MANCHESTER 1824

The University of Manchester

Working with big biomedical ontologies

Robert Stevens

BioHealth Informatics Group

University of Manchester

MANCHESTER 1824

The University of Manchester

The message

- Ontologies authored in the Web Ontology Language (OWL) can be tricky to author and understand
- But we're starting to find ways of managing the authoring and comprehension of these artefacts
  - Natural language versions of class descriptions;
  - Abstracting over ontologies to find "patterns" of axioms;
  - Scripting not hand-crafting axioms – exploiting patterns;
  - Making it easy to do the right thing – semantic spreadsheets
  - Hiding the ontology and hiding the OWL

MANCHESTER 1824

The University of Manchester

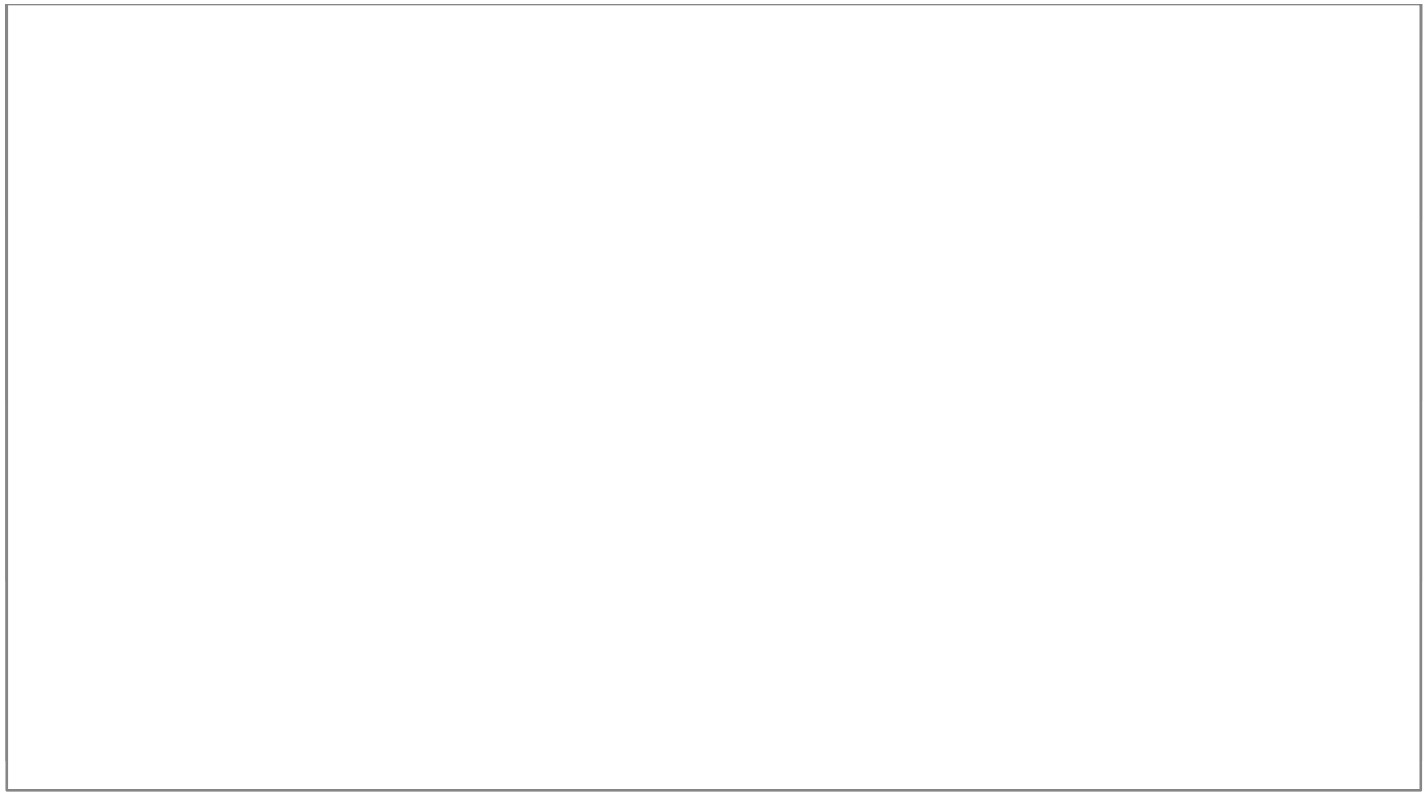
The need for descriptions of data

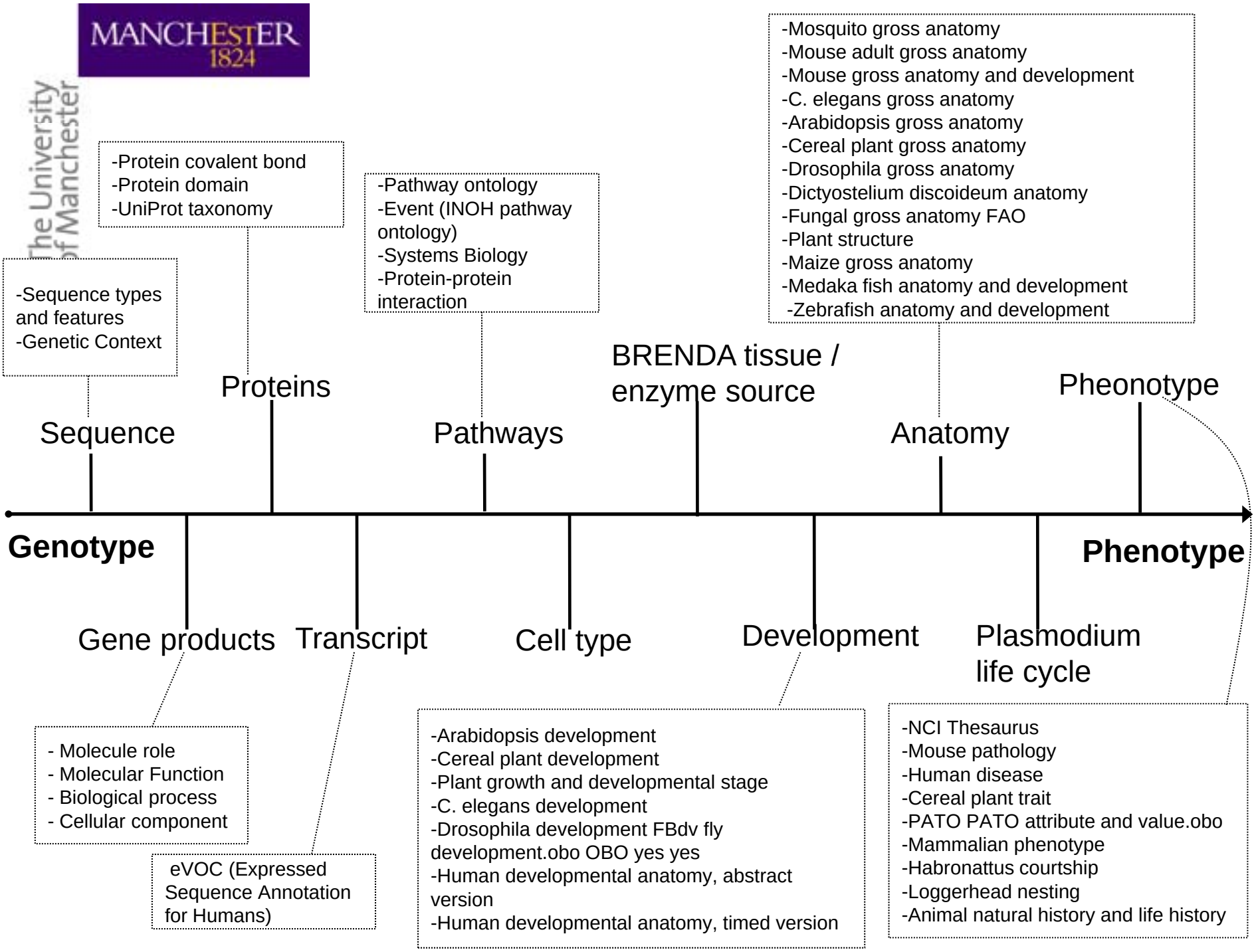
- 27 millimetres
- Tail of 27 millimetres
- Mouse tail of 27 millimetres
- Mouse of strain x that is 28 days old, tail that is 27 millimetres
- Data is only as good as its metadata

MANCHESTER 1824

The University of Manchester

Growth of interest in ontologies

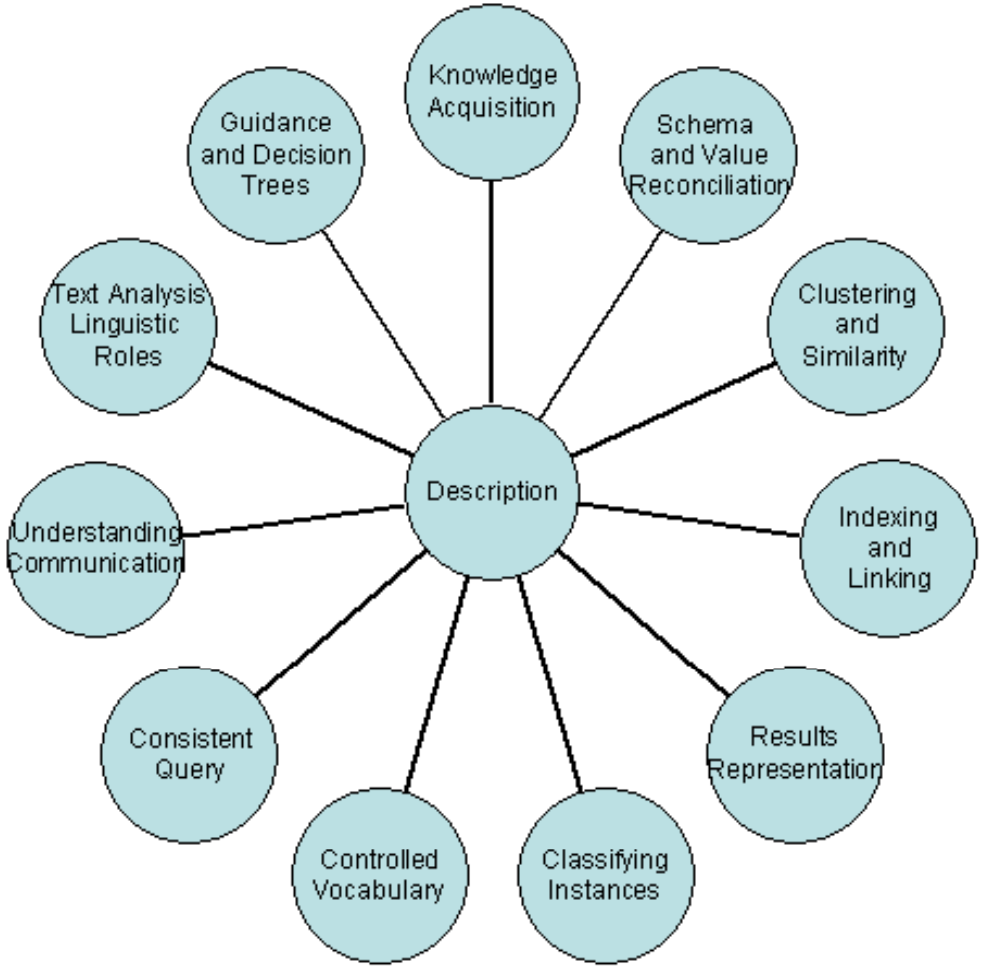



MANCHESTER 1824
The University of Manchester
-Protein covalent bond
-Protein domain
-UniProt taxonomy
-Pathway ontology
-Event (INOH pathway ontology)
-Systems Biology
-Protein-protein interaction
-Mosquito gross anatomy
-Mouse adult gross anatomy
-Mouse gross anatomy and development
-C. elegans gross anatomy
-Arabidopsis gross anatomy
-Cereal plant gross anatomy
-Drosophila gross anatomy
-Dictyostelium discoideum anatomy
-Fungal gross anatomy FAO
-Plant structure
-Maize gross anatomy
-Medaka fish anatomy and development
-Zebrafish anatomy and development
-Sequence types and features
-Genetic Context
Sequence
Proteins
Pathways
BRENDA tissue / enzyme source
Anatomy
Pheonotype
Genotype
Phenotype
Gene products
Transcript
Cell type
Development
Plasmodium life cycle
- Molecule role
- Molecular Function
- Biological process
- Cellular component
eVOC (Expressed Sequence Annotation for Humans)
-Arabidopsis development
-Cereal plant development
-Plant growth and developmental stage
-C. elegans development
-Drosophila development FBdv fly development.obo OBO yes yes
-Human developmental anatomy, abstract version
-Human developmental anatomy, timed version
-NCI Thesaurus
-Mouse pathology
-Human disease
-Cereal plant trait
-PATO PATO attribute and value.obo
-Mammalian phenotype
-Habronattus courtship
-Loggerhead nesting
-Animal natural history and life history

MANCHESTER 1824

The University of Manchester

Uses of Ontology in Bioinformatics



MANCHESTER 1824

The University of Manchester

OWL Axioms can be hard to understand

Disjointness

Making information explicit

Open world reasoning

- “Person **and** hasPet **some not** Cat”
- How many cats may a person have as pets?
- Cognitive complexity of OWL…

Coon logical constructs

Universal vs existential quantification

Domain and range constraints

Trivial satisfiability

Making definitions

Defined and primitive classes

MANCHESTER 1824

The University of Manchester

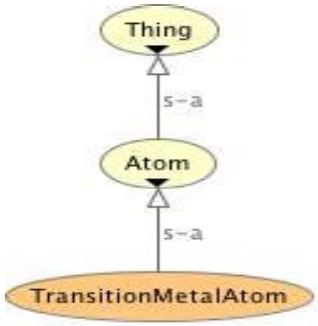
Measuring syntactic sophistication in an OWL ontology

$$n \sum_{i=1}^{i=n} \frac{a_c}{s} c$$

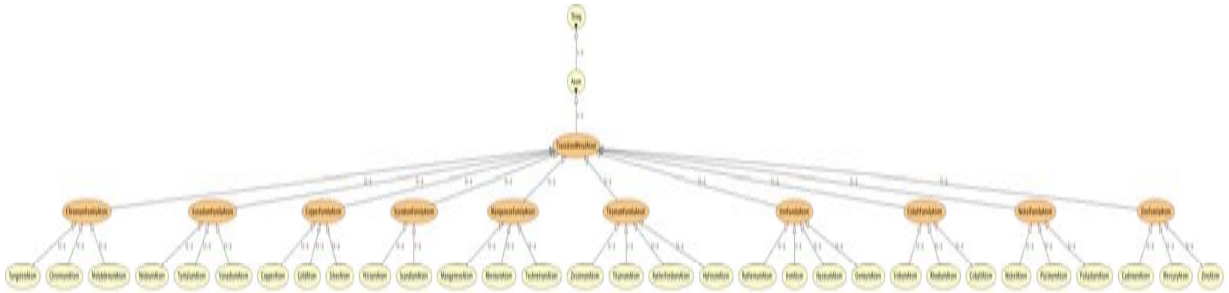
- n : number of patterns in an ontology
- c : is the complexity of a given pattern
- s : the size of the ontology in axioms
- a_c : number of axioms captured in a given pattern

The Elements Ontology (http://robertdavidstevens.wordpress.com/2011/05/05/an-ontology-of-the-periodic-table-using-electronic-structure-of-the-atom/)

Before reasoning



After reasoning



Class: PlatinumAtom
SubClassOf:
Atom
and (hasValenceElectronShell **some**
(DShell
and (contains **exactly** 9 Electron)
and (hasOrder **value** 5)))
and (hasValenceElectronShell **some**
(FShell
and (contains **exactly** 14 Electron)
and (hasOrder **value** 4)))
and (hasValenceElectronShell **some**
(SShell
and (contains **exactly** 1 Electron)
and (hasOrder **value** 6)))
and (hasValenceElectronShell **only**
((DShell
and (contains **exactly** 9 Electron)
and (hasOrder **value** 5))
or (FShell
and (contains **exactly** 14 Electron)
and (hasOrder **value** 4))
or (SShell
and (contains **exactly** 1 Electron)
and (hasOrder **value** 6))))
and (hasAtomicNumber **value** 78)

MANCHESTER 1824

The University of Manchester

How syntactically sophisticated are OWL ontologies?

- 25% have an S-mesure of 1
- About 50% has an S-mesure of 2
- Then it sort of tails off
- (the previous slide's ontology has an S mesaure of 580)

MANCHESTER 1824

The University of Manchester

Why can an OWL ontology be complex?

- We have potentially syntactically (and semantically) complex axioms, but most ontologies have axioms that are more or less simple in terms of form
- We have cognitive coplexity of even simmple syntactic constructs
- We have coplexity through size; lots and lots of simple things become complex
- We also have to understand the domain description too

MANCHESTER 1824

The University of Manchester

The consequence

- Is that OWL ontologies can be difficult to understand

MANCHESTER 1824

The University of Manchester

So, what do we do about it?

- Complex for whom?
- For the developer
  - Add comprehension tools; they have to interact with OWL, so make it easier
- For an ontology user:
  - Don’t show them the ontology!

MANCHESTER 1824

The University of Manchester

Hide the ontology

- OWL is horrid to look at
- When being used in a tool by the ultimate users, it just shouldn’t be seen

MANCHESTER 1824

The University of Manchester

iKUP http://www.kupkb.org

MANCHESTER 1824

The University of Manchester

Searching and Browsing the iKUP

Results View

The results table shows the KUPKB experiments that reference your search terms. You can sort the results table by clicking on the column headers. The navigation tree below gives you a summary of your results and can be used to filter the results table.(hint: hold down ctrl to select multiple filters)

Add Filter | Remove Filter

Visualize these results in KUPNetViz!

Cell (2)

- Somatic cell (2)
- Barrier cell (1)
- Kidney cell (2)
- Secretory cell (1)
- Contractile cell (1)
- Mesenchymal cell (1)
- Motile cell (2)
- Defensive cell (1)
- Leucocyte (1)
- Lower urinary tract cell (1)
- Electrically active cell (1)

Anatomy (3)

- Kidney (3)
- Loop of Henle, outer medullary portion (1)
- Pelvis (2)
- Lower urinary tract (2)

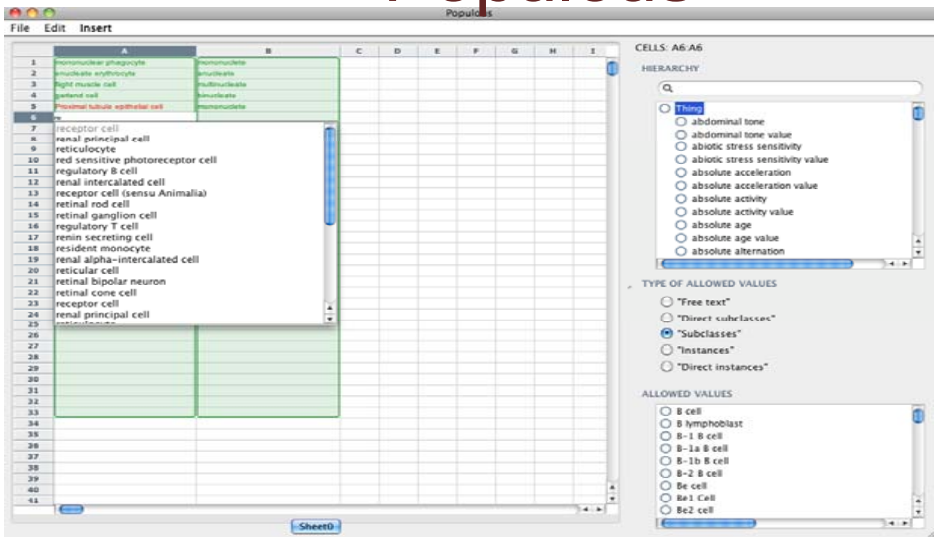
1-48 of 48

Entity id	Species	Anatomy	Disease/Model	Expression	Experiment	Type
ACTB	Human	Kidney proximal tubule epithelial cell	TGFbeta in vitro model	Up	Hills, Mol Endocrinol, 2010	mRNA
ACTB	Human	Kidney	Renal transplantation	Down	Head, GSE1563	mRNA
ACTB	Human	Kidney	Renal transplantation	Up	Head, GSE1563	mRNA
ACTB	Human	Kidney	Acute renal allograft rejection	Up	Head, GSE1563	mRNA
ACTB	Human	Glomerulus	Healthy	Present	Cuellar, Nephrology (Carlton), 2009	mRNA
ACTB	Human (adult)	Glomerulus	Healthy	Present	HKUPP Glomerulus database	Protein
ACTB	Human (adult)	Bladder urine	Healthy	Present	Mischak, Proteomics Clin Appl, 2010	Protein

MANCHESTER 1824

The University of Manchester

Populous



http://www.e-lico.eu/populous

- Generic tool for populating ontology templates
- Spreadsheet style interface
- Supports validation at the point of data entry
- Expressive Pattern language for OWL Ontology generation

MANCHESTER 1824

The University of Manchester

Generating natural language from OWL

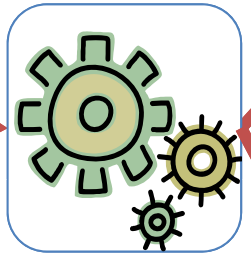
- An axiom is a sentence in OWL
- Transform it into natural language to give a familiar form for reading (and for input…)
- Body hasPart some Head
- All bodies have at least one part that is a head….
- Bodies have heads….
- A body has part a head
- A body has part x, has part y, has part z, has part ….

Input and output of OntoVerbal-M

Input: SNOMED-CT

1. **Class**: Diastolic hypertension
   SubClassOf: Hypertensive disorder
2. **Class**: Sustained diastolic hypertension
   SubClassOf: Diastolic hypertension
3. **Class**: Labile diastolic hypertension
   SubClassOf: Diastolic hypertension
4. **Class**: Secondary diastolic hypertension
   SubClassOf: Secondary hypertension
   and
   Diastolic hypertension

OntoVerbal-M



Output: English

Diastolic hypertension is a kind of *hypertensive disorder*. More specialised kinds of *diastolic hypertension* are
- *sustained diastolic hypertension*

and
- *labile diastolic hypertension*.

Another relevant aspect of *diastolic hypertension* is that: *secondary diastolic hypertension* is a kind of *secondary hypertension* and *diastolic hypertension*.

Output: Mandarin

*心臟舒張高血壓*屬於*高血壓失調*，它也包含了*持續性的心臟舒張高血壓*和*不安定的心臟舒張高血壓*。其他與*心臟舒張高血壓*相關的資訊為：*續發性的心臟舒張高血壓*屬於*續發性高血壓*和*心臟舒張高血壓*的交集。

MANCHESTER 1824

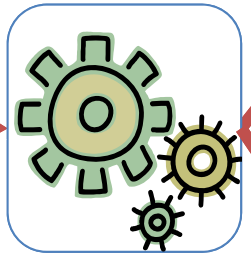
The University of Manchester

A complex axiom example

Input: SNOMED-CT

1. **Class**: Renal arterial hypertension
   SubClassOf: renovascular

OntoVerbal-M



Output: English

Renal arterial hypertension is a kind of *renovascular hypertension* that *has a finding site* in *a kidney*

Output: Mandarin

*腎臟(renal)動脈(artery)的(apostrophe)高血壓(hypertension)*是(is)一種(a kind of)*腎血管性(renovascular)高血壓(hypertension)*中(among)在(at)*腎臟(kidney)結構(structure)*上(upon)*有(has)病灶(finding site)*。

MANCHESTER 1824

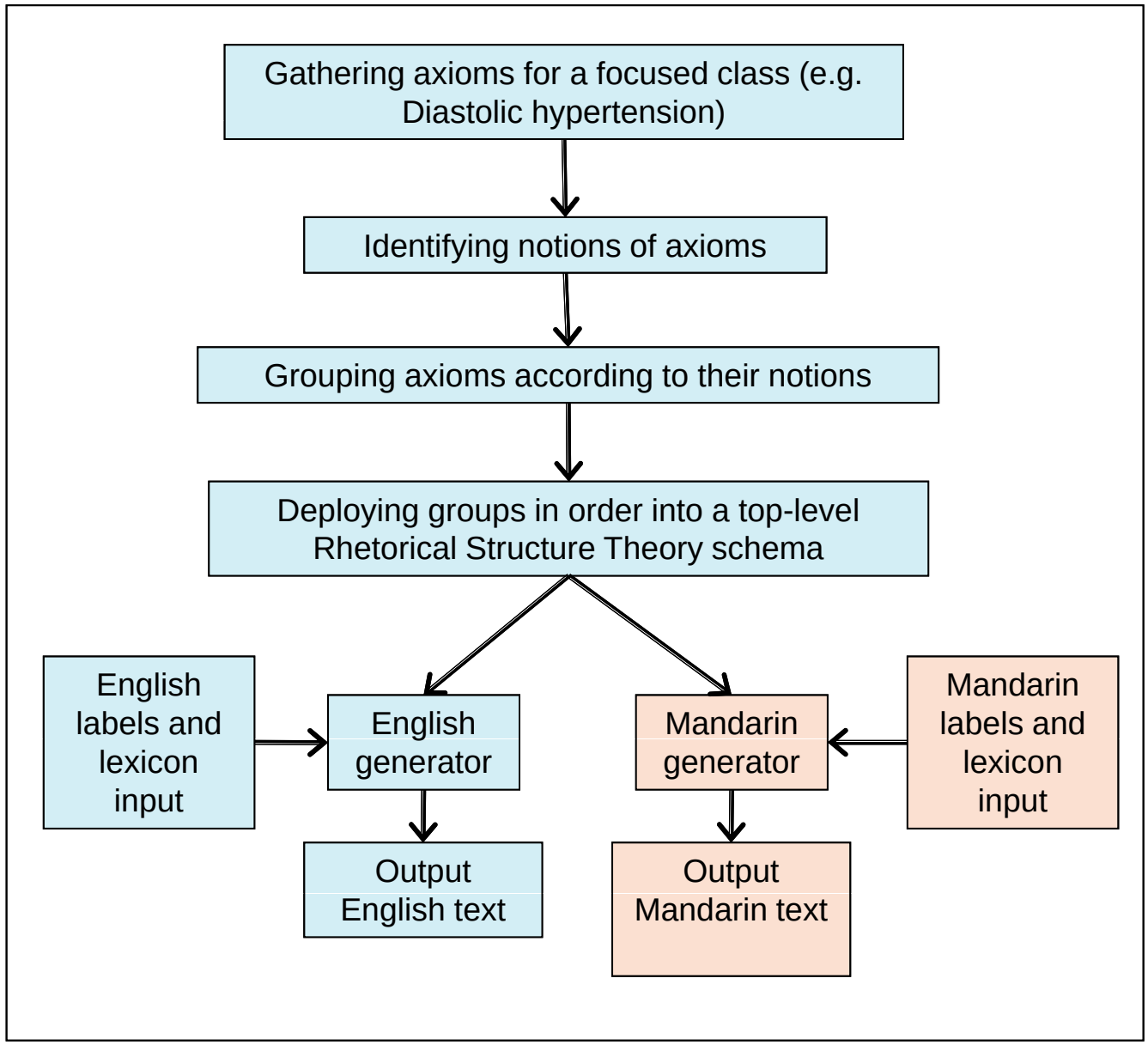
The University of Manchester

OntoVerbal

- Ontoverbal is a plugin for Protégé 4
- It generates natural language for classes in an ontology
- http://swatproject.org/demos.asp

- It groups axioms, aggregates repeating properties, organises axiom types according to rhetorical structure theory, does some msoothing of language
- People can round-trip back to OWL better than with human written natural language
- Finds favour as a way of generating natural languaage definitions
- See also http://jamesmaloneebi.blogspot.co.uk/2012/06/ontology-turing-test.html

System architecture of OntoVerbal-M



MANCHESTER 1824

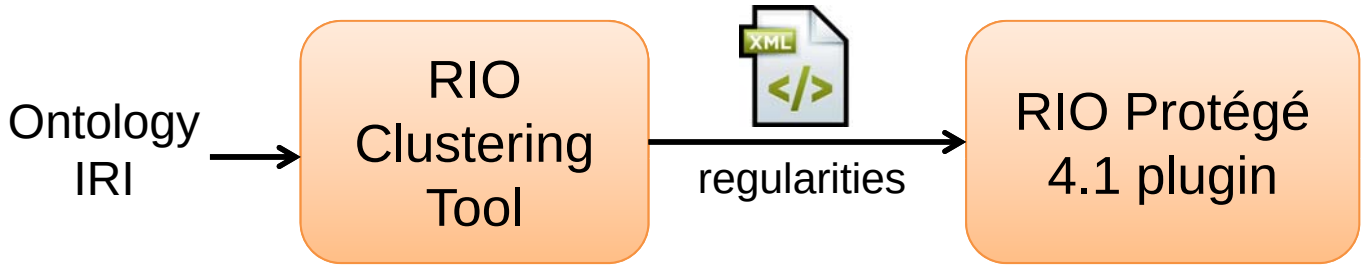
The University of Manchester

Patterns in OWL ontologies

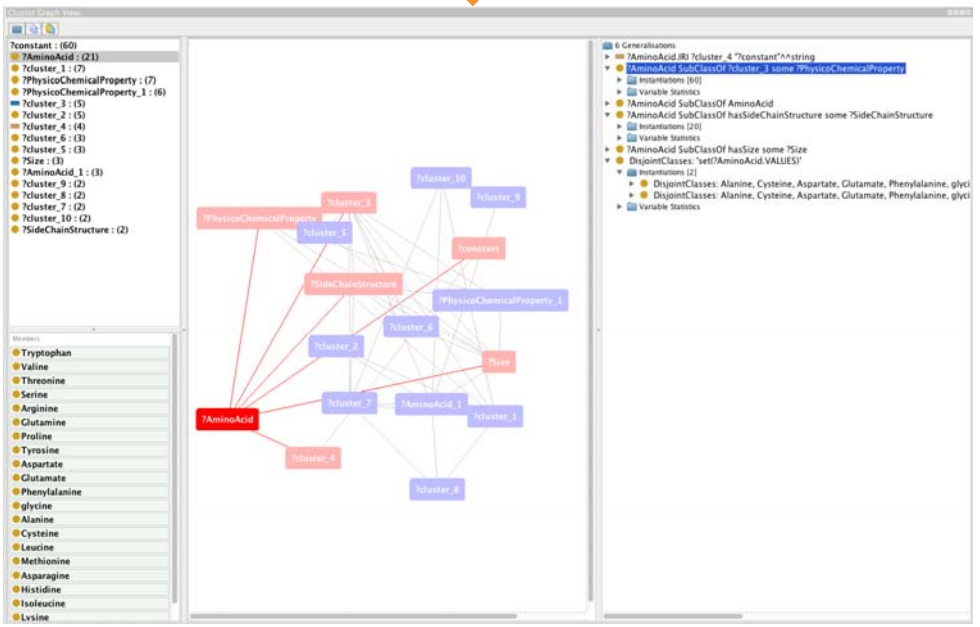
- Axioms (should) repeat themselves as patterns in an ontology

- Patterns that represent accepted solutions to a representation problem are *design patterns*
- A pattern doesn't have to be a design pattern
- They can just be *regularities* in the ontology

Regularity Inspector for Ontologies (RIO)



Regularity Views



RIO url:
http://riotool.sourceforge.net/

Examples of regularities in SNOMED-CT

1. Syntactic Regularity describing ‘Chronic findings’:

?Chronic_finding *EquivalentTo* **?Disorder** **and** (RoleGroup **some**
(‘Clinical course (attribute)’ **some** ‘Chronic (qualifier value)’))

2. Syntactic Regularity describing ‘Acute findings’:

?Acute_finding *EquivalentTo* **?Disorder** **and** (RoleGroup **some**
(‘Clinical course (attribute)’ **some**
'Sudden onset AND/OR short duration (qualifier value)'))

- Regularities expressed as axioms with ***variables*** holding similar entities:
  - E.g. **?Chronic_finding = [**Chronic pyonephrosis (disorder),
    Chronic pneumothorax (disorder)**]**

Some stats on snomed and RIO

	Acute findings	Chronic findings
Number of clusters describing the entities	34	34
Target entities not found in a cluster	12	11
Axioms instantiating naming pattern	76 (5%)	210 (11%)

- Entities not included in any cluster can be a starting point for tracing design defects
  - Eg. ’Chronic back pain (finding)’ was missing an existential restriction

MANCHESTER 1824

The University of Manchester

Summary of the message

- We’re getting more sophisticated in how we handle OWL ontologies

MANCHESTER 1824

The University of Manchester

Acknowledgements

- Simon Jupp
- Julie Klein
- Fennie Liang
- Donia Scott
- Eleni Mikroyannidi
- Azlinayati Manaf
- Sean Bechhofer