Technical Communicator as Taxonomist

Panning for gold: Mine term values in your organization

Sabine Ocker
Consultant
Comtech Services

Introductions

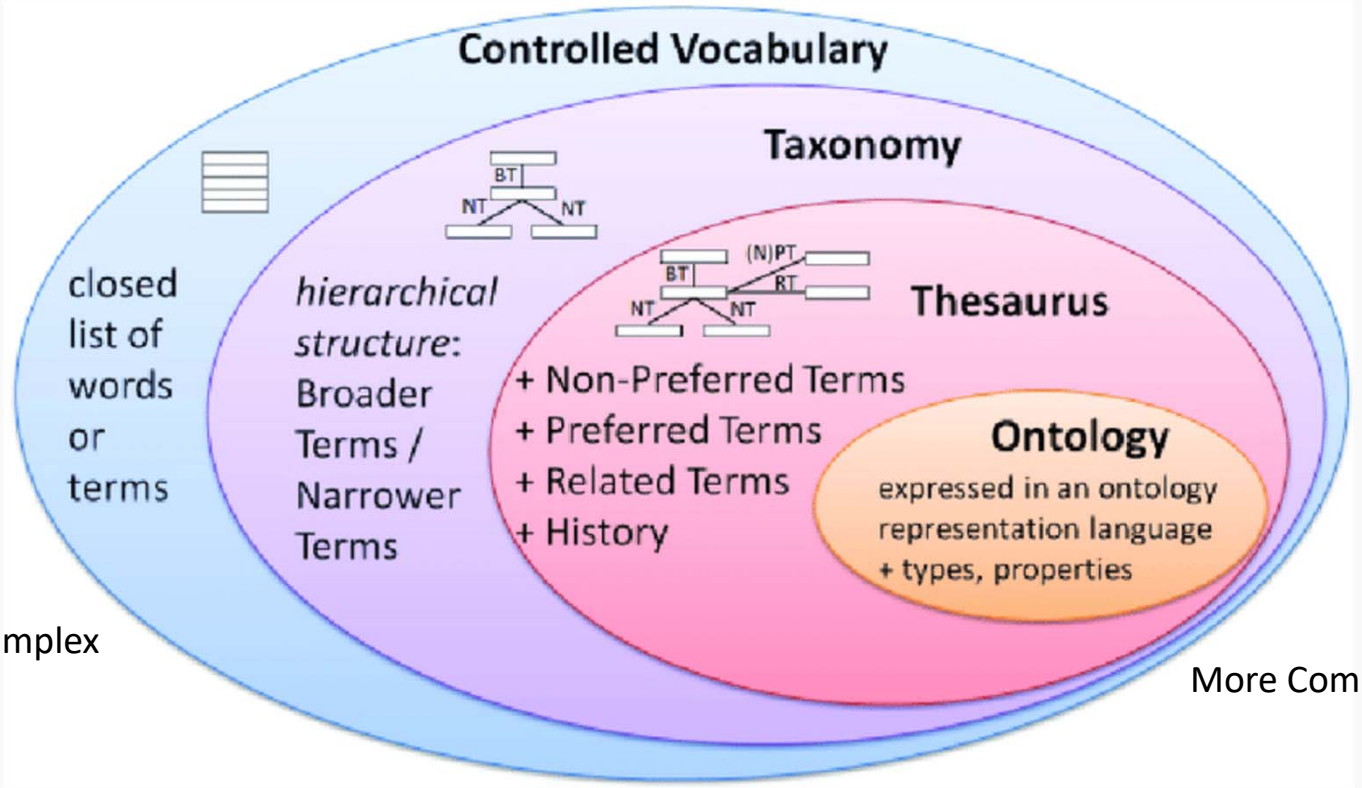
Over 40 years providing consulting and training services to information development organizations in all industries around the world

- Over 20 years as an XML Information Architect
- Worked as a customer and as a vendor prior to joining Comtech
- Passionate about the value of metadata and taxonomies
- Sabine.Ocker@Comtech-serv.com

What is a term?

- Word or phrase which describes a thing
- Uses:
  - a keyword for searches
  - a value in a taxonomy or DITA Subject Scheme
  - A part of a thesaurus
    - Synonyms
    - Variants
    - Competitor terminology
- Can have a label and an ID
- Usually expressed as metadata
  - DITA elements or attributes
  - CCMS

Taxonomy Complexity

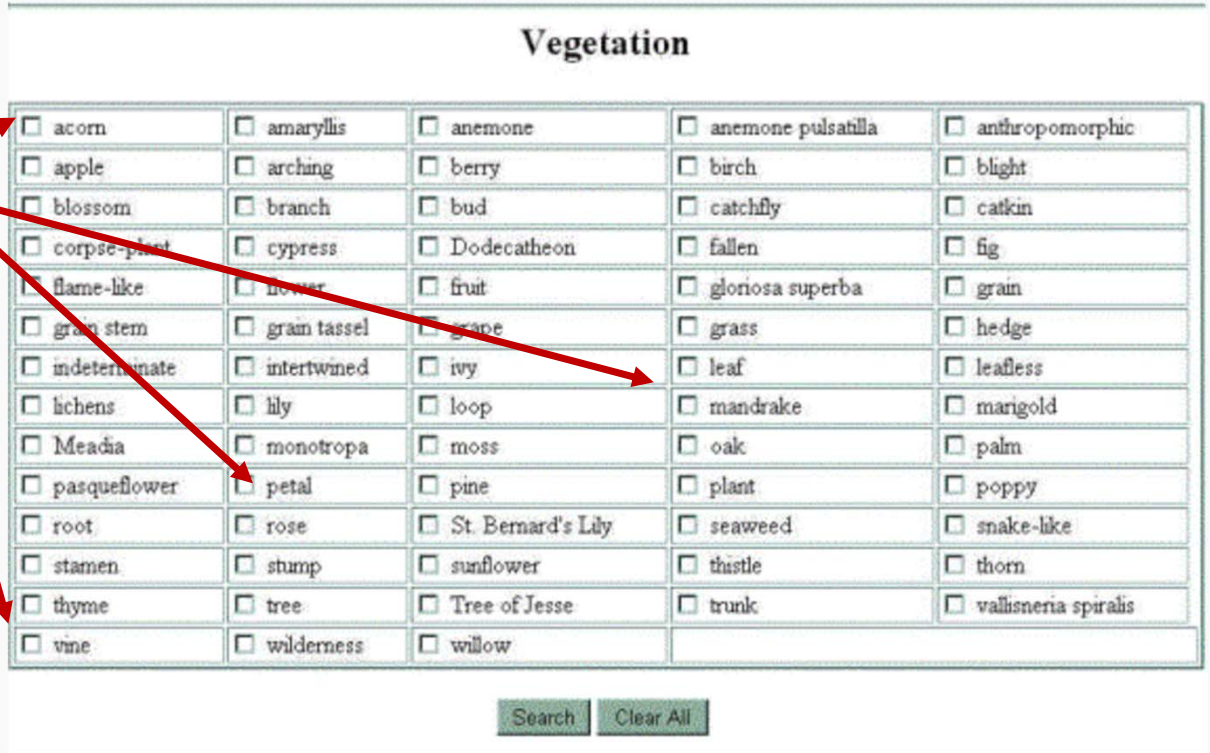


Less Complex

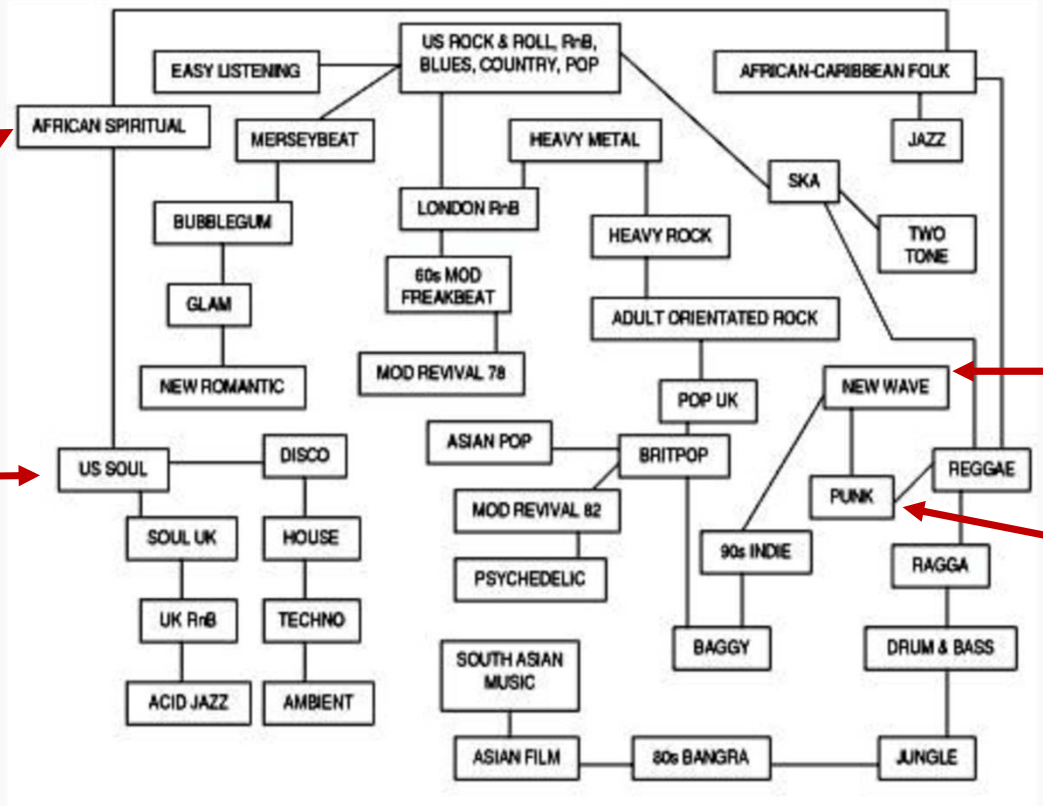
More Complex

Terms in a controlled vocabulary

These are all terms



Terms in a taxonomy



This is a term

This is a term

This is a term

This is a term

Terms in a DITA Subject Scheme

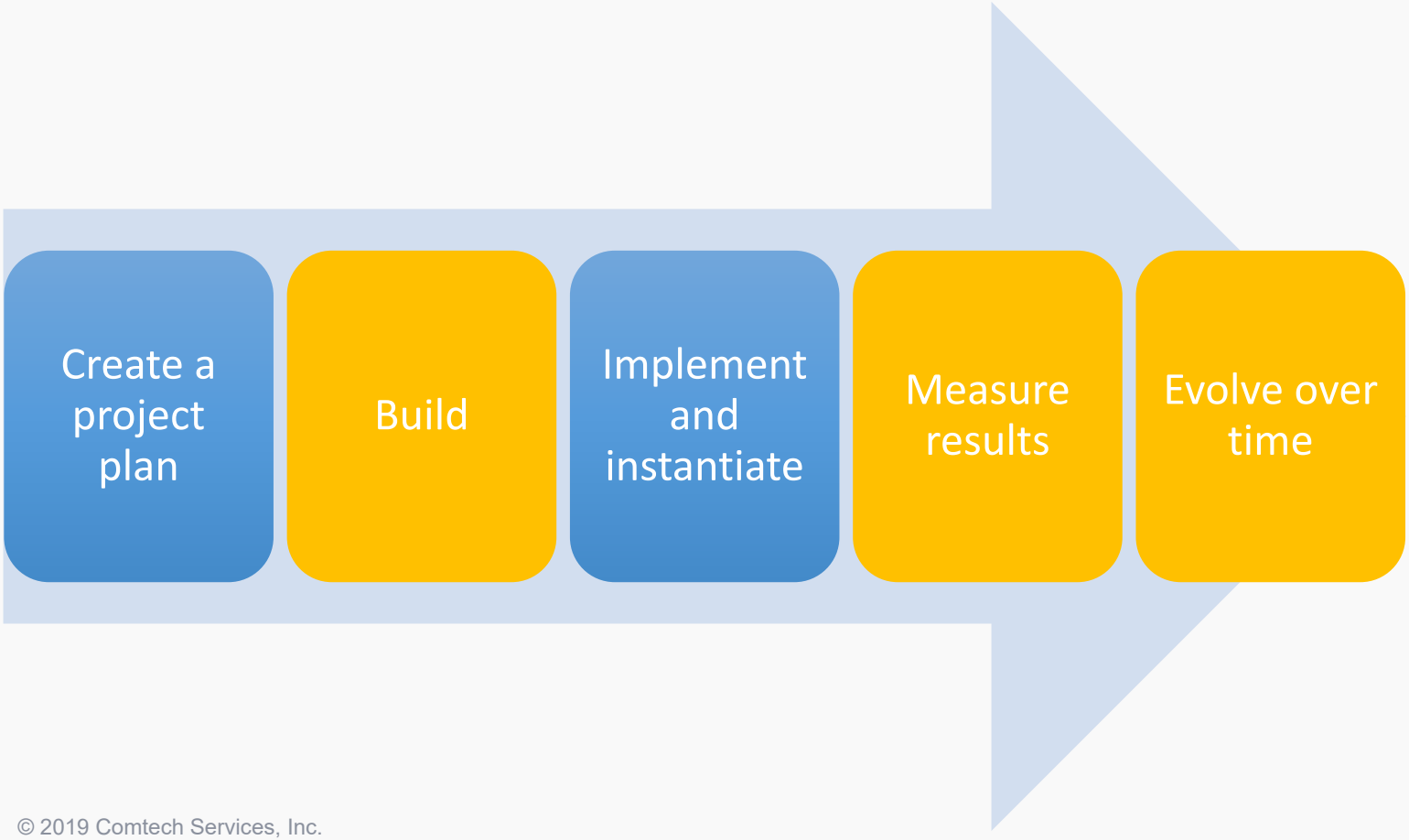
- Can specify a hierarchy via nested references
- Separate metadata value from the label via navtitle and keys
- Some delivery platforms can create search facets from subject schemes

This is a term

This is a term

```
<!-- This examples uses @navtitle rather than <navtitle> solely
to conserve space. Best practises for translate include using <navtitle>. -->
<subjectScheme>
  <subjectdef keys="os" navtitle="Operating system">
    <subjectdef keys="linux" navtitle="Linux">
      <subjectdef keys="redhat" navtitle="RedHat Linux"/>
      <subjectdef keys="suse" navtitle="SuSE Linux"/>
    </subjectdef>
    <subjectdef keys="windows" navtitle="Windows"/>
    <subjectdef keys="zos" navtitle="z/OS"/>
  </subjectdef>
  <enumerationdef>
    <attributedef name="platform"/>
    <subjectdef keyref="os"/>
  </enumerationdef>
</subjectScheme>
```

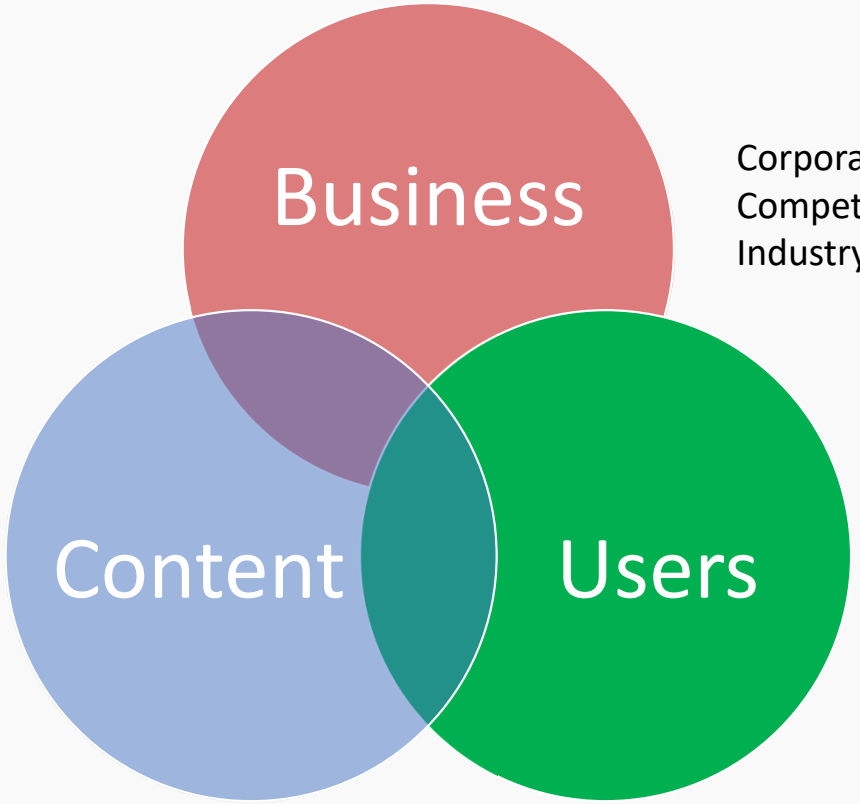
Taxonomy development steps



Methods for term generation

- Start from scratch
  - Stakeholders including product management, engineering
- Utilize existing
  - Corporate taxonomy
- Harvest from content
  - Tool

Source of terms



Corporate website
Competitor analysis
Industry experts

Content audit

Surveys and interviews
Task analysis
Social media
Search analytics

When generating your own terms

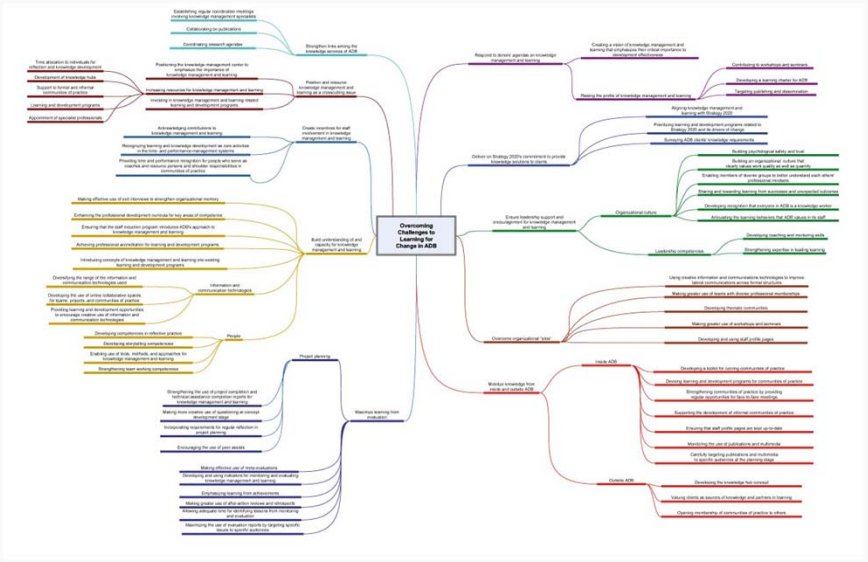
- Knowledge mapping methodology
  - Gather community of stakeholders to brainstorm terminology
  - Also consider:
    - What terms do your competitors use?
    - What terms do other content domains use, such as marketing or support?
    - What terms are in your corporate taxonomy or official term database?
    - What terms are in your content?
  - Create a visual diagram of terms and their relationships

When generating your own terms

- Method 1: Mass term generation and sorting
  - Everyone write terms on sticky notes
  - Sort and create categories
  - Count duplications
  - Vote on importance

When generating your own terms

- Method 2: Directed drilldown
  - Start with single term
  - Is a part of or is a parent to
  - Mind mapping software



Content Harvesting

- Some CCMS systems can crawl content stored in the repository and categorize content by metadata values or create a thesaurus
- Some delivery platforms can crawl content and expose metadata as either a facet or a tag
- Libraries and content aggregators have been using harvesting for scholarly and scientific journals

Term design principles

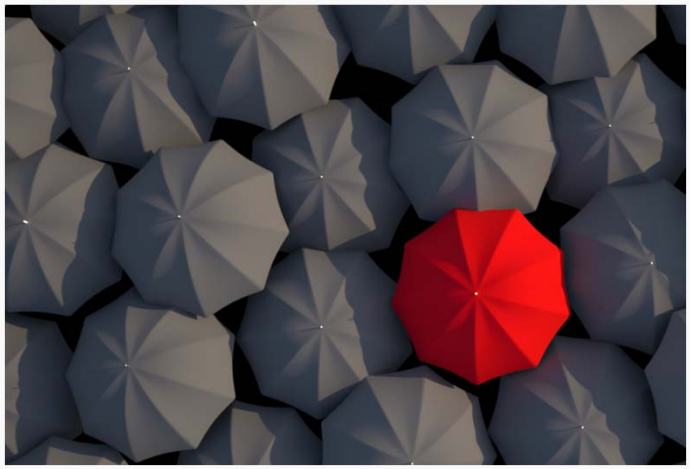
1. Intuitive
2. Unambiguous
3. Consistent
4. Predictable
5. Relevant
6. Meaningful
7. Durable

Term selection: Meaning

- Does the term accurately reflect the intended meaning to the intended audience?
- Does the term uniquely represent the concept to users without competing interpretations?
- Is the term appropriate for the intended user communities?

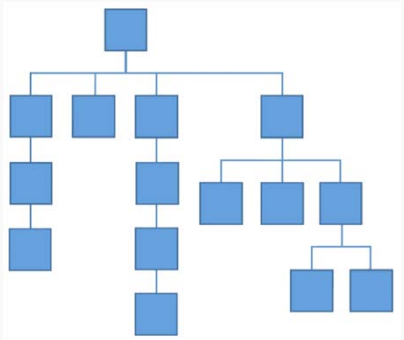
Term selection: Distinctiveness

- Does the term represent a distinction that is important to the audience?
- Does it overlap with another term?
- Does it need a qualifier to distinguish it from another term?



Term selection: Relationship

- Does the concept fit within the coverage of the intended taxonomy structure?
  - Is it a child of the parent element?
  - Does it have a defined relationship to another term?
  - Is it a clear member of the facet?
- Does the proposed term refer to a concept that is not already in the taxonomy?
- Does the term, along with others, warrant a new section in the taxonomy or a new facet?
- Does the term imply additional concepts or terms?



Term selection: Consistency

- Is the term stylistically consistent with the other terms in this taxonomy structure?
  - Part of speech
  - Tense
  - Form
  - Number

Term selection: Currency

- Does the term reflect a current common usage for the concept?
- Is the term dated or obsolete?

Term selection: Standardization

- Is the term part of an authorized standard vocabulary for which there is a compliance requirement?

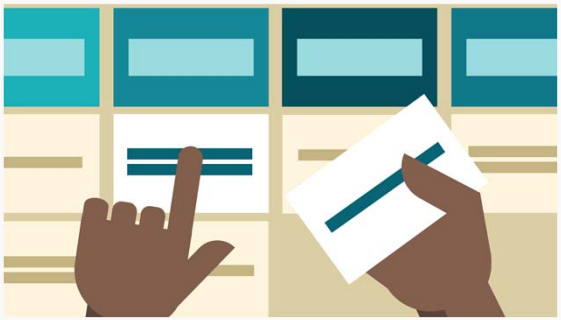
When to test

- Validate taxonomy design and organization after paper prototyping
- Validate taxonomy tagging after early document drafts are completed

Testing guidelines

- 6-8 people per round of testing
  - Represent target users
  - Same people or same types of people in each round of testing
  - 2-3 rounds of testing
- In-person as much as possible
  - However, possible to automate results (for example, //www.optimalworkshop.com/treejack.htm)

Card sorts



Open:

- Give participants cards showing taxonomy terms with no pre-established groupings.
- Ask them to sort cards into groups that they feel are appropriate and then describe each group.

Closed:

- Give participants cards showing taxonomy terms and set of primary groups.
- Ask them to sort cards into the group they feel works best

Managing your taxonomy

- Once you've deployed your taxonomy, it will need training and maintenance to continue to work properly.
- Terms change and domains grow and morph into new topics.

Taxonomy management tools

- Typical features
  - Data modeling
  - Editing
  - Import/export
  - Workflow
  - Integration interfaces
- Advanced features
  - Visualization of terms and relationships
  - Built-in auto-classification tools
  - Built-in term mapping tools
  - Robust versioning and archiving support
  - Multi-lingual support
  - Text analytics

Thank you!

Sabine.Ocker@Comtech-serv.com

Find me on: LinkedIn

Comtech Services