MANCHESTER 1824

The University of Manchester

Knowledge Elicitation

COMP30411

Sean Bechhofer
sean.bechhofer@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

…there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know.

MANCHESTER 1824

The University of Manchester

Knowledge Elicitation

- Knowledge Representation is about representing some knowledge
  - Unsurprising!
- We need to determine what that knowledge *is*
- This is the process of *Knowledge Acquisition* of *Elicitation*
- This is a non-trivial process though
  - The information is often locked away in the heads of domain experts
  - The experts themselves may not be aware of the implicit conceptual models that they use
- We have to draw out and make explicit all the known knowns, unknown knowns, etc….

MANCHESTER 1824

The University of Manchester

Steps in Modelling

1. Establish the purpose
   - Without purpose, no scope, requirements, evaluation,
2. Informal/Semiformal knowledge elicitation
   - Collect the terms
   - Organise terms informally
   - Paraphrase and clarify terms to produce informal concept definitions
   - Diagram informally
3. Refine requirements & tests

MANCHESTER 1824

The University of Manchester

Steps in Modelling

4. Implementation
   - Paraphrase and comment at each stage before implementing
   - Develop normalised schema and skeleton
   - Implement prototype recording the intention as a paraphrase
     - Keep track of what you meant to do so you can compare with what happens
       - Implementing logic-based ontologies is programming
   - Scale up a bit
     - Check performance
   - Populate
     - Possibly with help of text mining and language technology

MANCHESTER 1824

The University of Manchester

Steps in Modelling

5. Evaluate & quality assure
   - Against goals
   - Include tests for evolution and change management
   - Design regression tests and "probes"
6. Monitor use and evolve
   - Process not product!

MANCHESTER 1824

The University of Manchester

Eliciting Knowledge

- Most knowledge is in the heads of experts
- Experts have vast amounts of knowledge
- Experts have a lot of tacit knowledge
- They don't know all that they know and use
- Tacit knowledge is hard (impossible) to describe
- Experts are very busy and valuable people
- Each expert doesn't know everything

- Additional Reading:
  http://www.epistemics.co.uk/Notes/63-0-0.htm

MANCHESTER 1824

The University of Manchester

Elicitation Techniques

- Experts can be a scarce and valuable resource.
  - Techniques should take experts off the job for short time periods
- Focus on the essential knowledge
- Can capture tacit knowledge
- Allow knowledge to be collated from different experts
- Allow knowledge to be validated and maintained
- Allow non-experts to understand the knowledge

- To a certain extent these techniques can be independent of the actual KR formalism used
  - Frames, FOL, DL, Topic Maps, etc.

MANCHESTER 1824

The University of Manchester

Types of technique (1)

- Protocol-generation techniques
  - include various types of interviews (unstructured, semi-structured and structured), reporting techniques (such as self-report and shadowing) and observational techniques
- Protocol analysis techniques
  - are used with transcripts of interviews or other text-based information to identify various types of knowledge, such as goals, decisions, relationships and attributes. This acts as a bridge between the use of protocol-based techniques and knowledge modelling techniques.

MANCHESTER 1824

The University of Manchester

Types of technique (2)

- Hierarchy-generation techniques
  - such as laddering are used to build taxonomies or other hierarchical structures such as goal trees and decision networks.
- Matrix-based techniques
  - involve the construction of grids indicating such things as problems encountered against possible solutions. Important types include the use of frames for representing the properties of concepts and the repertory grid technique used to elicit, rate, analyse and categorise the properties of concepts.

MANCHESTER 1824

The University of Manchester

Types of technique (3)

- Sorting techniques
  - are used for capturing the way people compare and order concepts, and can lead to the revelation of knowledge about classes, properties and priorities.
- Limited-information and constrained-processing tasks
  - are techniques that either limit the time and/or information available to the expert when performing tasks. For instance, the twenty-questions technique provides an efficient way of accessing the key information in a domain in a prioritised order.

MANCHESTER 1824

The University of Manchester

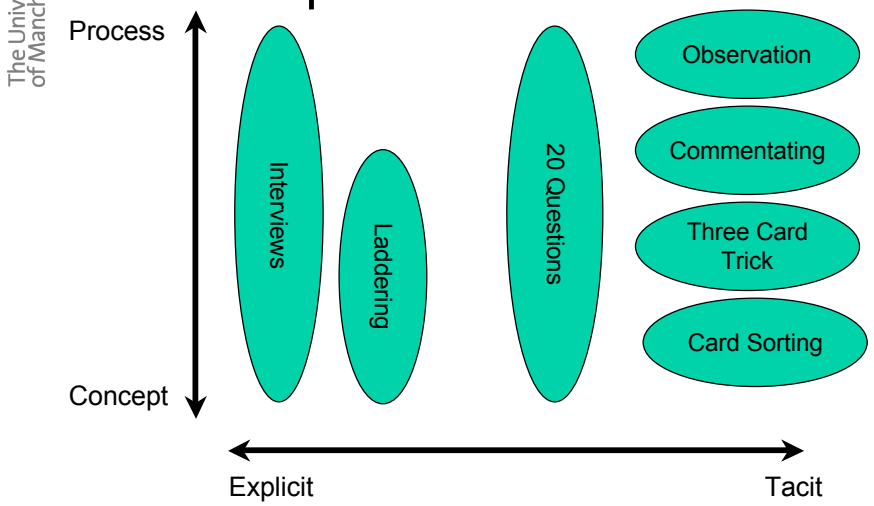
Types of technique (4)

- Diagram-based techniques
  - include the generation and use of concept maps, state transition networks, event diagrams and process maps. The use of these is particularly important in capturing the "what, how, when, who and why" of tasks and events.

MANCHESTER 1824

The University of Manchester

Process vs Concept
Explicit vs Tacit



[epistemics]

COMP30411: Knowledge Representation 13

MANCHESTER 1824

The University of Manchester

Initial Steps

- Collect the terms
- Gather together the terms that describe the objects in the domain.
- Analysis of relevant sources
  - Documents
  - Manuals
  - Web resources
  - Interviews with Experts

COMP30411: Knowledge Representation 14

MANCHESTER 1824

The University of Manchester

Card Sorting

- Card Sorting is a technique that helps us in identifying similarities between the classes of concepts that we might have in our model.
- A relatively informal procedure
- Works best in small groups
- Write down each concept/idea on a card
- Organise them into piles
- Link the piles together
  - Identify what the pile represents
  - Record the collections or groupings that have been identified and why these might be considered as groups
- Do it again, and again
  - Each time, note down the results of the sorting

MANCHESTER 1824

The University of Manchester

Card Sorting in Action

MANCHESTER 1824

The University of Manchester

Triadic Elicitation: "3 Card Trick"

- Select 3 cards at random
- Identify which 2 cards are the most similar?
  - Why?
  - What makes them different from the third card?
- Helps to determine the characteristics of our classes
- Picking 3 cards forces us into identifying differences between them
  - There will always be two that are "closer" together
  - Although which two cards that is may differ depending on your perspective

MANCHESTER 1824

The University of Manchester

Twenty Questions

- Twenty Questions is an approach similar to the game of the same name
- The Knowledge Engineer thinks of an object/concept in the domain
- The Domain Expert asks a series of yes/no questions
- Knowledge Engineer notes the questions and the order in which they are asked
  - Can provide an insight into the key aspects, properties or categories and their relative priorities.
- Note that the main purpose of this exercise is not really to try and find out what the Engineer is thinking of, but to determine the important properties!

MANCHESTER 1824

The University of Manchester

Other Techniques

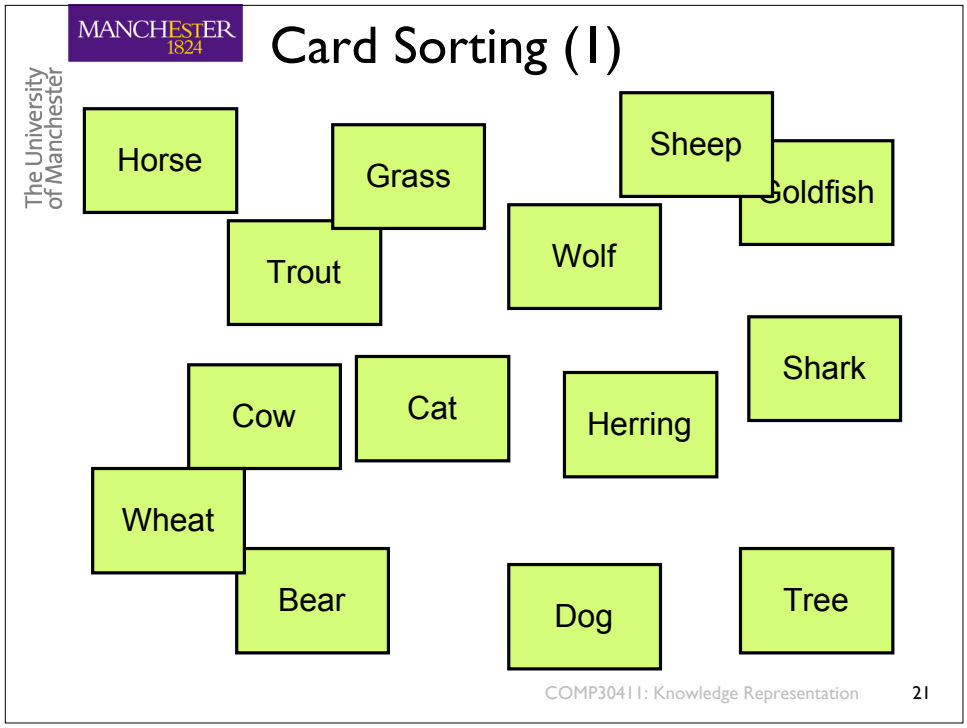
- Think Aloud or Commentary
- Interviews
  - Structured and unstructured
- Observational techniques
  - Observing experts at work
  - Videotaping
  - Can be costly in terms of time/effort

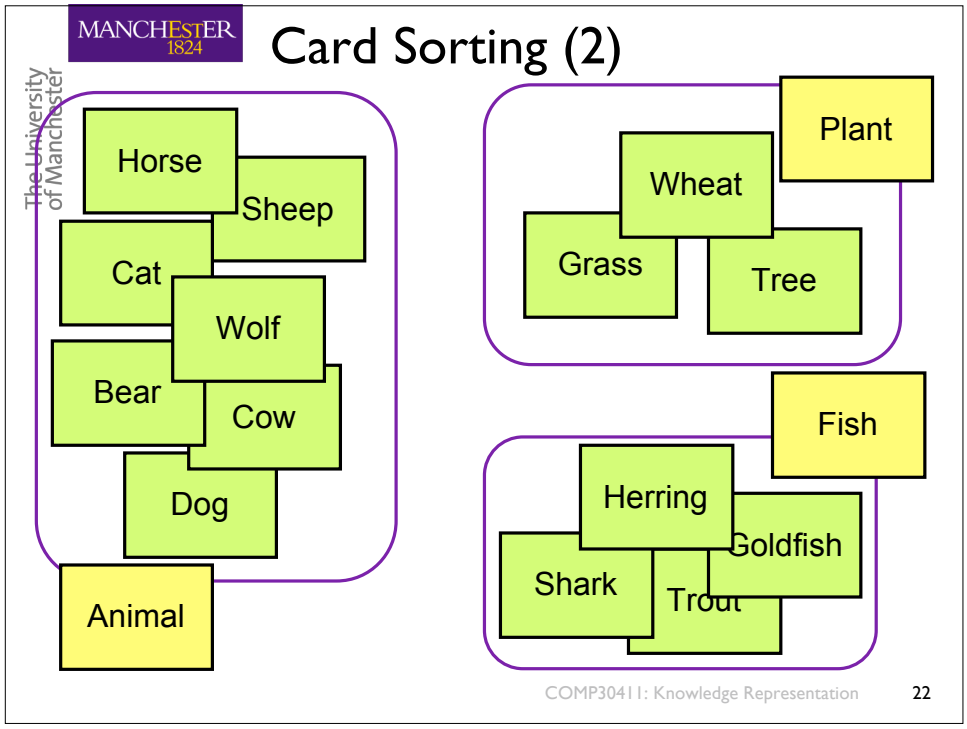
MANCHESTER 1824

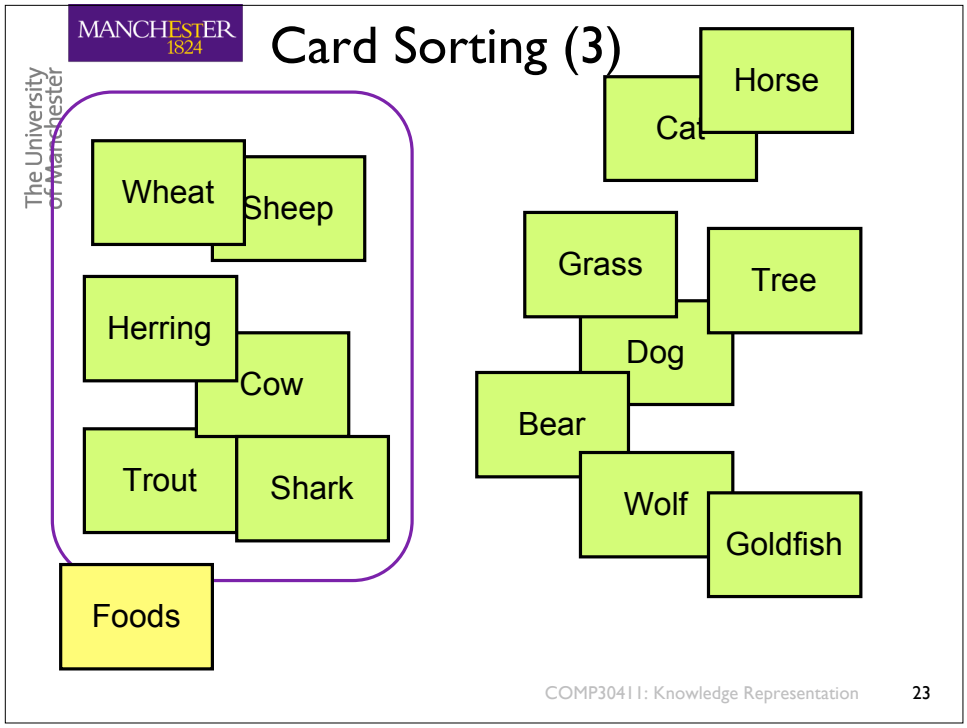
The University of Manchester

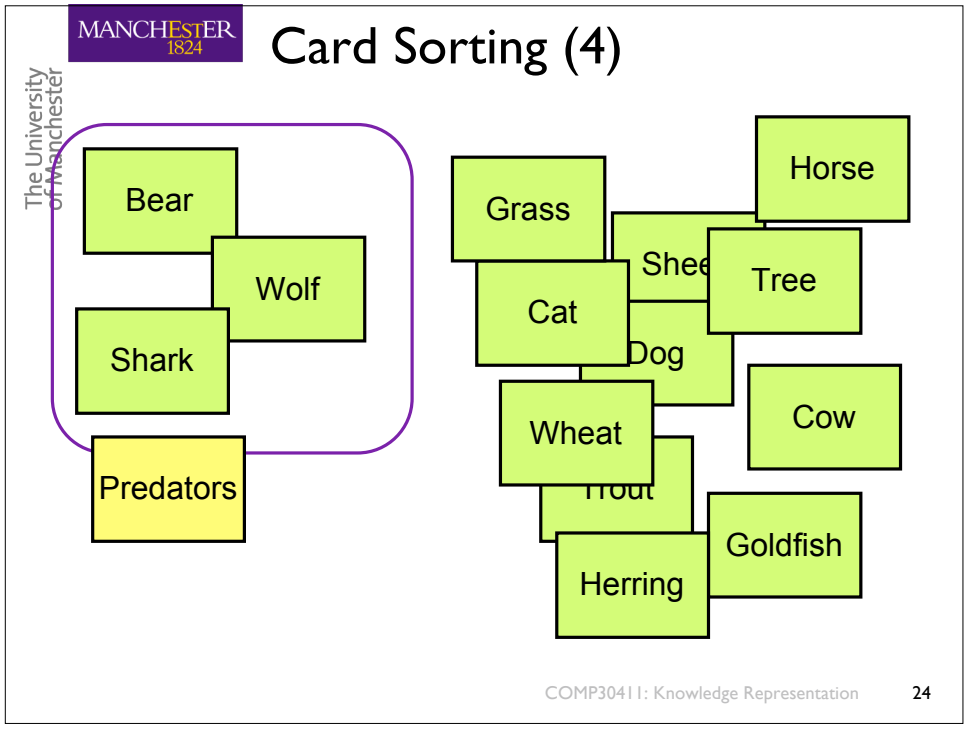
Example: An Animals Ontology

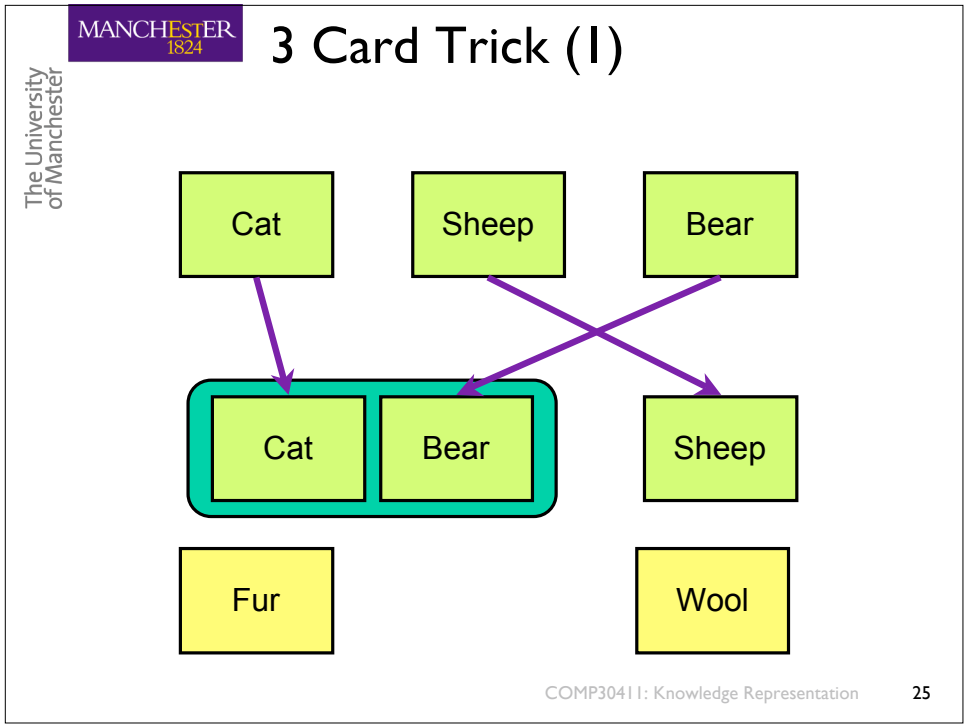
- Task: To provide an ontology for an index of a children's book of animals including
  - Where they live
  - What they eat
    - Carnivores, herbivores and omnivores
  - How dangerous they are
  - How big they are
  - A bit of basic anatomy
    - legs, wings, fins?
    - skin, feathers, fur?
  - …

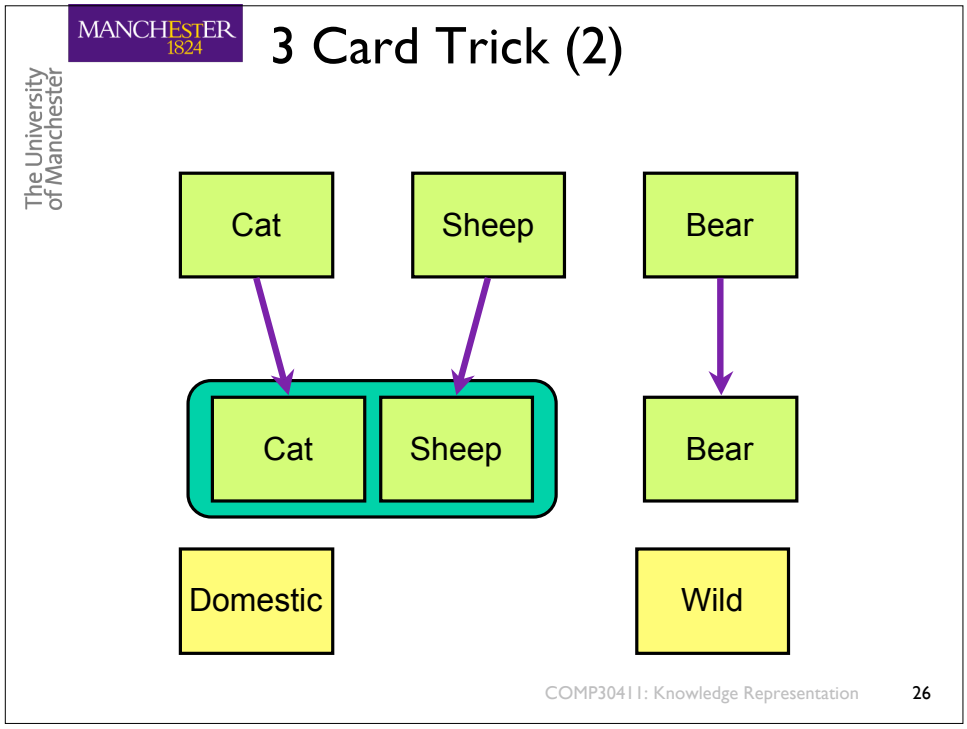

MANCHESTER 1824
The University of Manchester
Card Sorting (1)
Horse
Grass
Sheep
Goldfish
Trout
Wolf
Shark
Cow
Cat
Herring
Wheat
Bear
Dog
Tree
COMP30411: Knowledge Representation
21


MANCHESTER 1824
The University of Manchester
Card Sorting (2)
Horse
Sheep
Cat
Wolf
Bear
Cow
Dog
Animal
Plant
Wheat
Grass
Tree
Fish
Herring
Goldfish
Shark
Trout
COMP30411: Knowledge Representation
22


MANCHESTER 1824
The University of Manchester
Card Sorting (3)
Wheat
Sheep
Herring
Cow
Trout
Shark
Foods
Horse
Cat
Grass
Tree
Dog
Bear
Wolf
Goldfish
COMP30411: Knowledge Representation
23


MANCHESTER 1824
The University of Manchester
Card Sorting (4)
Bear
Wolf
Shark
Predators
Grass
Horse
Tree
Cat
Dog
Cow
Wheat
Goldfish
Herring
COMP30411: Knowledge Representation
24


MANCHESTER 1824
The University of Manchester
3 Card Trick (1)
Cat
Sheep
Bear
Cat
Bear
Sheep
Fur
Wool
COMP30411: Knowledge Representation
25


MANCHESTER 1824
The University of Manchester
3 Card Trick (2)
Cat
Sheep
Bear
Cat
Sheep
Bear
Domestic
Wild
COMP30411: Knowledge Representation
26

MANCHESTER 1824

The University of Manchester

Extend the concepts: Laddering

- Take a group of things and ask what they have in common
  - Then what other 'siblings' (brothers/sisters) there might be
- Examples
  - Plant, Animal ➡ Living Thing
    - Might add Bacteria and Fungi but not necessarily now
    - Depends on purpose/application
  - Cat, Dog, Cow, Horse ➡ Mammal
    - Others might be Goat, Sheep, Horse, Rabbit,…
  - Cow, Goat, Sheep, Horse ➡ Hoofed animal ("Ungulate")
    - What others are there? Do they divide amongst themselves?
  - Wild, Domestic ➡ Domestication
    - What other states – "Feral" (domestic returned to wild)

MANCHESTER 1824

The University of Manchester

Refining the Model

- Identify relations
  - e.g. "eats", "owns", "parent of"
- Identify self standing entities
  - Things that can exist on their own
  - People, animals, houses, actions, processes, …
    - Roughly nouns
- Identify modifiers
  - Things that modify ("inhere") in other things
    - Roughly adjectives and adverbs
- Identify definable things
  - e.g. "child", "parent", "Mother", "Father"
    - Things where you can say clearly what it means
    - Try to define a dog precisely – very difficult
      - A "natural kind"
- Add abstractions where needed to help in the organisation of the taxonomy
  - e.g. "Living thing"

MANCHESTER 1824

The University of Manchester

Resulting Skeleton

- Living Thing
  - Animal
    - Mammal
      - Cat
      - Dog
      - Cow
      - Person
    - Fish
      - Trout
      - Goldfish
      - Shark
  - Plant
    - Tree
    - Grass
    - Wheat
- Modifiers
  - Domestic
    - Pet
    - Farmed
      - Draft
      - Food
  - Wild
  - Health
    - Healthy
    - Sick
  - Sex
    - Male
    - Female
  - Age
    - Adult
    - Child
- Relations
  - eats
  - owns
  - parent-of
  - …
- Definable
  - Carnivore
  - Herbivore
  - Child
  - Parent
  - Mother
  - Father
  - Food Animal
  - Draft Animal

MANCHESTER 1824

The University of Manchester

Next Steps

- The results of the previous steps help us in gathering our knowledge.
- Once we have done this, we can set about formalising using some representation
  - E.g. writing definitions in FOL or OWL or defining frames.

MANCHESTER 1824

The University of Manchester

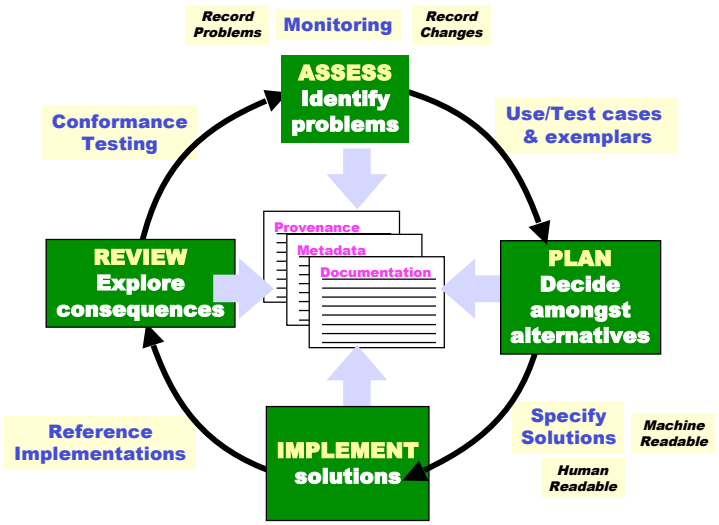
Engineering and Quality Assurance

- Conceptual Modelling and Software engineering share some characteristics
  - Requirements capture
  - Specification
  - Implementation
- How do we know we're building a good model?
- What are you going to do with it once you're finished?
  - A conceptual model isn't just an artefact to be built. It should have some intended purpose
  - Competency questions
- Ensure that the process you follow to build your model is
  - Principled
  - Well documented

MANCHESTER 1824

The University of Manchester

Engineering ontologies: A *process* not a product



MANCHESTER 1824

The University of Manchester

Next Session: An Exercise

- The next lecture will be a hands-on exercise on knowledge acquisition
- The domain will be one which is (hopefully) familiar to you

The University

- During the session we'll use some of the techniques we've described to help build an initial model of the domain
- Following the session, you will be expected to produce an initial model based on the work you do during the session
  - Attendance at and participation in the session will help you greatly in this task!
  - Make sure you bring something for taking notes during the session!

MANCHESTER 1824

The University of Manchester

KA Exercise

- **Before** the Session
  - Using the University and School web pages as a guide, try and identify at least 10 concepts that you think should appear in a model of the University.
- **During** the Session
  - Identify a core set of concepts
  - Identify a draft classification/hierarchy
  - Identify core properties
- **After** the Session
  - Following the session, you will produce a sketch of the model using Protégé.
  - Further details of the task are available on the moodle site.

MANCHESTER 1824

The University of Manchester

KA Exercise: Techniques

- Card Sorting
  - Try and arrange the cards into groups
    - Base the groupings on some kind of characteristic that you can identify: what do the concepts have in common?
    - Each time you form a group, note down *which* cards you placed in this group and *why*
      - Try and keep a note of any discussion you have while doing the card sorting
  - Do it again, and again, and again,…

MANCHESTER 1824

The University of Manchester

KA Exercise: Techniques

- 3 Card Trick
  - Using the 3 Card Trick technique, try and identify some properties of the concept that could be used to characterise them
- Twenty Questions
  - One of you play the role of domain expert, one of knowledge engineer and try twenty questions with a random card.