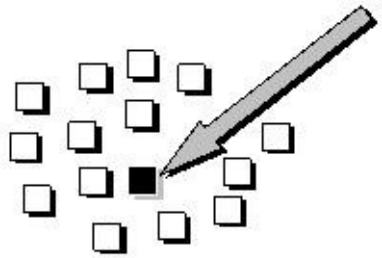
Creating a structured vocabulary



Leonard Will

L.Will@willpowerinfo.co.uk

ISKO UK Meetup

Creating a structured vocabulary

Basic steps in managing a vocabulary project

Following ISO 25964-1:2011:
Information and documentation — Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval (£361)

Creating a structured vocabulary

1. Why, who, what?
2. Collect concepts
3. Structure concepts
4. Combine concepts
5. Test and refine
6. Install and distribute
7. Maintain

Purpose

- What is the vocabulary for?
  - Is there an existing system?
  - Future plans?
  - Long term support?

What resources will it index?

- Things with words
  - Textual documents
  - Forms
- Things without words (may have captions)
  - Images
  - Products or objects
  - AV media
- Search existing text or add index terms as metadata?

What software will it use?

- Stand alone thesaurus / vocabulary system
- Part of a database system
- Shared or linked data
- Centralised or distributed
- Features and constraints
- Black box or interactive search refinement

Who will use it for searching?

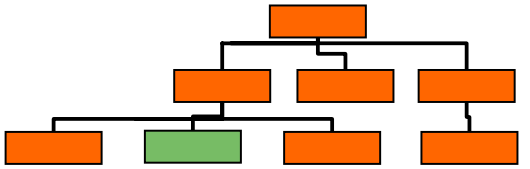
- Subject specialists
  - Internal
  - External
- Information specialists
  - Library / information / IT staff
- Public

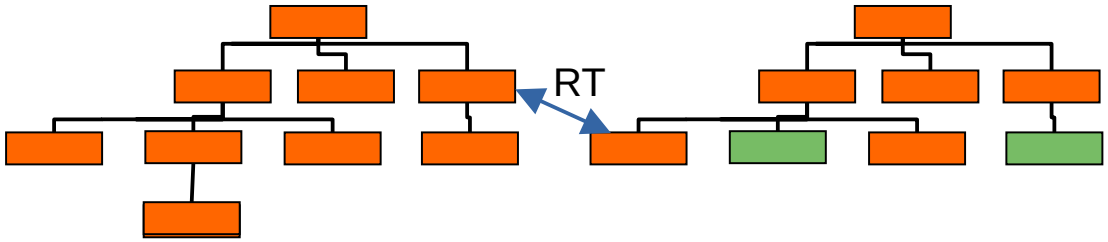
Who will create and maintain it?

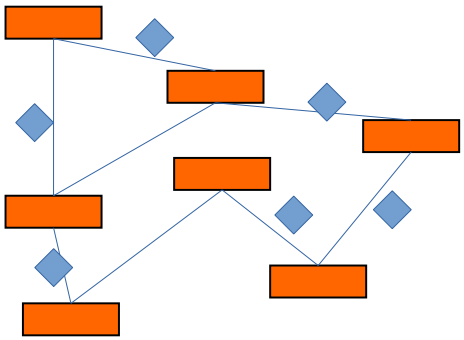
- Editors
  - Initial creation, subsequent maintenance
- Indexers
  - Suggested terms or changes found to be needed
- Users
  - Feedback, suggestions, query logs

What form will it take?

- **Taxonomy**
  - Single generic relationship only (monohierarchical)
  - Each concept at the “place of unique definition” – what something ***is*** rather than what properties it may have
- **Thesaurus**
  - Multiple generic relationships
  - Catch-all “related” relationship
- **Ontology**
  - Many kinds of relationship can be specified







- **Classification scheme**
  - Concepts combined according to a stated citation order to give useful linear sequence of compounds
- **Authority file**
  - Proper names e.g. of people, organizations or places

What exists already?

- Existing schemes
  - External: Basic Register of Thesauri, Ontologies & Classifications (BARTOC) lists 3400 vocabularies and many other registries and services (bartoc.org)
  - Internal: departmental and personal
- Use as they stand or as a source of concepts / terms

Creating a structured vocabulary

1. Why, who, what?
2. Collect concepts
3. Structure concepts
4. Combine concepts
5. Test and refine
6. Install and distribute
7. Maintain

Collecting concepts and terms

- If you don’t adopt an existing scheme, terms may come from:
  - Personal or departmental schemes or indexes
  - Log of previous queries
  - Reference works. Glossaries, dictionaries, textbooks, catalogues

Analysing concepts

- Clarify meaning / scope
  - Railway ticket clerk, with a limited number of categories for accompanying animals, said to an old lady travelling with a menagerie of pets. “Cats are “dogs” and rabbits are “dogs”, and so are parrots, but this here tortoise is an “insect” and there’s no charge for that.” [Punch, 1869]
- Check for duplication or overlap. Write scope notes
  - What are ships? boats? vessels? watercraft?
- Check whether within the scope of the scheme
  - Do we include cooking vessels?

Choose preferred terms

- Choose from various terms that may refer to the same concept
  - Synonyms
    - lochs USE lakes
  - Quasi-synonyms
    - cash USE money
- Compound concepts
  - To split or not to split?
    - coal mining USE coal + mining
  - or
    - mining NT coal mining

Creating a structured vocabulary

1. Why, who, what?
2. Collect concepts
3. Structure concepts
4. Combine concepts
5. Test and refine
6. Install and distribute
7. Maintain

Build hierarchies

- Sort into facets

  The choice of “fundamental” categories is not absolutely objective, but should be consistent within a knowledge organization scheme. Hierarchies within a facet must be genus/species = “is-a” relationship.

- Top down or bottom up?

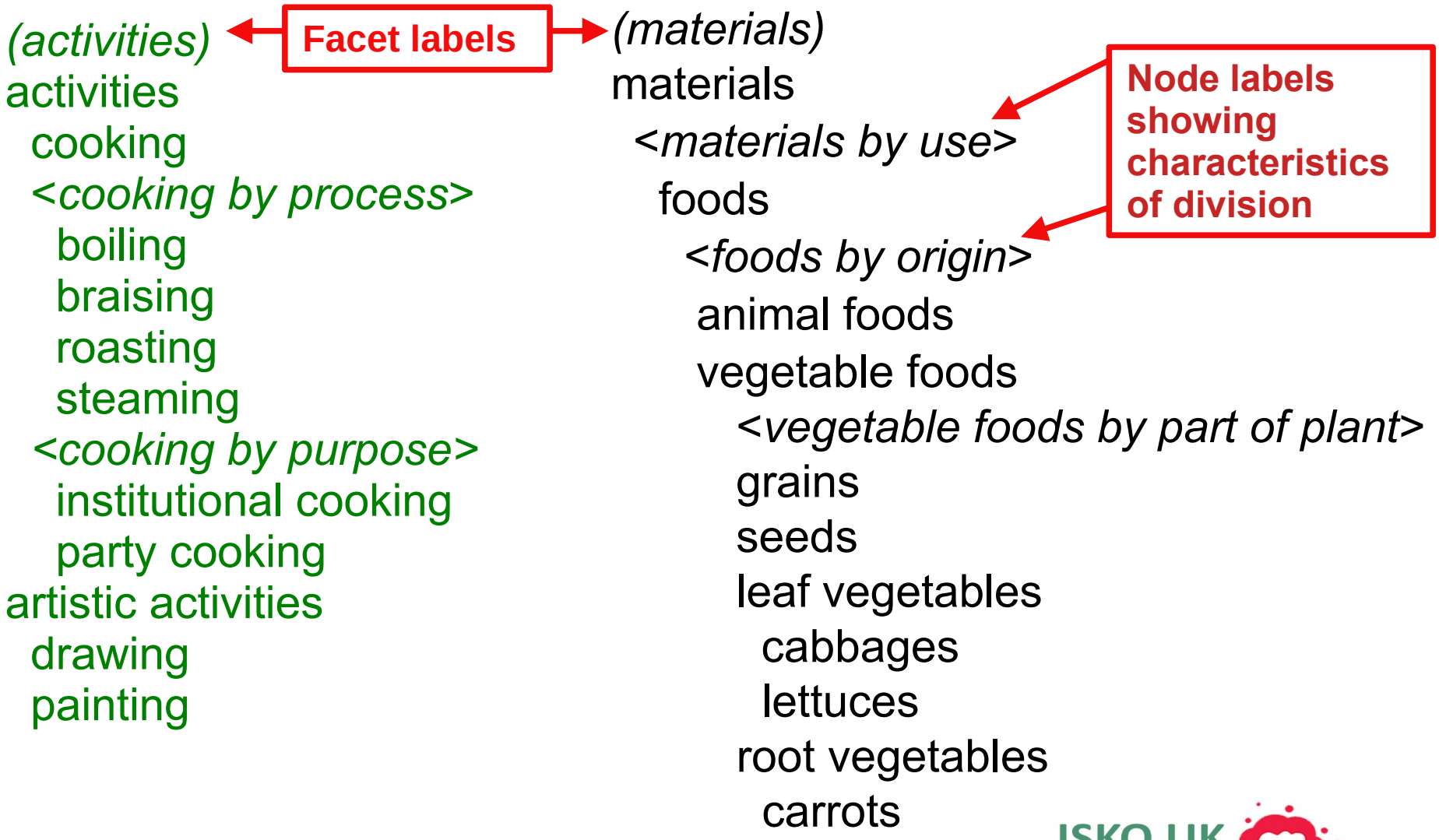
(persons)
persons
<persons by age>
babies
children
adults
old people
<persons by occupation>
cooks
pastry cooks
farmers
information scientists
librarians
students

Fundamental categories - facets

Distinct and mutually exclusive groups, e.g.

- **Things** - *ships, shoes, cabbages, power stations, heating systems*
- **Activities**, processes, disciplines - *cutting, thinking, dancing, rusting, physics*
- **Abstract concepts** - *love, war, costs, benefits, access, rights*
- **Places** - *continents, mountains, countries, political groupings, rivers, Europe*
- **Times** - *21st century, pre-war, 2012-12-25, mediaeval*
- **Materials** - *sealing wax, water, dirt, adhesives, polymers, aspirin*
- **Properties** - *size, colour, intelligence, plasticity*
- **People and organizations** - *kings, children, hospitals, ISKO*
- **Events** - *battles, conferences, wars, investigations, festivals*

Examples of facets



Create “related concept” relationships

- mining
  RT quarrying
  miners
  mines
- cabbages
  RT cabbage butterflies
  coleslaw
  sauerkraut
- If creating an ontology, use more specific relationship types, e.g. “source of”, “component of”, “made in”, etc.

Creating a structured vocabulary

1. Why, who, what?
2. Collect concepts
3. Structure concepts
4. Combine concepts
5. Test and refine
6. Install and distribute
7. Maintain

Combine concepts when searching

- Post-coordination: combine at search time
  - Specify several terms in a search statement
    - *cabbages cooking*
  - Combine with explicit or implicit Boolean operators (AND, OR, NOT)
    - *(cabbages AND cooking) NOT pickles*
  - Provide filters to refine searches

Combine using filters

Design/Finish

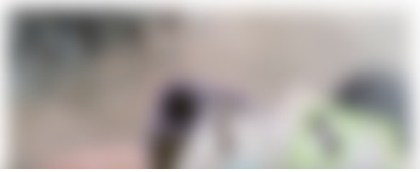
- [ ] Patterned (4,438,683)
- [ ] Pictorial (2,568,222)
- [ ] Glossy (963,072)
- [ ] Luxury (856,475)
- [ ] Matt (694,785)
- [ ] Marble (667,240)
- [ ] Slim (627,658)
- [ ] Transparent (606,082)

See all

Material

- [ ] Rigid Plastic (3,207,450)
- [ ] Silicone/Gel/Rubber (2,129,11
- [ ] Leather (1,125,445)
- [ ] Faux Leather (1,081,838)
- [ ] Synthetic Leather (897,430)
- [ ] Plastic (827,981)
- [ ] TPU (260,546)
- [ ] Acrylic (143,066)

See all



ISKO UK Meetup

Combine concepts when indexing

- Pre-coordination: to provide a useful sequence for compound concepts, concepts can be combined at the time of indexing rather than when searching
  - Specify a citation order for concepts
    - cultivating : cabbages
    - cultivating : carrots
    - cooking : cabbages
    - cooking : carrots
  - or
    - cabbages : cooking
    - cabbages : growing
    - carrots : cooking
    - carrots : cultivating

Faceted classification

(people)
people
 <people by age>
 babies
 children
 adults
 old people
 <people by occupation>
 cooks
 farmers
 information scientists
 librarians
 students

(activities)
activities
 cooking
 <cooking by process>
  boiling
  braising
  roasting
  steaming
 <cooking by purpose>
  institutional cooking
  party cooking
artistic activities
 drawing
 painting

(materials)
materials
 <materials by use>
  foods
   <foods by origin>
    animal foods
    vegetable foods
     <vegetable foods by part of plant>
     grains
     seeds
     leaf vegetables
      cabbages
      (activities)
      cooking
       <cooking by process>
        steaming

Creating a structured vocabulary

1. Why, who, what?
2. Collect concepts
3. Structure concepts
4. Combine concepts
5. Test and refine
6. Install and distribute
7. Maintain

Train and explain

- Write introduction explaining:
  - Structure
  - Scope
  - Conventions
  - Feedback and updating arrangements

Test and adjust

- Apply to sample documents
- Apply to sample queries
  - Compiler / editor
  - Indexers
  - Users
- Load trial to database, if not already integrated

Creating a structured vocabulary

- Why, who, what?
- Collect concepts
- Structure concepts
- Combine concepts
- Test and refine
- Install and distribute
- Maintain

Load and disseminate

- Load live version
  - Single shared network version, or several?
- Are paper prints needed?
- Within organization or wider?
- If you are willing to share, consider clearinghouse deposit or link

Creating a structured vocabulary

- Why, who, what?
- Collect concepts
- Structure concepts
- Combine concepts
- Test and refine
- Install and distribute
- Maintain

Updating

- Feedback, review and updating
  - Online suggestions while in use
  - Suggestions for new terms, new relationships, clarification of scope
  - Review of queries: failures, successes
  - Too many or too few postings show a need for more or less specificity

Recording changes

- Notification of changes
  - Record date of change
  - Deleted terms become non-preferred
- Updating dispersed systems
- Previously indexed material
  - Retrospective re-indexing?
  - Alerts to search on previous terms?

For life

- Like a dog, a thesaurus is for life, not just for Christmas
- Like a library, a thesaurus is a growing organism
- Good luck!