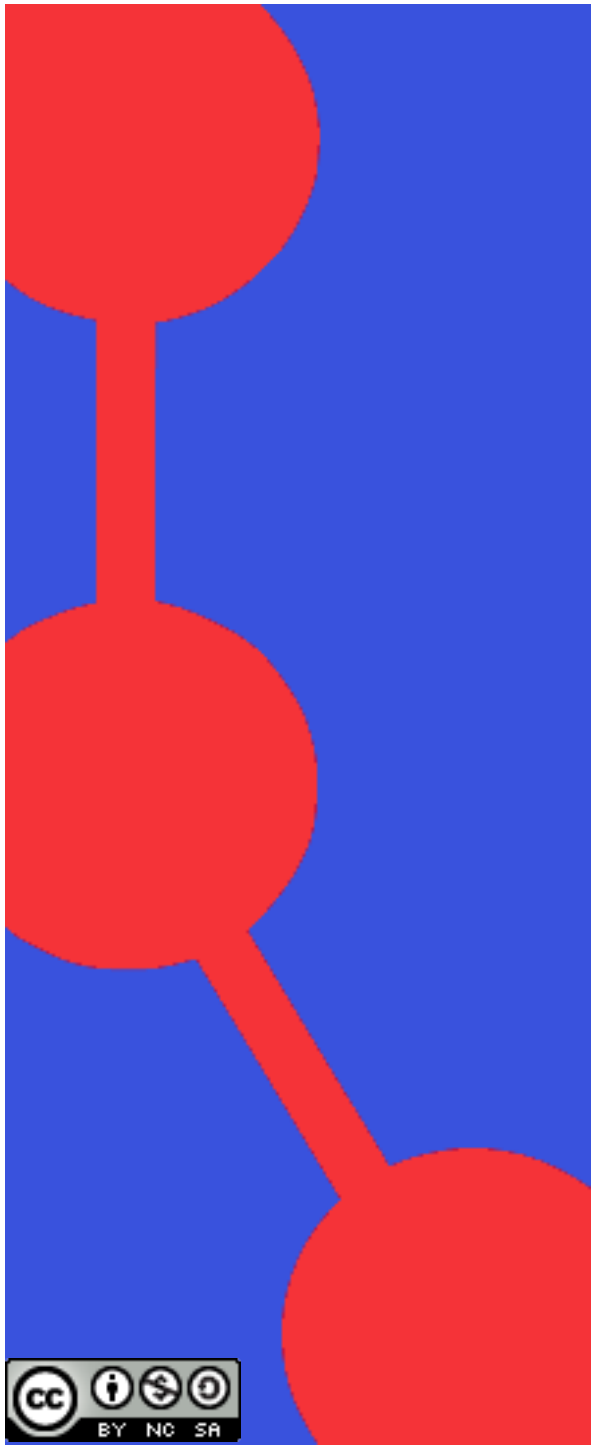


# 10 basic rules to overcome ontology engineering deadlocks in collaborative ontology engineering tasks

**Ontology Summit 2014**
**Track-C: Overcoming Ontology Engineering Bottlenecks – II**
**20/03/2014**


**Oscar Corcho**
ocorcho@fi.upm.es
@ocorcho
https://www.slideshare.com/ocorcho

# Disclaimers


- The contents of this slideset represent my own view on this topic
  - Not the views of all members of the Ontology Engineering Group
- They are based on some of my own experiences in ontology engineering
  - These are not necessarily generalisable
  - Specially not valid, probably, in ontology-savvy domains
- Actually, they are not 10 rules yet
  - Working on that (but the title looks nicer like that)

- And more important…
  - I am trying to be provocative here, to generate discussion

# Context

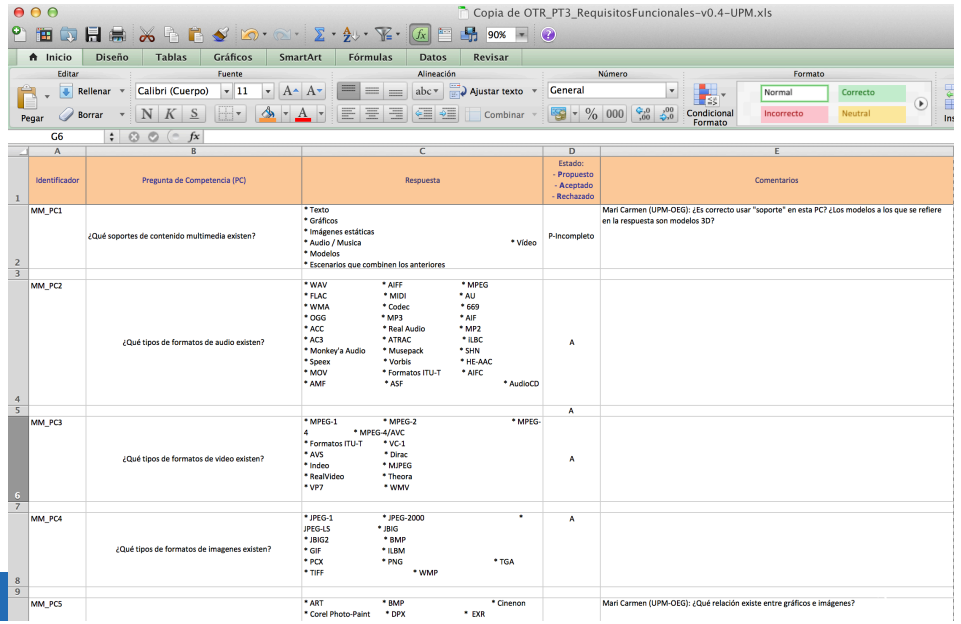
- Project BuscaMedia
  - http://www.cenitbuscamedia.es/
  - Already finished
  - A collaborative Spanish project to generate a search engine for multilingual multimedia multi-domain content
  - Multiple and very generic domains (sports, news, etc.)
  - Private tech companies, universities, content providers
- AENOR CTN 178 group on open data for smart cities
  - SC3 Gobierno/movilidad, GT1 Gobierno, SG1 Open Data
  - Starting now the real work on ontologies (to be finished by June 2014)
  - Provision of common vocabularies for datasets from open data portals
  - Private companies and public administrations (mostly cities and regions)

# Starting the ontology engineering process…

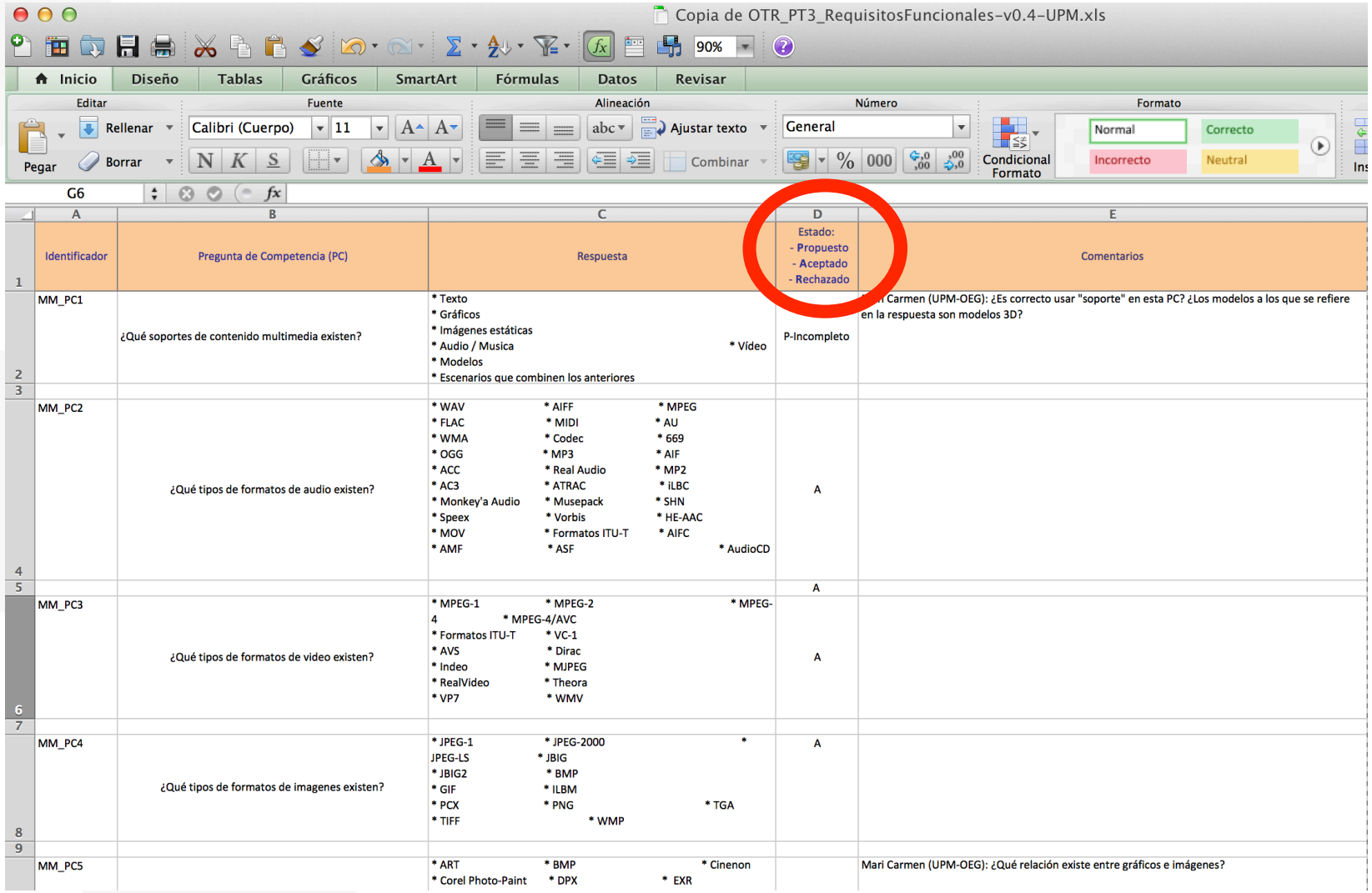
# Deadlock 1. Which ontology tool do we use?

- Which tool do we use?
  - Standalone
    - Protege / NeOn Toolkit / TopBraid Composer
  - Collaborative
    - WebProtege

- When starting, users are not worried about tools, but about discussing
  - Rec1: Use Excel (standalone or Google Docs)
  - Not just for requirements, also for documentation



# Deadlock 1. Which tool do we use?



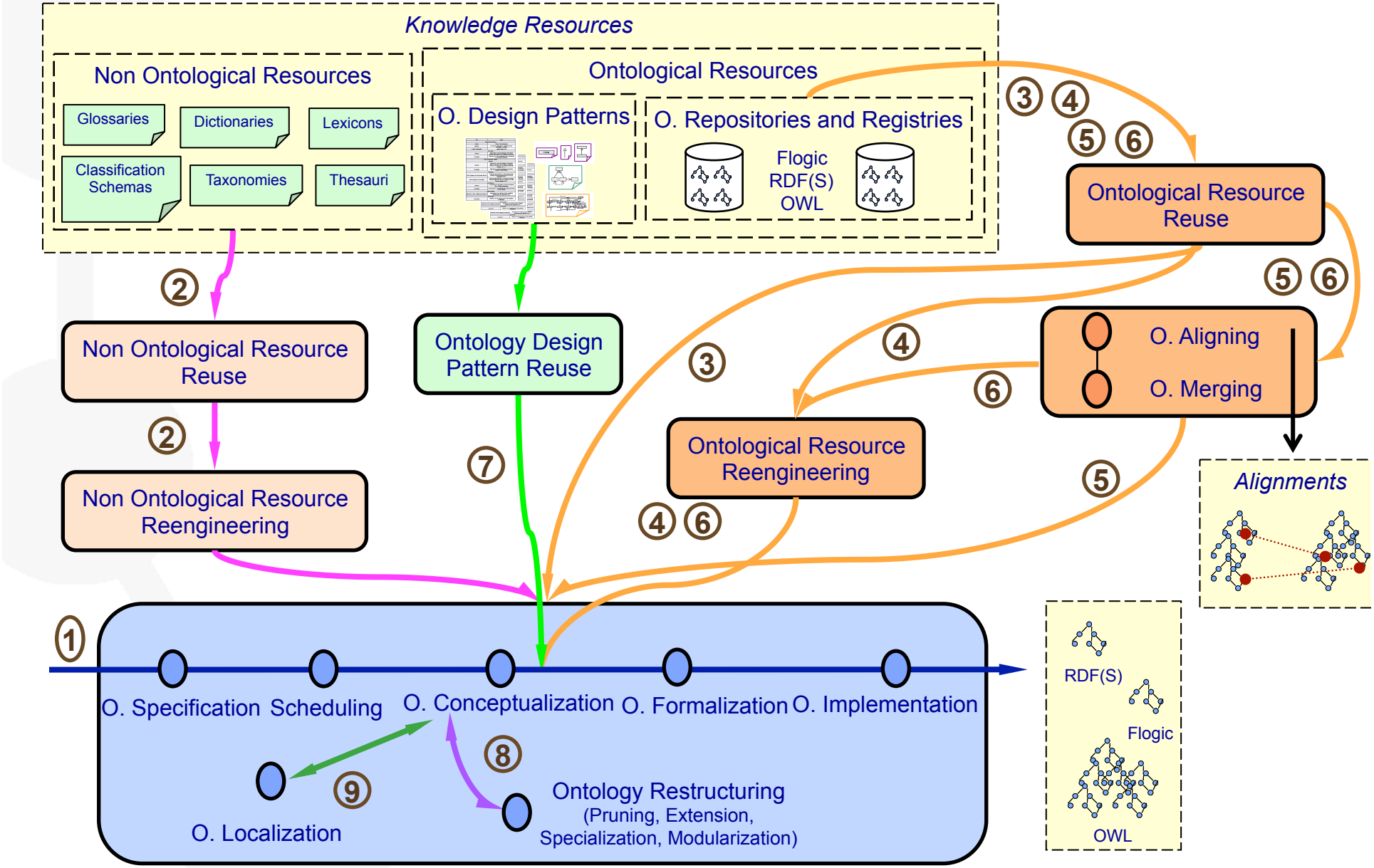

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Identificador | Pregunta de Competencia (PC) | Respuesta | Estado:<br>- Propuesto<br>- Aceptado<br>- Rechazado | Comentarios |
| 2 | MM_PC1 | ¿Qué soportes de contenido multimedia existen? | * Texto<br>* Gráficos<br>* Imágenes estáticas<br>* Audio / Musica<br>* Vídeo<br>* Modelos<br>* Escenarios que combinen los anteriores | P-Incompleto | Mari Carmen (UPM-OEG): ¿Es correcto usar "soporte" en esta PC? ¿Los modelos a los que se refiere en la respuesta son modelos 3D? |
| 3 | | | | | |
| 4 | MM_PC2 | ¿Qué tipos de formatos de audio existen? | * WAV * AIFF * MPEG<br>* FLAC * MIDI * AU<br>* WMA * Codec * 669<br>* OGG * MP3 * AIF<br>* ACC * Real Audio * MP2<br>* AC3 * ATRAC * iLBC<br>* Monkey'a Audio * Musepack * SHN<br>* Speex * Vorbis * HE-AAC<br>* MOV * Formatos ITU-T * AIFC<br>* AMF * ASF * AudioCD | A | |
| 5 | | | | A | |
| 6 | MM_PC3 | ¿Qué tipos de formatos de video existen? | * MPEG-1 * MPEG-2 * MPEG-4 * MPEG-4/AVC<br>* Formatos ITU-T * VC-1<br>* AVS * Dirac<br>* Indeo * MJPEG<br>* RealVideo * Theora<br>* VP7 * WMV | A | |
| 7 | | | | | |
| 8 | MM_PC4 | ¿Qué tipos de formatos de imagenes existen? | * JPEG-1 * JPEG-2000 * JPEG-LS * JBIG<br>* JBIG2 * BMP<br>* GIF * ILBM<br>* PCX * PNG * TGA<br>* TIFF * WMP | A | |
| 9 | | | | | |
| | MM_PC5 | | * ART * BMP * Cinenon<br>* Corel Photo-Paint * DPX * EXR | | Mari Carmen (UPM-OEG): ¿Qué relación existe entre gráficos e imágenes? |

# Deadlock 2. I have been told to reuse other ontologies

- We all recommend reusing other ontological and non-ontological resources
  - That's the basis of ontological engineering
- However, this should not be done at an early state of ontology development
  - It causes confusion
    - Should I use FOAF, or the Organization Ontology, or vCard, or schema.org?
    - And prevents people from being creative
  - It causes endless discussions about terms (and lots of problems with translations)

- Rec2: start working with experts so that they provide their definitions, and get agreements on those

# Deadlock 3. The methodology tells me to…



# Deadlock 3. The methodology tells me to…

- There are lightweight and heavyweight methodologies
  - The NeOn Methodology accounts for both cases
  - But do not explain much to domain experts…

- Rec3: use an agile approach, based on sets of competency questions for each sprint

Now we start implementing the ontologies…

# Deadlock 4. Large groups work more slowly

- Sprint mode (as I commented before)
- Prefer small groups of experts with a few ontology engineers
  - Personally I prefer groups of five people (4 experts + 1 ontology engineer)

- Rec4: Avoid non-experts, and use all experts from the same level
  - Avoid differences in ontology goals

# Deadlock 5. But these ontologies to reuse are in English

- Context:
  - Most of the ontologies that we have developed in these projects/initiatives are for Spanish users
  - But when it comes into reusing other ontologies, those are normally using English identifiers
- We all know that identifiers are not the most important element
  - e.g., labels and comments in different languages should be enough
- Main question:
  - Can we mix term identifiers in different languages?
  - Do we translate all terms to our language?

- Rec5: generate all terms in your language (extending existing ones in other languages)

# Deadlock 6. I want my ontology to do inferences…

- OWL (and description logics) is funny to teach at University
  - I enjoy teaching DL, showing the niceties of reasoning, consistency checking, etc.
  - It is useful in many domains
  - But developing such ontologies is a task for hardcore ontology engineers (not for public administration tech people)

- Rec6: Just work with text patterns, and guide them to write good term definitions
  - A district contains only neighbourhoods and census sections
  - A shop can have at most three economic activities associated to it

# Deadlock 7. I want my ontology to be ligthweight…

- I want my ontology to be used for Linked Data publishing
  - I have been said not to put domains or ranges
  - I have been said to create only light taxonomies
  - I have been said to use only RDF Schema

- Rec7: again, text patterns are the best option to follow here
  - Don’t make your experts worry about languages or formal aspects

And assuming that we have the OWL/RDFS files…

# Deadlock 8. The ontology is done, but is it good?

- We have been said that the ontology has to be consistent
  - And I load it in Protégé, run the reasoner and it says that everything is ok
  - So no more work is needed… Let's go home

- Rec8: it's ok to run the reasoner, but that won't tell you enough. Go for other non-logical checks (e.g., use the Oops! Pitfall scanner)
  - http://www.oeg-upm.net/oops/
  - But do not get obsessed with it

# Deadlock 9. How do I tell others how to use the ont?

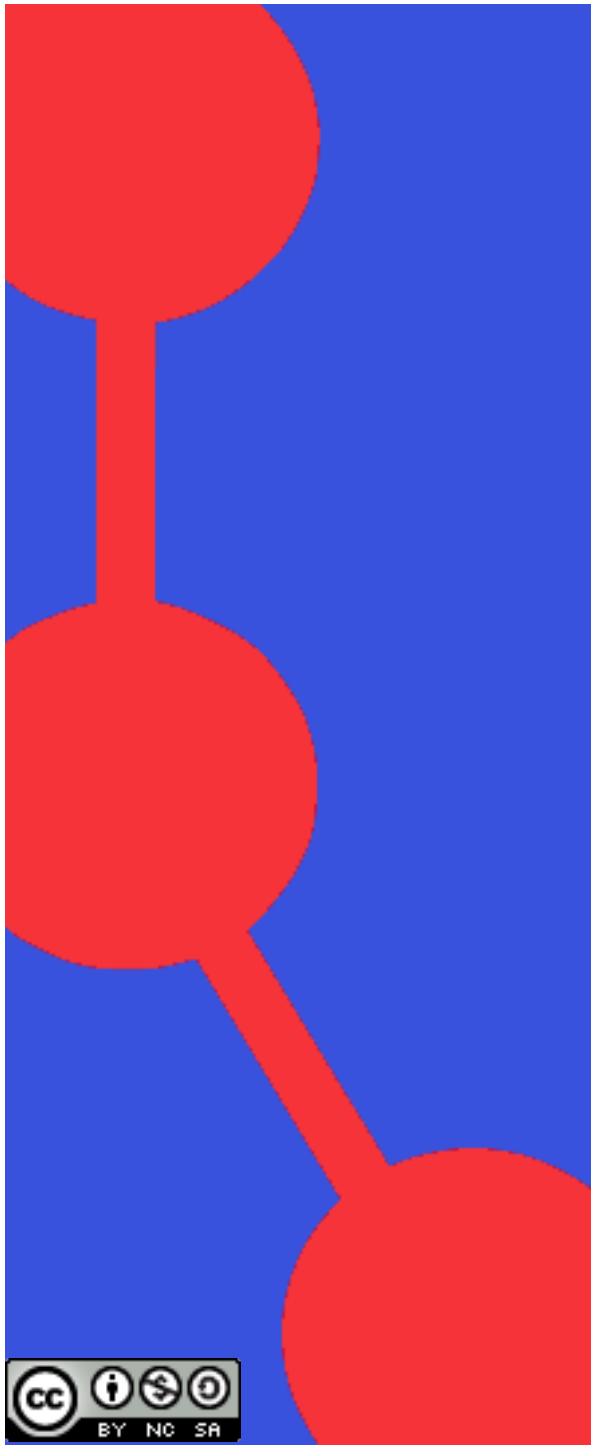
- The ontology is now done, it has been agreed by a group of people, but…
  - Should I put the OWL/RDFS ontology somewhere?
  - How do I explain people how to use it?

- Rec9: simple documentation (in HTML, in Word), with simple examples, with a link to the revised competency questions, and a simple diagram!!
  - E.g., HTMLTemplateGenerator based on LODE ( https://github.com/dgarijo/HtmlTemplateGenerator)
  - Examples at http://vocab.linkeddata.es/

# Again… Disclaimers


- The contents of this slideset represent my own view on this topic
  - Not the views of all members of the Ontology Engineering Group
- They are based on some of my own experiences in ontology engineering
  - These are not necessarily generalisable
  - Specially not valid, probably, in ontology-savvy domains
- Actually, they are not 10 rules yet
  - Working on that (but the title looks nicer like that)

- And more important…
  - I am trying to be provocative here, to generate discussion



# 10 basic rules to overcome ontology engineering deadlocks in collaborative ontology engineering tasks

**Ontology Summit 2014**
**Track-C: Overcoming Ontology Engineering Bottlenecks – II**
**20/03/2014**


**Oscar Corcho**
ocorcho@fi.upm.es
@ocorcho
https://www.slideshare.com/ocorcho