Linked Data as Method for Supporting DH Research on the Cultural Resources of Chinese Wooden Slips and the Interpretation of Ancient Chinese Characters

Sophy Shu-Jiun CHEN, Lu-Yen LU
Academia Sinica Center for Digital Cultures, Taiwan (R.O.C.)

Content

- **Research Plan on the "Wooden Slips Character Dictionary" (WCD) and functions to support DH research**

- **LOD-based Data Model Design in WCD**

- **Application of IIIF APIs on WCD and IIIF-based Union Catalog of WCD**

- **Integration of IIIF Manifest Data with LOD-based Resources**

- **Closing Remarks**

Research Plan on the "Wooden Slips Character Dictionary" (WCD) and functions to support DH research

WCD Research Plan

- Initiated by the Institute of History and Philology, Academia Sinica and ASCDC from 2019.

- Focus: Establishing image-based digital humanity platform, a retrieval system to research, which is structured by IIIF standards and linked data, to study the Chinese ancient wooden slips and characters written on them.

- Digitized objects in WCD: More than 10,000 Wooden Slips excavated in 1930 in Kansu Province, China (甘肅省).

- The written characters are living cultural evidence on the military and legal systems, educational practices, economy, beliefs, and everyday life of military personnel and civilians in the Han dynasty (206 BC-AD 220).

**Wooden Slips
Item H01637**

Structural and Contextual Views on the WCD System

- IIIF-based system of linked data structured database.

- Content: 848 Objects of Wooden Slips, 1,046 IIIF-based images, 21,265 data entries on annotated characters

- Core functions in WCD platform:
  * Metadata Search by object and by character
  * Image Research on objects and characters
  * Image annotation by classified categories
  * Cross-database query for Chinese characters

簡牘字典　史語所藏居延漢簡資料庫　　　　　操作說明　｜　登入

簡牘檢索 ▼　請輸入簡號、品名、釋文、時代、遺址……　🔍　顯示常用關鍵字 進階檢索

簡牘檢索
以關鍵字對簡牘後設資料欄位進行檢索，供快捷的查詢簡牘資訊及圖像。
開始檢索

單字檢索
搜尋簡牘釋文中是否包含特定單字，供快捷的查詢釋文內容及單字圖像比對。
開始檢索

標註檢索
以關鍵字對簡牘標註的評論內容進行查詢。
開始檢索

中央研究院歷史語言研究所收藏豐富的漢代簡牘與共出文物，數量約1萬3千餘枚。其中以1930年出土的「居延漢簡」為代表，此外也收藏部份羅布淖爾漢簡、敦煌漢簡，及武威剌麻灣漢簡。這些珍貴的漢代文書呈現當時邊塞地區移民墾殖、烽燧、郵驛、屯田戍衛的情形，具有極高的史料價值，被譽為二十世紀初中國學術界四大發現之一。

「簡牘字典—史語所藏居延漢簡資料庫」收錄歷年來史語所漢代簡牘數位典藏成果及《居延漢簡》出版內容。以開放的角度協助使用者查詢簡牘資料、字形圖像，是一個能檢索字形、書寫特徵、形態，文字與文字關係、簡牘與簡牘連結的數位人文工具。

歡迎您透過簡牘上率真流動的書跡，感受兩千年前漢代邊塞軍民日常生活的溫度。

中央研究院 歷史語言研究所 Institute of History and Philology, Academia Sinica

Digital Cultures Academia Sinica Center for 中央研究院 數位文化中心

歷史文字資料庫統合檢索系統
史的文字データベース連携検索システム

WCD Website: https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/index.php

WCD: Information Retrieval

Object-based retrieval

Character-based retrieval

WCD: Image Research

Image comparison of the characters "Jia chu" (甲渠), name of a military base

- The IIIF-based image viewer can make comparisons of the characters' writing style and identify the possible common handwriting between related characters

WCD: Image Annotation

- Images of objects and characters can be annotated under different motivations and sub-types as an extension to IIIF APIs.

- The method of classified annotation in different sub-categories makes the annotated text more reusable for scholars' needs by saving object data or interpreting the form and meaning of the characters.

- The annotated information will be saved in IIIF manifest and could be reusable, exchangeable as linked data.

- Images can be researched by crowdsourcing.

Motivation types
(1) Tagging
(2) Describing
(3) Commenting
(4) Classifying

The character "Jia" (甲) can be annotated under different motivations such as tagging, describing, commenting, or classifying (Item H04737)
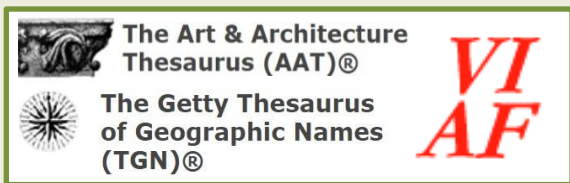
WCD: Image Annotation
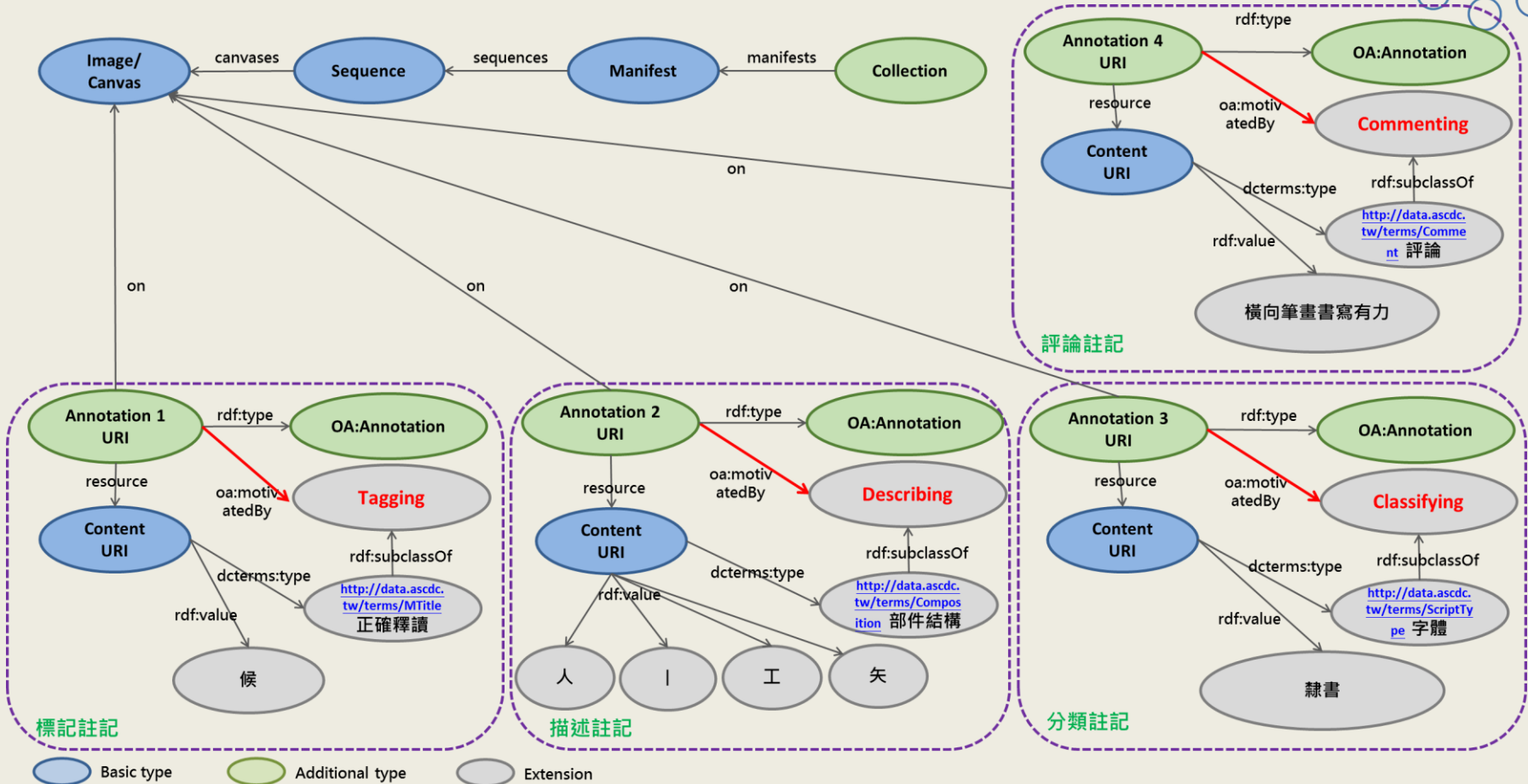
LOD-based Data Model Design in WCD

WCD Data Model

- **CIDOC-CRM based** WCD-data model: Describing information on the cultural object of the Chinese wooden slips.

- Core structures in WCD data model:
  * 23 classes in object-based and event-based model (CIDOC-CRM)
  * 47 properties from 9 semantic vocabularies (ascdc, bmo, CIDOC-CRM, dcterms, owl, rdf, rdfs, schema, skos)
 * 3 external resources (AAT, TGN, VIAF) reused for data enrichment

The Art & Architecture Thesaurus (AAT)®

The Getty Thesaurus of Geographic Names (TGN)®

VIAF

Classes [23]
crm:E4_Period , crm:E7_Activity, crm:E8_Acquisition, crm:E12_Production, crm:E14_Condiciton_Assessment, crm:E22_Man-Made_Object, crm:E31_Document, crm:E33_Linguistic_Object, crm:E34_Inscription, crm:E35_Title, crm:E38_Image, crm:E39_Actor, crm:E42_Identifier, crm:E52_Time-Span, crm:E53_Place, crm:E54_Dimension, crm:E55_Type, crm:E57_Material, crm:E58_Measurement_Unit, crm:E60_Number, crm:E62_String, crm:E73_Information_Object, crm:E79_Part_Addition

Properties [47]
ascdc:type, bmo:PX_inscription_position, crm:P1_is_identified_by, crm:P2_has_type, crm:P3_has_note, crm:P4_has_time-span, crm:P7_took_place_at, crm:P10_falls_within, crm:P14_carried_out_by, crm:P24i_changed_ownership_through, crm:P32_use_general_technique, crm:P33_use_specific_technique, crm:P34i_was_assessed_by, crm:P43_has_dimension, crm:P45_consists_of, crm:P48_has_preferred_identifier, crm:P50_has_current_keeper, crm:P52_has_current_owner, crm:P65_shows_visual_item, crm:P67_refers_to, crm:P70i_is_documented_in, crm:P90_has_value, crm:P91_has_unit, crm:P102_has_title, crm:P108i_was_produced_by, crm:P111i_was_added_by, crm:P115_finishes, crm:P116_starts, crm:P127_has_broader_term, crm:P128_carries, crm:P129i_is_subject_of, crm:P134i_was_continued_by, crm:P137_exemplifies, crm:P139_has_alternative_form, dcterms:identifier, dcterms:isPartOf, dcterms:type, owl:sameAs, rdf:type, rdfs:label, rdfs:seeAlso, schema:endDate, schema:identtfier, schema:startDate, skos:altLabel, skos:broadMatch, skos:related
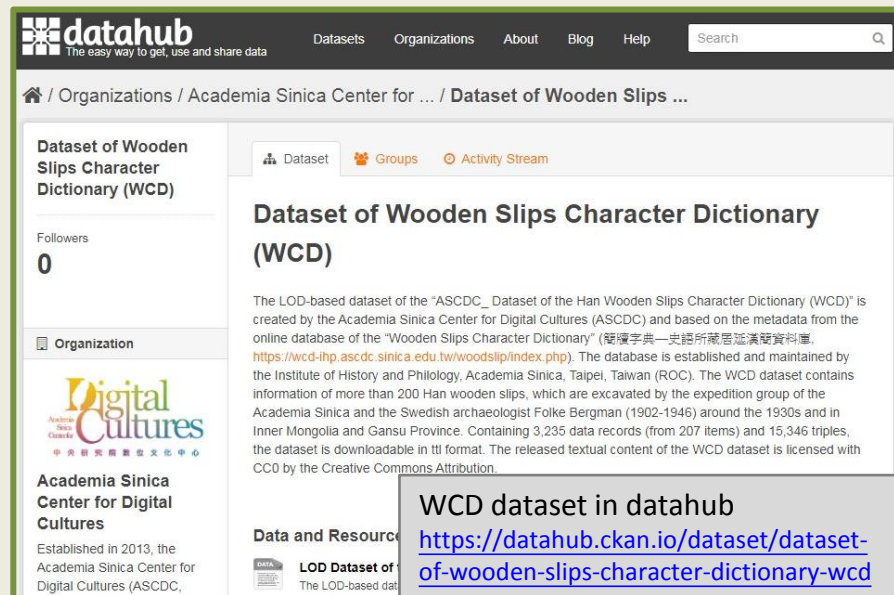
WCD Data Model

12
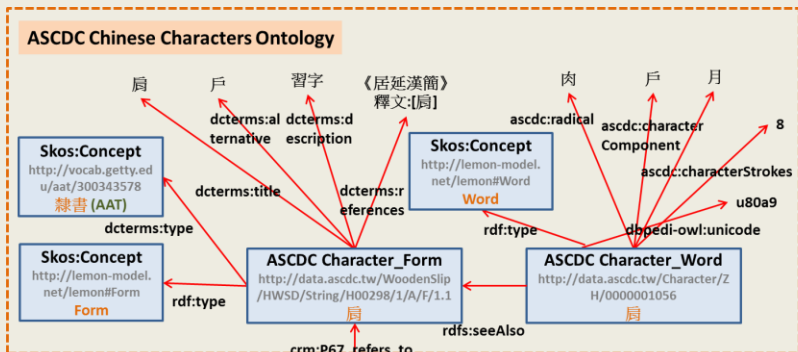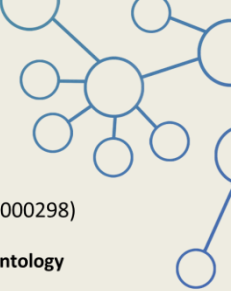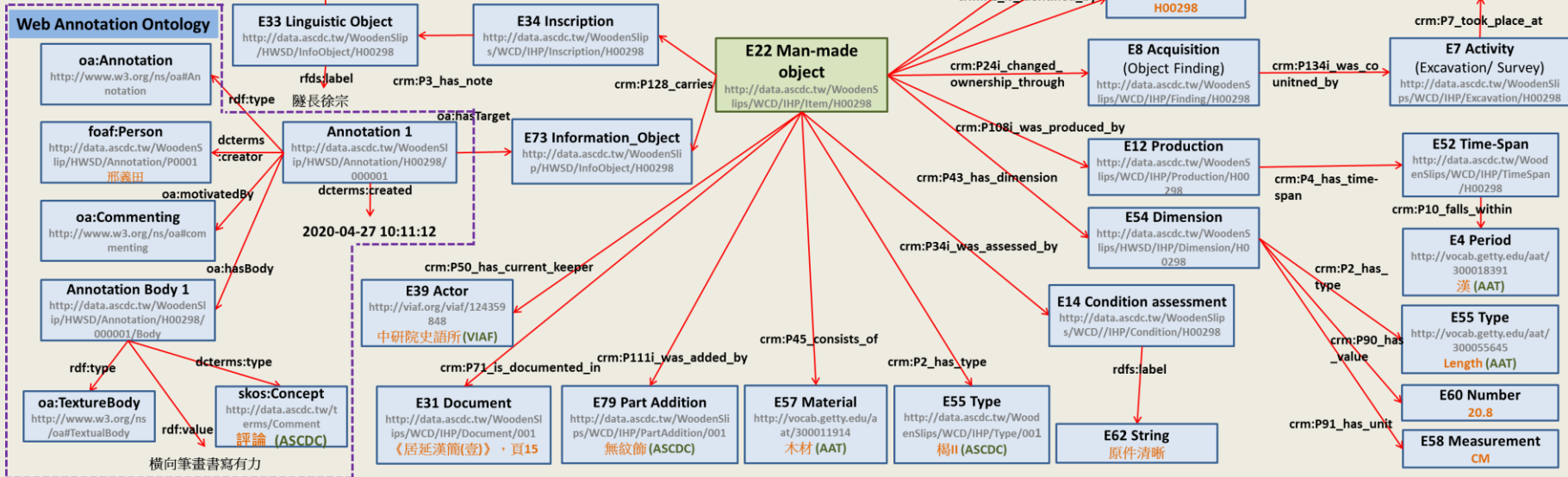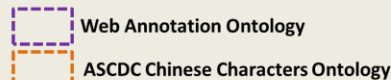
LOD Dataset of WCD

- **Based on CIDOC-CRM, metadata of 207 slips are converted into an LOD dataset with 15,346 triples. Released in Datahub.**

- **The concerted data contain information on the title, item number, registration number, excavation site, provenance, Creation data, Dynasty, Type, decoration, material, dimensions of work, references, related item, collection, and image information of a slip.**

WCD data record in RDF
Example: Item H00185

WCD dataset in datahub
https://datahub.ckan.io/dataset/dataset-of-wooden-slips-character-dictionary-wcd

Application of IIIF APIs on WCD and IIIF-based Union Catalog of WCD

Application of IIIF APIs on the WCD Platform

- **Image presentation in the WCD database functions based on the IIIF standard, especially in application of the IIIF Presentation API, IIIF Image API and IIIF Content Search API.**

	API type	Purpose of application	Application
1	**IIIF Presentation API**	Present an image or series of images in IIIF viewer by its correct image sequence, structure and layout based on the linked data and web page structure principles.	Image presentation, Image sharing, Structured retrieval system of WCD
2	**IIIF Image API**	Formulate an image's URI design, parameter specification, and attribution for data management to achieve static image retrieval and delivery.	Image zooming, Image comparison, Viewing form definition
3	**IIIF Content Search API**	Provides a content retrieval service mechanism to retrieve the annotation information embedded in the IIIF manifest JSON-LD file or information structured in the IIIF presentation API.	Character retrievals, Structured image annotation

IIIF-based Union Catalog of WCD

- Images of WCD characters are further interoperable and retrievable in the "**Union Catalog**" for searching historical Chinese Characters in cooperation with international research communities, such as the Nara National Research Institute for Cultural Properties, Historiographical Institute of the University of Tokyo, National Institute of Japanese Literature, National Institute for Japanese Language and Linguistics, and Institute for Research in Humanities at Kyoto University in Japan.

- Functions of WCD's Union Catalog
  * **Character retrieval across institutes**
  * Redirection to original database
  * Access to the IIIF Manifest structure of retrieved characters
  * Presentation of retrieved characters in Mirador viewer

- System functions based on **IIIF APIs** and **customized API for sharing of the search results**

WCD Union Catalog ：https://wcd-ihp.ascdc.sinica.edu.tw/union/

Design of Interoperable API for Character Retrievals

- In cooperation with 6 research institutes in Japan, an **API for sharing query results** was developed to enable query of a single character across different institute databases and enhance data interoperability.

- API is structured based on JSON format.

- 15 elements are regulated in the WCD-Union query API: status code, search results, list, identifier, id, **title**, delegate, unicode, source, thumbnail url, manifest url, subject, creator, rights, rights url.

```
(1) status_code: "200",
(2) search_results: "32",
(3) - list: [
     (4) identifer: "http://wcd-ihp.ascdc.sinica.edu.tw/api/GET/anno/mysql/body_55_215",
     (5) id: "H00185",
     (6) title: "廣",
     (7) delegate: "1",
     (8) unicode: "5ee3",
     (9) source: {
          general_name: "居延漢簡",
          site: "A27 查科爾帖",
          time_Chinese: "東漢",
          time_CE: "25 ~ 220",
          material: "木"
     },
     (10) thumbnail_url: "https://wcd-ihp.ascdc.sinica.edu.tw/parser/IIIF/H00185BIWA-6/9318,393,77,94/full/0/default.jpg",
     (11) manifest_url: "https://wcd-ihp.ascdc.sinica.edu.tw/api/GET/55/manifest",
     (12) subject: "簡牘字典—史語所藏居延漢簡資料庫",
     (13) creator: "中央研究院歷史語言研究所 | 中央研究院數位文化中心",
     (14) rights: "",
     (15) rights_url: "https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/terms.html"
     },
```

匯廣字典─中央研究院歷史語言研究所|中央研究院數位文化中心
查詢結果：32筆　　顯示全部圖檔

木匣座─奈良文化財研究所
查詢結果：65筆
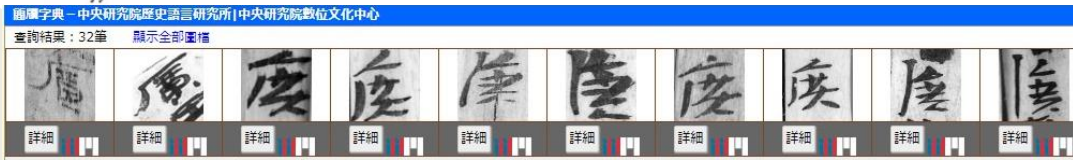
# Integration of IIIF Manifest Data with LOD-based Resources

Heterogeneous Data Linkage between LOD and IIIF

- **Using a linking property such as seeAlso regulated in IIIF Presentation API to link the LOD-based descriptive metadata of an object or a Chinese character with its image data embedded in the IIIF manifest and further enhance data distribution and reusability of WCD's LOD**
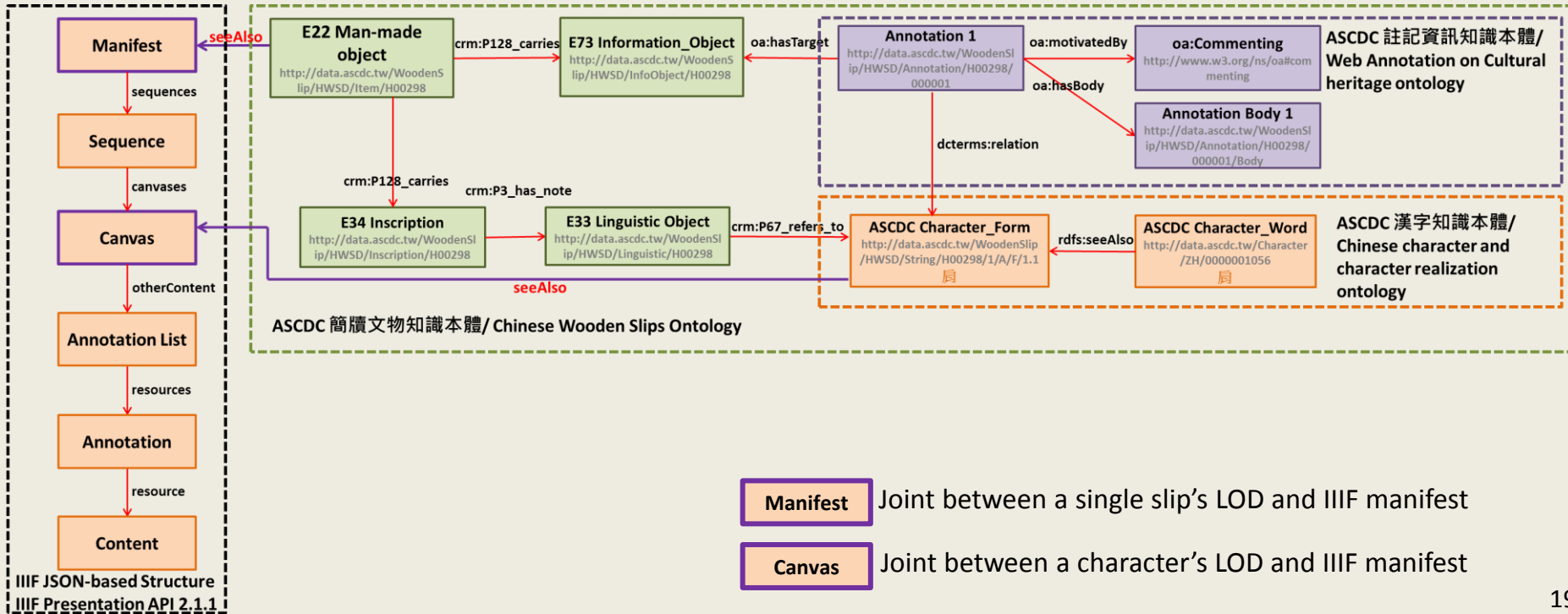
Data Linkage between LOD and IIIF: An Example


```
{
    @id: "https://wcd-ihp.ascdc.sinica.edu.tw/api/GET/43/manifest",
  - logo: {
        @id: "https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/web.png",
        @type: "dctypes:Image",
        width: 170,
        height: 90
    },
    @type: "sc:Manifest",
    label: "永光二年予候長鄭赦寧冊",
    license: "https://creativecommons.org/licenses/by-nc-nd/3.0/tw/",
    navDate: "2017-09-14T00:00:00Z".
  - seeAlso: {
        @id: "http://data.ascdc.tw/WoodenSlips/WCD/IHP/Item/H01637",
        format: "rdf"
    },
    @context: "http://iiif.io/api/presentation/2/context.json",
  - metadata: [
      - {
            label: "Published",
          - value: [
              - {
                    @value: "Institute of History and Philology, Academi
                    @language: "en"
                },
              - {
                    @value: "中央研究院歷史語言研究所|中央研究院數位文化中心
                    @language: "zh-TW"
                }
            ]
```

Manifest JSON-LD (H01637)

永光二年予候長鄭赦寧冊

Report on the bereavement leave of the officer Zheng She due to his father's death.

Title	居延漢簡
Registration ID	H01637
Item ID	057.001
Type	冊
Broader Type	簡牘及竹木器
Material	木/ Wood AAT
Date	西漢元帝永光二年
Place	A8 破城子
Item Owner	中央研究院歷史語言研究所/ IHP VIAF

Links to external resources (as AAT, VIAT)

LOD (H01637)

seeAlso as key property to integrate IIIF Manifest JSON-LD with LOD-based descriptive data

Closing Remarks

Closing Remarks

- **LOD is not only an effective method to make results of DH study more reusable, accessible, and enhance information retrieval across different datasets or institutes. Based on its essence as linked data, LOD can also integrate into other heterogeneous data structures or platforms such as IIIF.**

- **By using semantic properties, the LOD-based metadata can be connected with the data structure of the IIIF manifest. Thus, a model to support DH studies, both on their descriptive component and image component, is possible.**

Wooden Slips Item H00052

中央研究院數位文化中心
Academia Sinica Center for Digital Cultures

A   5F, Research Building, Institute of History and Philology

    No.128, Sec. 2, Academia Rd., Nangang Dist., Taipei City 11529, Taiwan (R.O.C.)

E   ascdc@sinica.edu.tw

P   +886-02-2652-1885

F   http://ascdc.sinica.edu.tw/en/

ASCDC

Q & A

THANK YOU !