

# LexWiki

**Harold Solbrig**

**Guoqian Jiang, PhD**

**Division of Biomedical Statistics and Informatics**

**Mayo Clinic College of Medicine**

**January 22, 2009**

# The Situation

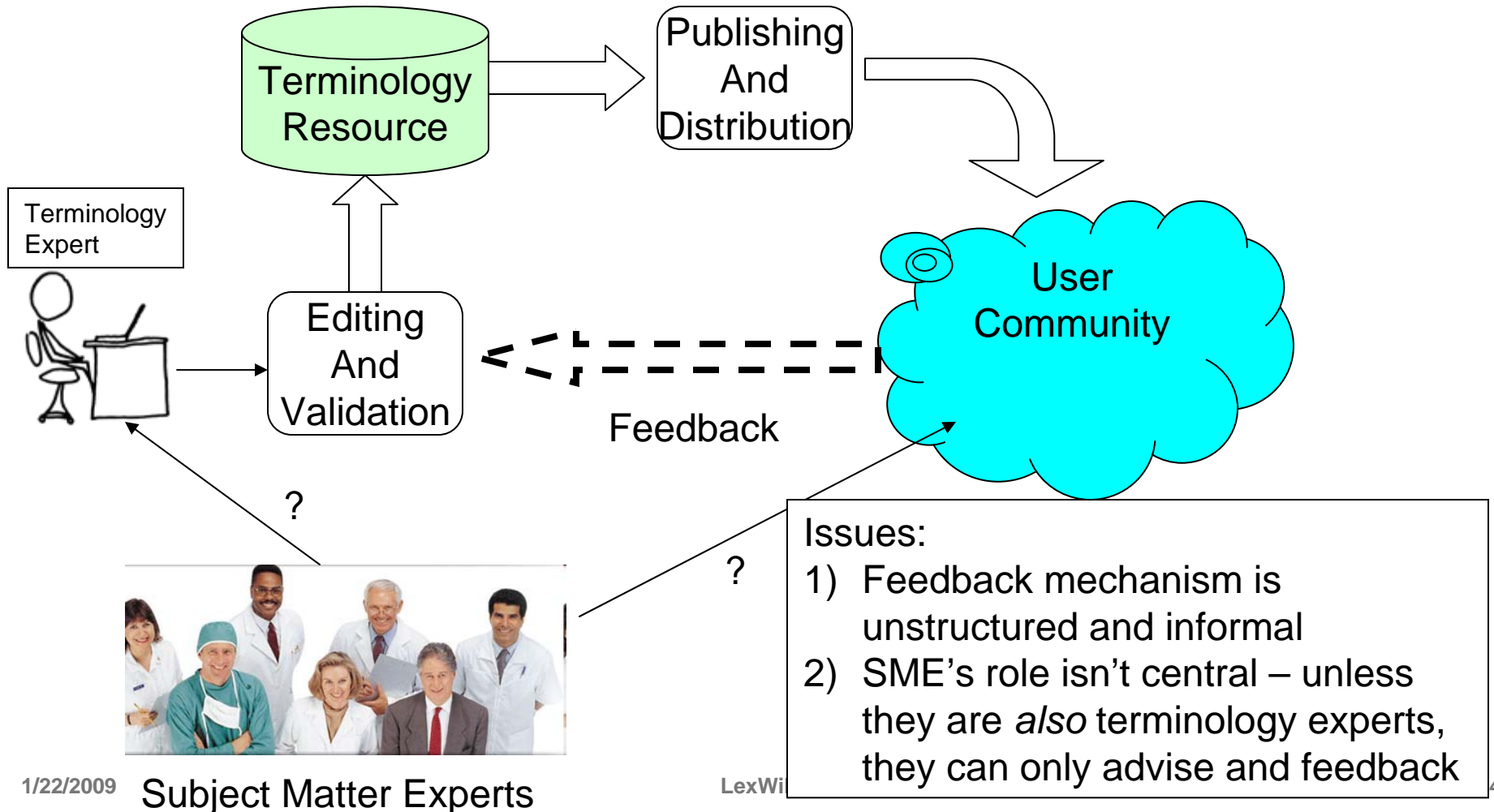
**Semantic MediaWiki was to be used for multiple terminology curation projects, including:**

- **BiomedGT – The National Cancer Institute’s Biomedical Grid Terminology**
- **ICD-11 – The World Health Organization’s International Classification of Diseases Revision 11**

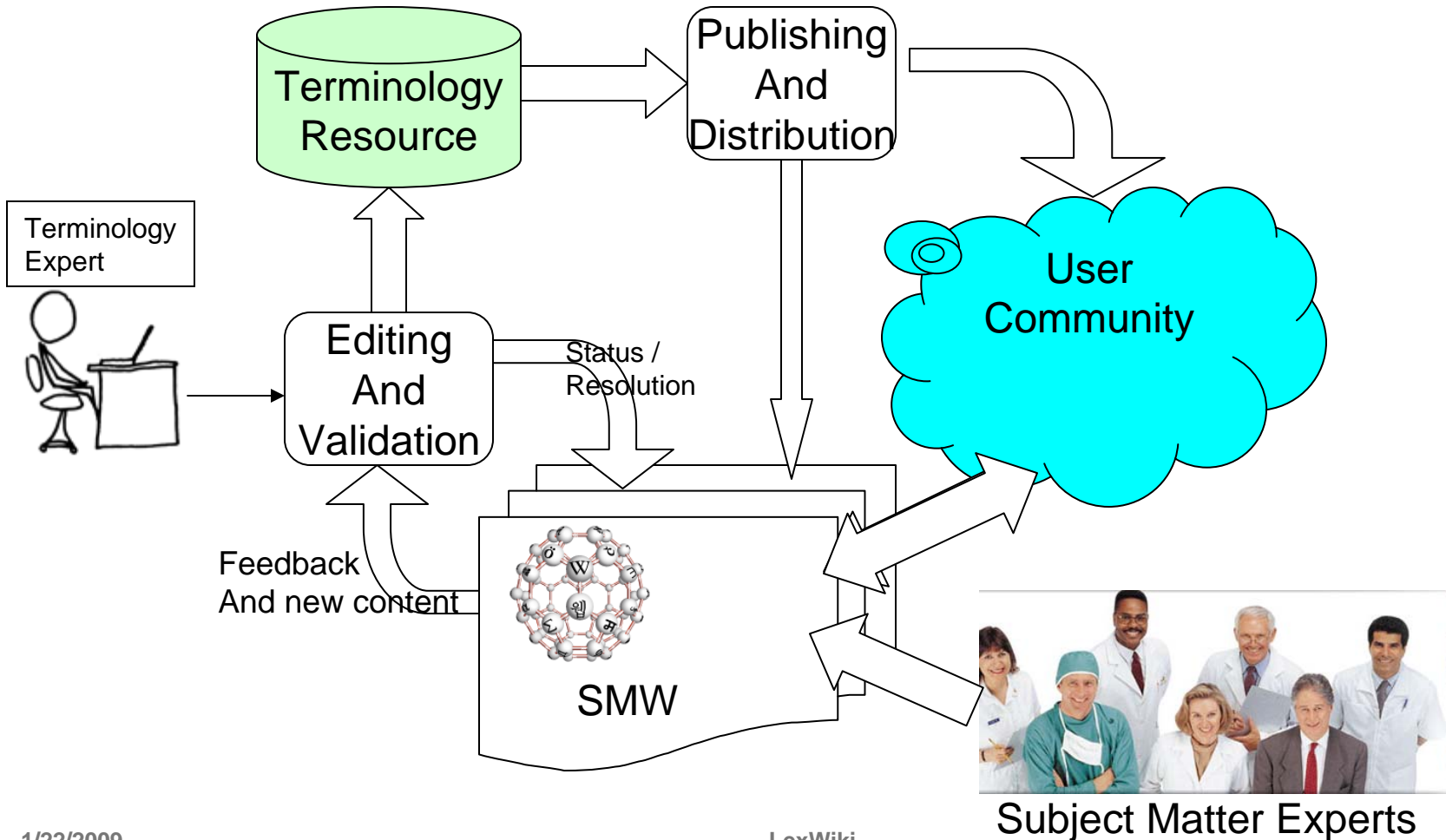
## The Situation (continued)

- **Both of these projects required the ability to reference and potentially integrate multiple terminological resources**
- **Neither of these projects had sufficient funding that a complete “stand-alone” solution would be practical... Collaboration was necessary.**
- **Both of these projects needed to expose content to a broad audience.**

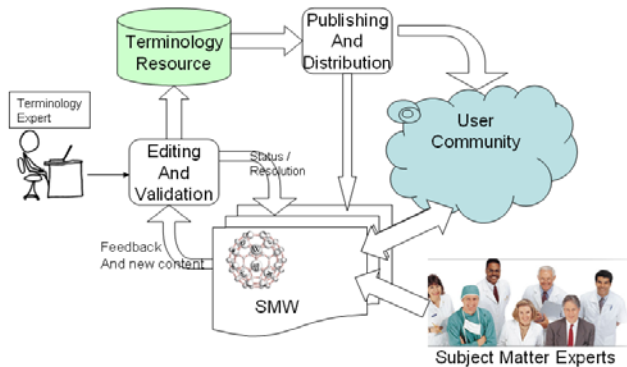
# Workflow Before SMW



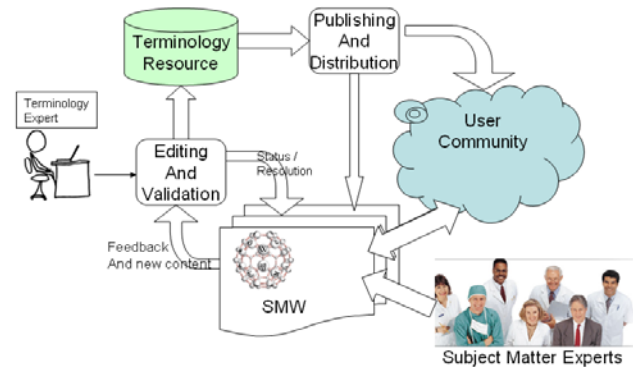
# Workflow with SMW



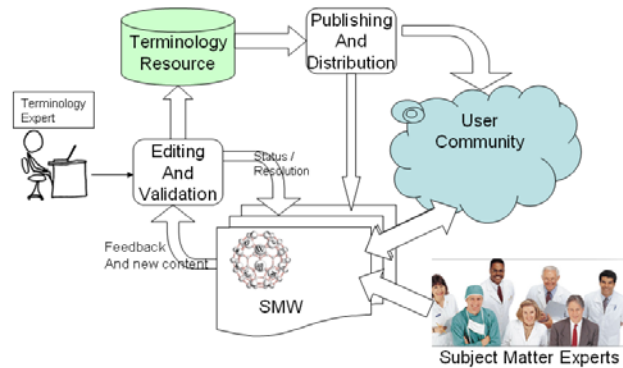
# Requirement: Reuse



BiomedGT

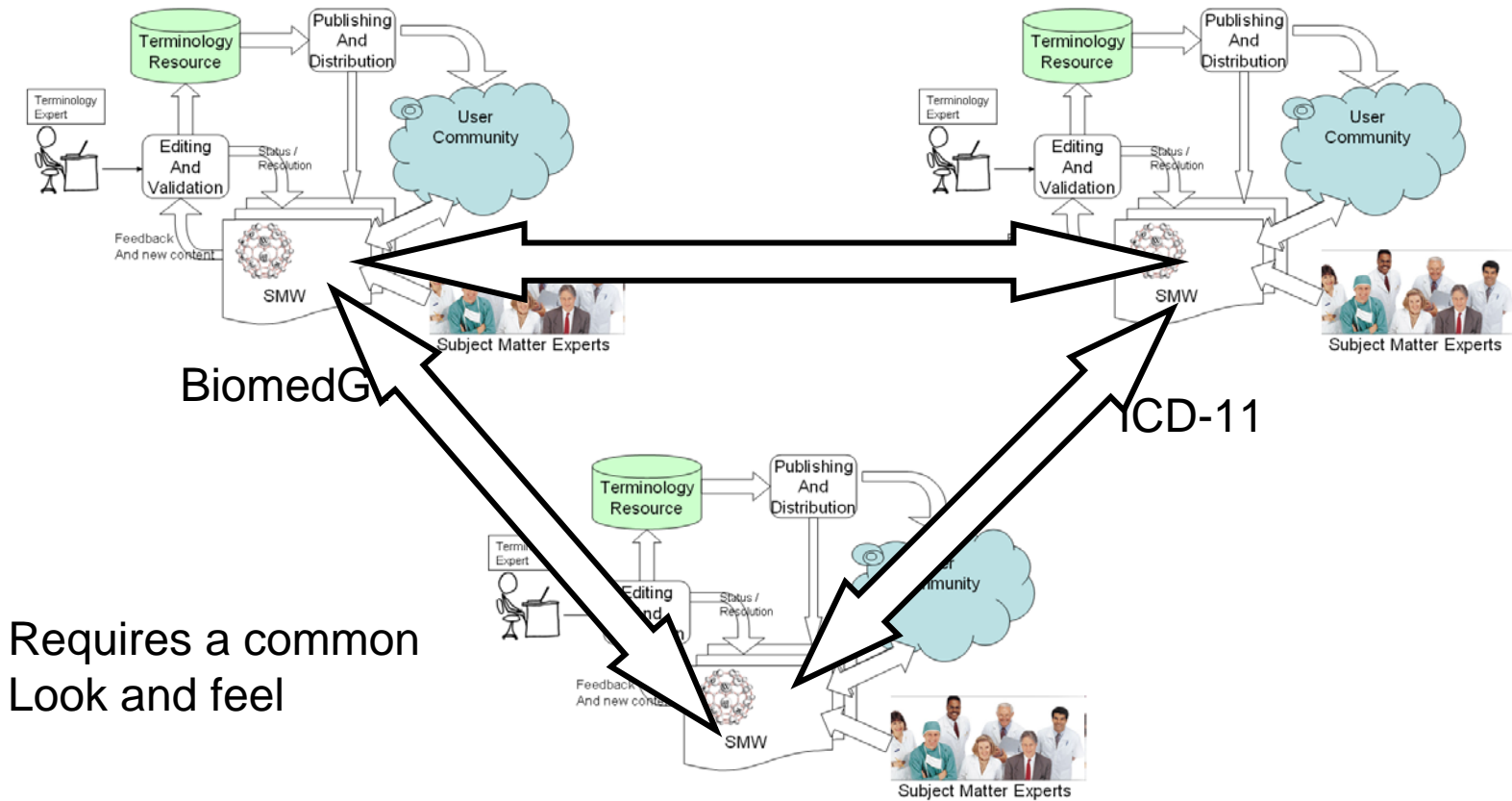


ICD-11



....

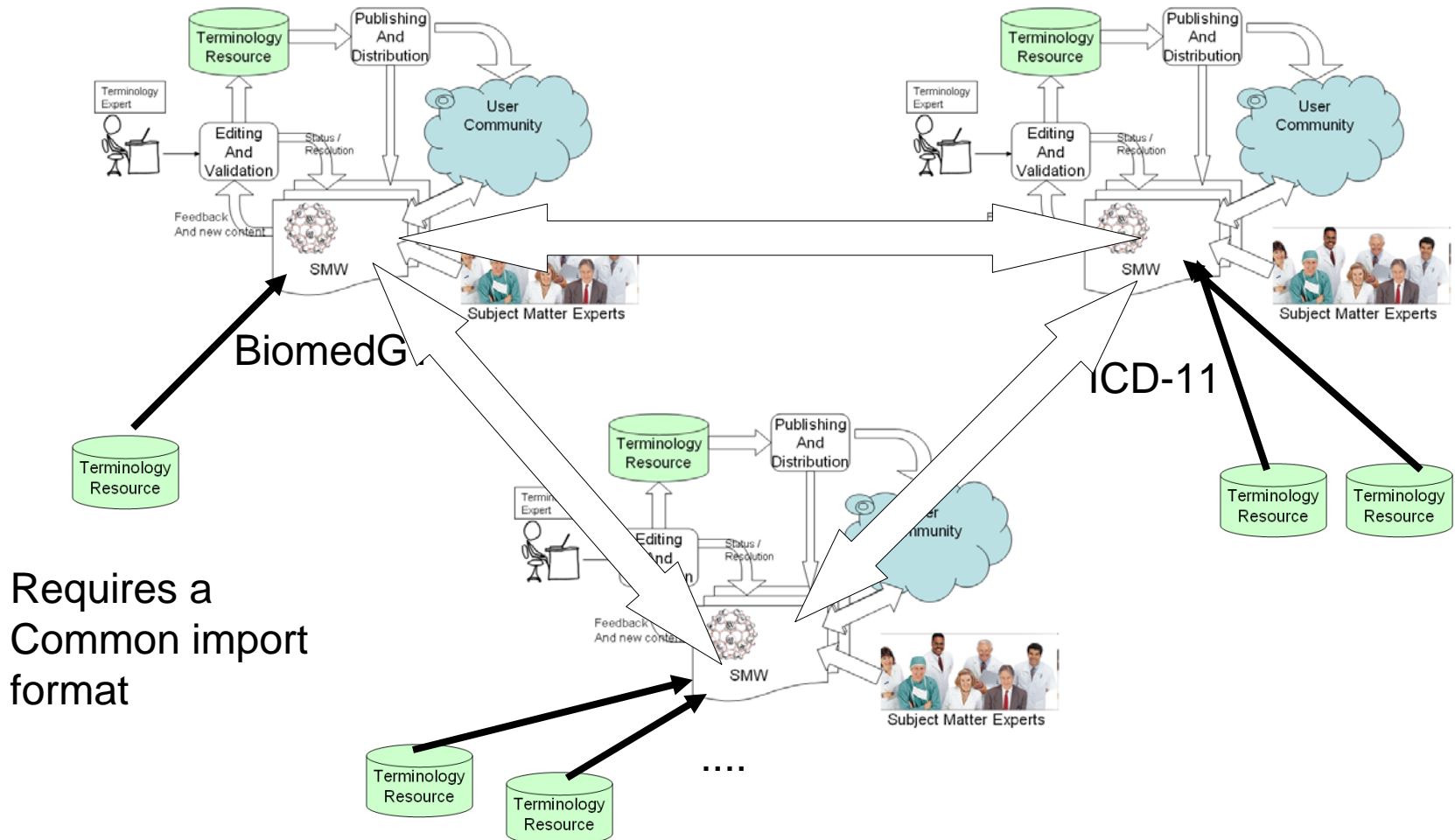
# Requirement: Inter-wiki References



Requires a common  
Look and feel

....

# Requirement: Multiple External Terminologies

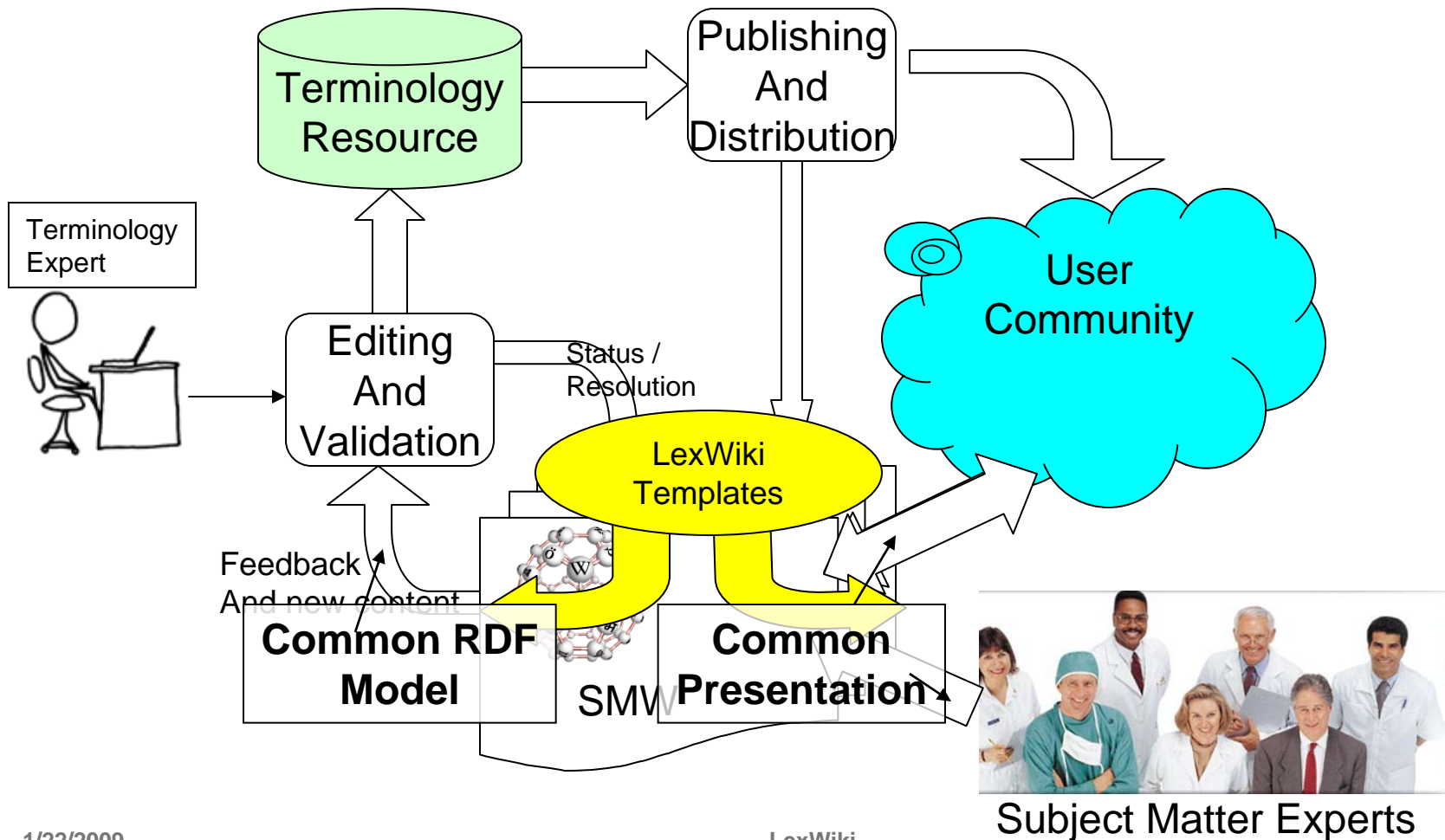




# Meeting the Requirements Common Input Format

- A set of MediaWiki templates that**
- A) Identify the semantics of the components (e.g. “Definition”, “Designation”, “Usage Note”, “Relationship”...)**
  - B) Create a common rendering**
  - C) Map to standard RDF model for output**

# Meeting the Requirements



# LexWiki Templates (Input into SMW)

```
{{MetaLexWiki_Title|1=BGT_Protein_Splicing_Abnormality(B45768)}}
{{LexWiki_Basic Data Header}}
{{LexWiki_Concept Code|1=B45768}}
{{LexWiki_BGT_Concept_Type|1=N}}
{{LexWiki_Preferred Name|1=Protein Splicing Abnormality}}
{{LexWiki_inScheme|1=BGT|2=01.01}}
{{LexWiki_Presentation|1=BGT_Synonym(P104)|2=Protein Splicing Abnormality|source=NCI|type=PT}}
{{LexWiki_Definition|1=A change in the amino acid sequence of a wild-type protein, resulting from aberrant excision of in-frame protein sequences (inteins) from a precursor protein followed by ligation of the flanking sequences (exteins).|source=NCI}}
{{LexWiki_editorialNote|1=Definition revised by CHK in 07/05.}}
{{LexWiki_URI|1=urn:oid:2.16.840.1.113883.3.26.1.2:B45768}}
{{LexWiki_Basic Data Trailer}}
{{LexWiki_Concept Property Header}}
{{LexWiki_Concept Property|1=BGT_NCI_META_CUI(P208)|2=CL335100}}
{{LexWiki_Concept Property|1=BGT_Semantic_Type(P106)|2=Cell or Molecular Dysfunction}}
{{LexWiki_Concept Property Trailer}}
{{LexWiki_Association Header}}
{{LexWiki_Parent|1=BGT_Protein_Sequence_Abnormality(B45767)|view=Defined|def=partial}}
{{LexWiki_Association Trailer}}
<noinclude>[[has default form::Form:LexWiki BGT Form]]</noinclude>
```

# Common Presentation



search

navigation

- [Main Page](#)
- [Current events](#)
- [Recent changes](#)
- [TODO List](#)
- [Help](#)

terminologies

- [BiomedGT](#)
- [CTCAE Subject Terms](#)
- [NanoTech Domain](#)
- [CRCH Nutrition Terms](#)
- [Biospecimen](#)
- [CTCAE\\_Category](#)
- [CTCAE\\_SOC \(Draft 1.0\)](#)

content search

- [Query ConceptCode](#)
- [Query Preferred Name](#)

workflow

- [Create / Update Package](#)
- [Bug Reporting](#)

[view source](#)

Category: [BGT Protein Splicing Abnormality\(B45768\)](#)

[BGT\\_Protein\\_Splicing\\_Abnormality\(B45768\)](#)

## Lexical

**Concept Code:** B45768 (N)

**Preferred Name:** Protein Splicing Abnormality

**Coding Scheme:** BGT (01.01 )

**Synonym:** [Protein Splicing Abnormality \(P7\)](#) (Source: NCI)

**Definition:** A change in the amino acid sequence of a wild-type protein, resulting from aberrant excision of in-frame protein sequences (inteins) (NCI)

**Editorial Note:** Definition revised by CHK in 07/05.

**URI:** urn:oid:2.16.840.1.113883.3.26.1.2:B45768

## Properties

**NCI\_META\_CUI:** CL335100

**Semantic\_Type:** Cell or Molecular Dysfunction

## Associations

**Parent:** [Protein\\_Sequence\\_Abnormality \(primitive\)](#)

Form: [LexWiki BGT Form](#)

### Facts about [BGT Protein Splicing Abnormality\(B45768\)](#) ⓘ

<a href="#">BGT Concept Type(P365)</a>	N + ⓘ
<a href="#">BGT NCI META CUI(P208)</a>	CL335100 + ⓘ ⓘ
<a href="#">BGT Semantic Type(P106)</a>	Cell or Molecular Dysfunction + ⓘ ⓘ
<a href="#">BGT Synonym(P104)</a>	Protein Splicing Abnormality + ⓘ

# Common RDF Model (Output from SMW)

```
- <rdf:RDF>
  <!-- Ontology header -->
  - <owl:Ontology rdf:about="">
    <smw:exportDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2009-01-20T20:30:39-05:00</smw:exportDate>
    </owl:Ontology>
    <!-- exported page data -->
  - <owl:Class rdf:about="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/Category:BGT_Protein_Splicing_Abnormality-28B45768-29">
    <rdfs:label>BGT Protein Splicing Abnormality(B45768)</rdfs:label>
    <smw:hasArticle rdf:resource="http://biomedgt.nci.nih.gov/wiki/index.php/Category:BGT_Protein_Splicing_Abnormality(B45768)"/>
    <rdfs:isDefinedBy rdf:resource="http://biomedgt.nci.nih.gov/wiki/index.php/Special:ExportRDF/Category:BGT_Protein_Splicing_Abnormality(B45768)"/>
    <rdfs:subClassOf rdf:resource="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/Category:BGT_Protein_Sequence_Abnormality-28B45767-29"/>
    <property:BGT_Concept_Type-28P365-29 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">N</property:BGT_Concept_Type-28P365-29>
    <property:BGT_NCI_META_CUI-28P208-29 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">CL335100</property:BGT_NCI_META_CUI-28P208-29>
    <property:BGT_Semantic_Type-28P106-29 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Cell or Molecular Dysfunction</property:BGT_Semant
    <property:BGT_Synonym-28P104-29 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Protein Splicing Abnormality</property:BGT_Synonym-28P1
    <property:LexWiki_Concept_Code rdf:datatype="http://www.w3.org/2001/XMLSchema#string">B45768</property:LexWiki_Concept_Code>
    <property:LexWiki_Preferred_Name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Protein Splicing Abnormality</property:LexWiki_Preferred_
    <property:LexWiki_URI rdf:resource="urn:oid:2.16.840.1.113883.3.26.1.2:B45768"/>
  - <property:LexWiki_Definition rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    A change in the amino acid sequence of a wild-type protein, resulting from aberrant excision of in-frame protein sequences (inteins) from a precursor protein followed by
    </property:LexWiki_Definition>
    <property:SKOS_editorialNote rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Definition revised by CHK in 07/05.</property:SKOS_editorialNot
    <property:DCTerms_hasVersion rdf:resource="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/BGT-2FVersions-2F01.01"/>
    <property:Has_default_form rdf:resource="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/Form:LexWiki_BGT_Form"/>
    <property:SKOS_inScheme rdf:resource="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/BGT"/>
    </owl:Class>
    <!-- auxilliary definitions -->
  - <smw:Thing rdf:about="http://cbvapp-s1009.nci.nih.gov/wiki/index.php/Special:URIResolver/BGT">
    <rdfs:label>BGT</rdfs:label>
    <smw:hasArticle rdf:resource="http://biomedgt.nci.nih.gov/wiki/index.php/BGT"/>
```

# LexWiki and LexGrid

## LexGrid is a model of terminology:

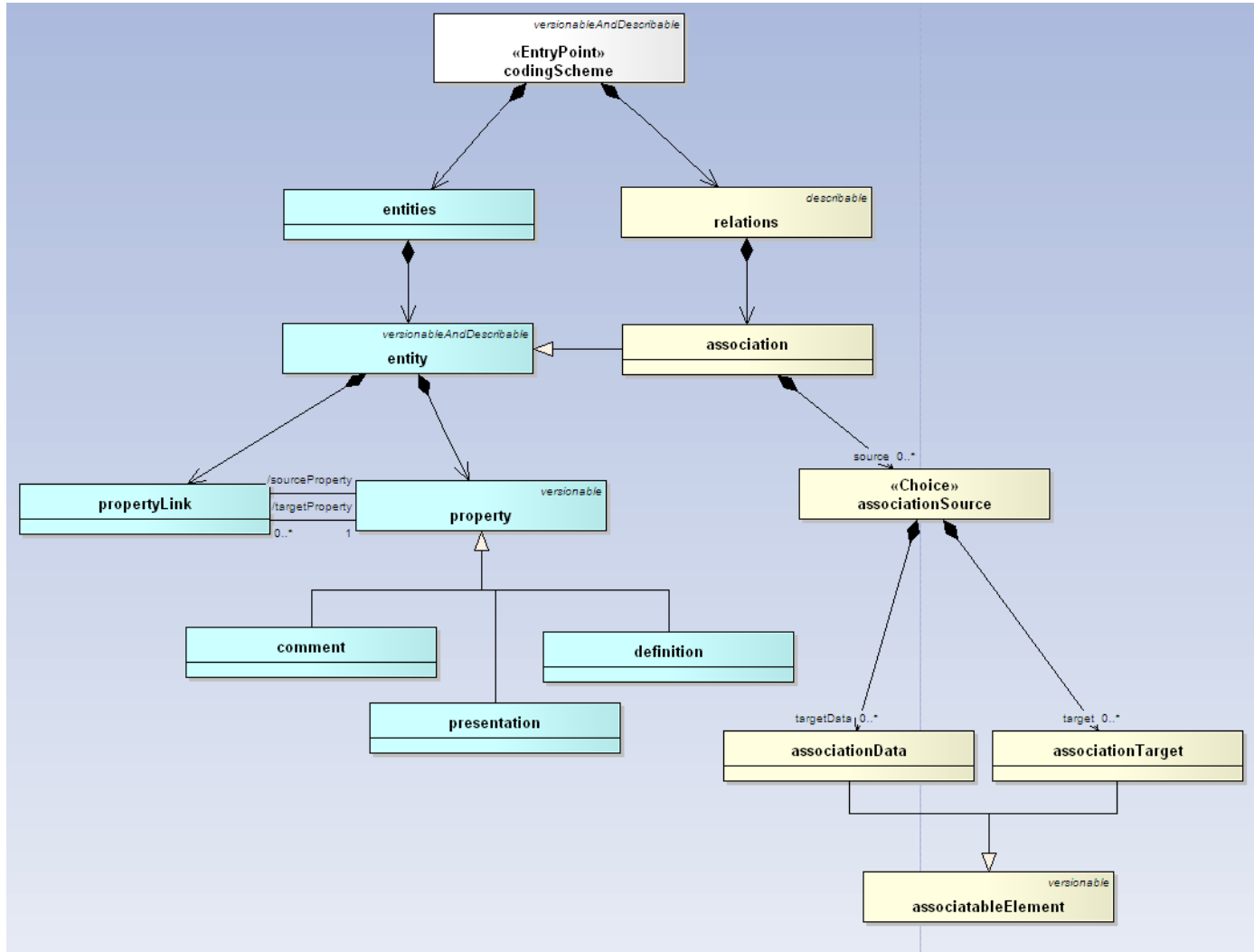
- **Designed to cover the spectrum from code/value tables to Thesauri to Ontologies**
- **Purpose includes:**
  - a) Standardization of the human/terminology interface**
  - b) Standard label set for terminological resources (e.g. Definition, Presentation, Comment, Language, ...)**

# LexGrid

## **A model of terminology:**

- **developed in the early 2000's**
- **Evolved from OMG's Lexicon Query Services Specification (1998)**
- **Compatible with the HL7 Common Terminology Services (CTS) specification**
- **Component of the NCI Cancer Bioinformatics Grid, NCVS CancerGrid, NCBO and others**

# LexGrid Model Overview



**Lexical  
Component**

**Logic  
Component**



# The LexGrid Model

## Available as:

- XML Schema
- UML Model
- Castor / EMF Binding

## Includes an import and export toolkit for:

- RDF
- OWL
- UMLS
- CSV

# LexWiki

- **A mapping of the LexGrid elements into MediaWiki Templates**
- **Tools to transform LexGrid content into template rendering**
- **Page naming tools**
- **Support infrastructure**

# LexWiki

## Page Naming

**Emerging terminology resources have been encouraging the use of “non-semantic” identifiers**

- **If the identifier consists of words...**
  - **... if you don't name it correctly at the start, how do you change it without breaking the link?**
  - **... if you speak the language, it is non-semantic anyway.**

# LexWiki

## Page Naming

- **If the identifier represents a nodes position in a hierarchy ...**
  - **... how do you change the position and still maintain the meaning?**
  - **... what do you do when you run out of “slots”?**

# LexWiki Page Naming

**The problem with non-semantic identifiers, however, is that they have no meaning (duh!)...**

**... and web (and Wiki!) links just plain don't work without some embedded semantics.**

# Page Naming Example

## “Mulberry Tree”


- Code as *Category:Mulberry* ?
  - Whoops – forgot about Mulberry Bush... *Category:Mulberry\_Tree*?
  - But, to an expert...  
*Category:Morus...*
  - But, the expert is Chinese...  
*Category:桑属 ...*

# Page Naming Example (Continued)

- With all of these changes, what do I do if I've got an external reference (another Wiki, a data record, ...)?

play.species\_encounter: 0 records total

	Location	Species
<input type="checkbox"/> *	Rochester	Mulberry



UVOL Morus(c 4948) - AgroWiki - Mozilla Firefox

History Bookmarks Tools Help

<http://wiktolog.com/agrowiki/index.php?title=Category:Morus>

# LexWiki Name

**Consists of**

- 1. A “non-semantic” identifier (when available)...**
- 2. The current, language specific, preferred designation... (mangled where necessary...)**
- 3. A *namespace* identifier in which the non-semantic identifier is rendered unique**



# LexWiki Name

**Format:**

**<Namespace>\_<name>( <nsid>)**

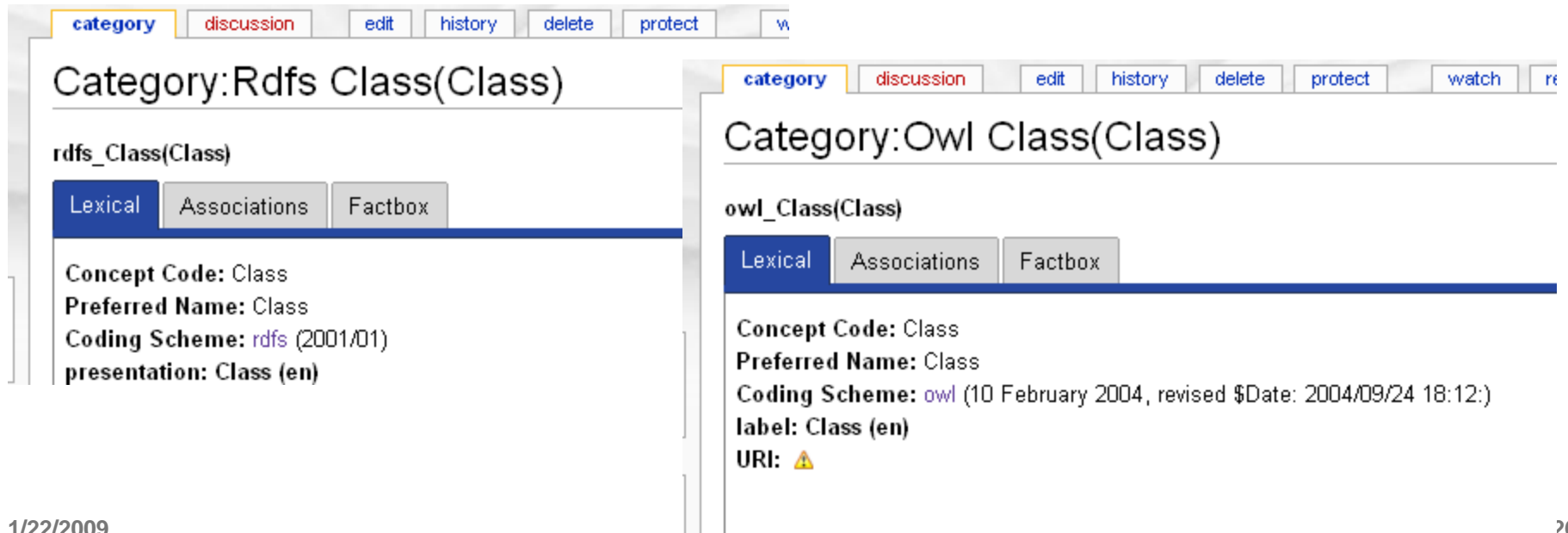
**Example:**

**AGROVOC\_Morus(c 4948)**

# LexWiki Name

The advantage of this structure is:

- 1) You can have multiple terminology resources in the same wiki w/o possibility of name clash...



The image shows two side-by-side screenshots of a LexWiki interface. The left screenshot displays the page for 'Category:Rdfs Class(Class)'. It has a navigation bar with 'category', 'discussion', 'edit', 'history', 'delete', and 'protect'. Below the title, there are tabs for 'Lexical', 'Associations', and 'Factbox'. The 'Lexical' tab is active, showing the following information: 'Concept Code: Class', 'Preferred Name: Class', 'Coding Scheme: rdfs (2001/01)', and 'presentation: Class (en)'. The right screenshot displays the page for 'Category:Owl Class(Class)'. It has a similar navigation bar with 'category', 'discussion', 'edit', 'history', 'delete', 'protect', 'watch', and 'revert'. Below the title, there are tabs for 'Lexical', 'Associations', and 'Factbox'. The 'Lexical' tab is active, showing: 'Concept Code: Class', 'Preferred Name: Class', 'Coding Scheme: owl (10 February 2004, revised \$Date: 2004/09/24 18:12:)', 'label: Class (en)', and 'URI: ⚠'.

# LexWiki Name

- **Links can survive name changes**
  - **If AGROVOC\_Mulberry(c\_4948) is not found...**
  - **... look up c\_4948 in AGROVOC tables...**
  - **... and, if found, initiate redirect process...**

You have followed a link to a page that does not exist yet.

To create the page, start typing in the box below (see the [help page](#) for more info).

If you are here by mistake, click your browser's **back** button.



The latest finding included several species of [[[:Category:AGROVOC\_Mulberry(c\_4948)]]], which ...

# LexWiki

## Status:

- **In use in BiomedGT and ICD11**
- **RDF, RDFS, SKOS, DC, OWL, ...  
Renderings have been prototyped**
- **Still a work in process...**

# Links

- **LexWiki**

<https://cabig-kc.nci.nih.gov/Vocab/KC/> (bottom of page)

- **LexGrid**

<http://LexGrid.org>

- **BiomedGT**

<http://biomedgt.nci.nih.gov/>

## **Agrovoc / RDF**

<http://wiktolog.com/agrowiki> (out of date...)