Gardens of Meaning –
the development and evolution of vocabularies for indexing and retrieval

Alistair Miles
CCLRC Rutherford Appleton Laboratory

Ontogenesis Network Meeting, October 2006, Manchester

Outline

1. A quick word on SKOS

2. Gardens of meaning ...

SKOS

- W3C Recommendation track
- Semantic Web Deployment WG
- Use cases and requirements

SKOS Scope

- **RDF representation** of ...

  ... **controlled structured vocabularies** ...

  ... used as **indexing languages** ...

  ... i.e. purpose is **information retrieval**.

- Controlled vocabularies with **simple conceptual structures**
  - broader/narrower hierarchies
  - associative links

SKOS Applications

- Retrieval applications ...

... exploiting metadata ...

... and **structure** of controlled vocabularies ...

... search and browse functionalities.

- Vocabulary data: **SKOS**

- Index data (metadata): **DC + SKOS**

- For more on my suggested direction for SKOS see "**SKOS: Requirements for Standardization**" presented at Dublin Core 2006 ...

http://dc2006.ucol.mx/papers/miercoles/10.30/presentation.pdf

http://isegserv.itd.rl.ac.uk/public/skos/press/dc2006/camera-ready-paper.pdf

SKOS Use Cases

- Do you have a use case?

SKOS Requirements ... Interoperability with OWL

- 2 main scenarios ...

  - **Hybrids**, i.e. SKOS + OWL

    - e.g. swed.org.uk

  - **Migration**, i.e. ? → SKOS → OWL

    - Methodology?
    - Use cases?

Outline

1. A quick word on SKOS
2. **Gardens of meaning** ...

Gardens of Meaning

- A metaphor for the development and evolution of controlled structured vocabularies ...

Less Formal Vocabularies

- **As an end in themselves**

- As a point on a migration path

As an end ...

- Challenges in the application of (less formal) controlled vocabularies to information retrieval.
  - Themes ... cost/benefit, minimising overheads, collaboration, change management, versioning, methodologies ...
- A way forward ... ?

What is the end?

- High precision, high recall service for the "retrieval" of objects (usually documents) from some collection of objects ...

- ... where performance is maintained over medium- to long- term (5 – 15+ years) ...

- ... at low initial and ongoing cost ...

- ... with significant initial and ongoing ROI.

Strategy

- A controlled vocabulary is created (by hand).

- Metadata is created (by hand), using the controlled vocabulary, for some set of objects (e.g. Documents).

- Queries are formed, using the controlled vocabulary.

- Information is "retrieved" by evaluating queries w.r.t. metadata.

Challenges ... COST!

- Cost of creating the vocab (bootstrap)
- Cost of maintaining currency of the vocab
- Cost of maintaining dependencies

Typical approach to change management ...

- "Discrete" versioning (iterative)
  - i.e. Periodic release of a new "version" or "edition"
- Changes are described as prose (if at all)

➔ Updating metadata ("re-indexing") is a huge intellectual effort
➔ Serious barrier to scalability
➔ **Serious barrier to viability!**

Real world ...

- Disposable vocabularies (build, use, throw away)
- Very poor ROI

A Better World ... ?

3 ─────────────

2 ─────────────

1 ─────────────

0 ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─
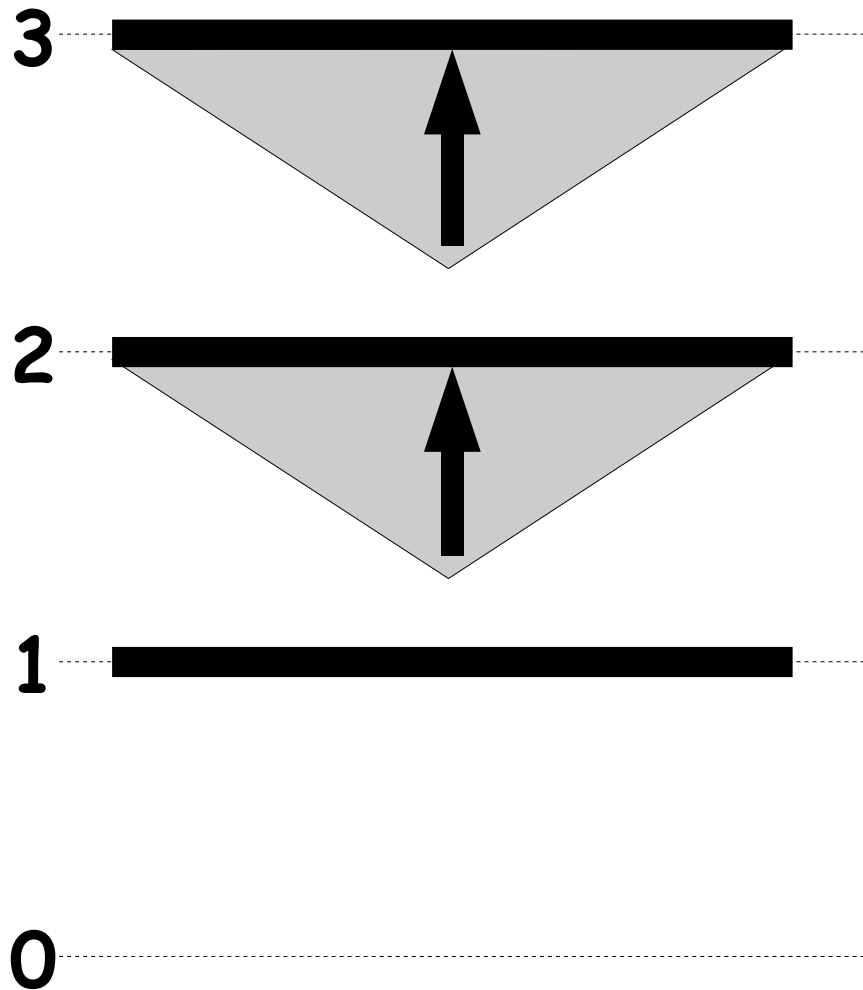
- First iteration, publish vocabulary.

- Second and subsequent iterations, publish vocabulary plus mapping from previous edition.

Mapping & Translation

- Mapping is sufficient to translate either metadata "forwards" or queries "backwards" ...

... via translation algorithm (i.e. "automatic") ...

... with predictable consequences on retrieval applications in terms of recall and precision.

N.B. See **purl.org/net/retrieval** esp. chapter 7.

Overcoming the Costs ...

- High bootstrap and currency costs?
    - Work concurrently
    - Real-time collaboration
- High cost of maintaining dependencies?
    - Capture minimal information about the nature of changes in meaning at the point of change
    - ➔ Generate mappings at time of publication with little or no additional effort

The Challenge!

Design a **process model** (i.e. a workflow) that ...

... gets **people working together** and ...

... **structures the development process** in such a way as to make it possible to ...

... **capture the right information about change** to **generate mappings between versions**, whilst ...

... providing the **smallest possible impediment to the developer**?

Gardens of Meaning?

- A **metaphor** for the development of controlled vocabularies for information retrieval ...

Cultivating a Garden ...

- The goal of a vocabulary development project is to **cultivate a garden**.

- Within each garden there are two main areas of work ... **the greenhouse** and ... **the vineyard**.

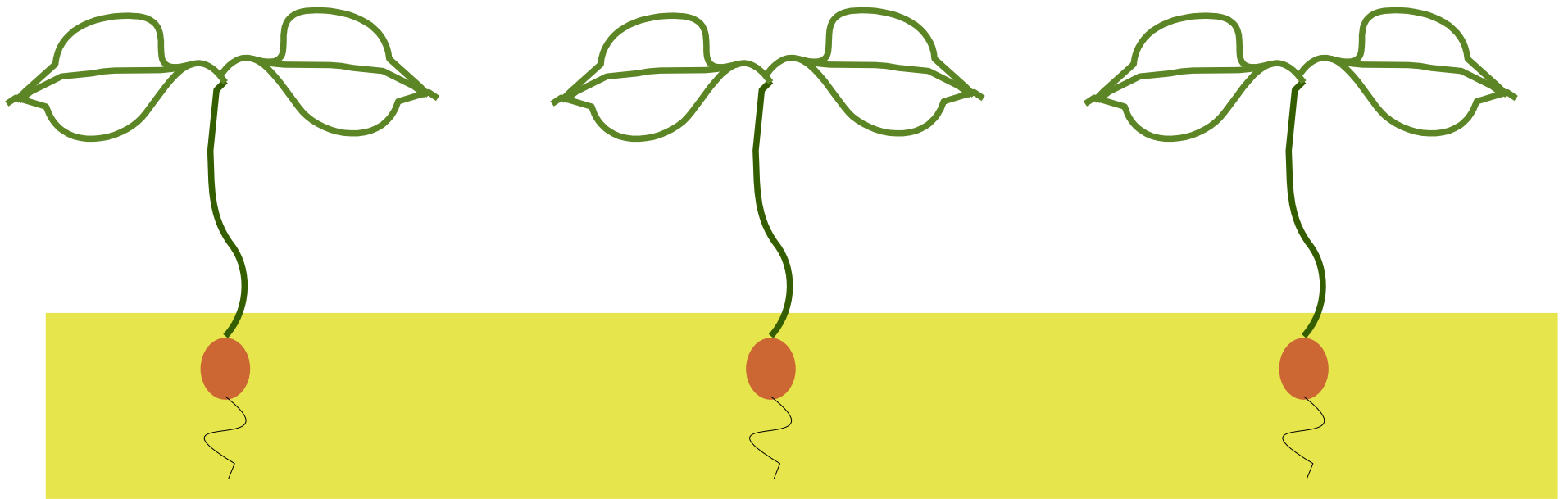# The Greenhouse ...

- Forming the **seeds of meaning**

The Vineyard ...

- **Cultivating meanings** through the **growth of vines**
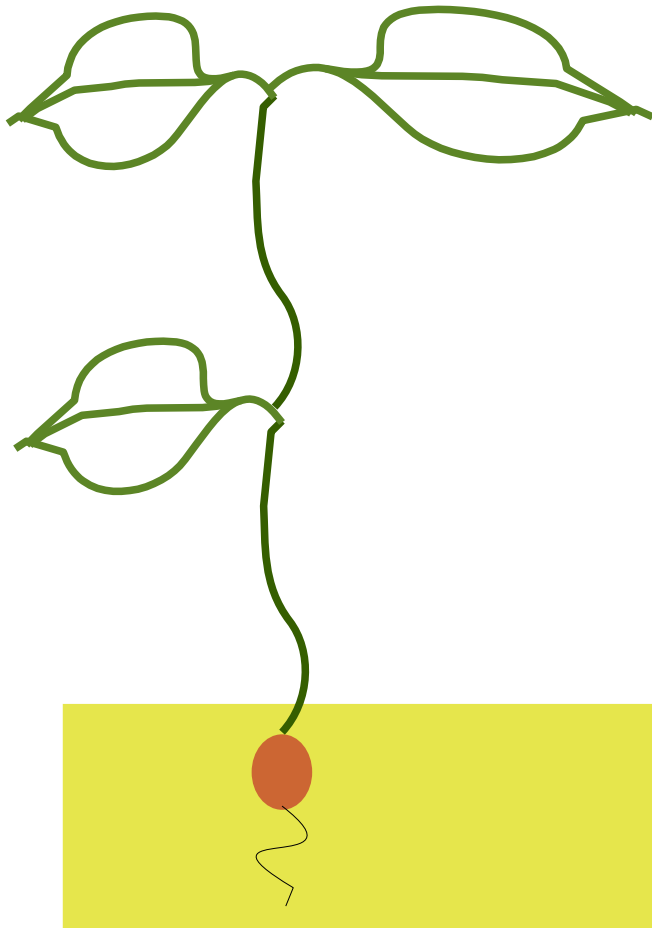- (Each vine captures a thread of meaning)

The Greenhouse Layout ...

Tag list

Work surface

love
affection

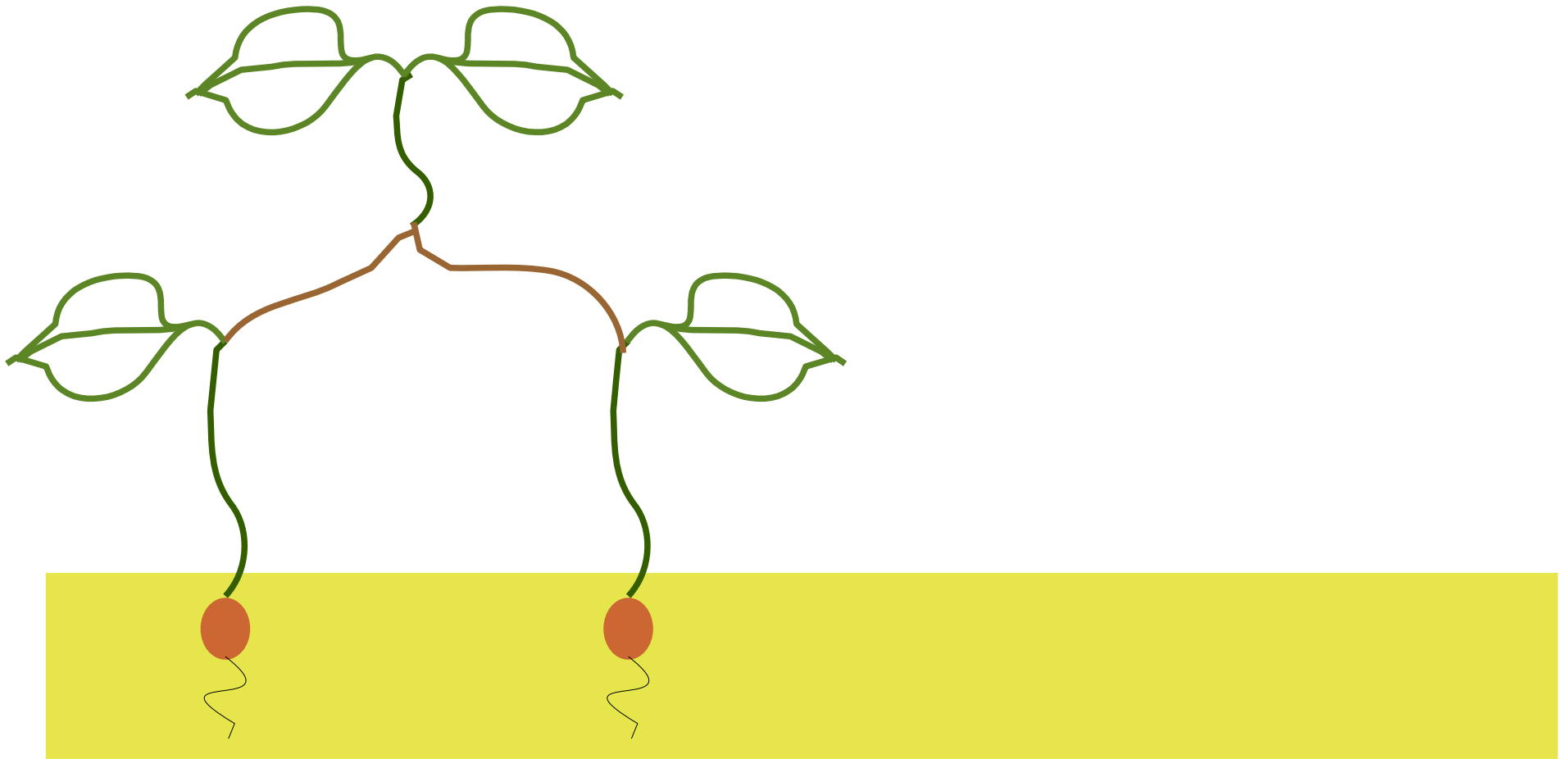Seed bed
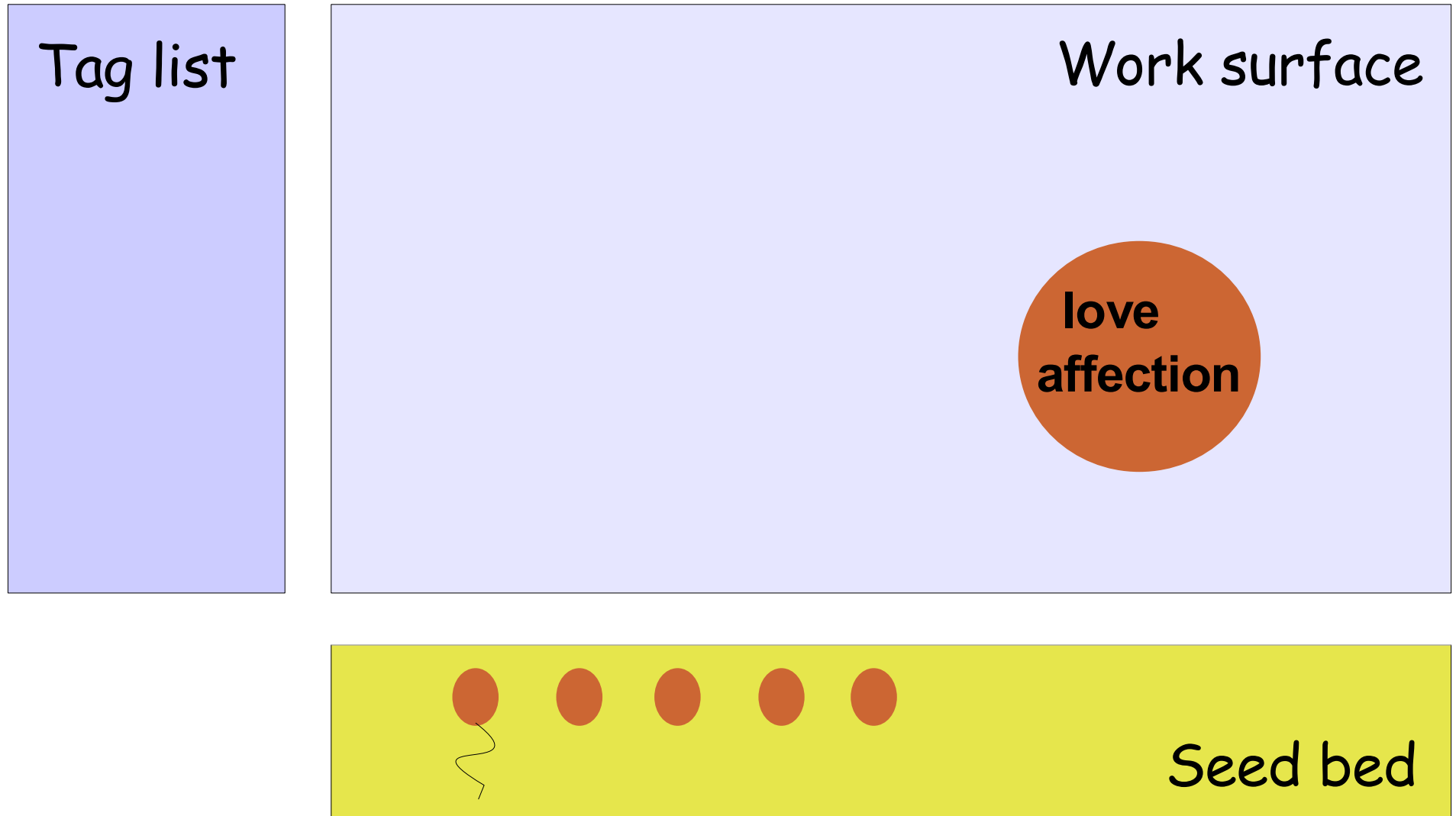
The Vineyard ...

Vines Grow in Stages ...

- Make changes by "growing" vines.
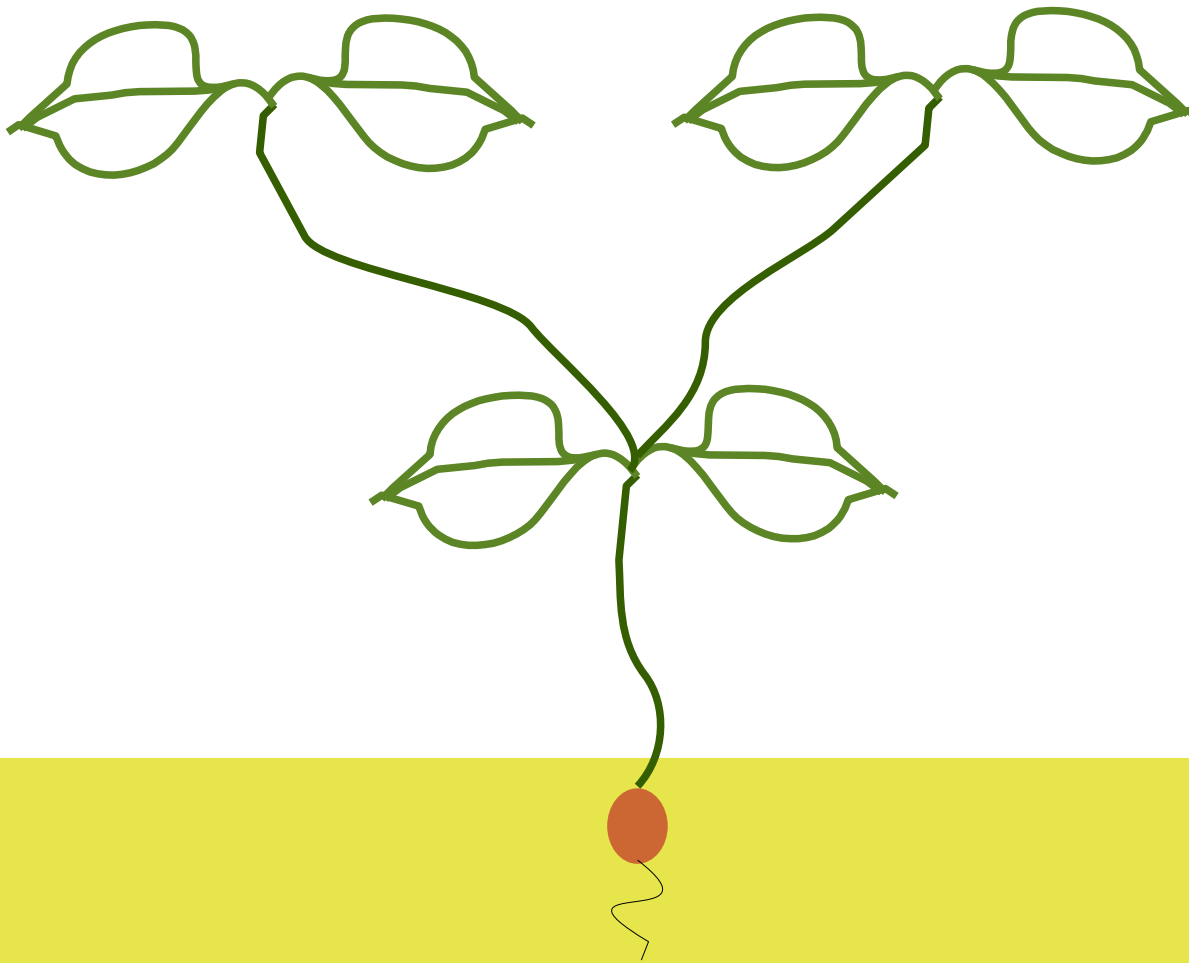- "Growing sessions" to manage concurrency.

E.g. Add a label ...

E.g. Join vines...

E.g. Split vines...

Neighbourhood (1)

- Use "neighbourhood" as metaphor for conceptual structure.

- To change the neighbourhood, must stimulate growth stage ...

... which also triggers growth stage in all new (added) neighbours and all old (removed) neighbours.

→ Each growth stage has a **unique neighbourhood.**

Neighbourhood (2)

- However, vines can grow in-situ, without forcing growth stages in all neighbours.

- (otherwise you could never change one vine without changing them all ... i.e. concurrency would be impossible).
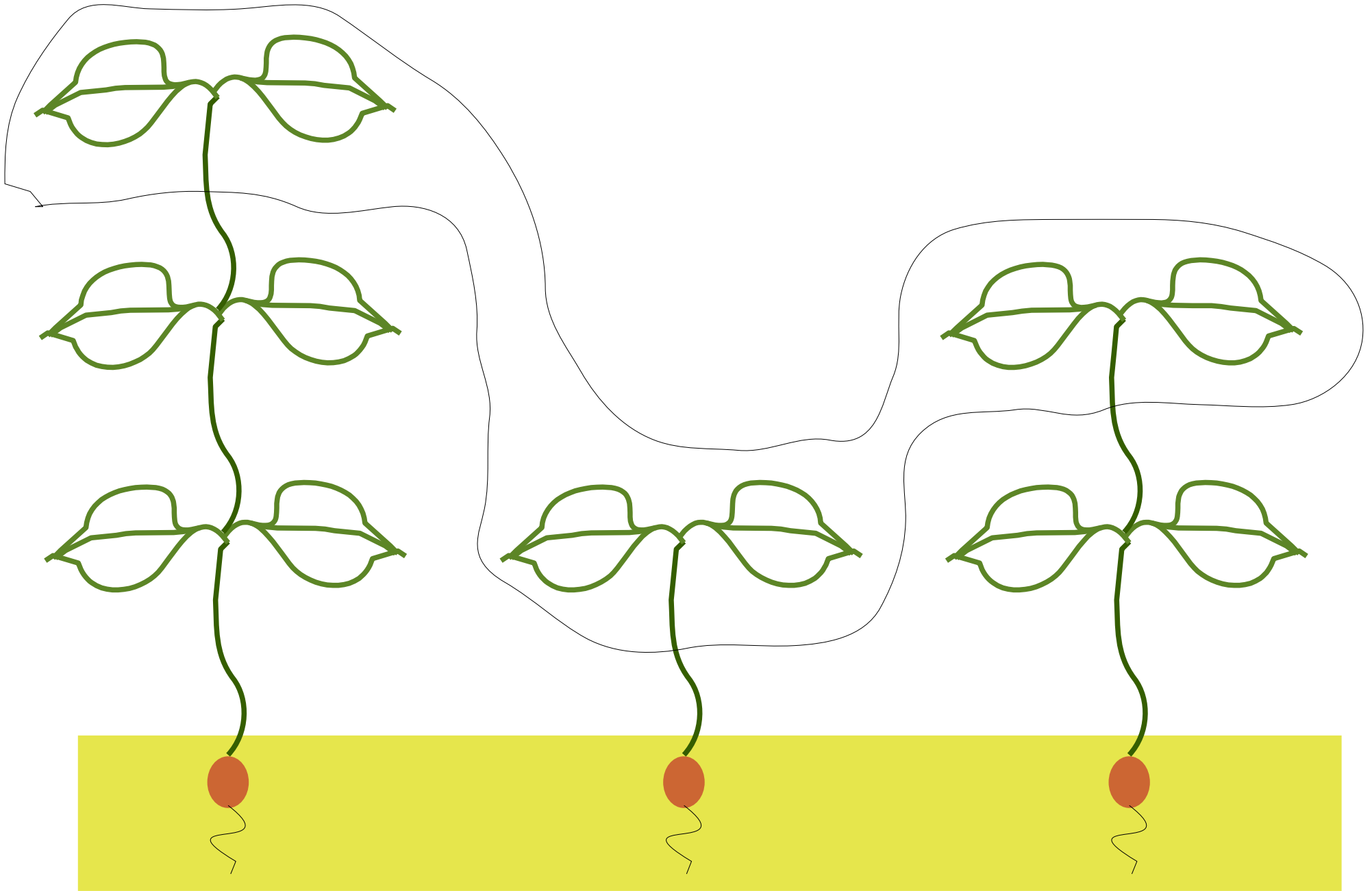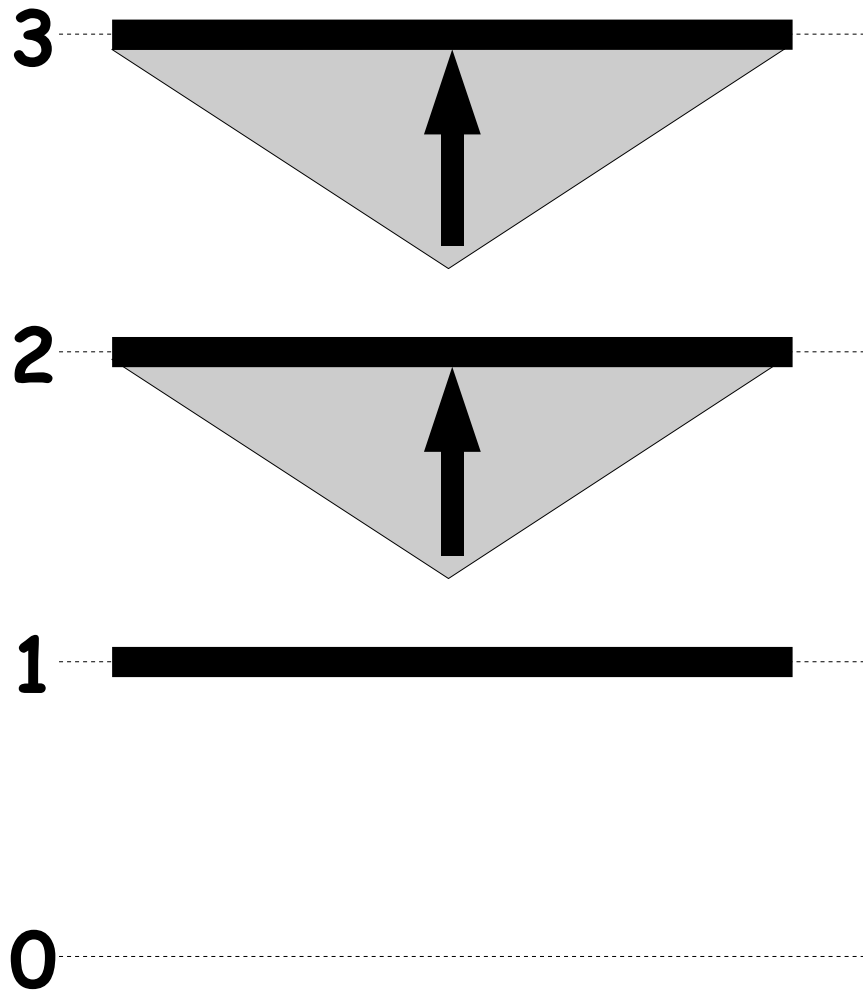
Versioning

- Create a "version" ("edition") by capturing most recent growth stage of all living vines.

- A bit like taking an "aerial photograph".

Summary so far ...

- Use metaphor to provide model for managing concurrency and capturing history of change.
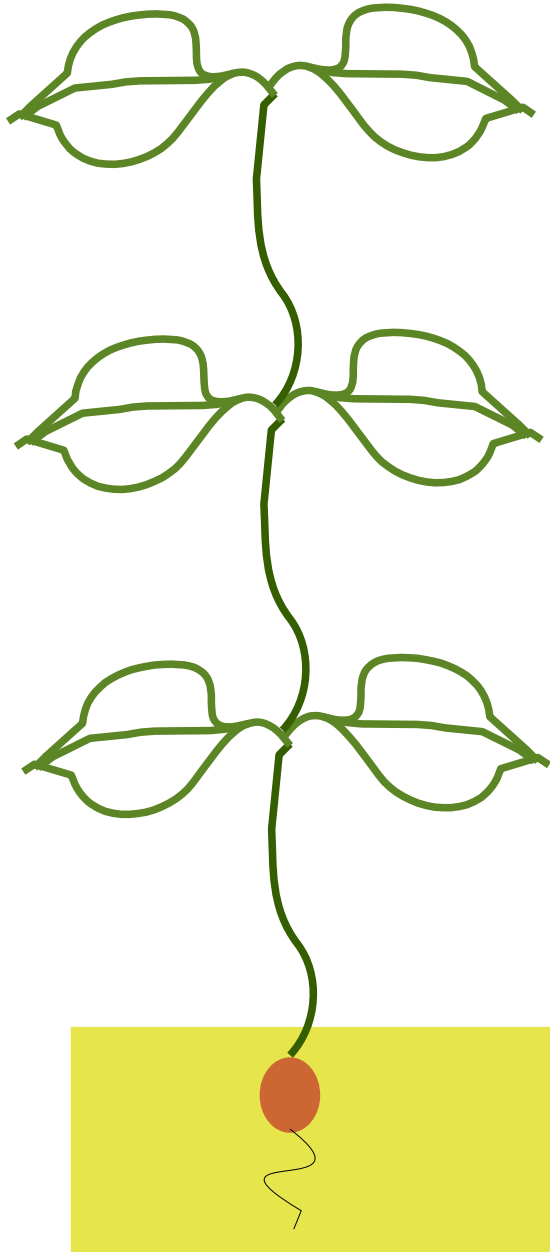
- What about the idea of mappings between versions?
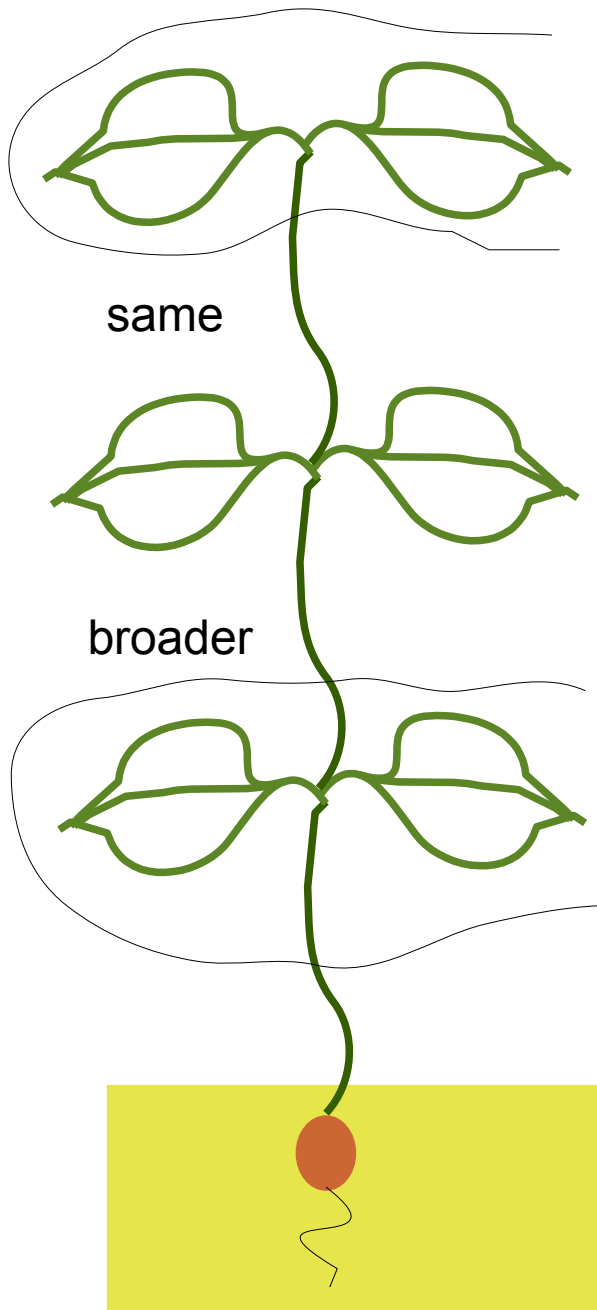
Phases of Development

3

2

1

0

- "Bootstrap" - up to publication of first edition.

- "Evolution" - from publication of first edition onwards.

Bootstrap...

- Allow gardeners to grow vines, without worrying about how the meaning is changing.

- Lowest impediment to growth.

Evolution...

same

broader

- For each growing session, gardeners must specify **whether or not the meaning has changed**.

- If meaning has changed, then specify how i.e. broader, narrower, associated?

Capturing Change Info

- By capturing simple nature of meaning change with each growth stage, can infer mapping between editions in most cases.

- Some cases will be ambiguous, need manual verification.

- I.e. Obtain mapping with minimal human effort.

Summary

- SKOS – have you got a use case?

- Challenges in use of "less formal" vocabularies for information retrieval.

- A metaphor for the development of structured vocabularies, to minimise costs of creation, currency and maintaining dependencies.

- What do you think?

Further Issues ...

- Strategies for multilingual vocabularies
- Suggestions and support
- Gardening "rights"
- Chat, messaging and comments
- Implementation ...