RELISH (Rendering Endangered Languages Lexicons Interoperable through Standards Harmonization)

Justin Petro

The LINGUIST List

Overview

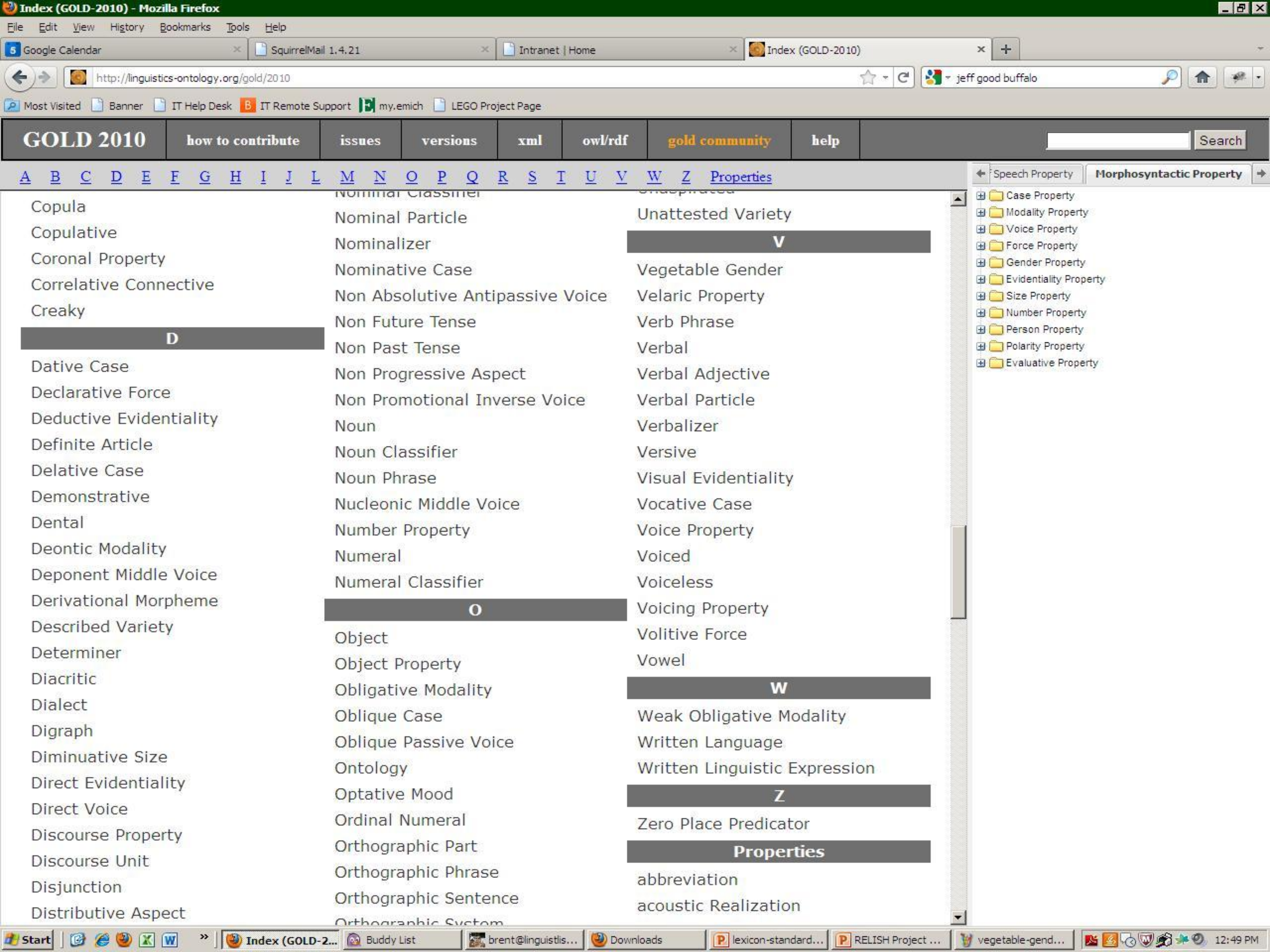
- Differences in Digital Standards for Lexicon Creation
  - U.S. (LEGO Project)
    - GOLD
    - LIFT/LEGO
  - Europe (LEXUS)
    - ISOcat
    - LMF
- Scope and Goals of RELISH Project
  - How RELISH harmonizes these standards
  - Rendering six lexicons into RELISH format (as proof of concept)
  - Ultimate goal of the project

Part 1: Diverging Standards

GOLD, ISOcat, LIFT, LMF

Introduction to GOLD

- General Ontology for Linguistic Description
- Formalized account of generally agreed upon linguistic concepts – “codifies general knowledge of the field”
- As an ontology, it allows complex relationships among concepts, although the only relationship currently instantiated is the hierarchical one of parent-child
- Can be used to promote interoperability and searching amongst disparate data sets (as seen in LEGO presentation)
- GOLD Website


Index (GOLD-2010) - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
Google Calendar
SquirrelMail 1.4.21
Intranet | Home
Index (GOLD-2010)
http://linguistics-ontology.org/gold/2010
jeff good buffalo
Most Visited
Banner
IT Help Desk
IT Remote Support
my.emich
LEGO Project Page
GOLD 2010
how to contribute
issues
versions
xml
owl/rdf
gold community
help
Search
A B C D E F G H I J L M N O P Q R S T U V W Z Properties
Speech Property
Morphosyntactic Property

Copula
Copulative
Coronal Property
Correlative Connective
Creaky

D

Dative Case
Declarative Force
Deductive Evidentiality
Definite Article
Delative Case
Demonstrative
Dental
Deontic Modality
Deponent Middle Voice
Derivational Morpheme
Described Variety
Determiner
Diacritic
Dialect
Digraph
Diminuative Size
Direct Evidentiality
Direct Voice
Discourse Property
Discourse Unit
Disjunction
Distributive Aspect

Nominal Classifier
Nominal Particle
Nominalizer
Nominative Case
Non Absolutive Antipassive Voice
Non Future Tense
Non Past Tense
Non Progressive Aspect
Non Promotional Inverse Voice
Noun
Noun Classifier
Noun Phrase
Nucleonic Middle Voice
Number Property
Numeral
Numeral Classifier

O

Object
Object Property
Obligative Modality
Oblique Case
Oblique Passive Voice
Ontology
Optative Mood
Ordinal Numeral
Orthographic Part
Orthographic Phrase
Orthographic Sentence
Orthographic System

Unaspirated
Unattested Variety

V

Vegetable Gender
Velaric Property
Verb Phrase
Verbal
Verbal Adjective
Verbal Particle
Verbalizer
Versive
Visual Evidentiality
Vocative Case
Voice Property
Voiced
Voiceless
Voicing Property
Volitive Force
Vowel

W

Weak Obligative Modality
Written Language
Written Linguistic Expression

Z

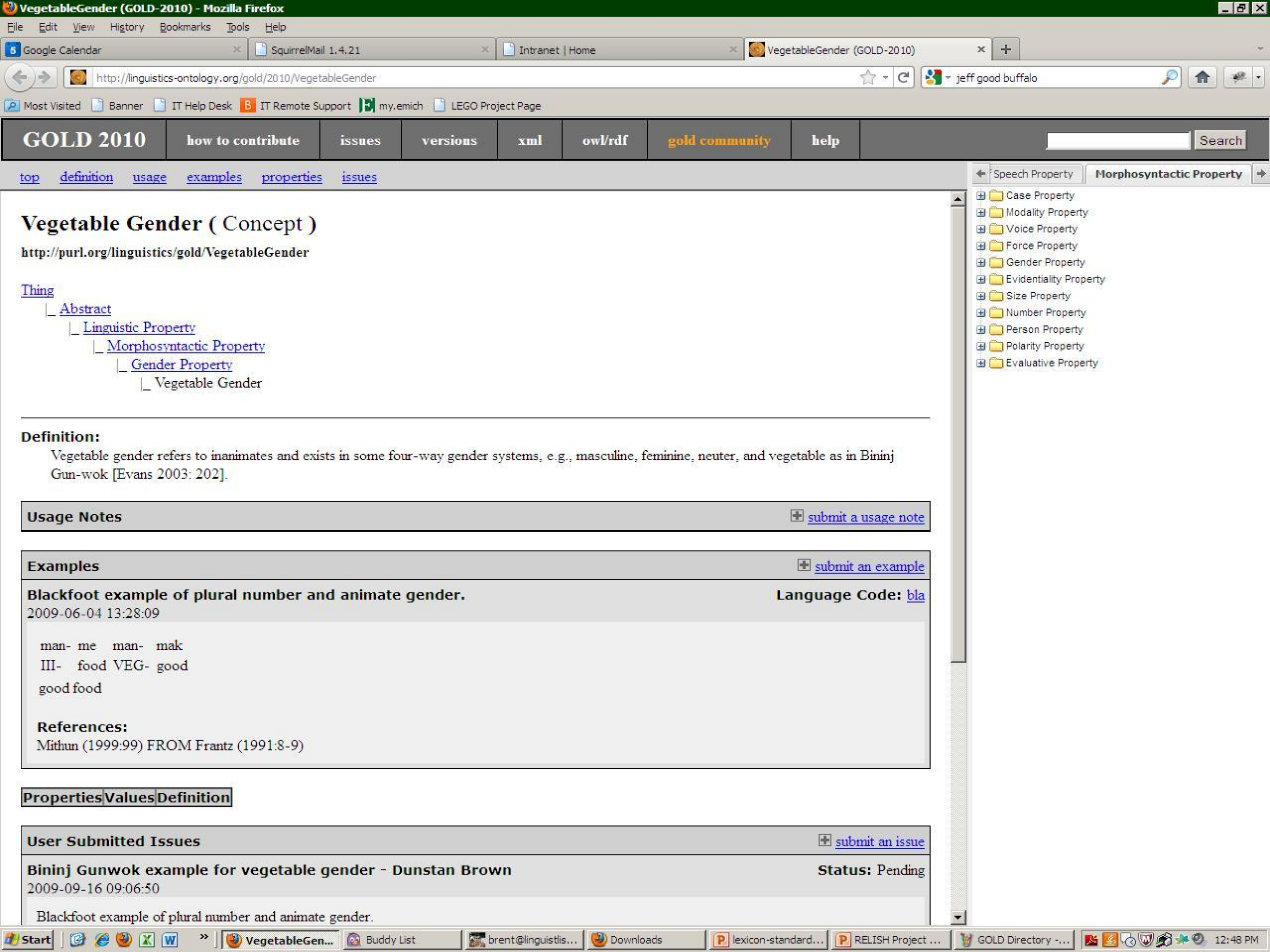
Zero Place Predicator

Properties

abbreviation
acoustic Realization

- Case Property
- Modality Property
- Voice Property
- Force Property
- Gender Property
- Evidentiality Property
- Size Property
- Number Property
- Person Property
- Polarity Property
- Evaluative Property

Start
Index (GOLD-2...
Buddy List
brent@linguistlis...
Downloads
lexicon-standard...
RELISH Project ...
vegetable-gend...
12:49 PM


VegetableGender (GOLD-2010) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Google Calendar
SquirrelMail 1.4.21
Intranet | Home
VegetableGender (GOLD-2010)
http://linguistics-ontology.org/gold/2010/VegetableGender
jeff good buffalo
Most Visited
Banner
IT Help Desk
IT Remote Support
my.emich
LEGO Project Page
GOLD 2010
how to contribute
issues
versions
xml
owl/rdf
gold community
help
Search
top
definition
usage
examples
properties
issues
Speech Property
Morphosyntactic Property
Case Property
Modality Property
Voice Property
Force Property
Gender Property
Evidentiality Property
Size Property
Number Property
Person Property
Polarity Property
Evaluative Property
Start
VegetableGen...
Buddy List
brent@linguistlis...
Downloads
lexicon-standard...
RELISH Project ...
GOLD Directory -...
12:48 PM

Vegetable Gender (Concept)

http://purl.org/linguistics/gold/VegetableGender

Thing
- Abstract
  - Linguistic Property
    - Morphosyntactic Property
      - Gender Property
        - Vegetable Gender

Definition:

Vegetable gender refers to inanimates and exists in some four-way gender systems, e.g., masculine, feminine, neuter, and vegetable as in Bininj Gun-wok [Evans 2003: 202].

Usage Notes

submit a usage note

Examples

submit an example

Blackfoot example of plural number and animate gender. **Language Code:** bla

2009-06-04 13:28:09

man-	me	man-	mak
III-	food	VEG-	good

good food

References:

Mithun (1999:99) FROM Frantz (1991:8-9)

Properties	Values	Definition

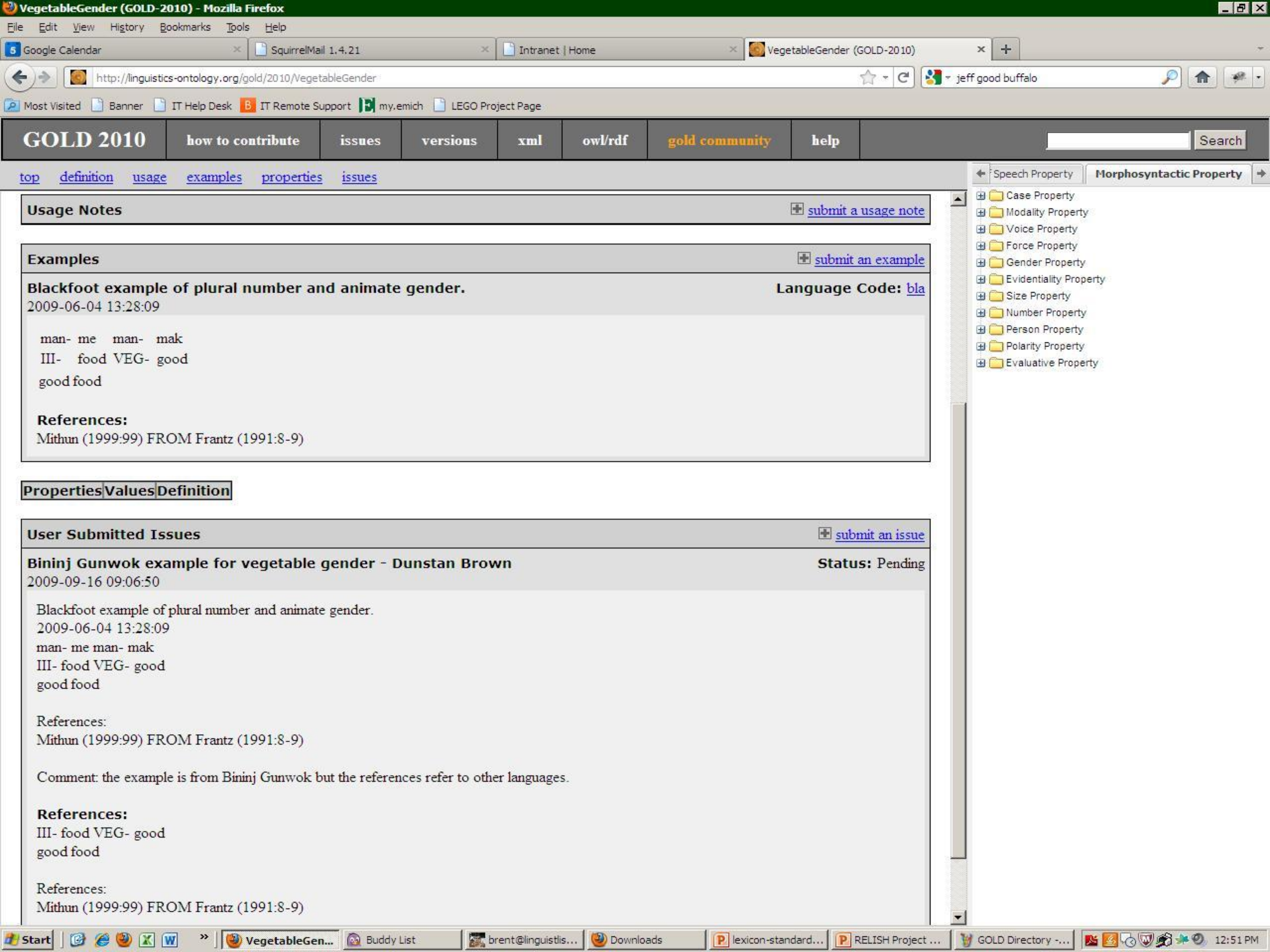
User Submitted Issues

submit an issue

Bininj Gunwok example for vegetable gender - Dunstan Brown **Status:** Pending

2009-09-16 09:06:50

Blackfoot example of plural number and animate gender.


VegetableGender (GOLD-2010) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Google Calendar
SquirrelMail 1.4.21
Intranet | Home
VegetableGender (GOLD-2010)
http://linguistics-ontology.org/gold/2010/VegetableGender
jeff good buffalo
Most Visited
Banner
IT Help Desk
IT Remote Support
my.emich
LEGO Project Page
GOLD 2010
how to contribute
issues
versions
xml
owl/rdf
gold community
help
Search
top
definition
usage
examples
properties
issues
Speech Property
Morphosyntactic Property
Case Property
Modality Property
Voice Property
Force Property
Gender Property
Evidentiality Property
Size Property
Number Property
Person Property
Polarity Property
Evaluative Property
Start
VegetableGen...
Buddy List
brent@linguistlis...
Downloads
lexicon-standard...
RELISH Project ...
GOLD Directory -...
12:51 PM

Usage Notes submit a usage note

Examples submit an example

Blackfoot example of plural number and animate gender. **Language Code:** bla

2009-06-04 13:28:09

man-	me	man-	mak
III-	food	VEG-	good

good food

References:

Mithun (1999:99) FROM Frantz (1991:8-9)

Properties Values Definition

User Submitted Issues submit an issue

Bininj Gunwok example for vegetable gender - Dunstan Brown **Status:** Pending

2009-09-16 09:06:50

Blackfoot example of plural number and animate gender.
2009-06-04 13:28:09
man- me man- mak
III- food VEG- good
good food

References:
Mithun (1999:99) FROM Frantz (1991:8-9)

Comment: the example is from Bininj Gunwok but the references refer to other languages.

References:
III- food VEG- good
good food

References:
Mithun (1999:99) FROM Frantz (1991:8-9)

ISO Data Category Registry (ISOcat)

- Developed by Max Planck Institute for Psycholinguistics, (Nijmegen, The Netherlands.)
- Terminology Registry for Linguistic Concepts, currently without hierarchy (work is being done to change this)
- Maps divergent data to a common term set: Specific terms can be associated with data element names using a namespace reference and a datcat PID
- Used in LMF-compliant digital lexicons

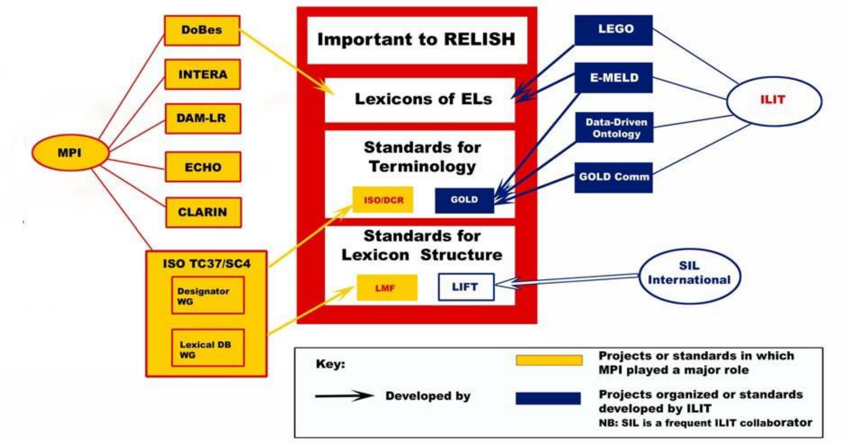
Digital Lexicon Standards

- LIFT (Lexicon Interchange FormaT)
  - Developed by SIL as a digital interchange format between lexicons created using different software tools
  - Schema restricted XML format
  - LL-LIFT: further restricted version of LIFT used by LEGO project
- LMF (Lexical Markup Framework)
  - ISO/24613:2008 standard for natural language processing (NLP) and machine-readable dictionary (MRD) lexicons.
  - Template for multilingual lexicon schemas
  - Used by LEXUS

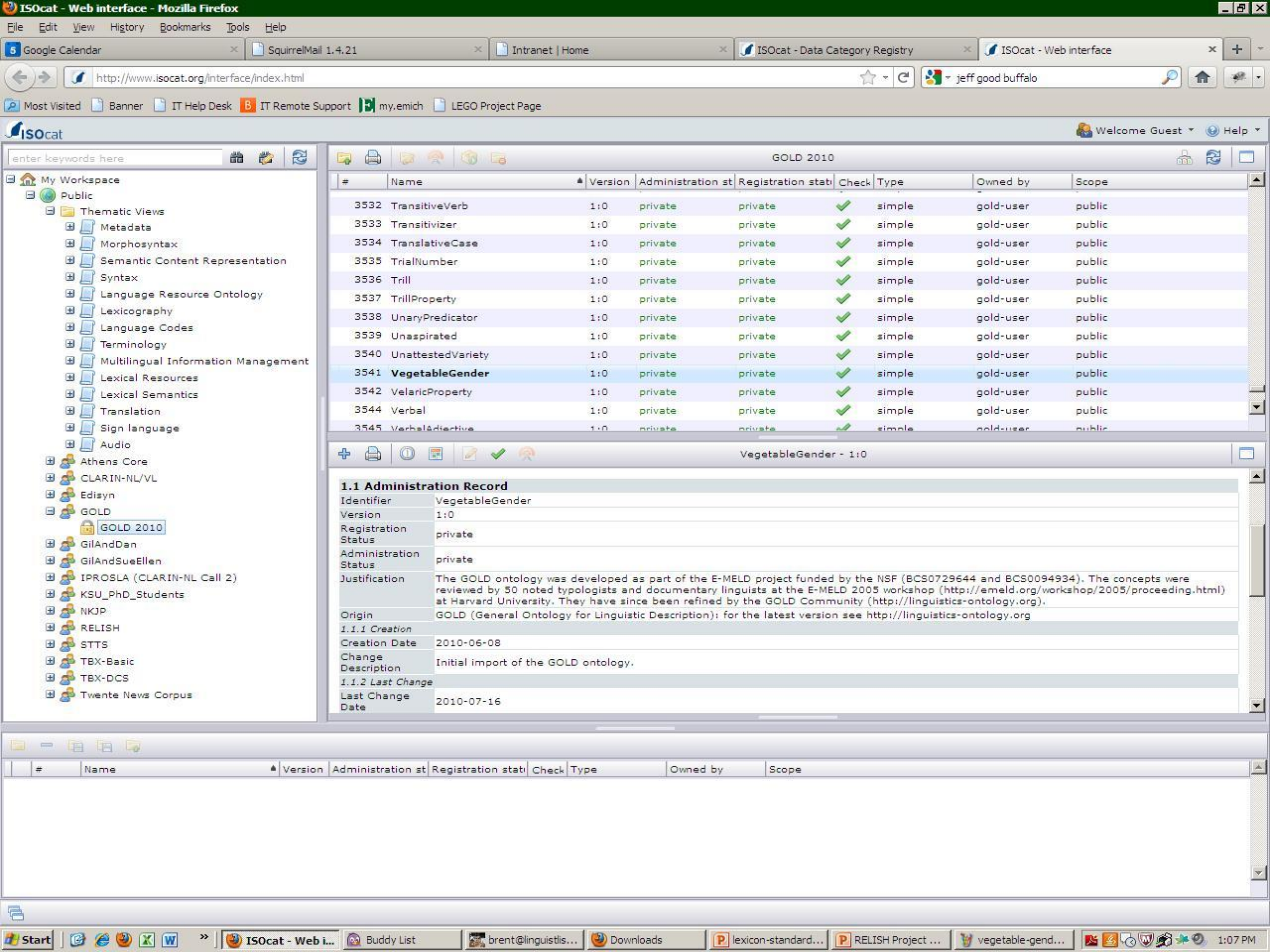
Part 2: The RELISH Project

Its Scope and Ultimate Goals

Scope of RELISH Project: Integration of Digital Standards


DoBes
INTERA
DAM-LR
MPI
ECHO
CLARIN
ISO TC37/SC4
Designator WG
Lexical DB WG
Important to RELISH
Lexicons of ELs
Standards for Terminology
ISO/DCR
GOLD
Standards for Lexicon Structure
LMF
LIFT
LEGO
E-MELD
Data-Driven Ontology
GOLD Comm
ILIT
SIL International
Key:
Developed by
Projects or standards in which MPI played a major role
Projects organized or standards developed by ILIT
NB: SIL is a frequent ILIT collaborator

- Integrate GOLD Concepts into ISOcat Registry
  - GOLD in ISOcat
- Create interchange format between lexicons output by LEXUS and those by LEGO
  - LMF serialization based on LIFT
  - LEGO and LEXUS will make RELISH output of lexicons available


ISOcat - Web interface - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Google Calendar
SquirrelMail 1.4.21
Intranet | Home
ISOcat - Data Category Registry
ISOcat - Web interface
http://www.isocat.org/interface/index.html
jeff good buffalo
Most Visited
Banner
IT Help Desk
IT Remote Support
my.emich
LEGO Project Page
ISOcat
Welcome Guest
Help
enter keywords here
My Workspace
Public
Thematic Views
Metadata
Morphosyntax
Semantic Content Representation
Syntax
Language Resource Ontology
Lexicography
Language Codes
Terminology
Multilingual Information Management
Lexical Resources
Lexical Semantics
Translation
Sign language
Audio
Athens Core
CLARIN-NL/VL
Edisyn
GOLD
GOLD 2010
GilAndDan
GilAndSueEllen
IPROSLA (CLARIN-NL Call 2)
KSU_PhD_Students
NKJP
RELISH
STTS
TBX-Basic
TBX-DCS
Twente News Corpus
GOLD 2010
VegetableGender - 1:0
Start
ISOcat - Web i...
Buddy List
brent@linguistlis...
Downloads
lexicon-standard...
RELISH Project ...
vegetable-gend...
1:07 PM

#	Name	Version	Administration st	Registration stat	Check	Type	Owned by	Scope
3532	TransitiveVerb	1:0	private	private	✓	simple	gold-user	public
3533	Transitivizer	1:0	private	private	✓	simple	gold-user	public
3534	TranslativeCase	1:0	private	private	✓	simple	gold-user	public
3535	TrialNumber	1:0	private	private	✓	simple	gold-user	public
3536	Trill	1:0	private	private	✓	simple	gold-user	public
3537	TrillProperty	1:0	private	private	✓	simple	gold-user	public
3538	UnaryPredicator	1:0	private	private	✓	simple	gold-user	public
3539	Unaspirated	1:0	private	private	✓	simple	gold-user	public
3540	UnattestedVariety	1:0	private	private	✓	simple	gold-user	public
3541	**VegetableGender**	1:0	private	private	✓	simple	gold-user	public
3542	VelaricProperty	1:0	private	private	✓	simple	gold-user	public
3544	Verbal	1:0	private	private	✓	simple	gold-user	public
3545	VerbalAdjective	1:0	private	private	✓	simple	gold-user	public

1.1 Administration Record

Identifier	VegetableGender
Version	1:0
Registration Status	private
Administration Status	private
Justification	The GOLD ontology was developed as part of the E-MELD project funded by the NSF (BCS0729644 and BCS0094934). The concepts were reviewed by 50 noted typologists and documentary linguists at the E-MELD 2005 workshop (http://emeld.org/workshop/2005/proceeding.html) at Harvard University. They have since been refined by the GOLD Community (http://linguistics-ontology.org).
Origin	GOLD (General Ontology for Linguistic Description): for the latest version see http://linguistics-ontology.org
1.1.1 Creation	
Creation Date	2010-06-08
Change Description	Initial import of the GOLD ontology.
1.1.2 Last Change	
Last Change Date	2010-07-16

#	Name	Version	Administration st	Registration stat	Check	Type	Owned by	Scope

RELISH Project Goals

- Grant Project Goal: Harmonize Standards
- Proof of Concept for this goal: Take 6 Lexicons and convert them into RELISH format
  - From LEGO: Mocovi, Western Sisaala, and Fulfulde
  - From LEXUS: Udi, Wichita, and Tuva
  - Demonstrates utility of RELISH format
- Ultimate Goal of Project:
  - Provide widest accessibility to lexicons of endangered languages
  - Make lexical data interoperable for not only linguists, but historians, anthropologists, geneticists, and other scientists interested in language
  - Participate in the trend to encourage data interoperability and integrated cyber-infrastructure.