

Language Technology WG Work In Progress Report

Agenda

- Remit
- Per CV and Schema Review
- Current CV and Schema Status
- WG Formation
- Timetable
- Issues

Remit: Vocabularies and Schemas

- Subject.Functionality
- Format
- Format.OS
- Format.CPU
- Format.Source-Code

- All are resource type independent

Subject.Functionality

- "the function of a resource" (doing-ness)
- LT domain focus
- Relocated from Type.Functionality
- CV and Schema
- 3 level CV hierarchy

Subject.Functionality Structure



Format

- "the physical or digital manifestation of a resource"
- eg. text/xml
- Domains beyond LT
- CV and Schema
- 2 level "Content Type" + "Subtype"
- Based on IETF RFC 1521/2077

Format.OS

- "the OS required to use a resource"
- eg. `unix/linux`
- LT domain focus
- CV and Schema
- re-usable by higher level ontologies

IRCS Workshop on Open Language Archives, 1/2/02

7

Format.CPU

- "the CPU required to utilise a resource"
- eg. `cpu/i386`
- CV and Schema
- reusable by higher level ontologies

IRCS Workshop on Open Language Archives, 1/2/02

8

Format.Source-Code

- "the programming language of a resource"
- eg. `sourcecode/perl`
- CV and Schema
- reusable by higher level ontologies

IRCS Workshop on Open Language Archives, 1/2/02

9

CV and Schema Status

CV/Schema	Status
Subject.Functionality	Editorial Process
Format	Under Development
Format.OS	Draft
Format.CPU	Draft
Format.Source-Code	Draft

IRCS Workshop on Open Language Archives, 1/2/02

10

WG Formation

- >3 Core Members
- OLAC WG web page
- WG document workspace
- LANGTECH Mailing List

IRCS Workshop on Open Language Archives, 1/2/02

11

Timetable ?

- Dec 2002: develop draft CV/Schema
- Jan 2003: review draft CV/Schema
- Feb 2003: proposals ?
- April 2003: recommendations ?

IRCS Workshop on Open Language Archives, 1/2/02

12

Issues

- Cross CV compatibility
 - Linguistic-Data-Type
 - Extensions for HLT Survey
- Orphan CV and Schema
 - Format.Encoding
 - Format.Markup
 - Format.Source-Status