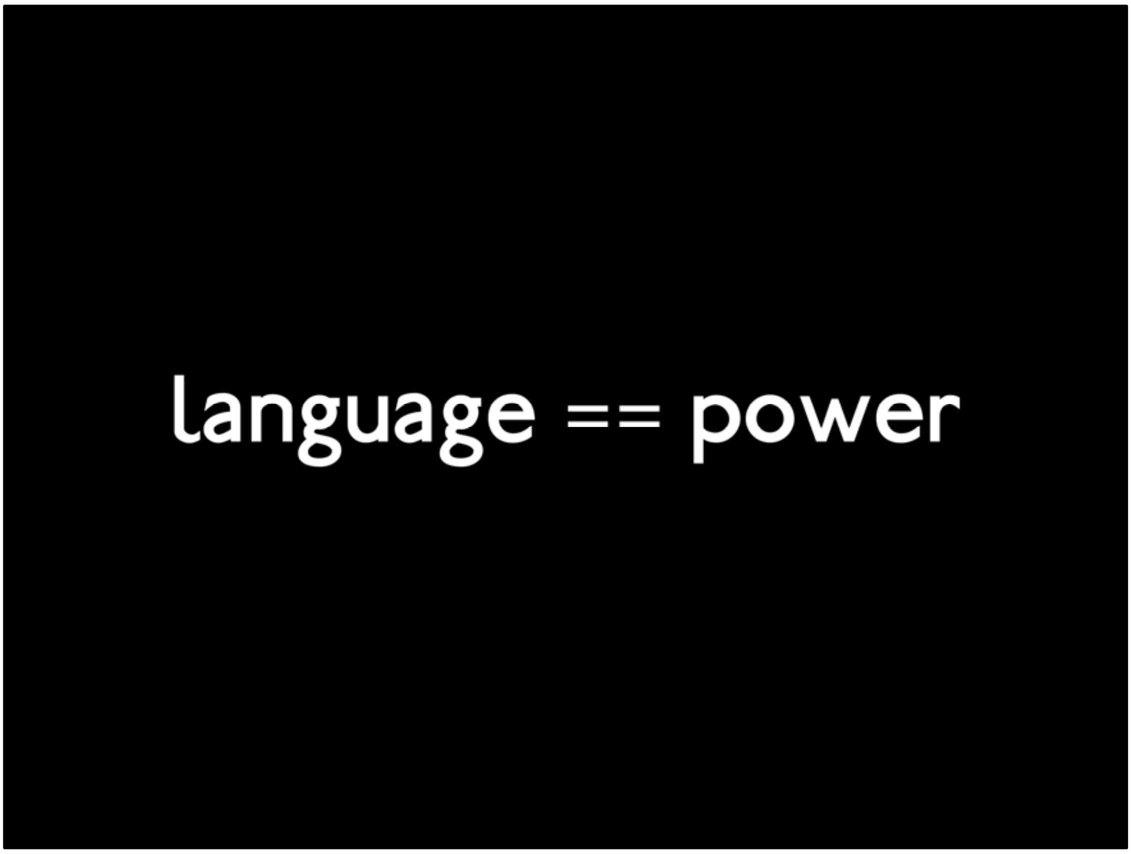
DAVID EARL

david@frankieandshadow.com

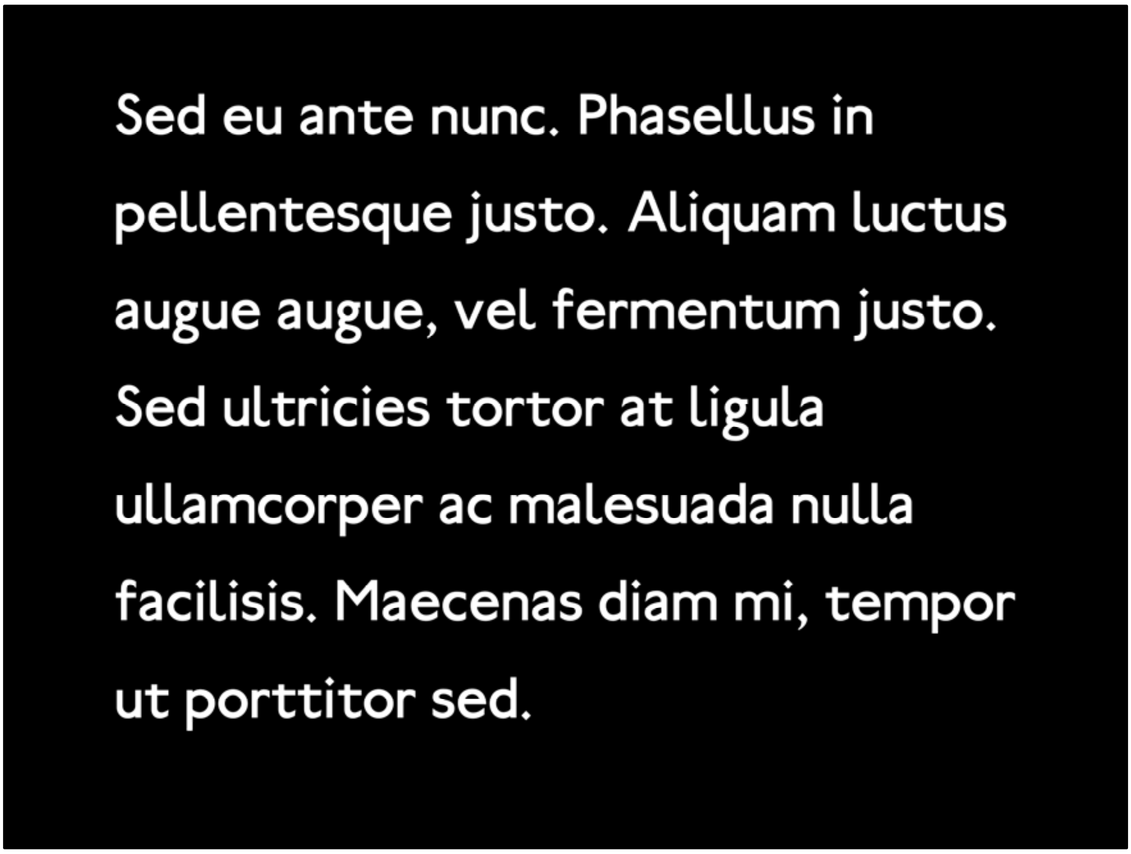
www.frankieandshadow.com

http://www.flickr.com/photos/[illegible]black/3506[illegible]534/

This slide represents this talk in oh so many ways, but initially I am using it to illustrate that…



language == power

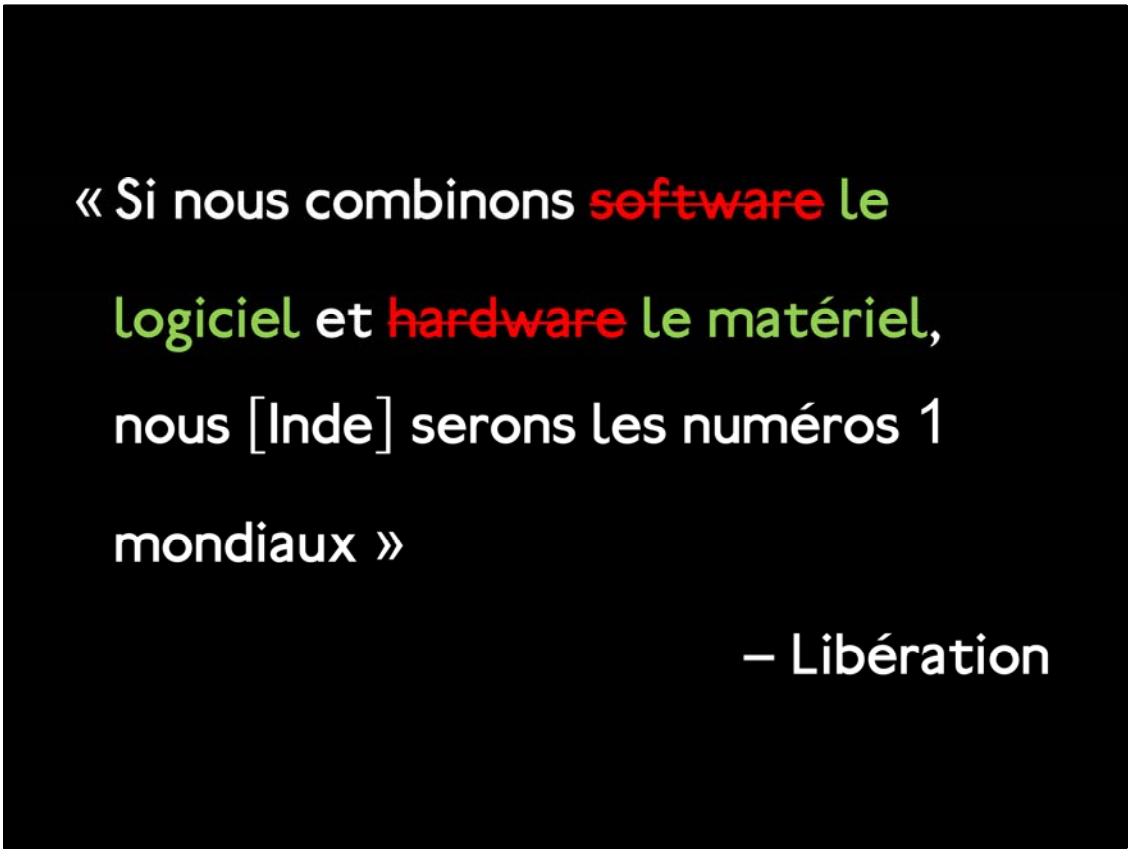


Sed eu ante nunc. Phasellus in pellentesque justo. Aliquam luctus augue augue, vel fermentum justo. Sed ultricies tortor at ligula ullamcorper ac malesuada nulla facilisis. Maecenas diam mi, tempor ut porttitor sed.

OK, happy with that? You all understand me of course. Just thought I'd choose my own words – no problem with that surely, and if other people follow my lead, we'll get a new language.

http://www.flickr.com/photos/ari/3151751561/

Welcome to the Anarchist Camp who advocate a completely free-for-all vocabulary for OpenStreetMap



« Si nous combinons ~~software~~ le logiciel et ~~hardware~~ le matériel, nous [Inde] serons les numéros 1 mondiaux »

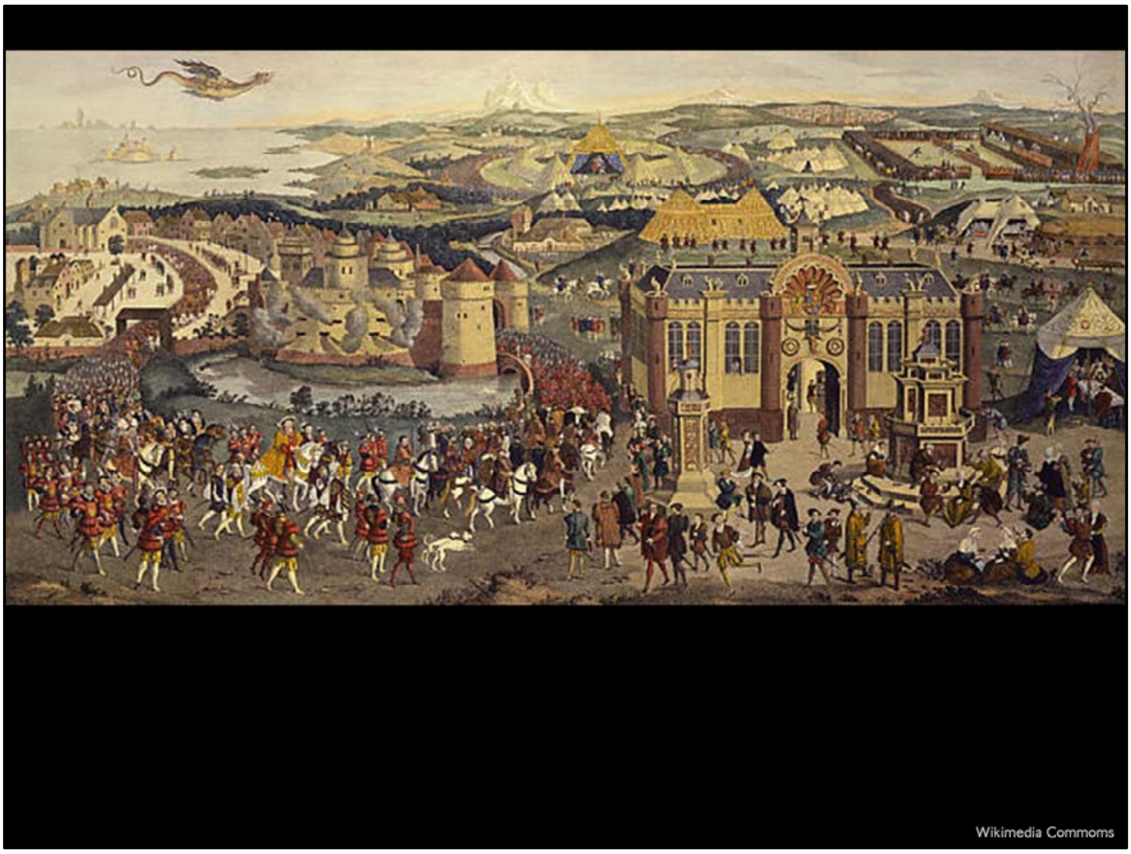
– Libération

On the other hand, the French paper Libération was taken to task for not using the approved French words for hardware and software, instead borrowing them from the English.

Wikimedia Commoms

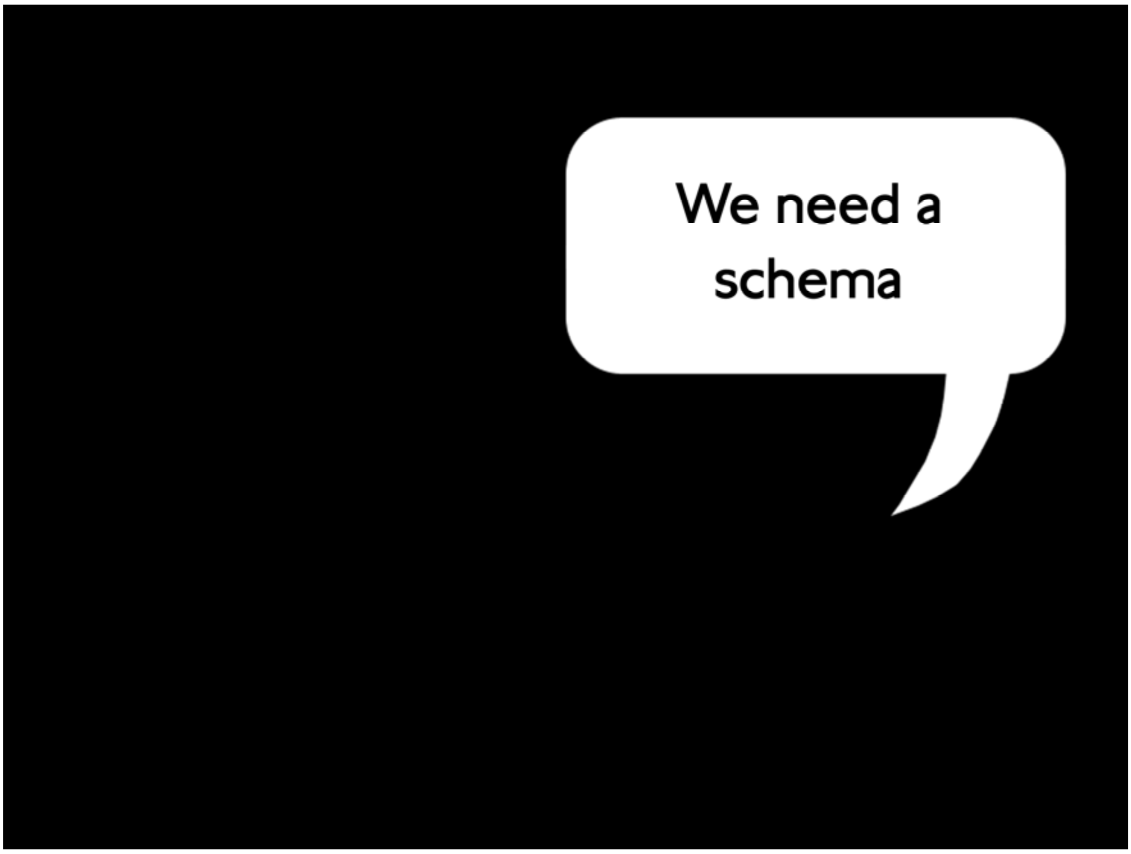
The “immortals” of the Academie Francaise live here

Welcome to the Autocrat Camp who want to dictate what vocabulary everyone should use within OpenStreetMap.


Wikimedia Commoms

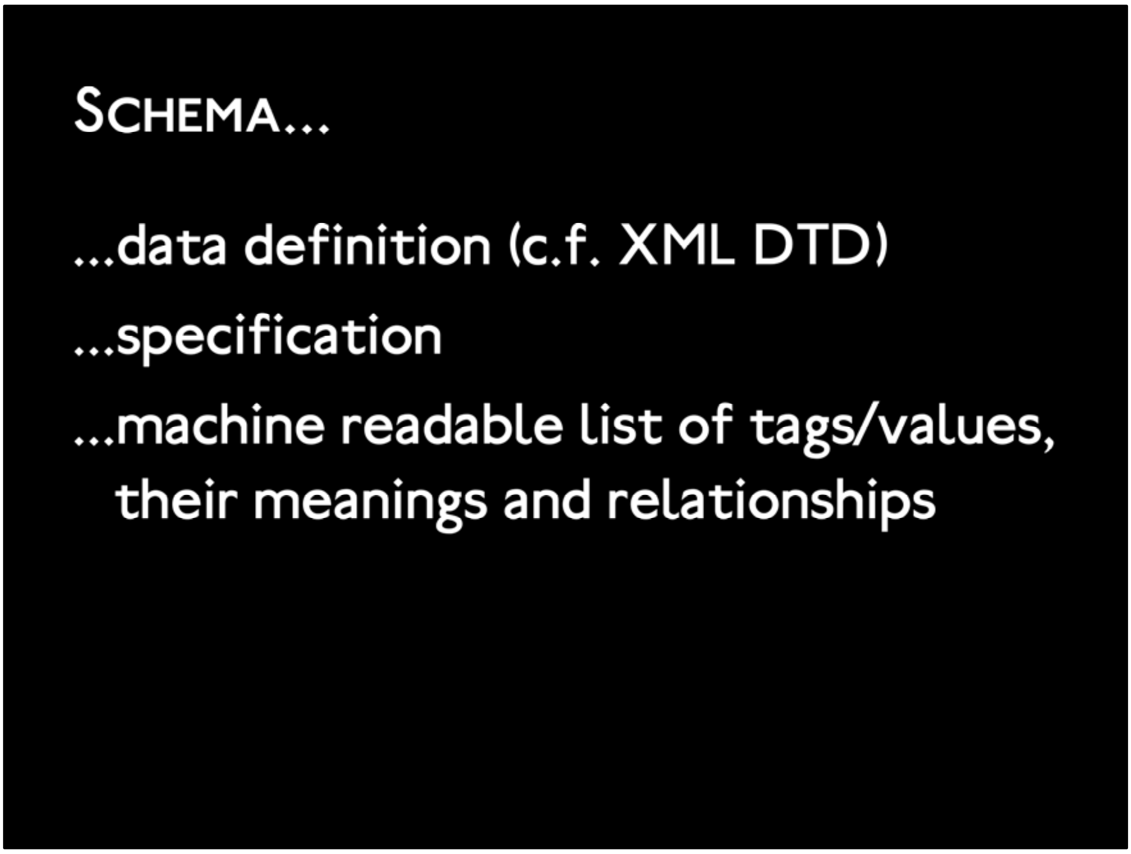
In the sixteenth century, two camps met at the Field of the Cloth of Gold near Calais in France to promote harmony between nations. Henry VIII of England and Francis I of France. That’s what I’d like to achieve with this proposal…

To create a half-way-house between our two camps that enables applications and people to understand our vocabulary without restricting the freedom to innovate – indeed promoting innovations.



We need a schema

So, I share the Autocrat view that “we need a schema”…



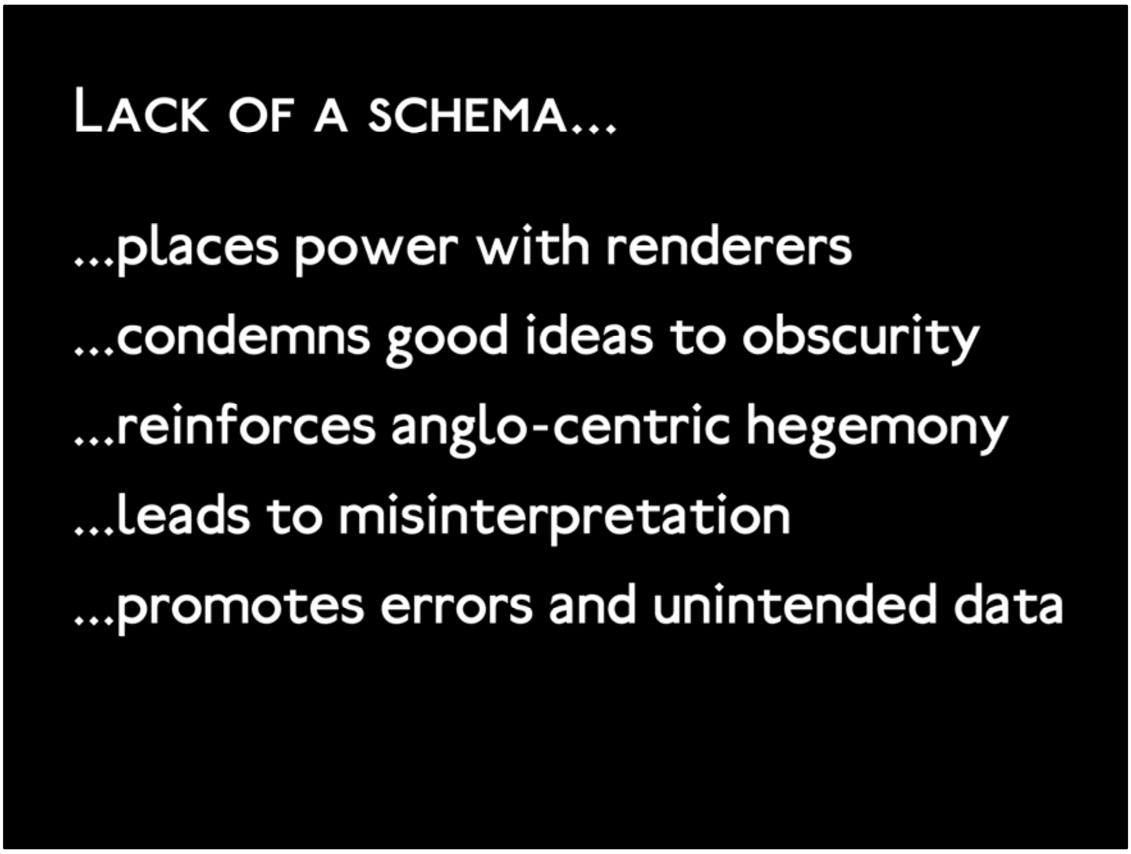
## Schema…

…data definition (c.f. XML DTD)

…specification

…machine readable list of tags/values, their meanings and relationships

(I’d better define what I mean by schema)



## Lack of a schema...

...places power with renderers

...condemns good ideas to obscurity

...reinforces anglo-centric hegemony

...leads to misinterpretation

...promotes errors and unintended data

In my opinion, not having one doesn't empower the contributor (surveyor, mapper etc) – it puts the power primarily in the hands of the renderers, mostly the default mapnik rendering. That's not to say they are malicious, merely that what shows up depends on what they notice. Of course, there are conscious choices of things not to render, but new tags or rare tags may well lie in obscurity because they are just not noticed. The raw tagging scheme is in (British) English and as such makes it hard to express concepts which are a bit different in other cultures and languages. Translations don't always help when trying to express a concept that is really not present in English or England: while autobahn and motorway are the same except in minor detail and therefore translatable, the Phllipino Barangay is much less like a village than translation might suggest. In any case, why are (or should we be) translating these things separately in every application, producer and consumer.

Without a schema, we only have two (English) words or brief phrases to describe something, which if that's all someone else sees, will often be misinterpreted or applied to things that were never intended. We don't have an authoritative source to check spellings against or (unintentionally) inappropriate combinations of tags. We can suggest what appropriate combinations might be except by building this knowledge (and where does that knowledge come from?) into each application that wants to know.

We do actually have schemas already – lots of them. But they are distributed all over the place to a greater or lesser extent in each application and in scant documentation. They contradict each other, duplicate effort, are dramatically incomplete and don't update except when manual effort is put into updating, they reflect the prejudices of the application writers not those who invented the concepts, and they aren't really machine readable on the whole.

We need a schema

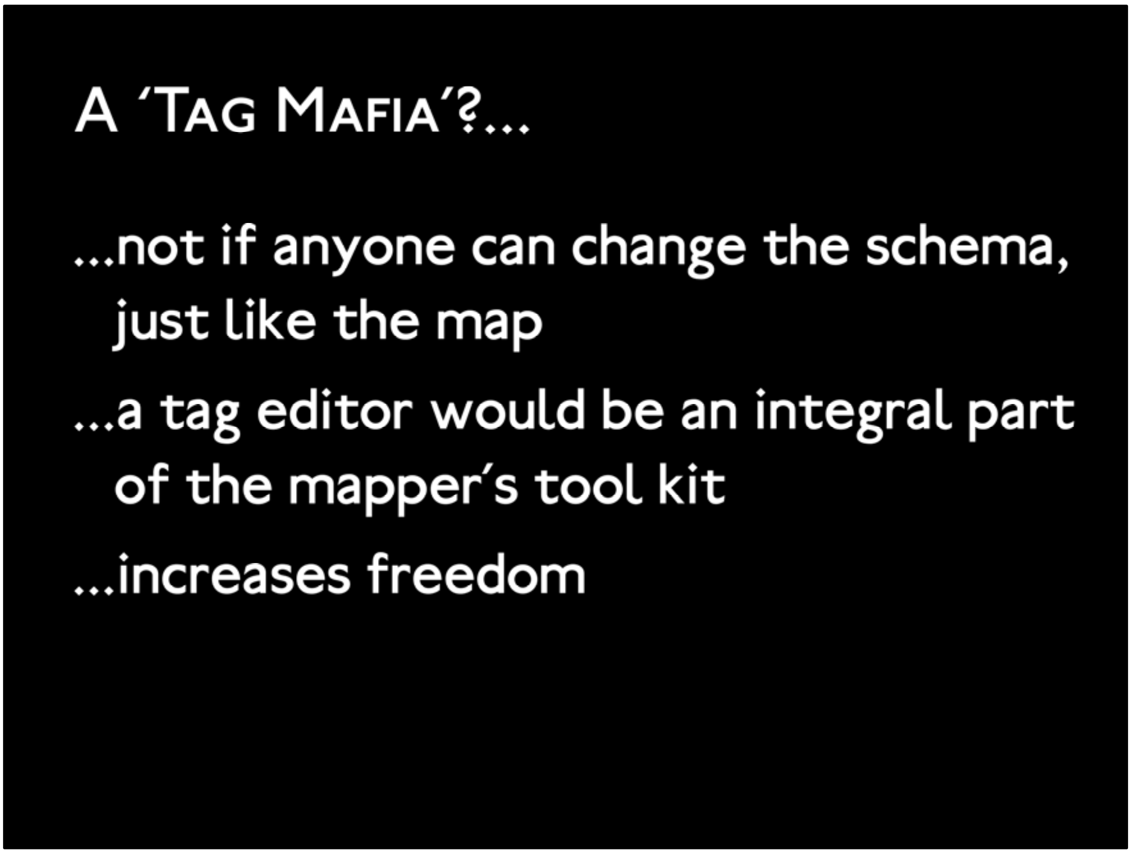
But we have to be able to do what we want

But, says that Anarchist camp, “we have to be able to do what we want”.  And I agree with them too. So wouldn’t a schema…

http://www.flickr.com/photos/msspider66/133189856/

http://www.flickr.com/photos/clinger/89294454/

…just lead to a `tag mafia’?



## A ‘Tag Mafia’?…

…not if anyone can change the schema, just like the map

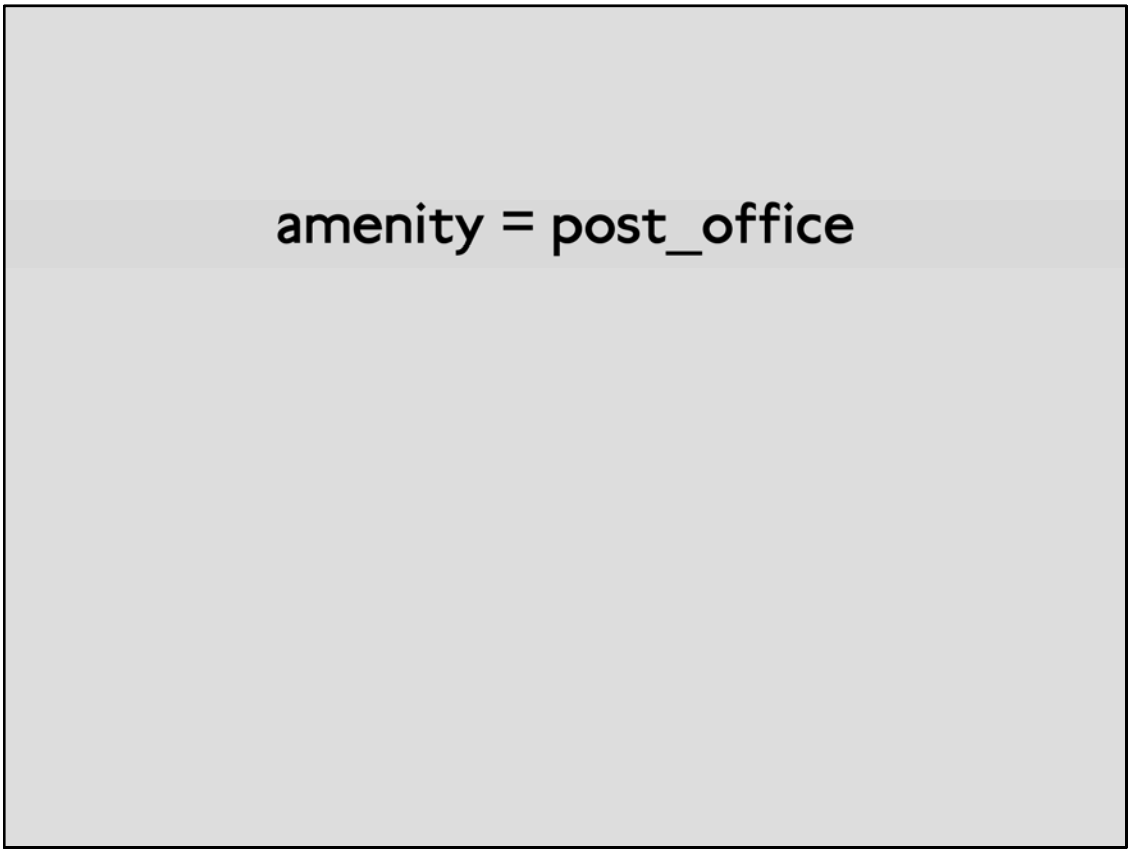
…a tag editor would be an integral part of the mapper’s tool kit

…increases freedom

I don’t think so. On the contrary. The theory that says you can use whatever tags you like to describe things is all very well, but it just isn’t true. That’s because the knowledge of those tags is embedded, mostly manually, in dozens of applications and web pages which you don’t have access to. You can’t promote, explain, add or change your model unless you have access to change all the documentation, consumer applications, and editor applications out there. I think the ability to add or change both a tag and its context as part of the editing environment *increases* not decreases ones freedom.

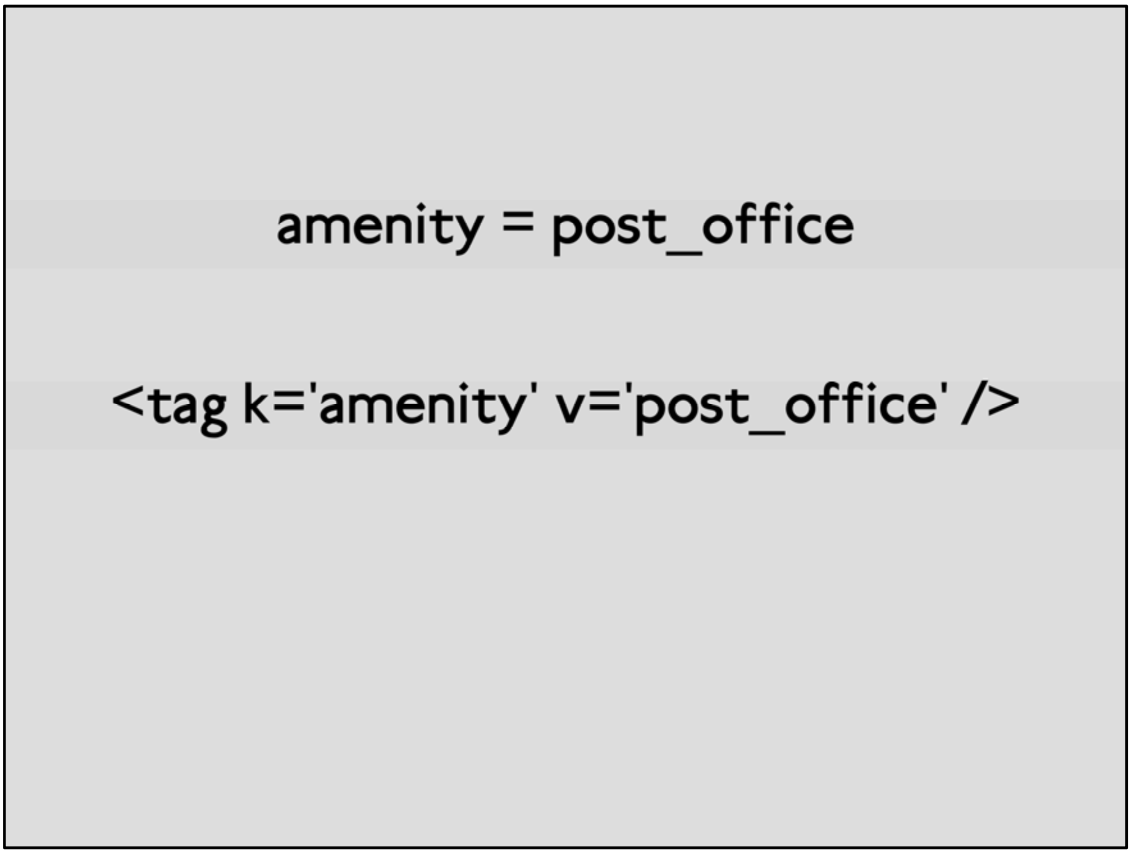
http://www.flickr.com/photos/fotologic/1119391517/

A quick diversion on terminology…



amenity = post_office

We are used to writing things like this. But this is in fact a shorthand for the “tag key has tag value” concept that is properly written down in a variety of ways…

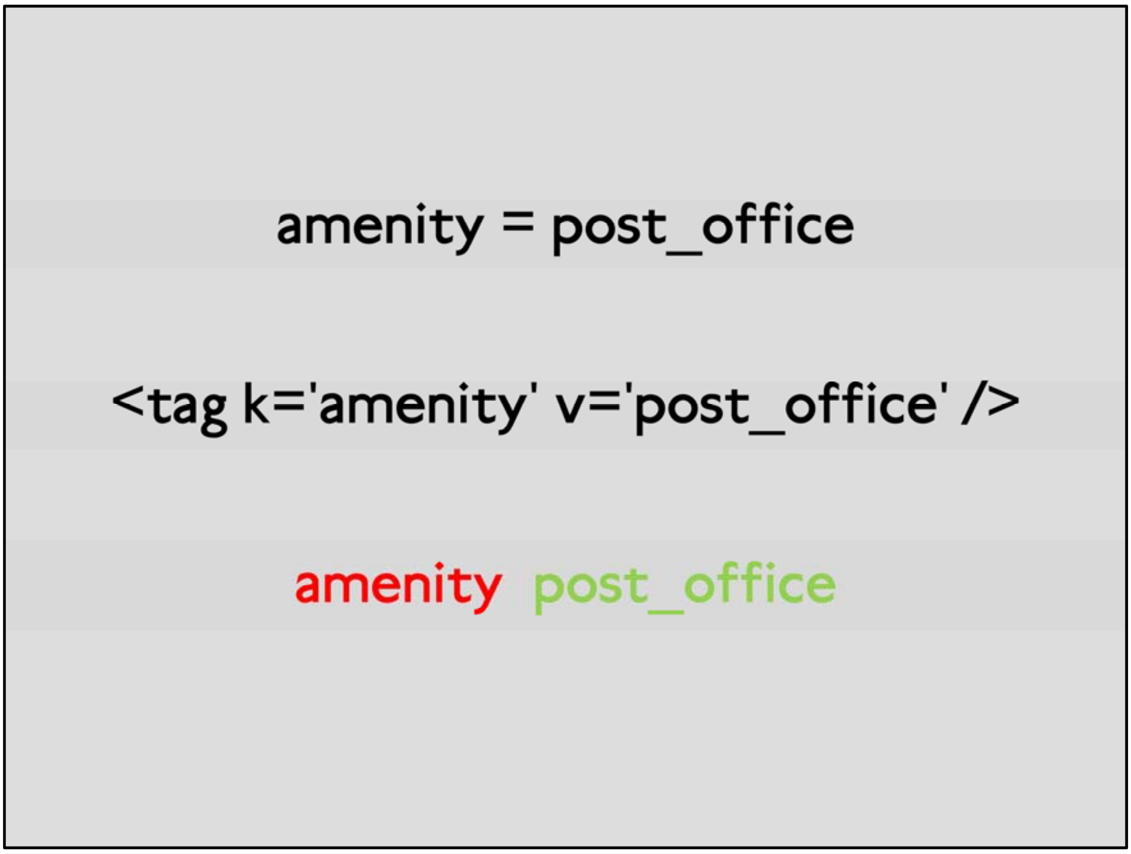


amenity = post_office

```
<tag k='amenity' v='post_office' />
```

… such as in the XML produced by the OSM API.

But for the purposes of this talk, the tag=value notation is going to bet confusing,…



amenity = post_office

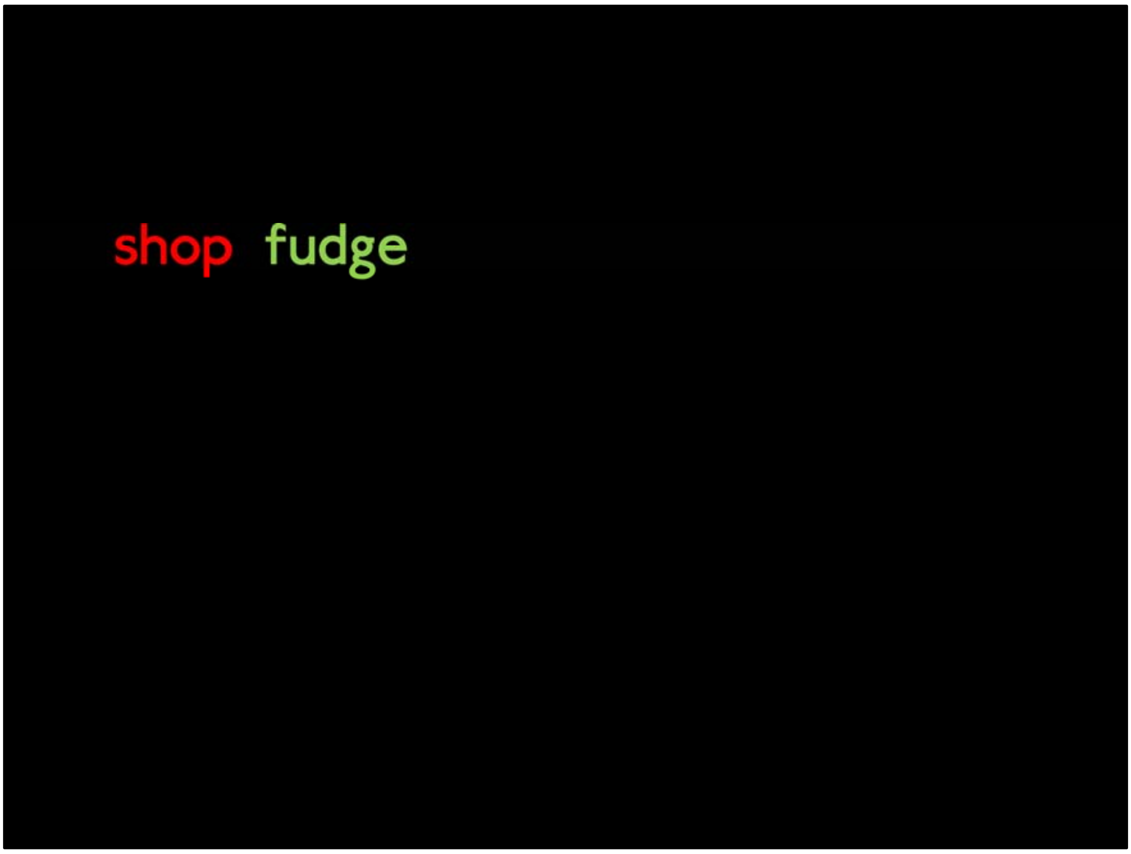
<tag k='amenity' v='post_office' />

amenity post_office

… so I shall write tag keys in red and tag values in green, like this.

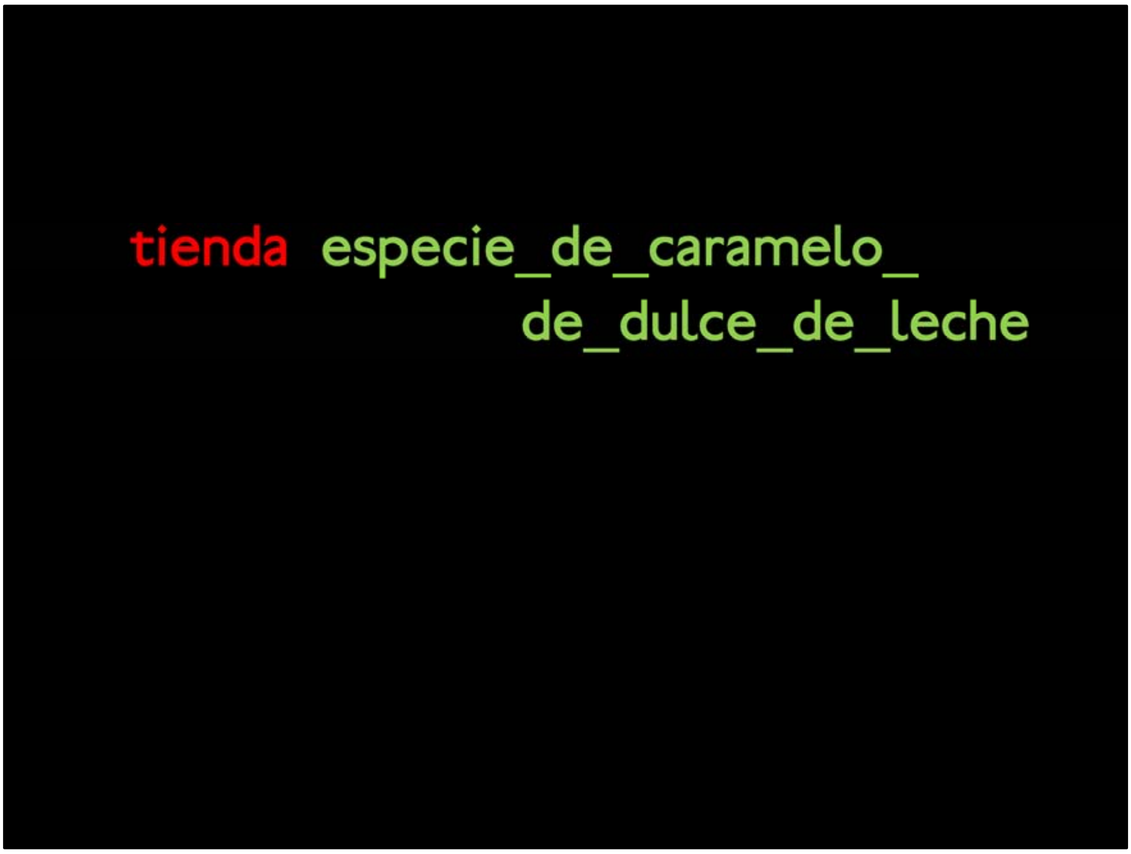
http://www.flickr.com/photos/lender/483011475/

Right, so we come across something unusual in Cambridge: a fudge shop.



shop fudge

Following a pattern, and not knowing how (and not knowing how to find out how to find out how) others may have tagged this unusual feature, if indeed I am not the first, I might choose a tag key ‘shop’ and tag value `fudge’. I may have considered shop=confectioner or shop=sweet_shop or some such but dismissed it because this shop sells (and makes) *only* fudge.



tienda especie_de_caramelo_
de_dulce_de_leche

If I was Spanish, I might have chosen to write this. Hmm, the rather wordy, descriptive text suggests that perhaps its an unfamiliar concept in Spain, so …

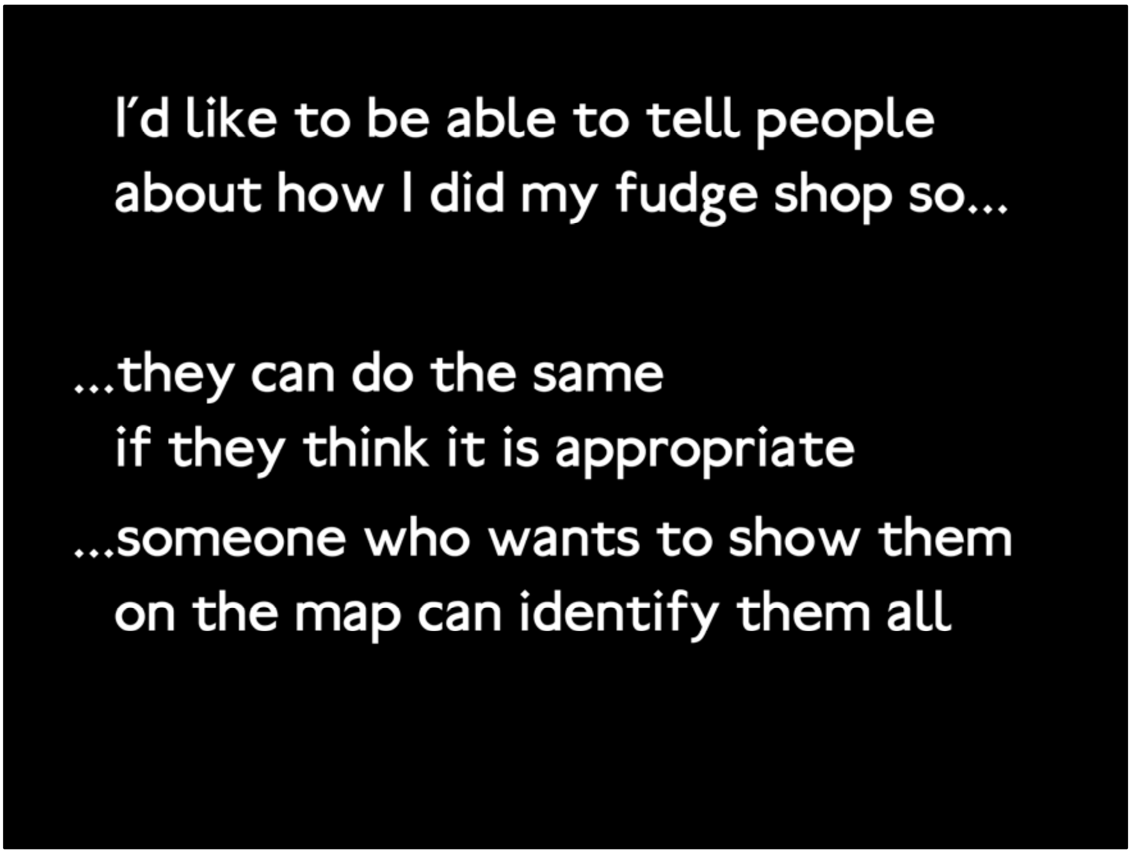
tienda especie_de_caramelo_
de_dulce_de_leche

http://www.flickr.com/photos/generalnoir/3336965724/

… here’s what it looks like – a sugary confection made with condensed milk, a bit like soft toffee.

http://www.flickr.com/photos/mugsy/189670350/

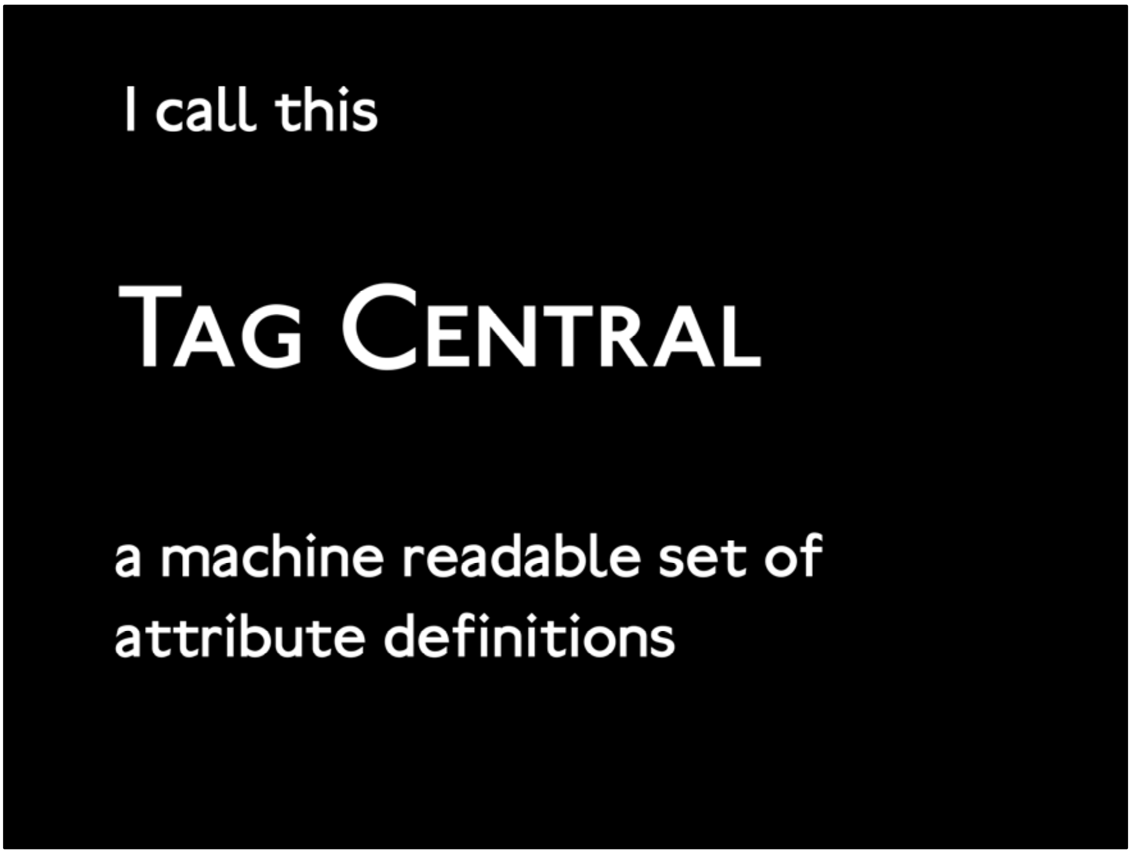
Ah, John Doe also came across a fudge shop while on holiday in Wisconsin.



I’d like to be able to tell people
about how I did my fudge shop so…

…they can do the same
if they think it is appropriate

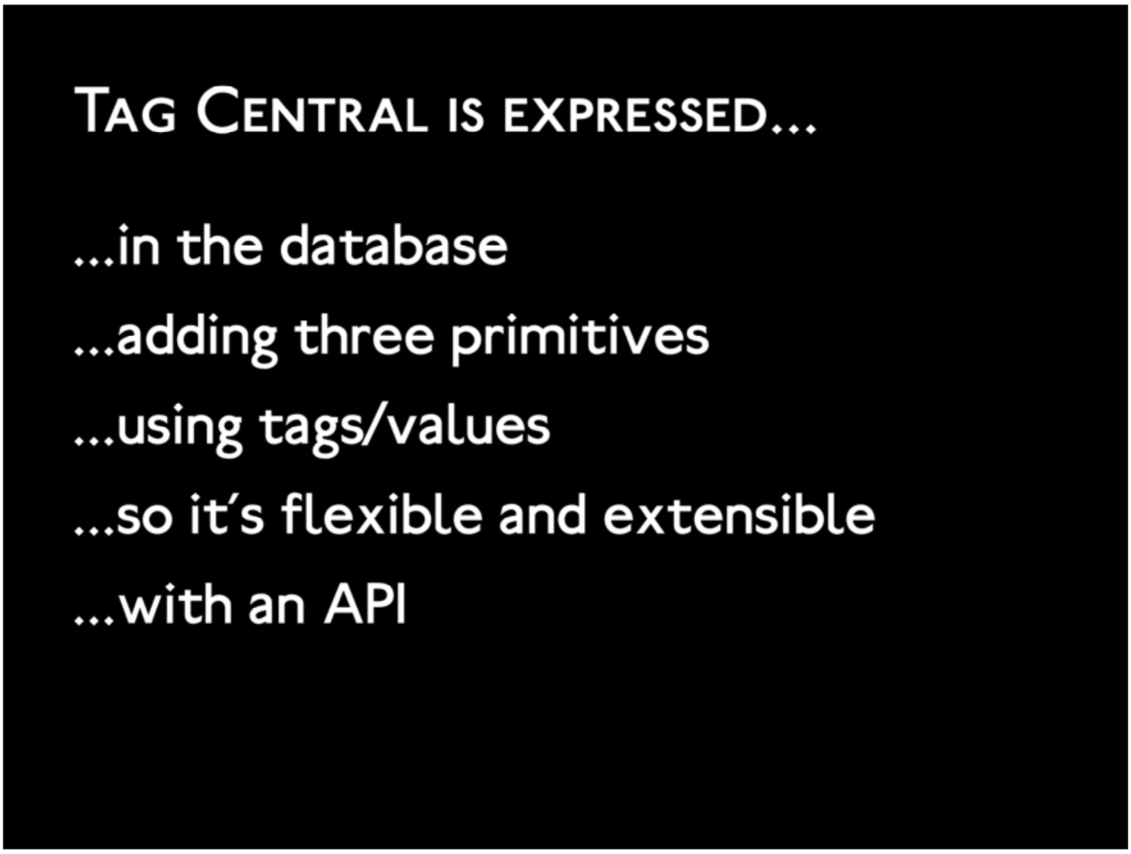
…someone who wants to show them
on the map can identify them all



I call this

# Tag Central

a machine readable set of attribute definitions



## Tag Central is expressed...

...in the database

...adding three primitives

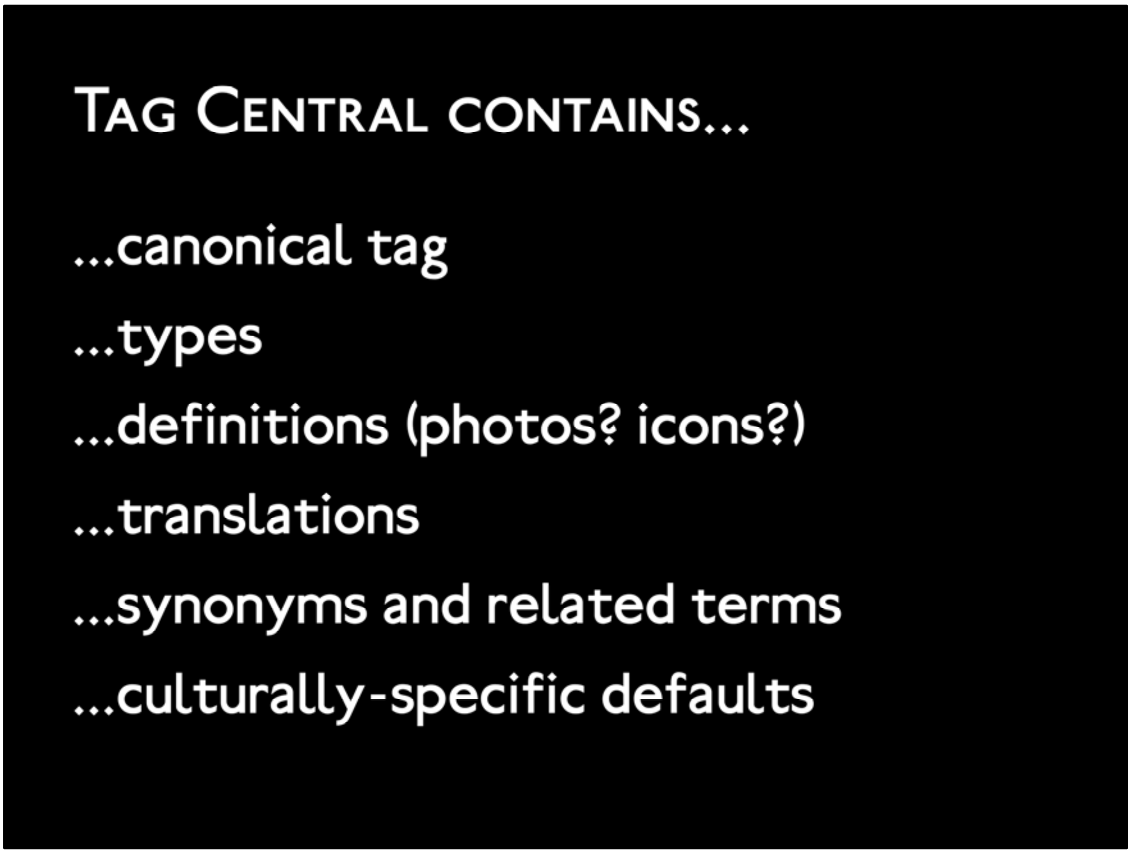
...using tags/values

...so it´s flexible and extensible

...with an API

in the database along with the map content
...using the same tag/value mechanism
...adding maybe three primitives to the existing node, way, relation
...and as flexible and extensible
...with a similar API

I think it has to be conceptually part of the database (though it need not actually reside in the same physical location, so long as it can share credentials), otherwise it just becomes another application to add to the many others that store information about tags and values.



## Tag Central contains...

...canonical tag

...types

...definitions (photos? icons?)

...translations

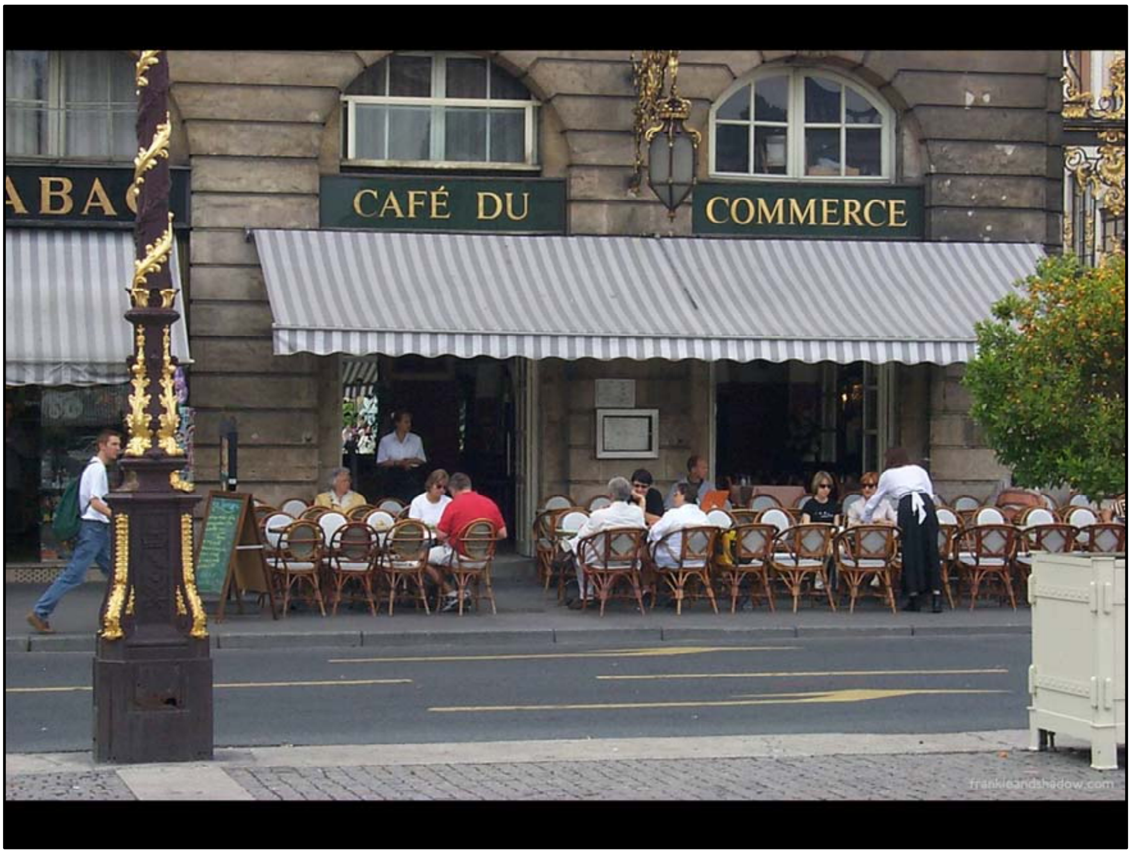
...synonyms and related terms

...culturally-specific defaults

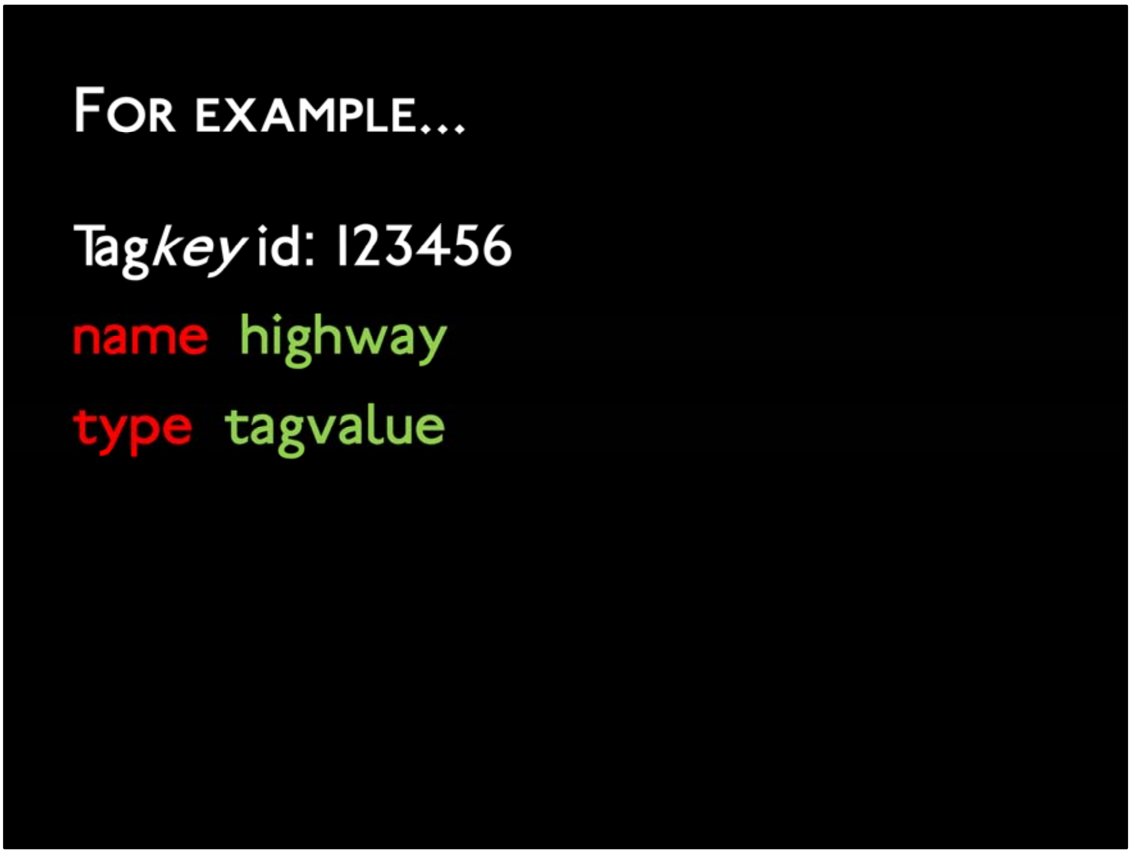
It tells us what words we should use for a concept, what the types of values we might expect are (including their units where relevant). We can explain in many languages what this concept means. We can offer similar concepts for consideration (so someone trying confectioner on encountering a fudge shop can discover others have tagged as fudge). Perhaps most important of all, we can offer defaults which differ from country to country, so that we can say that something is a motorway and know in machine readable as well as descriptive ways what this means specifically in each country. For example, country-specific defaults allow us to write routers which take maximum speeds into account without the streets having to be tagged with that explicitly everywhere, only when they differ from the default.

frankieandshadow.com

It might allow us to solve a couple of discussions that have come up on mailing lists recently. For example, “football”. Yes, it’s universally a team game, but does it mean soccer, American football , rugby union football, …



And café? I was very surprised to discover that French cafés are perceived more like bars by the French, even though I had visited them and felt they were akin to what I would call a café in England, notwithstanding the same word.



## For example…

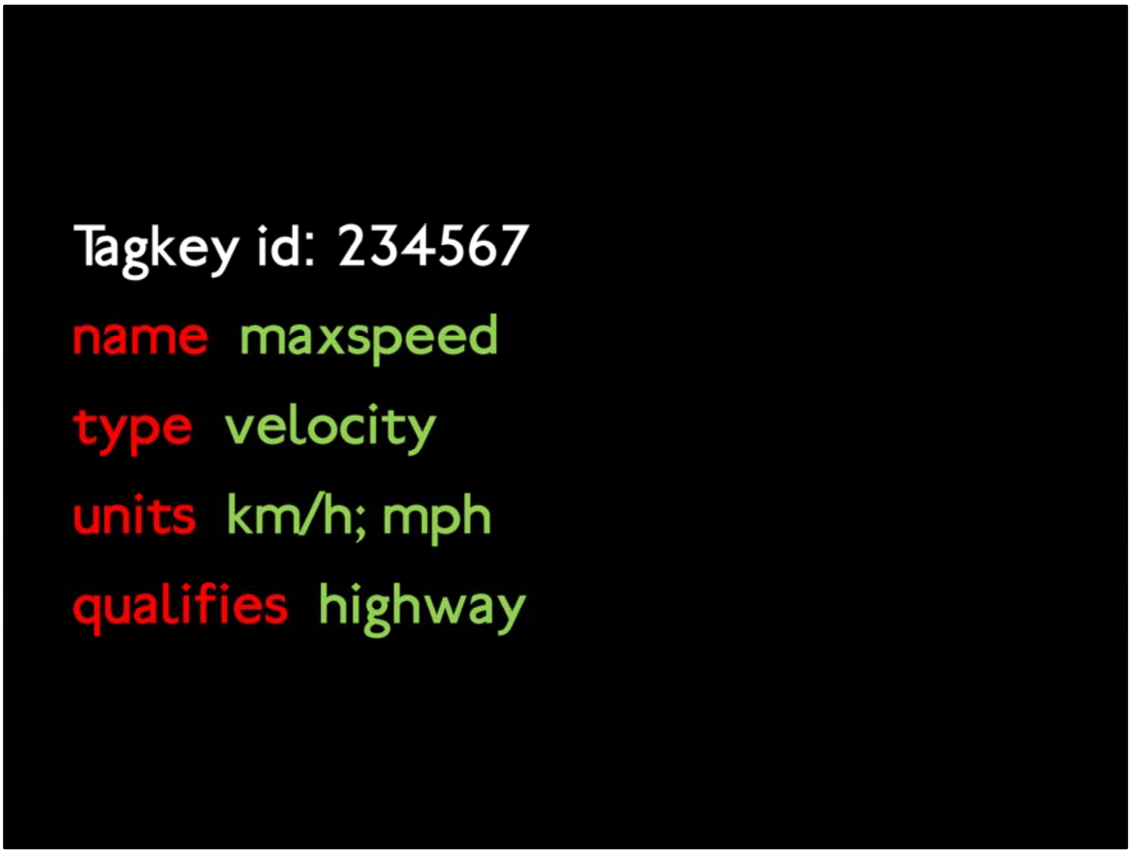
Tag*key* id: 123456

name highway

type tagvalue

So some examples. But please note I am trying to illustrate the concept here, not to provide every aspect of the syntax and semantics of the schema.

The tagkey primitive can introduce the tag key known as ‘highway’ and say what we expect to find as its value are tag values (rather than, say, numbers, speeds, arbitrary text or whatever). In other words, we should expect to find tag value primitives saying they can be used as the value of tag key ‘highway’.



Tagkey id: 234567

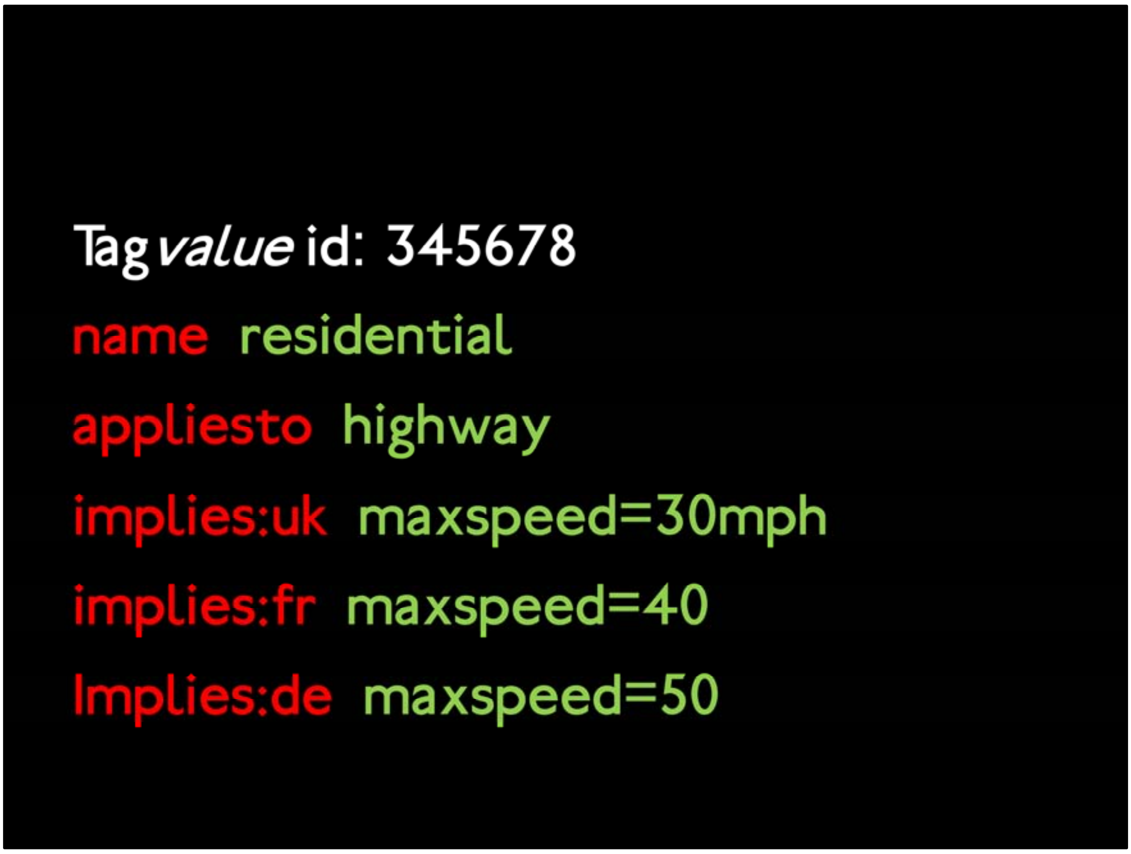
name maxspeed

type velocity

units km/h; mph

qualifies highway

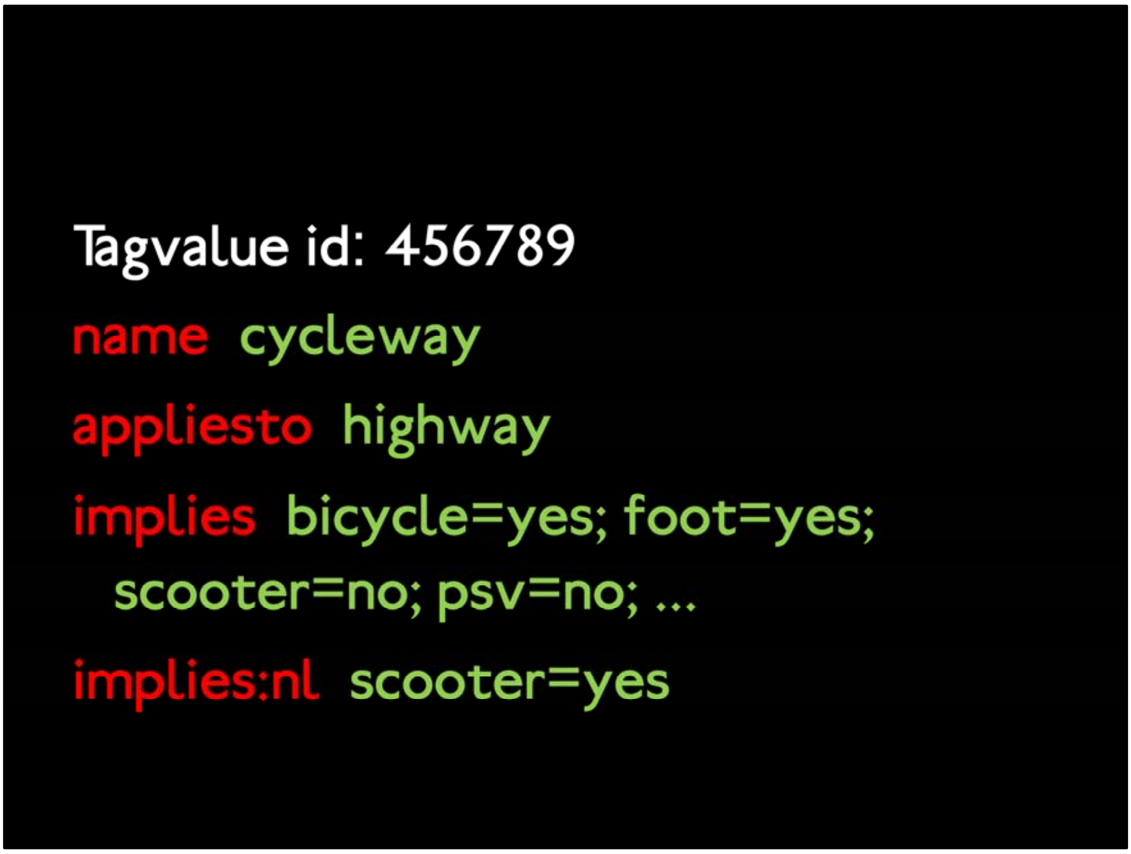
Here’s another key, ‘maxspeed’. This one expects to find velocity values, i.e. decimal numbers optionally followed by appropriate units (which are also listed, presumably the first is the default, or we could say that separately), representing a velocity, and that we would expect to find such tags applied to items which are tagged with the ‘highway’ key.



Tag *value* id: 345678

| | |
|---|---|
| name | residential |
| appliesto | highway |
| implies:uk | maxspeed=30mph |
| implies:fr | maxspeed=40 |
| Implies:de | maxspeed=50 |

Now a tag value, ‘residential’, which we‘d expect to find as a value for key ‘highway’ (as shown by the appliesto tag). The main point of this example, though, is that we can determine from this entry the maximum speed that would be applied *in different countries* if the street does not explicitly have a maxspeed tag.



Tagvalue id: 456789

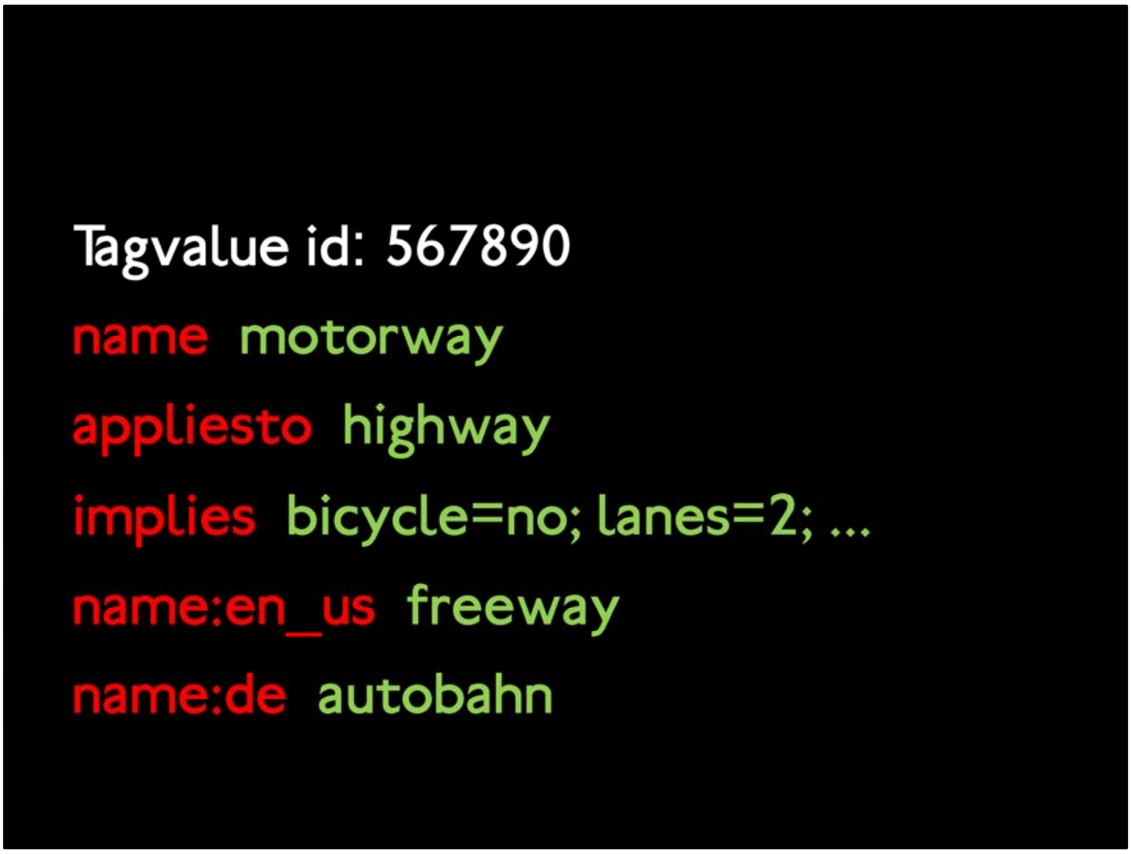
name cycleway

appliesto highway

implies bicycle=yes; foot=yes;
scooter=no; psv=no; …

implies:nl scooter=yes

Likewise, the common cultural feature of a cycleway is that it allows bicycles to use it. In most places, people on foot can use it but motor scooters cannot. In the Netherlands, where scooters can use cycleways, we can make an exception. (I am told they used to be allowed to but may not be the case now – I may be out of date– in which case a simple scheme change can reflect the updated situation countrywide and scooter routers would automatically update themselves not to route scooters via cycleways – well written applications adapt).



Tagvalue id: 567890

name motorway

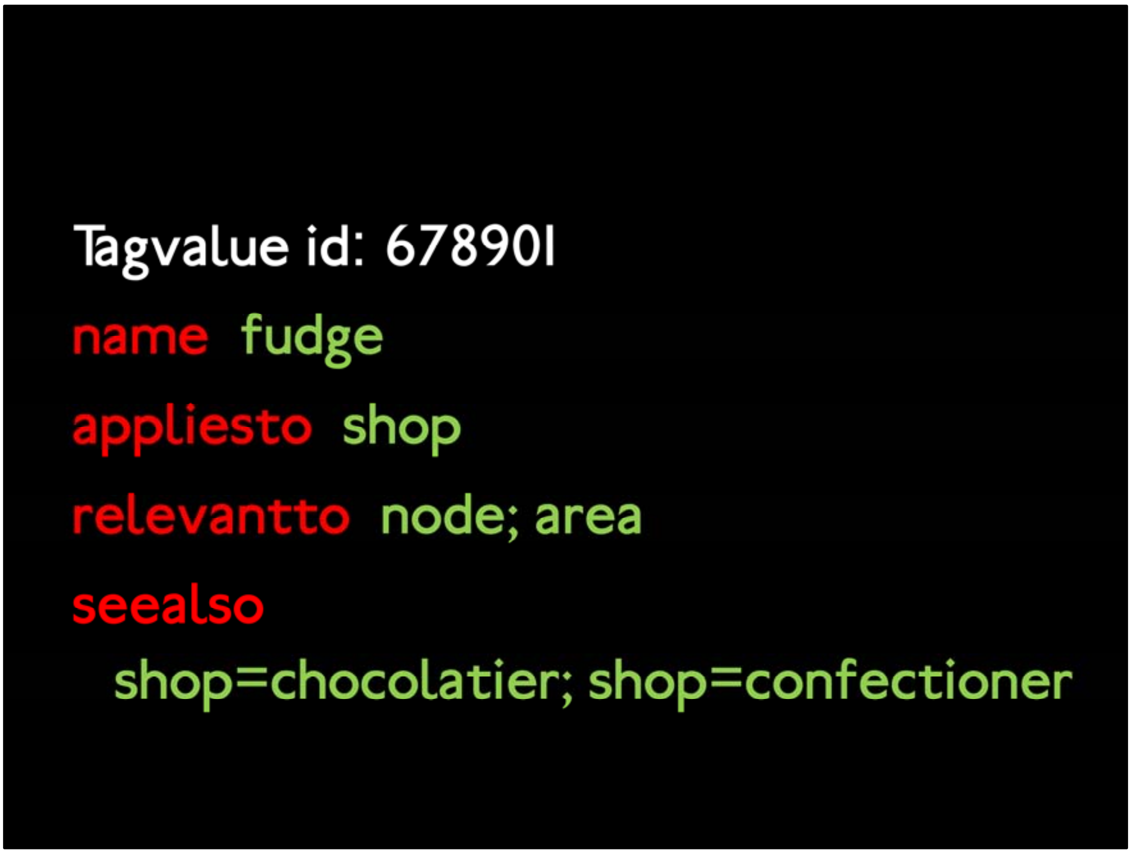
appliesto highway

implies bicycle=no; lanes=2; …

name:en_us freeway

name:de autobahn

By including linguistic alternatives we can also avoid the need for every application to do their own translations.



Tagvalue id: 678901

name fudge

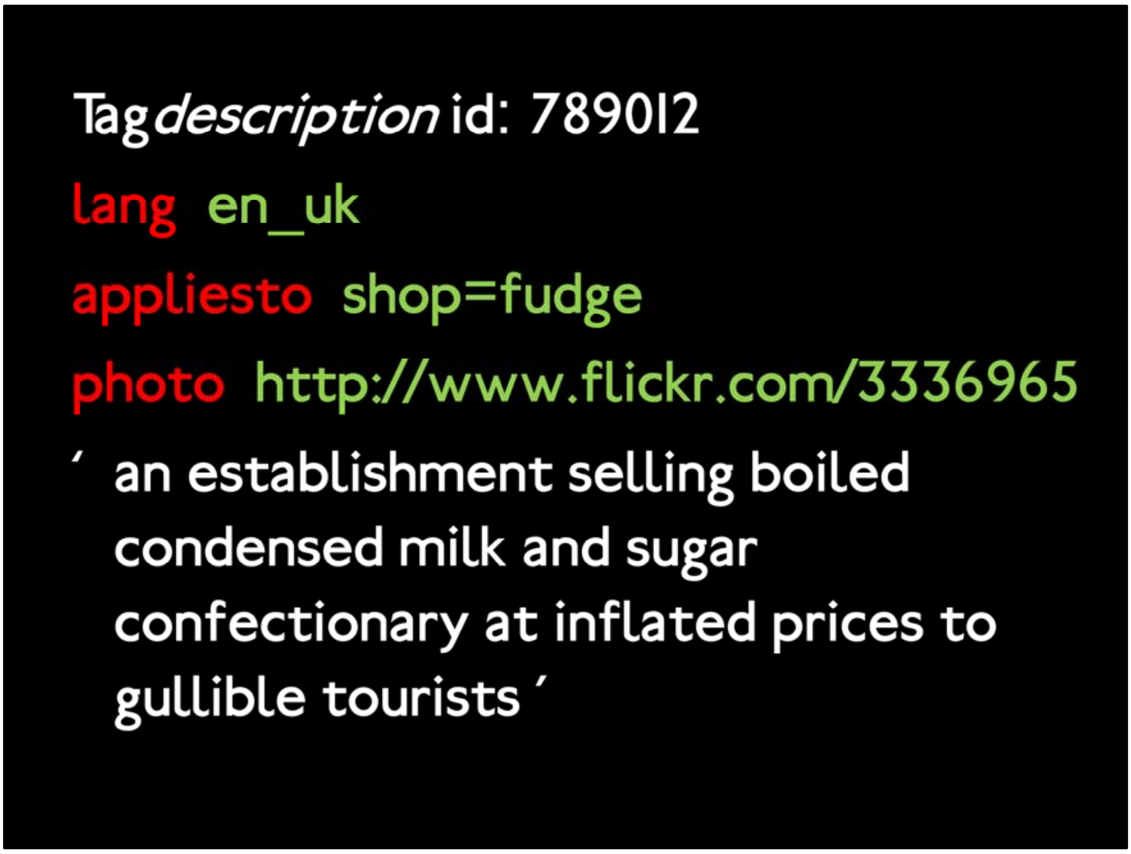
appliesto shop

relevantto node; area

seealso

shop=chocolatier; shop=confectioner

And back to our fudge shop. Note I’m now telling you that you wouldn’t normally expect to see this tag on ways, only on areas and nodes.



Tag*description* id: 789012

lang en_uk

appliesto shop=fudge

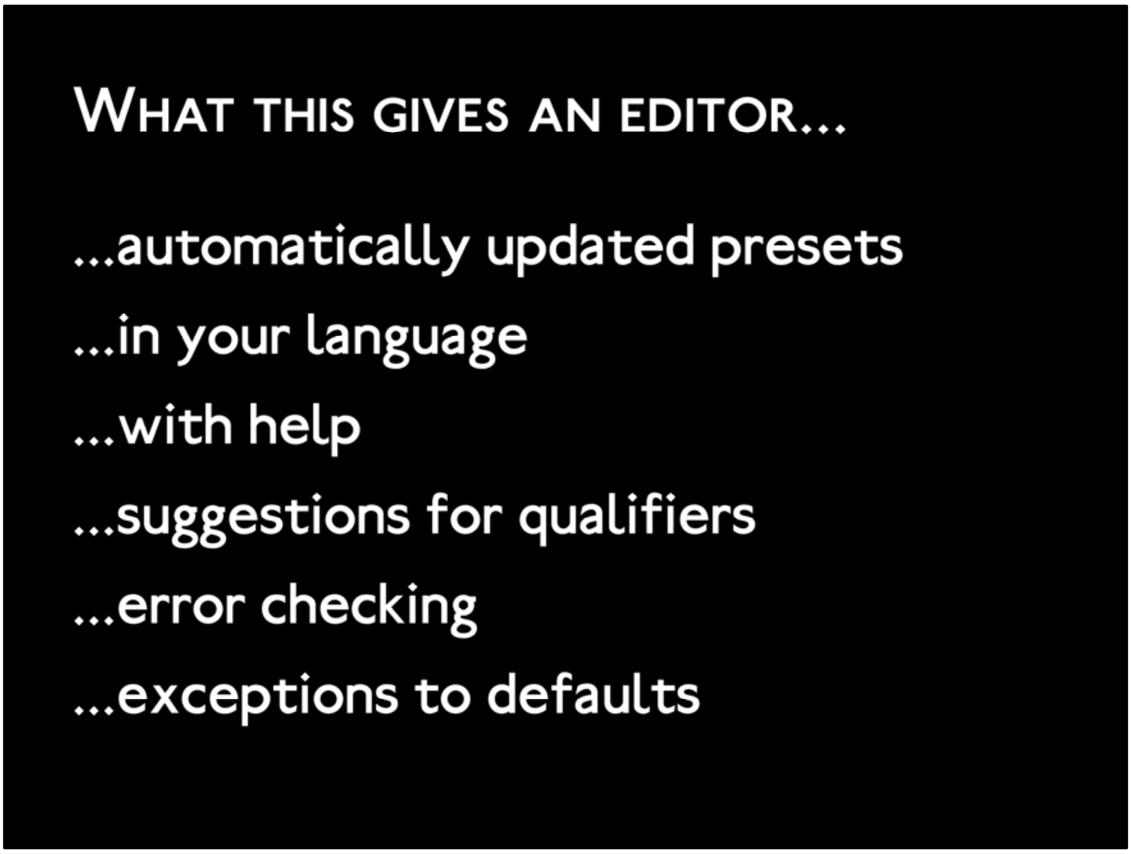
photo http://www.flickr.com/3336965

‘ an establishment selling boiled condensed milk and sugar confectionary at inflated prices to gullible tourists ’

Finally, descriptions. These could have been attached to the tag value or tag key primitives, but they are somewhat larger and dividing them up like this allows the set for each language to be accessed independently – so you only need the German tagdescriptions if you are working in German.

http://www.flickr.com/photos/st3f4n/4420831134/

Would this be compulsory? I think it comes in stages. We can get a basic level of documentation to start with. We can later start checking for errors against it. Editors can ask you to write a description and add the context for new tags. We could in the end reject tags for which a schema entry has not been provided, but that may be a step too far for many people. Bear in mind though, this is not preventing anyone inventing new tags, merely requiring you to document what you’ve invented.



## WHAT THIS GIVES AN EDITOR…

…automatically updated presets

…in your language

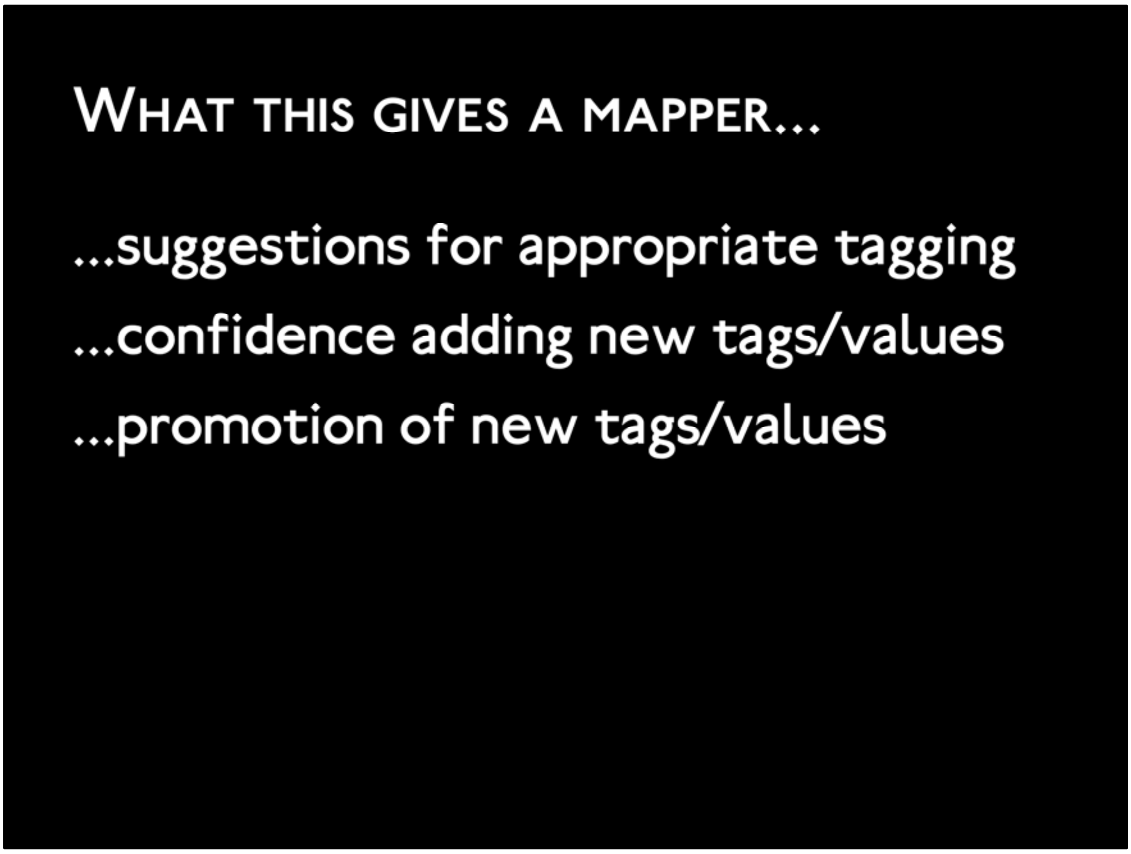
…with help

…suggestions for qualifiers

…error checking

…exceptions to defaults

This all means that editors can derive their presents centrally (obviously, they wouldn't show them all, but could offer a search for more obscure ones), central translations (including of raw tag values with automatic translation into canonical form, which no editor does today), offering help at ones fingertips on what tags mean beyond the raw tokens, making suggestions as to how you might want to flesh out or qualify objects (including country-relevant defaults – after all, the editor knows where you are editing), so you know when you can just leave the tag out. And it can cross check for errors in a way that isn't just a reflection of the opinion of the writer of the error checker.



## What this gives a mapper…

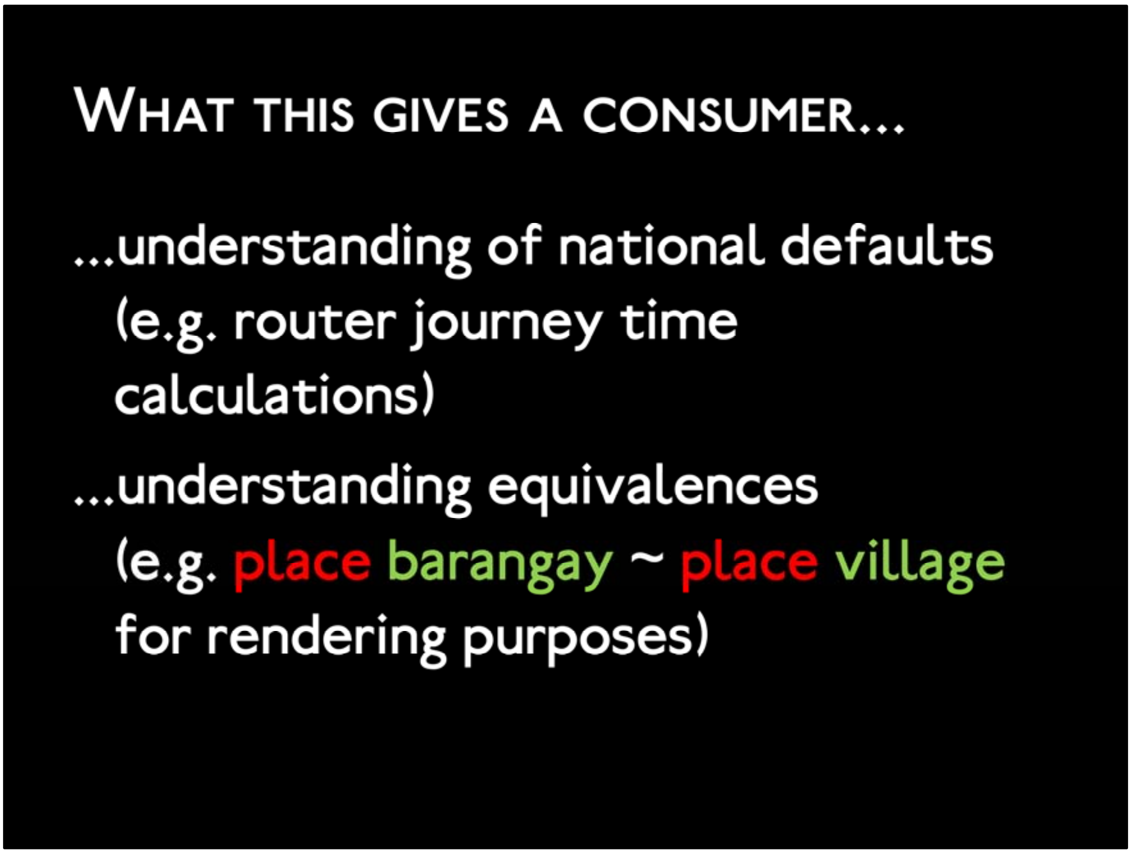
…suggestions for appropriate tagging

…confidence adding new tags/values

…promotion of new tags/values

The mapper can easily find out more about what other people have done. The problem with things like tagwatch (apart from its invisibility to most users) is that it reflects popular objects. Fudge shops are rare by definition, so will never climb to the top of the popularity tables, But it’s the rare cases I’ll not be sure how to tag.

It lets me introduce new things into the system with much more confidence that I’m not duplicating effort, and because what I do is then available to all editors and consumer applications, it promotes and in some cases brings into immediate use, the information I have provided.

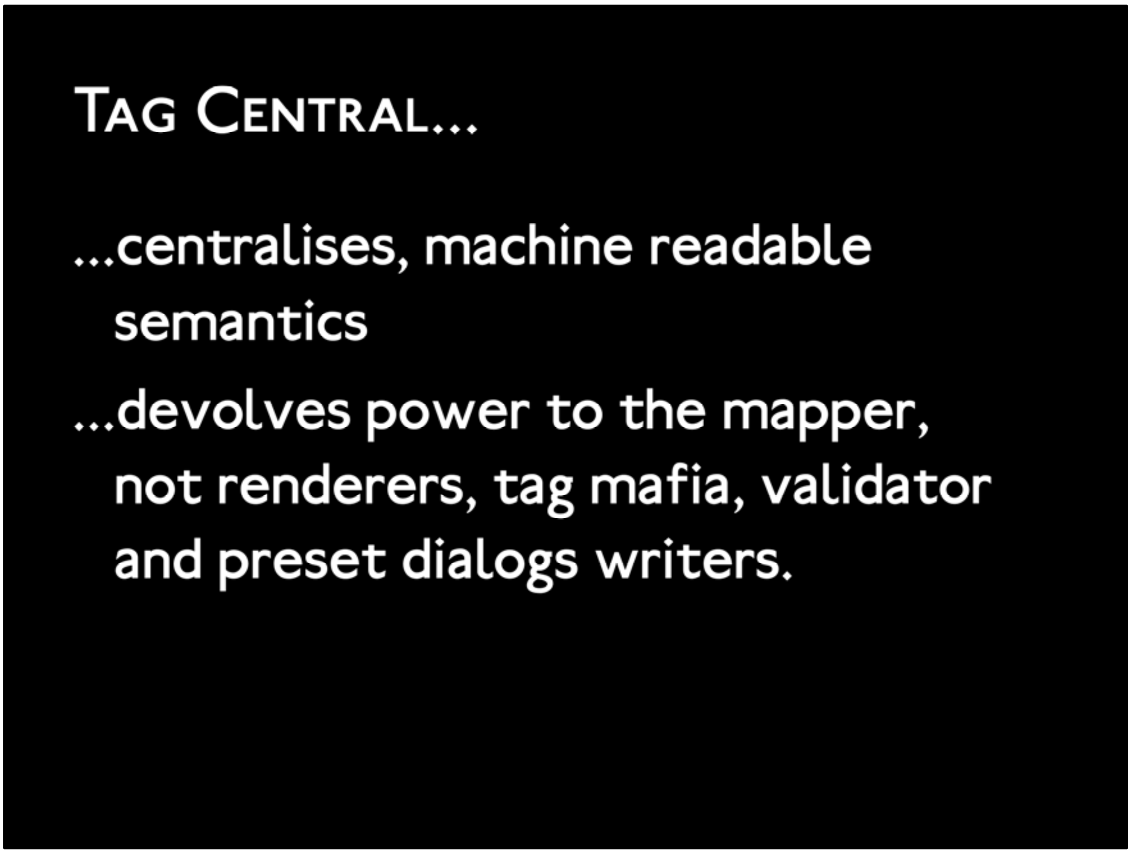


## WHAT THIS GIVES A CONSUMER…

…understanding of national defaults (e.g. router journey time calculations)

…understanding equivalences (e.g. place barangay ~ place village for rendering purposes)

For the consumer (the application that uses the data, like renderers of pictorial maps and routing programs), the defaults because much more amenable. We can also make items much more culturally specificso that we can retain the culturally-specific concept, while telling a renderer that it is sufficiently similar to something else that it can be rendered as such. It doesn’t have to be cultural either: a renderer may render confectioner’s shops with a specific icon. It we say our fudge shop is similar to a confectioners, we can get it on the map in a general way rather than just ignoring it, while retaining the conceptual specificity (and, crucially, without having to change the renderer!)



## Tag Central…

…centralises, machine readable semantics

…devolves power to the mapper, not renderers, tag mafia, validator and preset dialogs writers.

To summarise…

In essence, all this is asking of the mapper is to *document* your tags. All the rest is about communicating that documentation efficiently.

http://www.flickr.com/photos/tobanblack/3506012534/