HOW TO DOCUMENT ONTOLOGY DESIGN PATTERNS

Nazifa Karima, Karl Hammar and Pascal Hitzler

OVERVIEW

- Motivation and Background.
- Pattern Documentation Components Survey.
- Ontology Engineering Methods and Tooling Surveys.
- Summary of Findings.

MOTIVATION

- For the purpose of this talk and paper, ODP == CODP
- ODPs are being created or extracted by the community, but:
  - ODP reuse in practice is still rather limited.
  - ODP evaluation is even more limited.
- By reflecting on the obstacles to reuse (including evaluating and/or improving existing ODPs), we can hopefully overcome those obstacles.
  - A key obstacle (among several!) is the lack of knowledge regarding how to best document ODPs.

BACKGROUND

- Karima and Hitzler initiated survey on ODP documentation to adress aforementioned knowledge gap.
- Hammar had independently worked on similar issues using (among other methods) similar surveys.
- Excellent opportunity for cross-validation of results!
- Reservation: the latter work is still ongoing - some response sets are still rather small.

PATTERN DOCUMENTATION COMPONENTS SURVEY

- Ten ODP documentation components listed:
  - Schema diagram
  - Instantiation example
  - Competency questions
  - Axiomatisation
  - OWL file
  - Related patterns
  - Metadata
  - Example SPARQL queries
  - Example datasets
  - SHACL constraints

- Question: How essential is each component for good ODP documentation?

- Scale from 5 ('essential') through 1 ('not important').

PDC SURVEY RESPONSES (1)

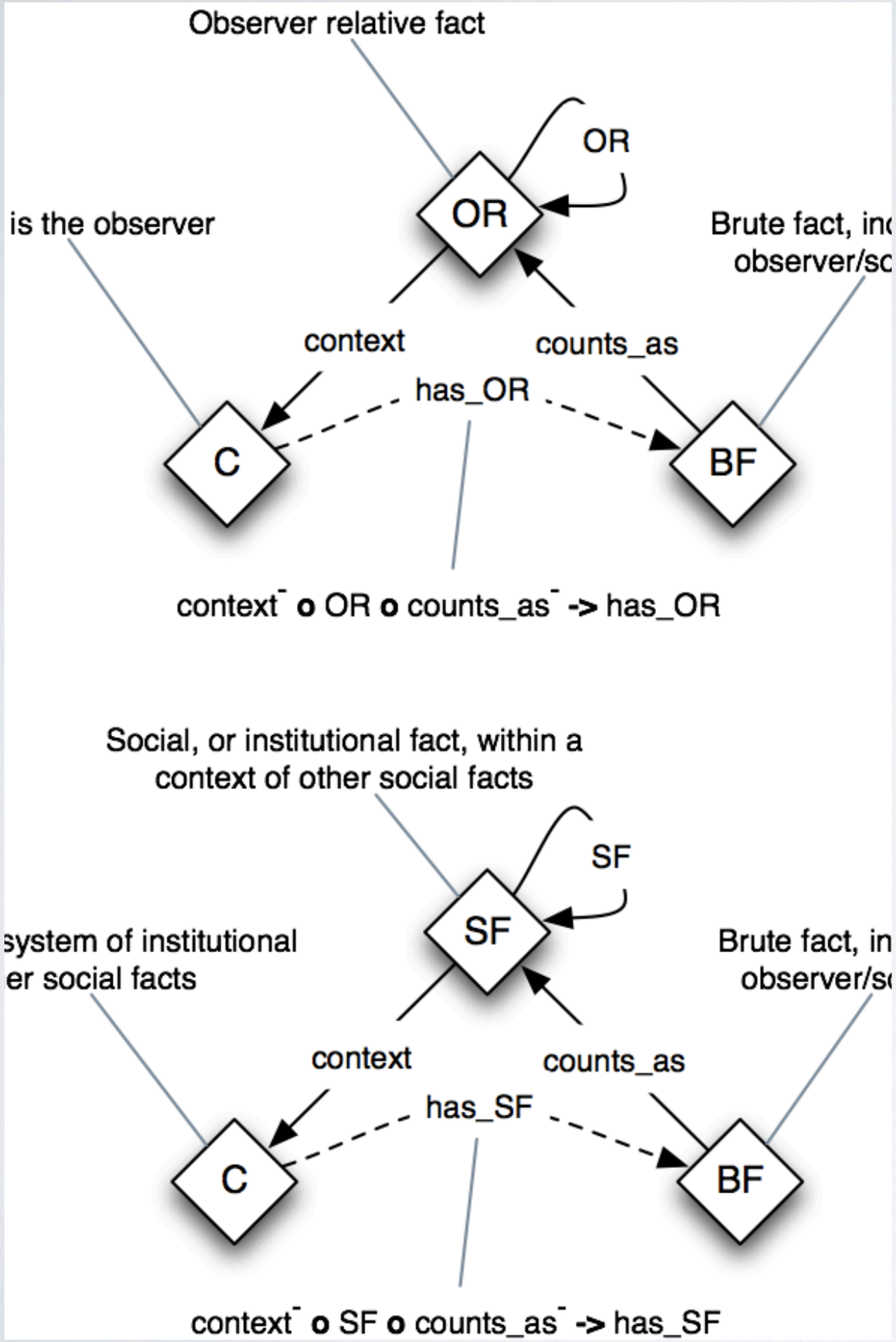
- 35 responses.
- Approximately half faculty at universities — other half from institutes or industry.
- Ranking computed by mean score for each component.

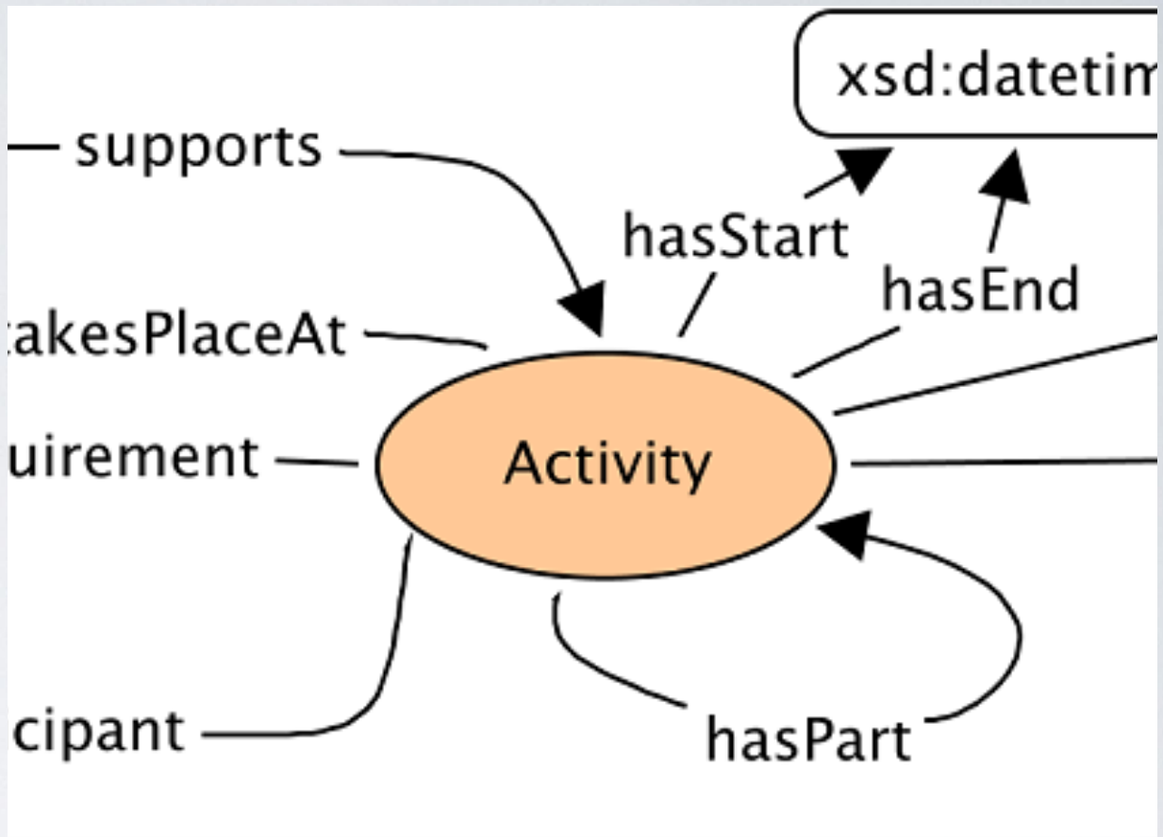
PDC SURVEY RESPONSES (2)

Component	Score (percentage of responses)					Mean score
	5	4	3	2	1	
Schema diagram	74,3 %	20,0 %	5,7 %	0,0 %	0,0 %	4,7
Instantiation example	34,3 %	48,6 %	11,4 %	2,9 %	2,9 %	4,1
Competency questions	40,0 %	34,3 %	14,3 %	8,6 %	2,9 %	4,0
Axiomatisation	40,0 %	22,9 %	31,4 %	5,7 %	0,0 %	4,0
OWL file	28,6 %	40,0 %	20,0 %	8,6 %	2,9 %	3,8
Related patterns	17,1 %	48,6 %	20,0 %	14,3 %	0,0 %	3,7
Metadata	22,9 %	34,3 %	31,4 %	11,4 %	0,0 %	3,7
Example SPARQL queries	17,1 %	45,7 %	20,0 %	11,4 %	5,7 %	3,6
Example datasets	2,9 %	32,4 %	44,1 %	17,7 %	2,9 %	3,1
SHACL constraints	3,4 %	17,2 %	41,4 %	24,1 %	13,8 %	2,7

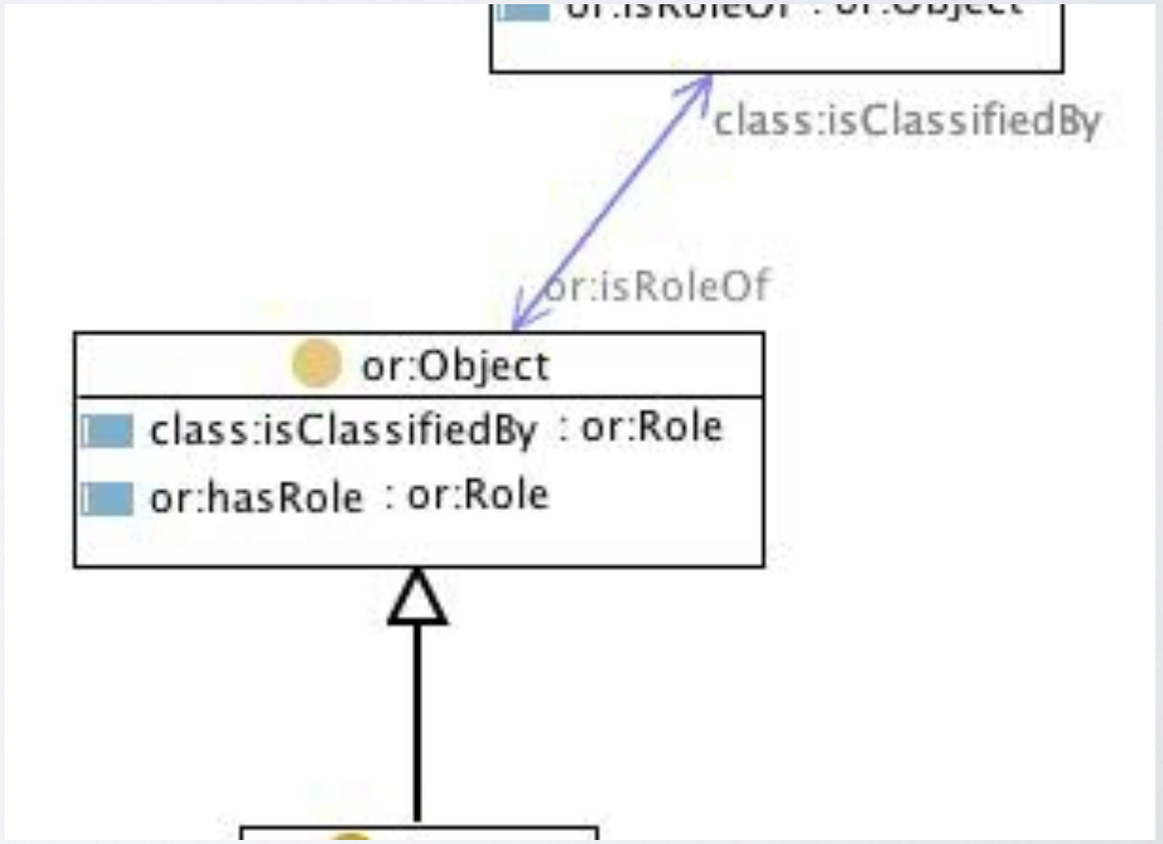
PDC REFLECTIONS (1)

- Schema diagram clearly essential — but:
- Only 75 % of ODPs in portal have diagrams.
- Diagram style and quality varies widely — likely to decrease understandability and appropriateness recognizability.


Observer relative fact
OR
OR
is the observer
Brute fact, inc
observer/sc
context
counts_as
has_OR
C
BF
context- o OR o counts_as- -> has_OR


xsd:datetim
supports
hasStart
hasEnd
takesPlaceAt
uirement
Activity
cipant
hasPart

Social, or institutional fact, within a
context of other social facts
SF
SF
system of institutional
er social facts
Brute fact, in
observer/s
context
counts_as
has_SF
C
BF
context- o SF o counts_as- -> has_SF


class:isClassifiedBy
or:isRoleOf
or:Object
class:isClassifiedBy : or:Role
or:hasRole : or:Role

PDC REFLECTIONS (2)

- Instantiation example also essential.

- Instantiation examples ought to exist, since the ODPs must have come from some use case (whether made from clean slate or extracted from existing ontology), but:

  - Only 7 of 130 ODPs have KnownUse field filled (and only 60 % of ODPs have Scenarios field filled).

  - There's no automatic mapping from ODP to its reuse, whether inside portal (i.e., through other ODPs) or outside it.

PDC REFLECTIONS (3)

- CQ:s also considered essential, but:
  - Only 73 % of ODPs have CQ:s formalized.
  - CQ content structure varies widely among these: line breaks or no line breaks, numbering or not, embedded HTML content or bullet points in some. This makes reuse of ODP portal content in other systems difficult.
  - CQ language varies: most are free-standing questions that an ontology using the ODP can answer, but several are not: some are phrased as statement of fact, some consist of sentences that have dependencies on one another, etc. This makes search for ODPs (whether automated by some search engine or simply human browsing) needlessly difficult.

PDC REFLECTIONS (4)

- Axiomatization considered essential by many respondents.
- I.e., some human-readable logic representation other than the OWL file itself.
- No such field in the portal today.

OE METHODS & TOOLING SURVEYS

- OEM: Performed at and around ISWC 2014. 33 questions on OE method and tooling preferences. 81 respondents (28 for ODP-specific portion).

- OET: Performed in projects Hammar works within. 13 questions on OE tooling preferences. 14 responses.

- Subset of these surveys overlap with the survey by Karima & Hitzler.

OE METHODS SURVEY

- Question: "*How important are each of the following factors in enabling industry adoption of semantic web ontology technology?*" (5 point scale ranging from "Not important" through "Critically important").
    - Factors to choose from: Tooling Quality, Documentation Quality, Method Support.
- Question: "*In an ODP search engine or an ODP portal/catalogue, which fields or metadata about an ODP is most important when ascertaining the suitability of that ODP for reuse?*" (5 point scale ranging from "Not important" through "Critically important").
    - Fields to rank: Title, Description, Graphical illustration, Competency questions, Example uses, Size in classes, Size in axioms, OWL 2 Profile adherence.

OE METHODS RESPONSES (1)

	Some experience	Confident	Expert	All
Tooling Quality	64 %	100 %	92 %	90 %
Documentation Quality	50 %	79 %	69 %	70 %
Method Support	62 %	68 %	64 %	67 %

Percentage of 81 respondents considering each factor *Very important* or *Critically important* in enabling industry adoption of SemWeb tech.

OE METHODS RESPONSES (2)

	Some experience	Confident	Expert	All
Example uses	100 %	75 %	88 %	87 %
Description	75 %	50 %	81 %	73 %
CQ:s	75 %	29 %	71 %	63 %
Graphical Illustration	75 %	50 %	53 %	55 %
Title	0 %	13 %	56 %	37 %
OWL 2 Profile	25 %	13 %	13 %	14 %
Size in Classes	0 %	0 %	18 %	10 %
Size in Axioms	0 %	0 %	13 %	7 %

Percentage of 28 respondents considering each component *Very important* or *Critically important* in evaluating ODP for reuse.

OE METHODS REFLECTIONS

- Title is not considered important - possibly effect of question phrasing? (”*when ascertaining the suitability of <u>that</u> ODP for reuse*”).

- Three of top four components are identical to those in Karima & Hitzler’s survey. The differing component in each is lacking in the other survey.

OEM FREE-FORM COMMENTS

- *"Much of the documentation is included in research papers; more information with direct sentences should be available outside research publications."*
  – Expert with 10 years of experience.

- *"In many cases, the documentation is non existent or incomplete."*
  – Application developer with 2+ years of experience.

- *"IMO, one of the main issue with the re-use of ontology patterns (e.g. those defined at http://ontologydesignpatterns.org/) is the lack of concrete documentation. Generally, the pattern is described at a very generic level and explained based on a particular use case. Furthermore, the graphical representation (when available) are not consistent across patterns making it difficult to adopt."*
  – Expert with 9+ years of experience

- *"Documentation: some patterns could be ambiguous as, therefore they should be documented in a way the user can be sure they are using the right pattern."*
  – PhD student with 7 years of experience

OE TOOLING SURVEY

- Intended to study the relative importance of the documentation component fields within the ”General description” section of the ODP portal.
- Question: *”Please rank the following ODP documentation fields in terms of how important you think they are to understanding whether an ODP is suitable for reuse in your project”*.
    - Documentation fields: Name, Intent, Solution, Consequences, Competency Questions, Scenarios, Domains.

OE TOOLING RESPONSES

Documentation field	Ranking score
Intent	5.91
Competency Questions	4.64
Name	4.09
Solution	3.82
Scenarios	3.45
Domains	3.18
Consequences	2.91

Average ranking of documentation fields (first place scores 7, last place scores 1)

OE TOOLING REFLECTIONS

- Competency questions still considered important.
- *Intent* scoring rather high. In the portal, 114 of 130 CODPs have this field set, though in the majority of cases, the field is filled with a simple sentence that does not give much more guidance than the ODP name.
  - Example: *"To represent time intervals"*.
- Bonus finding: users consistently prefer VOWL graphical notation for illustrations over the representations available in other Protégé plugins.

IN SUMMARY

- Most important ODP documentation components or fields: *graphical illustration*, *example instantiations*, *competency questions*.

- Other important fields: *textual description*, particularly *intent*.

- Candidate for inclusion: *axiomatization*.

- Community portal ODPs too often lack the above information.

- Troubling in light of the importance of documentation quality to Semantic Web adoption. Respondents' free-text responses indicate the same.