

# The Workflows of Ontology Authoring: Controlled vs. Naturalistic Settings

Markel Vigo, Nicolas Matentzoglou, Caroline Jay, Robert Stevens

UKON 2016

School of Computer Science  
University of Manchester



# Motivation

- Four years ago we knew very little about
  - How ontologists go about authoring
  - Typical authoring workflows
  - The effectiveness of current tool support
- WhatIf Project set out to start fill in the blanks
  - Among others: Wang et al 2015, Walk et al 2015, Horridge et al 2013, and more

# Motivation

- Four years ago we knew very little about

– H File Edit View Reasoner Tool

– T untitled-ontology-2006 (http://www.semanticweb.org/nico/ontologies/2016/3/untitled-ontology-2006)

– T Active Ontology x Entities x Classes x Object Properties x Data Properties x Annotation Properties x Class mat

– T Class hierarchy Class hierarchy (inferred) Usage

Class hierarchy:

2. Run the reasoner

3. Check the inferred hierarchy

- Wh

– A Thing

– 2 DomainEntity

- Person
  - Ancestor
  - Descendant
  - Man
  - Woman
- Sex
  - Femaleness
  - Maleness

1. Define a class

inks

idge et al

Description:

Equivalent To +

SubClass Of +

# Some recap...

- Interview study with ontologists (UKON 2014)
  - Problems and strategies for sense-making, exploration, search, building, debugging and evaluation

Vigo, Jay and Stevens (2014) *Design insights for the next wave ontology authoring tools*. Proc. of CHI 2014

Vigo, Bail, Jay and Stevens (2014) *Overcoming the pitfalls of ontology authoring: strategies and implications for tool design*. Int. J. of Human-Computer Studies

- User study in the lab (UKON 2015)
  - Identification of authoring workflows for exploration, editing and reasoning

Vigo, Jay and Stevens (2015) *Constructing Conceptual Knowledge Artefacts: Activity Patterns in the Ontology Authoring Process*. Proc. of CHI 2015

- Implications for tool design

# External validity of lab study at risk

## Lab study

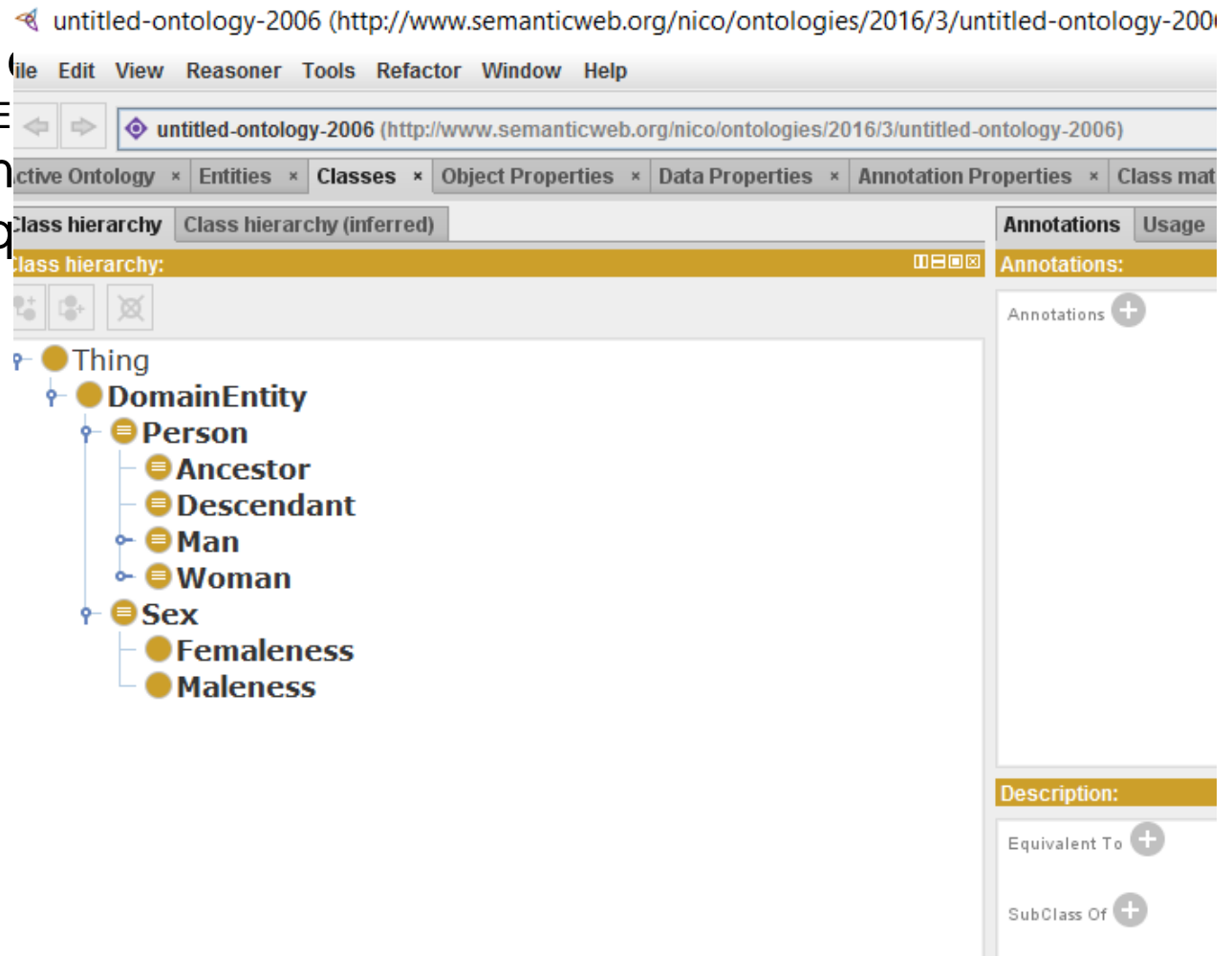
- Tasks were predefined
- Toy ontology given
- Controlled environment with lab apparatus
- No time limit but under pressure
- 16 users
- 30-75 min.
- Protégé4US + eye tracker
- Data driven analysis of event and gaze data

## Remote study

- Their own tasks
- Their own ontologies
- On their own working environment
- No time constraints
- 7 users
- 1 day-1 month
- Protégé4US
- Data driven analysis of event data

# Analysis: identifying workflows from log data

1. Collect raw event data
2. Clean data
3. Merge same (HierarchyE
4. Workflow min
5. Workflow seq



# Analysis: identifying workflows from log data

1. Collect raw event data
2. Clean data
3. Merge same consecutive events:  
HierarchyExpanded, HierarchyExpanded → M\_HierarchyExpanded
4. Workflow mining through N-gram analysis
5. Workflow sequence mining

```
P4.csv — 4 (git: master)
12727 1430317189737,1,Entities,MouseExited,,OWLEditorKit,HIERARCHY
12728 1430317190450,1,Entities,MouseEntered,,OWLEditorKit,HIERARCHY
12729 1430317190594,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5220,HIERARCHY
12730 1430317190618,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_7319,HIERARCHY
Q 12731 1430317191480,1,Entities,HExpanded,asserted,http://purl.obolibrary.org/obo/CHEBI\_59132,HIERARCHY
12732 1430317192026,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_3328,HIERARCHY
12733 1430317192066,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_7319,HIERARCHY
12734 1430317192130,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5220,HIERARCHY
12735 1430317192324,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5300,HIERARCHY
12736 1430317192370,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5091,HIERARCHY
12737 1430317192705,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5300,HIERARCHY
12738 1430317192850,1,Entities,EntitySelected,asserted,http://purl.obolibrary.org/obo/CHEBI\_53000,HIERARCHY
12739 1430317194305,1,Entities,MouseExited,,OWLEditorKit,HIERARCHY
12740 1430317195658,1,Entities,MouseEntered,,OWLEditorKit,HIERARCHY
12741 1430317195970,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5221,HIERARCHY
12742 1430317196082,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5221,HIERARCHY
Q 12743 1430317196862,1,Entities,HExpanded,asserted,http://purl.obolibrary.org/obo/CHEBI\_52210,HIERARCHY
12744 1430317197474,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_4870,HIERARCHY
Q 12745 1430317197951,1,Entities,HExpanded,asserted,http://purl.obolibrary.org/obo/CHEBI\_48705,HIERARCHY
12746 1430317198418,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_7119,HIERARCHY
12747 1430317198578,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_7231,HIERARCHY
12748 1430317198626,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_6356,HIERARCHY
12749 1430317198658,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_6458,HIERARCHY
12750 1430317198762,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_6855,HIERARCHY
Q 12751 1430317199255,1,Entities,HExpanded,asserted,http://purl.obolibrary.org/obo/CHEBI\_24621,HIERARCHY
12752 1430317199498,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_3784,HIERARCHY
12753 1430317199506,1,Entities,EntityHovered,asserted,http://purl.obolibrary.org/obo/CHEBI\_5011,HIERARCHY
12754 1430317199962,1,Entities,MouseExited,,OWLEditorKit,HIERARCHY
12755 1430317200850,1,Entities,MouseEntered,,OWLEditorKit,HIERARCHY
```

# Analysis: identifying workflows from log data

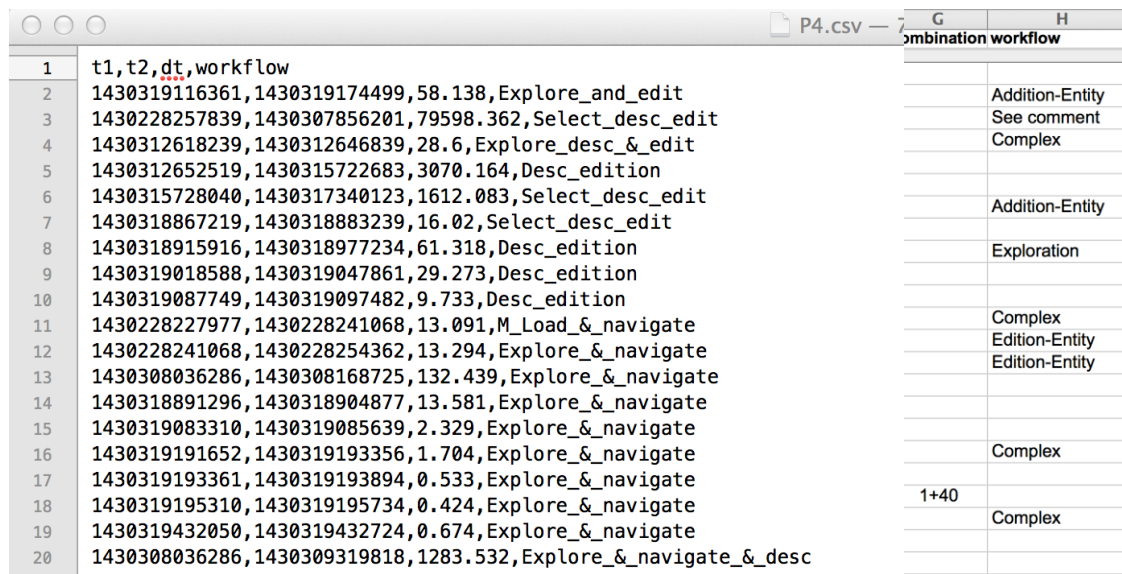
1. Collect raw event data
2. Clean data
3. Merge same consecutive events:  
     HierarchyExpanded, HierarchyExpanded → M\_HierarchyExpanded
4. Workflow mining through N-gram analysis
5. Workflow sequence mining

A	B	C	D	E	F	G	H
id	size	sequence	frequency	permutation	extension/mod	combination	workflow
34	8	Mdesc_select&modify Mdesc_select&modify	64	1			
35	2	desc_select&modify_e	62		1		Addition-Entity
36	2	entityedited m_entity_selected	62				See comment
37	3	m_entity_selected desc_select&modify	62				Complex
38	2	entityselected m_hierarchy	60	15			
39	2	entityselected individualadded	55				
40	3	dslect&mod&save	54		1		Addition-Entity
41	3	select&modify entityselected	54		21		
42	2	loadontology m_hierarchy	52				Exploration
43	3	desc_select&modify entityselected	50		1		
44	4	select&addclass select&addclass	50	6			
45	2	m_entity_selected m_entity_edited	46				Complex
46	3	w_edition	46				Edition-Entity
47	2	m_description_selected entityedited_e	44				Edition-Entity
48	4	classaddition select&addclass entityselected	44	6			
49	2	exploration2	42	23			
50	2	save entityselected	42				
51	2	save loadontology	42				Complex
52	4	entityedited dslect&mod&save	42		40		
53	5	desc_select&modify dslect&mod&save	42			1+40	
54	2	entityselected m_entity_edited	40				Complex
55	4	m_entity_selected desc_select&modify descriptionselecte	40		37		
56	2	annotationedited descriptionselected	38	16			



# Analysis: identifying workflows from log data

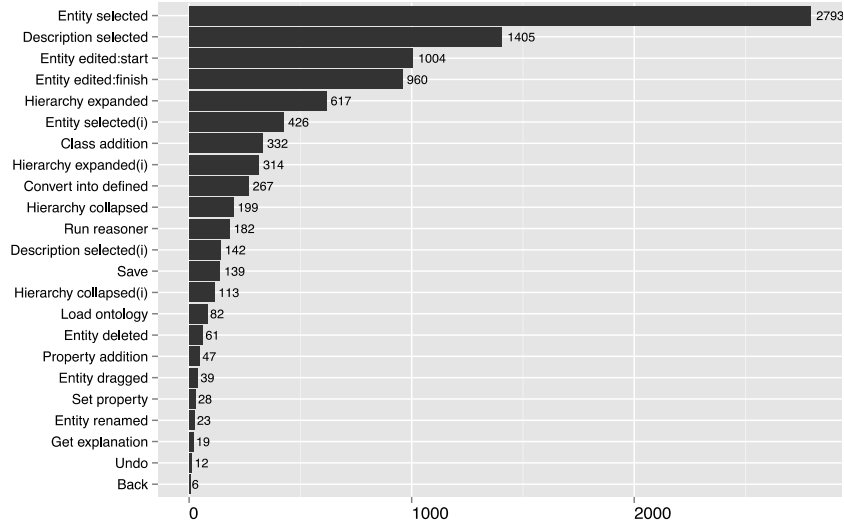
1. Collect raw event data
2. Clean data
3. Merge same consecutive events:  
HierarchyExpanded, HierarchyExpanded → M\_HierarchyExpanded
4. Workflow mining through N-gram analysis
5. Workflow sequence mining



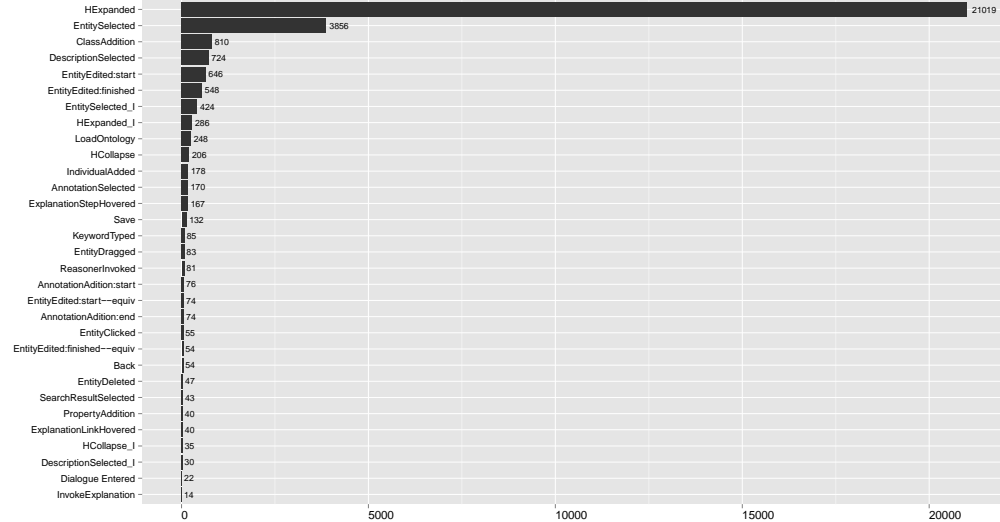
	G	H
	combination workflow	
1	t1, t2, dt, workflow	
2	1430319116361, 1430319174499, 58.138, Explore_and_edit	
3	1430228257839, 1430307856201, 79598.362, Select_desc_edit	Addition-Entity
4	1430312618239, 1430312646839, 28.6, Explore_desc_&_edit	See comment
5	1430312652519, 1430315722683, 3070.164, Desc_edition	Complex
6	1430315728040, 1430317340123, 1612.083, Select_desc_edit	
7	1430318867219, 1430318883239, 16.02, Select_desc_edit	Addition-Entity
8	1430318915916, 1430318977234, 61.318, Desc_edition	Exploration
9	1430319018588, 1430319047861, 29.273, Desc_edition	
10	1430319087749, 1430319097482, 9.733, Desc_edition	
11	1430228227977, 1430228241068, 13.091, M_Load_&_navigate	Complex
12	1430228241068, 1430228254362, 13.294, Explore_&_navigate	Edition-Entity
13	1430308036286, 1430308168725, 132.439, Explore_&_navigate	Edition-Entity
14	1430318891296, 1430318904877, 13.581, Explore_&_navigate	
15	1430319083310, 1430319085639, 2.329, Explore_&_navigate	
16	1430319191652, 1430319193356, 1.704, Explore_&_navigate	Complex
17	1430319193361, 1430319193894, 0.533, Explore_&_navigate	
18	1430319195310, 1430319195734, 0.424, Explore_&_navigate	1+40
19	1430319432050, 1430319432724, 0.674, Explore_&_navigate	Complex
20	1430308036286, 1430309319818, 1283.532, Explore_&_navigate_&_desc	

# Results: descriptive statistics

## Lab



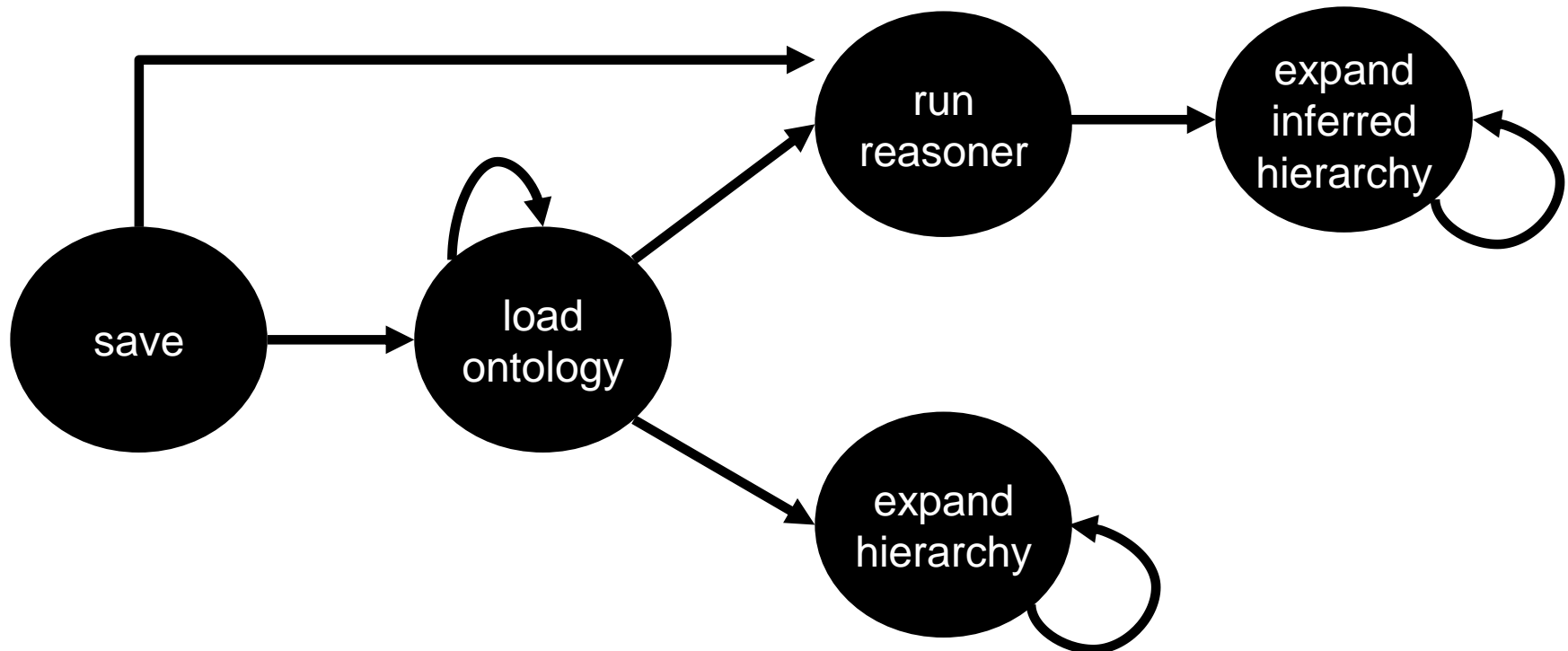
## Remote



- 9K events (lab) vs 30K (remote) – not including mouse hovering events
- Similar distribution, same most frequent events
- More variety in the remote setting
  - More undo events: *back*, *undo*, *delete*
  - More search
  - More Individuals and annotations

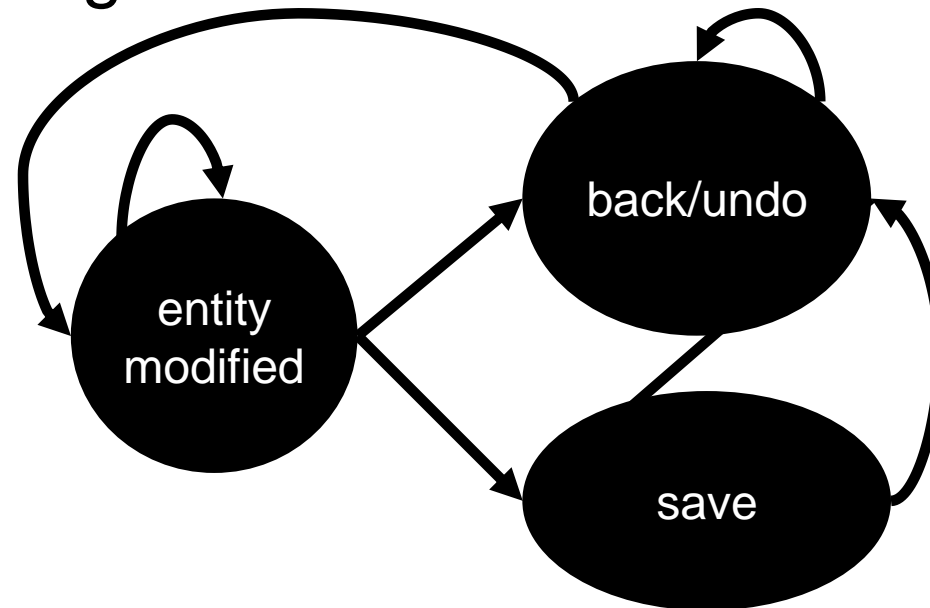
# Results: identified workflows I

- Load an ontology & explore, *what's new?*
  - Saving before loading an ontology
  - Multiple ontology loading
  - Running the reasoner right after loading ontologies

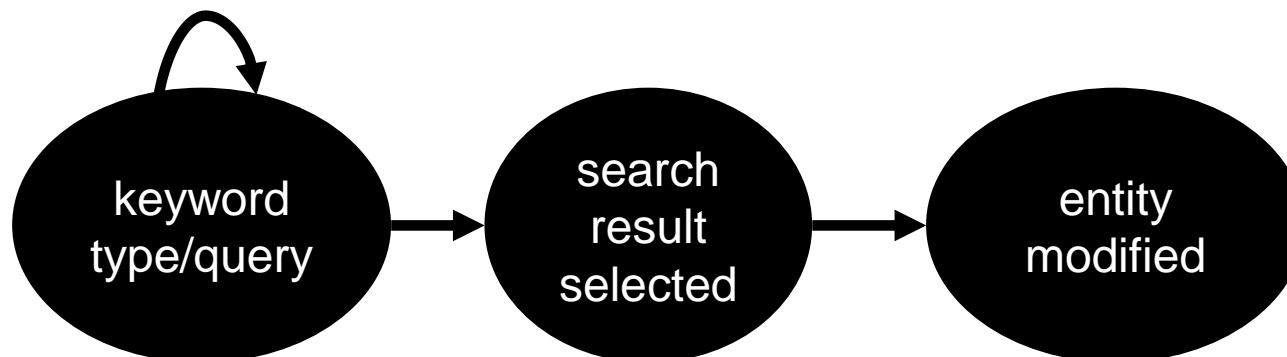


# Results: identified workflows II

- Heavy editing

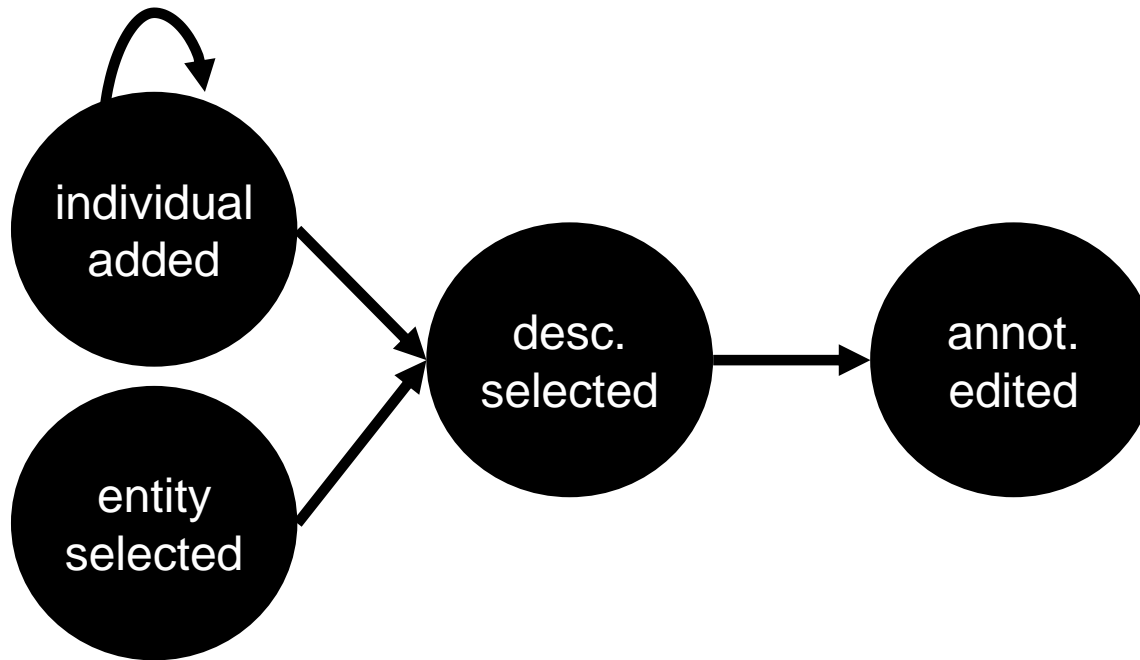


- Search & modify entity



# Results: identified workflows III

- Annotations





# Evidence-based development

- Horridge et al 2013
- Key insight @WhatIf: Understanding the consequences of modelling actions critical
  - Denaux et al 2011
  - Interview Study
  - Lab and Remote study

# Evidence-based development

Whatif Entailment Inspector: Done! 100%

Keep critical inferences in view  
 Show only selected entities

### New Inferences +

P	OWL	Info	G
P2	Ancestor $\sqsubseteq$ Descendant	?	
P2	Ancestor $\sqsubseteq$ Person	?	
P2	Descendant(William_Bright_1970)	?	>
P2	Descendant $\equiv$ Person	?	
P2	Descendant $\sqsubseteq$ DomainEntity	?	
P2	FemaleAncestor(Eileen_Mary_Rever_1929)	?	>
P2	FemaleAncestor $\sqsubseteq$ Ancestor	?	
P2	FemaleAncestor $\sqsubseteq$ FemaleDescendant	?	
P2	FemaleAncestor $\sqsubseteq$ Woman	?	
P2	FemaleDescendant $\equiv$ Woman	?	
P2	FemaleDescendant $\sqsubseteq$ Descendant	?	
P2	FemaleDescendant $\sqsubseteq$ Person	?	
P2	MaleAncestor(Charles_Herbert_Rever_1895)	?	>
P2	MaleAncestor $\sqsubseteq$ Ancestor	?	
P2	MaleAncestor $\sqsubseteq$ MaleDescendant	?	
P2	MaleAncestor $\sqsubseteq$ Man	?	
P2	MaleDescendant $\equiv$ Man	?	
P2	MaleDescendant $\sqsubseteq$ Descendant	?	
P2	MaleDescendant $\sqsubseteq$ Person	?	
P2	Man $\sqsubseteq$ Descendant	?	

### Lost inferences -

P	OWL	Info	G

### Consequences of new assertions

P	OWL	Type	Info	G
P4	Ancestor $\equiv$ Person $\sqcap$ ( $\exists$ isAncestorOf.Person)	OWL RL Profile Violation	?	
P4	Ancestor $\equiv$ Person $\sqcap$ ( $\exists$ isAncestorOf.Person)	OWL QL Profile Violation	?	
P4	Descendant $\equiv$ Person $\sqcap$ ( $\exists$ hasAncestor.Person)	OWL RL Profile Violation	?	
P4	Descendant $\equiv$ Person $\sqcap$ ( $\exists$ hasAncestor.Person)	OWL QL Profile Violation	?	



# Conclusions

- The remote study
  - Corroborates the lab findings
  - Extends our findings
- We make contributions to
  - Knowledge:
    - We know more about how ontologists go about authoring ontologies
    - User stories/ontologists personas: we can identify and profile the type of ontologists.
  - Evidence-based SE:
    - Inference Inspector
    - Diff-tool based on ECCO, Goncalves et al 2012
- Next
  - Evaluating Inference Inspector
  - Exploration of new axis of analysis: Task difficulty estimation using pupillometry
  - Cross-comparing data from more than 6 independent studies

# Questions?

## The Workflows of Ontology Authoring: Controlled vs. Naturalistic Settings

Markel Vigo, Nicolas Matentzoglou, Caroline Jay, Robert Stevens

School of Computer Science

University of Manchester