Building Controlled vocabularies/taxonomies, terminologies, Ontologies

Robert Stevens
Bio-Health Informatics Group
School of Computer Science
University of Manchester
Robert.Stevens@manchester.ac.uk

Options

- Structured controlled vocabulary
- Tags clouds
- Minimum information models
- They're all vocabulary or use vocabulary
- How do you choose?

Building these things can be hard

- … particularly sociologically
- Also difficult to get things right technically:
  - Consistent in naming
  - Facts explicit and non-contradictory,
  - Terms in the right place
  - Answering the right questions

Never really complete or correct

… but we can make things considerably less difficult

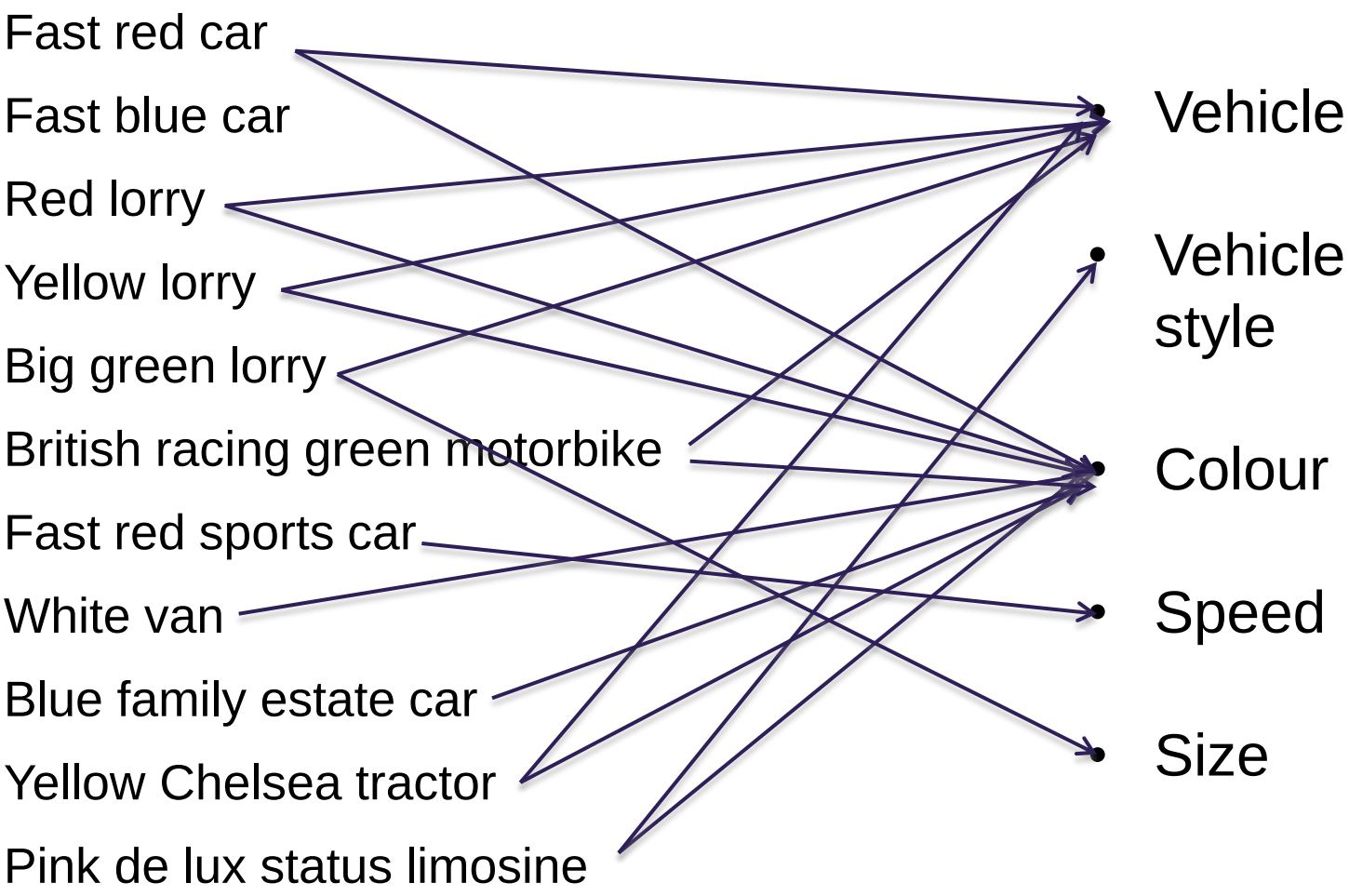
Know your users

- Scientists:
  - Academics;
  - Experimental officers;
  - Research associates;
  - Ph.D. students;
- University administrators
- Funders

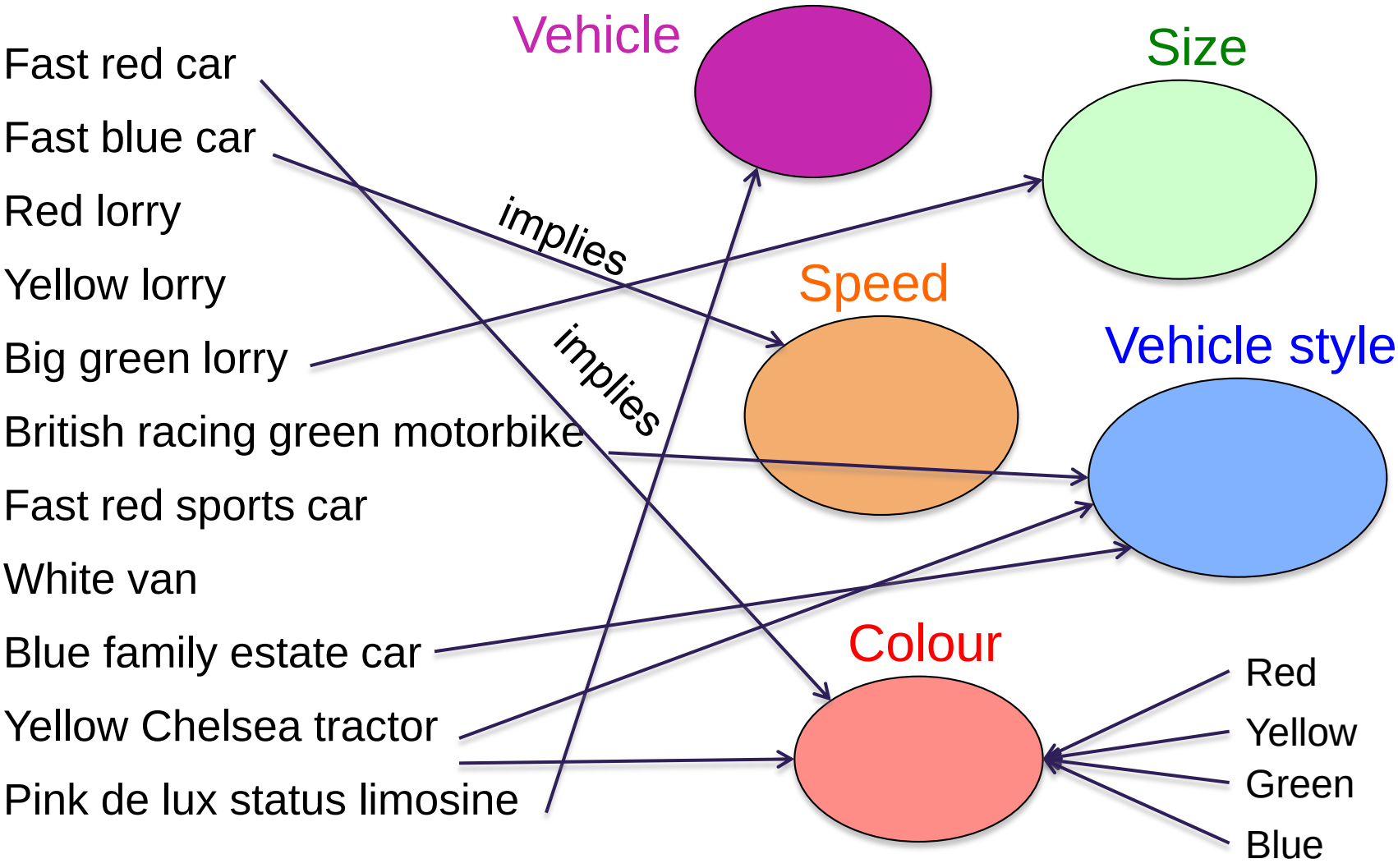
Competency questions

- What software works best with my dataset?
- Does it do what I want or need it do data e.g. render a gif?
- Which software tool created this data?
- What software can perform task x?
- Is it appropriate software for my task?
- What are the primary inputs and outputs?
- Is this software available as a web service?
- What open source, maintained software can I use to process these in this format?
- Where can I get the software?
- Is there a mailing list?
- How and where has this software been used successfully in the past?
- http://softwareontology.wordpress.com/2011/04/01/user-sourced-competency-questions-for-software/

Vehicle types



Vehicle types (2)



Describing a Fast red sports car

Class: FastRedSportsCar
SubClassOf: Car,
hasColour some Red,
hasSpeed some Fast,
hasStyle some Sports

Building the taxonomy

Class: RedCar
    EquivalentTo: Car
        and hasColour some Red

Class: SportsCar
    EquivalentTo: Car
        and hasStyle some Sports

Building a terminology of techniques

Terms:

- Thin Film Deposition (techniques)
  - Pulsed Laser Deposition
  - Chemical Vapor Deposition
  - Electrodeposition
  - Ion Beam Deposition
  - Molecular Beam Epitaxy

Establishing some dimensions

Thin Film Deposition

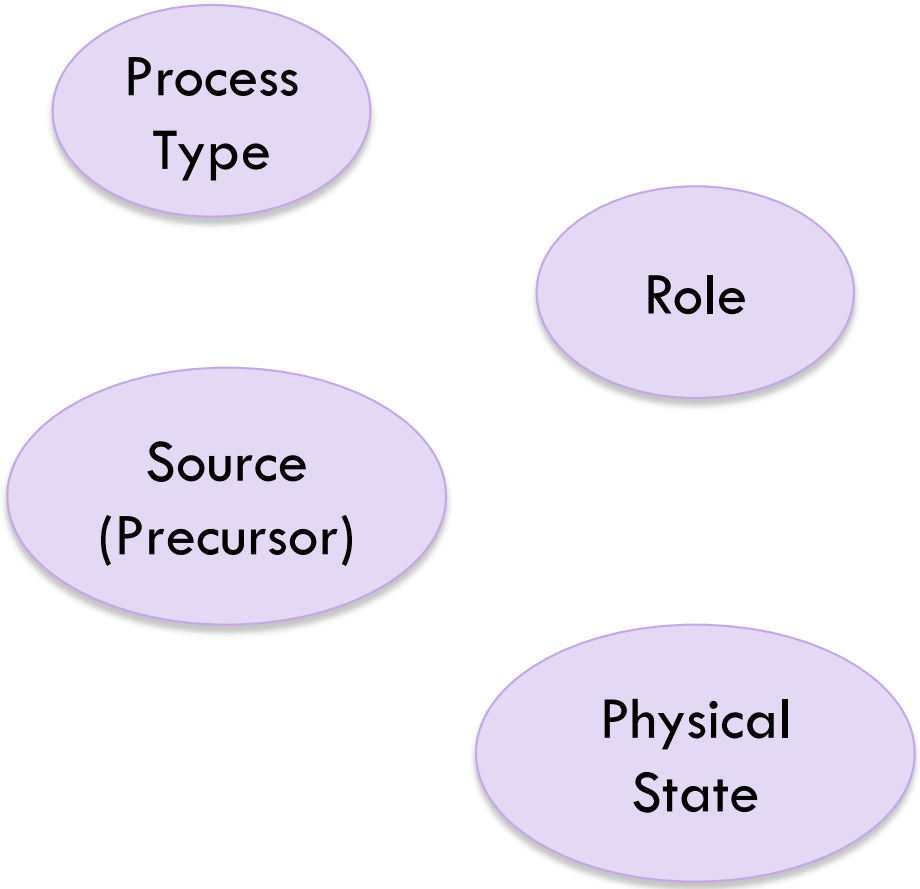
Pulsed Laser Deposition

Chemical Vapor Deposition

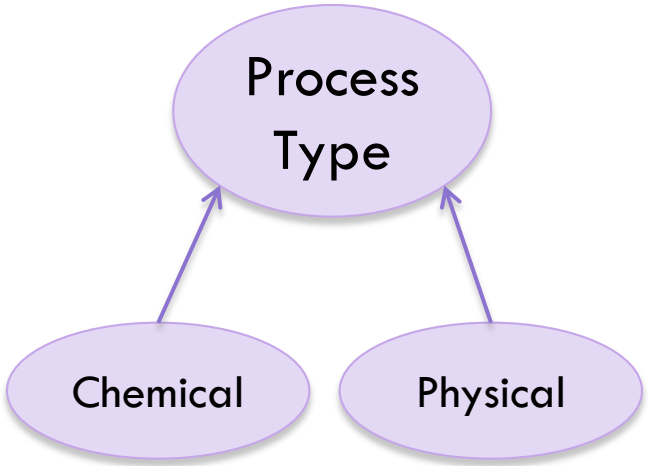
Electrodeposition

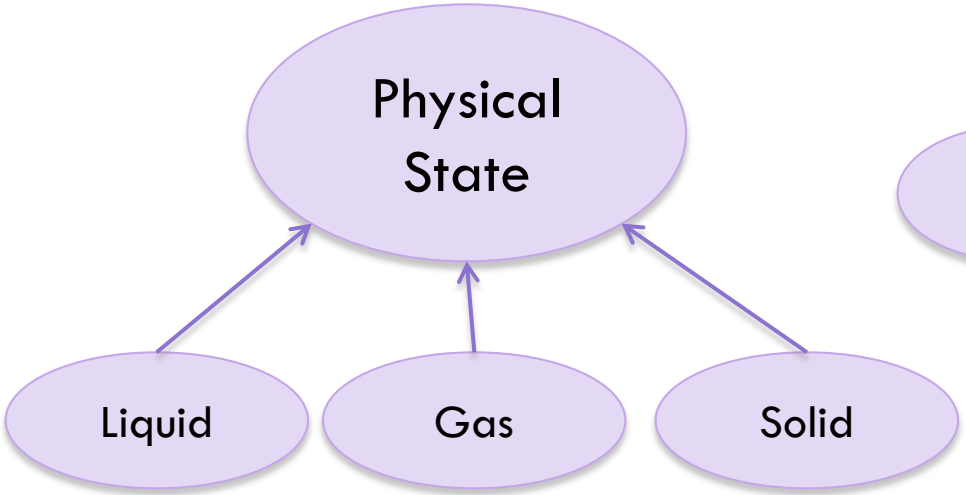
Ion Beam Deposition

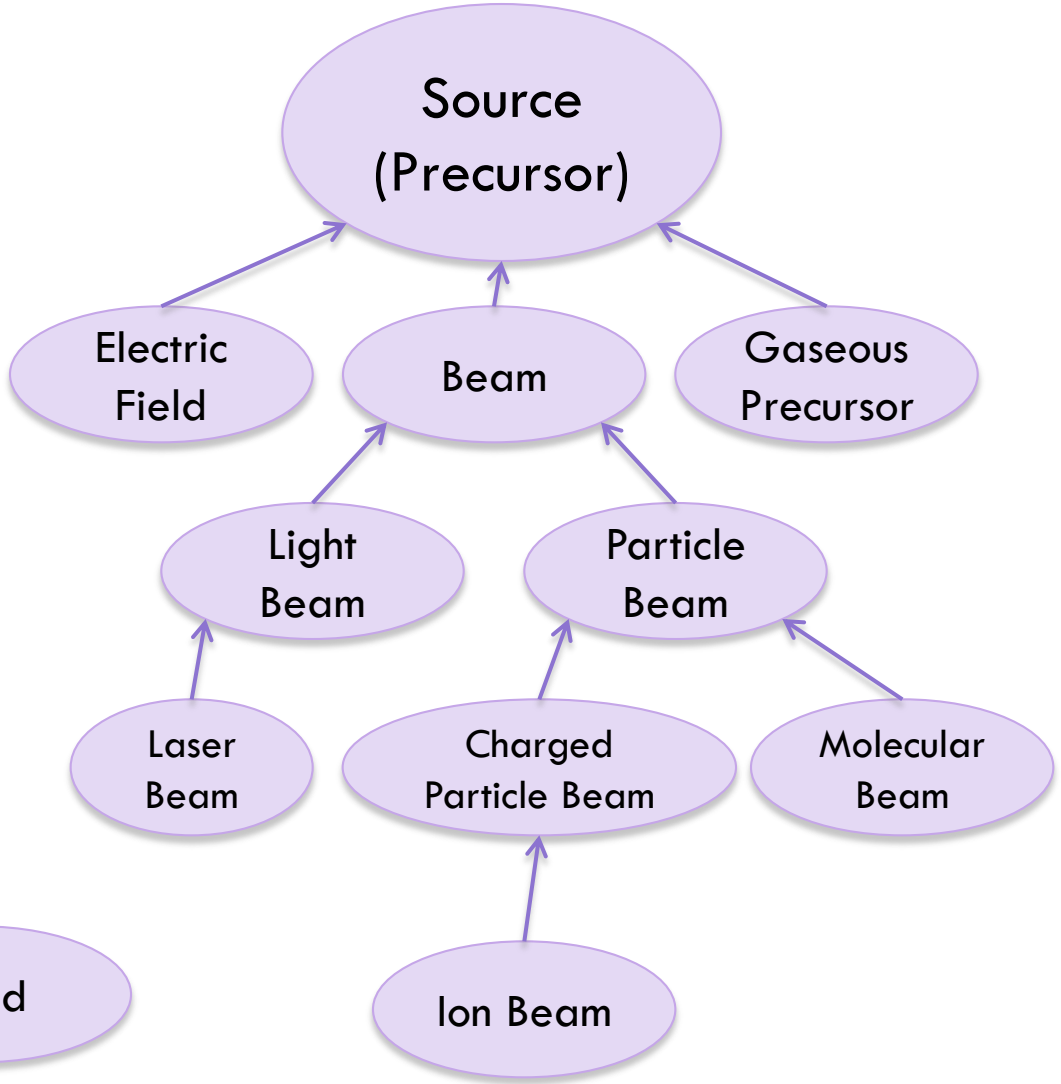
Molecular Beam Epitaxy



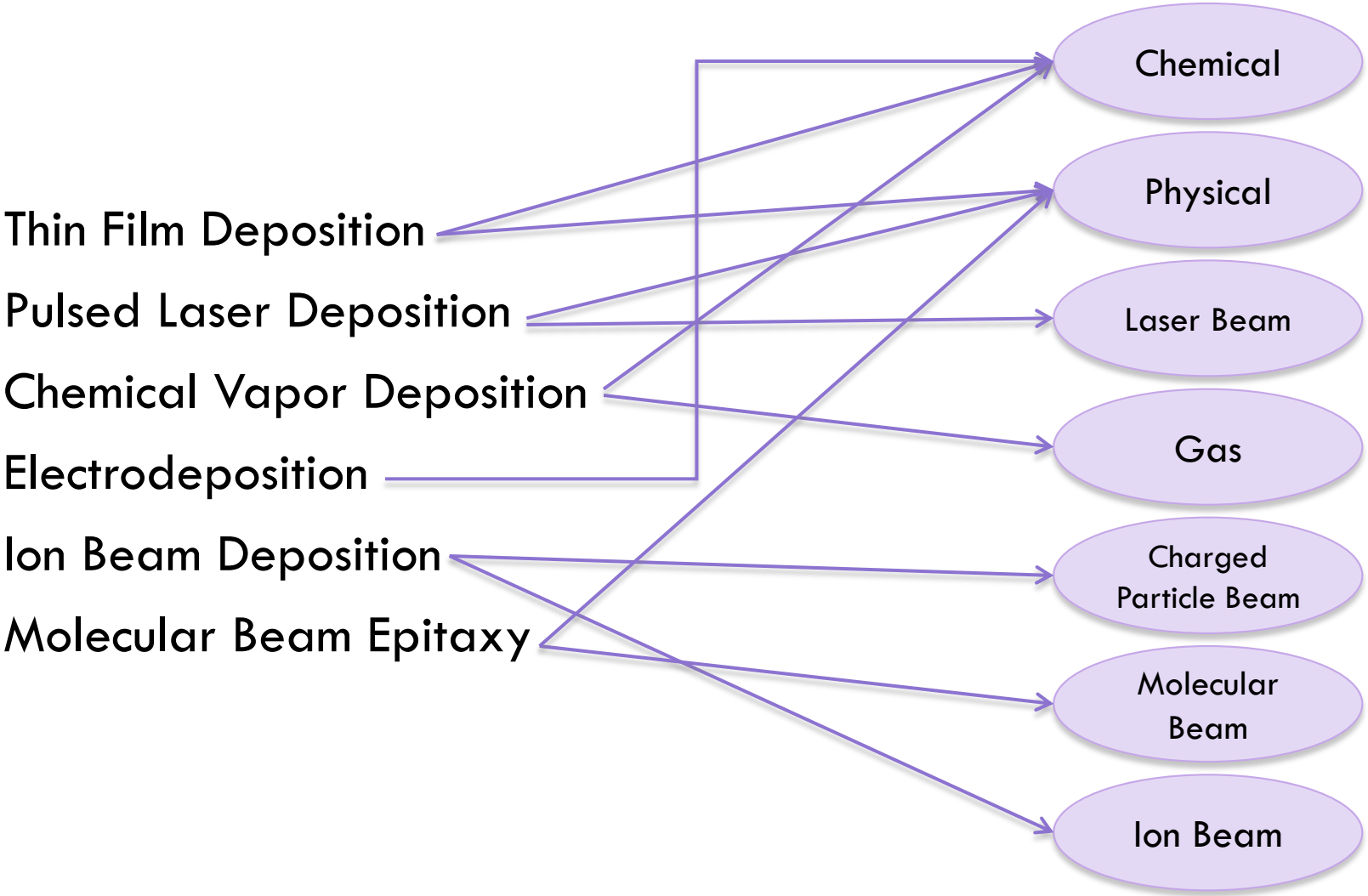
Expanding on those dimensions



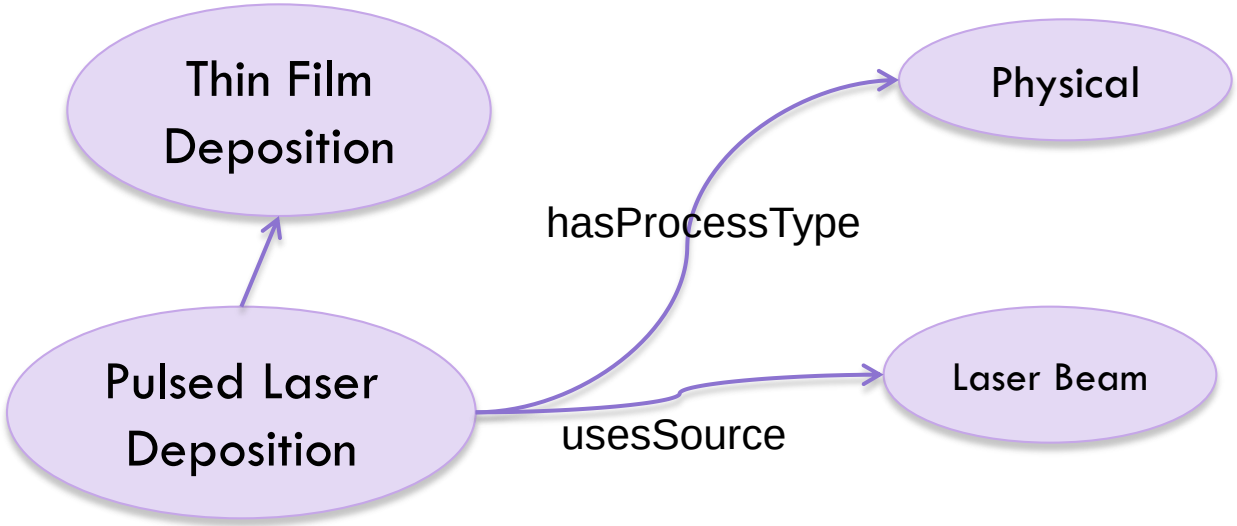




(non-exhaustively) Linking our terms



Describing Pulsed Laser Deposition



Class: PulsedLaserDeposition

SubClassOf: ThinFilmDeposition,

hasProcessType some Physical,

usesSource some LaserBeam

Lessons

- Know your users;
- Know the questions that need to be asked;
- Pull things apart and build them up again
  - Makes things explicit
  - Easier to discuss simple taxonomies;
  - Easier to maintain simple taxonomies;
  - Can build complex taxonomies automatically…