

語料庫 Metadata 計畫第二次會議記錄

2001/9/3

- 會議地點：計算中心四樓會議室
- 出席人員：語料庫 - 羅鳳珠老師、張如瑩、盧秋容、余清華
Metadata 工作組 - 陳亞寧、陳淑君、張怡婷、翁翠霞
- 會議事由：討論語料庫 Metadata 計畫和 Metadata 工作組合作事宜
- 討論事項與共識：
 - 一、語料庫提出「Metadata 需求確認表單」和根據 OLAC 所填寫的元素需求表(參見「語言典藏 Metadata 標誌(OLAC)」)和根據 MICI 所填寫的元素需求表(參見「語言典藏 Metadata 標誌(CIMI)」)
 1. 語料庫認為 OLAC 無法提供語料庫所需要的欄位，是否可採用其他標準或在 OLAC 內自行增加其他欄位？
 2. Metadata 工作組對「元素需求表」裡欲著錄的最小單位提出疑問

共識：1. 為使語料庫 Metadata 加入 OLAC，現階段仍採用 OLAC 的標準。如以後發現其他標準更適合描述語料庫計畫，將作其他標準與 OLAC 的 mapping

 2. 請語料庫提供滿足語料庫描述需求的元素需求表，並附上釋例。可以根據工作組所提供過去主題計畫所作出的相關資料作建議或修改。
 3. 語料庫張如瑩小姐將一同參與日後 Metadata 工作組的小組會議。
 - 二、語言資源範圍的討論

語言資源可因類別、層次的不同，所需欄位即不同，採用的 Metadata 標準也就不同

共識：1. 語料庫先提供幾種類型的語料範本，例如先是文章、再是詩，供 Metadata 工作組分析

 2. Metadata 工作組根據這幾種類型的語料範本作著錄欄位分析。
 - 三、古漢語語料庫內羅鳳珠老師的計畫
 1. 羅老師的計畫第一階段以「詩」為範圍，選錄目標是蘇軾某一時期的詩，著錄的最小單位是「一首詩」，已有訂定「詩」的著錄標準。
 2. 第二階段將以「小說」為範圍，著錄目標是<<紅樓夢>>裡的「韻文」，包含「詩、詞、

- 曲、賦、樂府」等類型，詩之下可分成「絕句、律詩」，絕句之下可分為「五絕、七絕」
3. 建議「人名權威檔」增加「及第年」欄位；要能表現人名和「官銜」的引申關係。
 4. 建議「地名權威檔」要能表現中國文學裡特有的地名和人名的引申關係，例如：韋應物又名「韋蘇州」。
 5. 建議建立「主題權威檔」。

共識：1. Metadata 工作組將根據之前羅老師計畫訂定的「詩」的著錄標準作分析評估

2. 羅老師的計畫可採用 Metadata 工作組已制定出的「人名權威檔」和「地名權威檔」為基礎，提出修改建議。

四、南島語語料庫的討論

1. 南島語的語料庫是以採用 IPA 輸入，而漢語語料庫是採用 Uni-code 輸入，在呈現方式上是否會產生問題？
2. 南島語語料庫現有兩種版本，Word 版和 XML 版，要提供給 Metadata 工作組哪一種版本的範例？

共識：請語料庫提供給 Metadata 工作組 Word 版和 XML 版兩種版本的語料範例。

五、請語料庫提供的資料

1. 語料庫的 Metadata 和 OLAC 彙整後的元素需求表與釋例
2. 羅老師計畫先前已制定出的「詩」的標準
3. 兩種版本(Word 和 XML)的南島語語料範例

六、請 Metadata 工作組提供的資料

1. 先前主題計畫所產生的 DTD 或相關資料的規格書
2. 人名權威檔

■ 結論：

1. 語料庫計畫現階段仍採用 OLAC 的標準。如以後發現其他標準更適合描述語料庫計畫，將作其他標準與 OLAC 的 mapping。
2. 語料庫先提供幾種類型的語料範本，例如先是文章、再是詩，供 Metadata 工作組作欄位分析。
3. Metadata 工作組將根據之前羅老師計畫訂定的「詩」的著錄標準作分析評估。
4. 請語料庫提供的資料：

- a. 語料庫的 Metadata 和 OLAC 彙整後的元素需求表與釋例
 - b. 羅老師計畫先前已制定出的「詩」的標準
 - c. 兩種版本(Word 和 XML)的南島語語料範例
5. 請 Metadata 工作組提供的資料：
- a. 先前主題計畫所產生的 DTD 或相關資料的規格書
 - b. 人名權威檔