

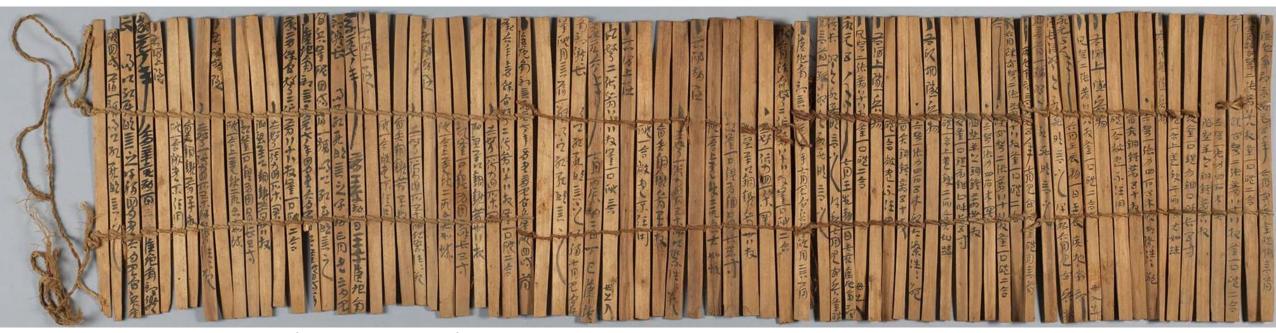
Sophy Shu-Jiun Chen, Lu-Yen LU
Academia Sinica Center for Digital Cultures (ASCDC)

December 02, 2022

SWIB 2022, the 14th Semantic Web in Libraries Conference, 28.11-02.12 2022

Research on the Juyan Han Wooden Slips

- The traditional writing medium in the Early Chinese Medieval period, especially in the Han Dynasty (206 BC-AD 220).
- Composed of woods or bamboos and binding together, the Chinese characters on the slips are living evidence to study the military and legal systems, educational practices, economy, beliefs, and everyday life of the society at that time.

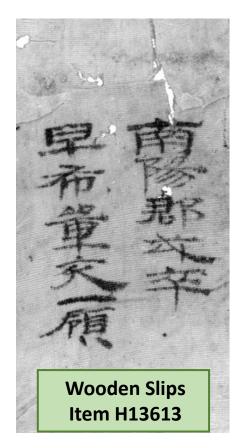


Monthly and Seasonal Records of Military Supplies from the Kuang-ti South Platoon in the Yung-yüan Era 廣地南部永元五年至七年官兵釜磑月言及四時簿, ca. AD 97, Item H00185

Research on the Juyan Han Wooden Slips

Around 1930s, more than 10,000 Han wooden slips, called as Juyan Han Wooden Slips (居延漢簡), were excavated by the Chinese and Swedish scholars of the Sino-Swedish Scientific Expedition, in the watershed of the Edsen-Gol (額濟納河) of the Kansu Province (甘肅省), China.







Wooden Slips Item H07851



Wooden Slips Item H02544



Wooden Slips Item H01637

Digital Archive of the Han Wooden Slips

 To preserved these treasurable archaeological objects, the Institute of History and Philology in Academia Sinica has established Digital Archive of Han Wooden Slips since the 1990s.



Interface of Digital Archive of Han Wooden Slips (Item H01637)
http://rub.ihp.sinica.edu.tw/~woodslip/index.htm

Challenges in traditional Database for Wooden Slips

Restriction of the traditional database for DH study of the Han Wooden Slips

	Problems of the traditional database system	Solution methods
1	Database focused on the information of an entire wooden slips and can not provide information on the characters written on them	Enhancing the granularity for information retrieval methods embedded in system
2	Database established as an isolated silo, can not connect with other related data resources	Constructing RDF-formatted Linked Data for cross-database query
3	Textual content and Image can not be mapped to each other on the interface	Applying IIIF APIs combined with Linked Data-based image annotation function

Wooden Slips Character Dictionary (WCD)

- Content: 7,377 Objects of Wooden Slips, 8,491 IIIF-based images, 65,736 data entries on annotated characters
- Core functions in WCD platform:
 - * Information retrieval by object and by character
 - * Image research on objects and characters
 - * Image annotation by classified categories
 - * Cross-database query for Chinese characters
- Structure of Metadata designed based on CIDOC-CRM in order to further apply as linked data



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

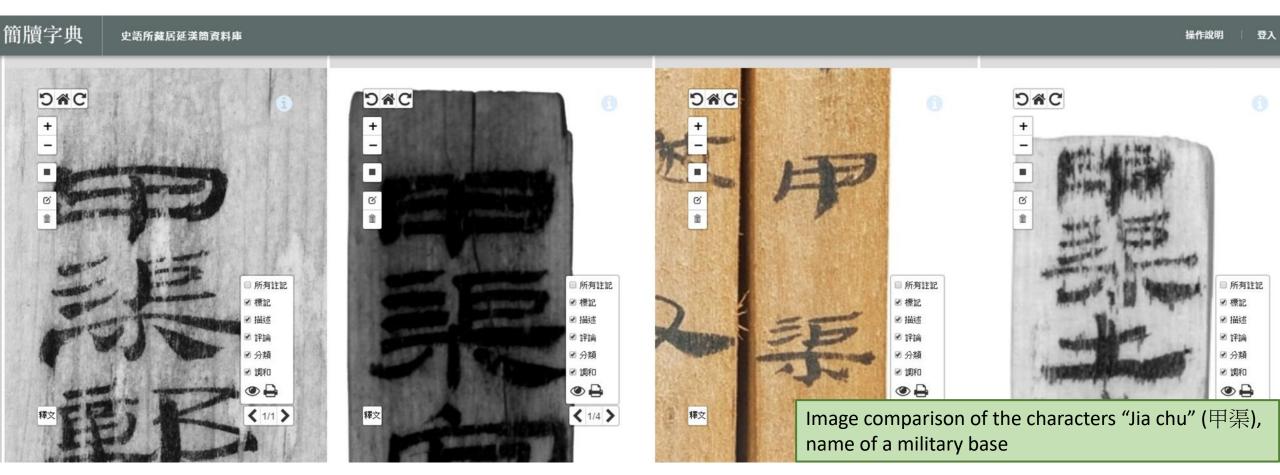
WCD: Information Retrieval



Object-based retrieval

Character-based retrieval

WCD: Image Research



• The IIIF-based image viewer can make comparisons of the characters' writing styles and identify the possible writing contributor between related characters.

Data Model of WCD

- 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



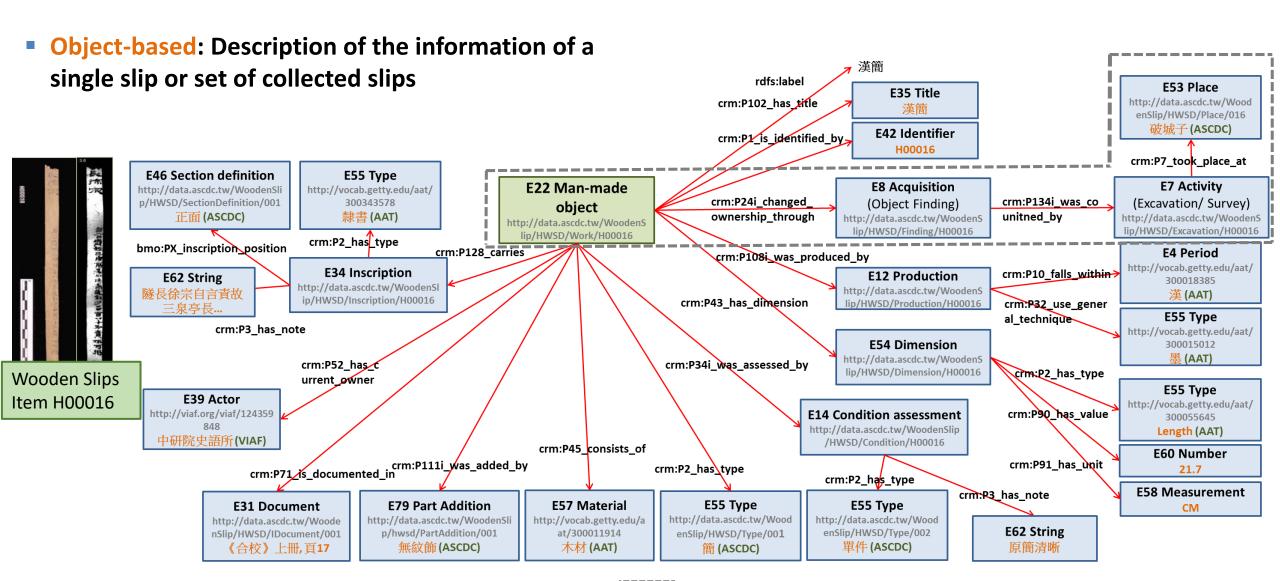
Classes [23]

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

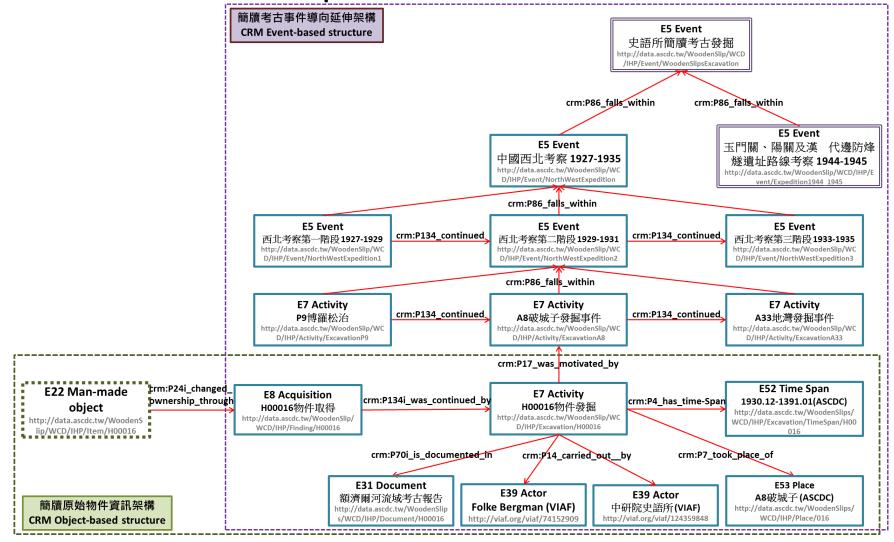
WCD Data Model: Object-based Design



Event-based: information on archaeological events

WCD Data Model: Event-based Design

 Event-based: Description of the hierarchical, archaeological events, in which the slips are excavated during the Sino-Swedish Scientific Expedition in the 1930s

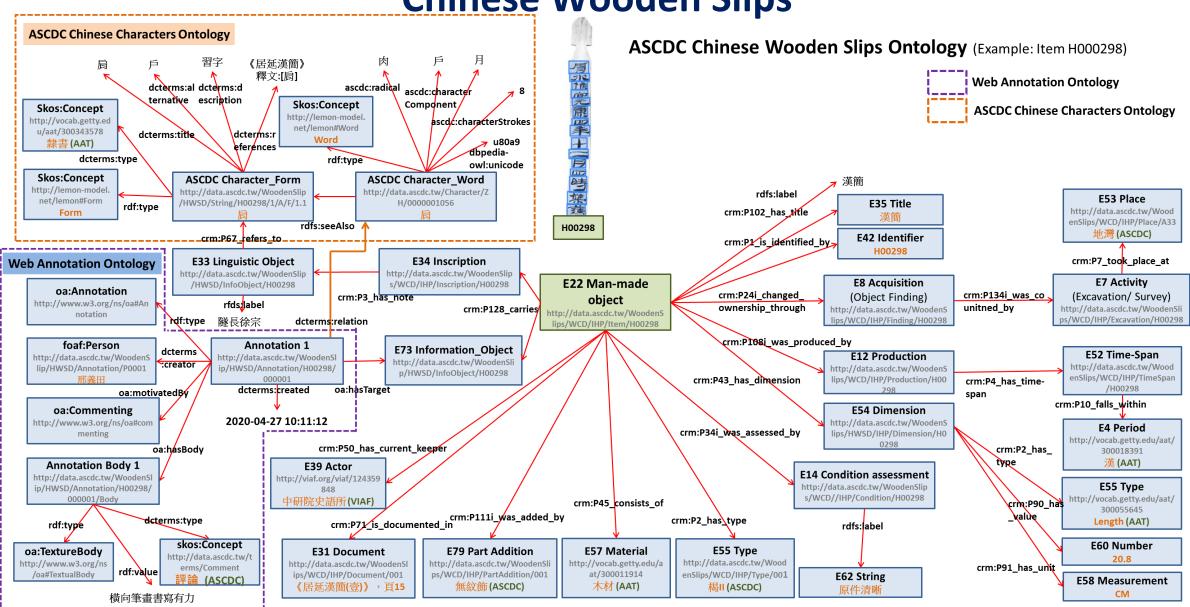


The Integrative Ontological Design of the Chinese Wooden Slips

- Purpose: Integration of the classified text information of IIIF-based image annotation with LOD-based metadata of Han Wooden Slips to enhance the interoperability of IIIF JSON-LD files with LOD data resources and exchange the researcher's opinion between domain experts.
- The current integrative ontological design for Chinese wooden slips contains:

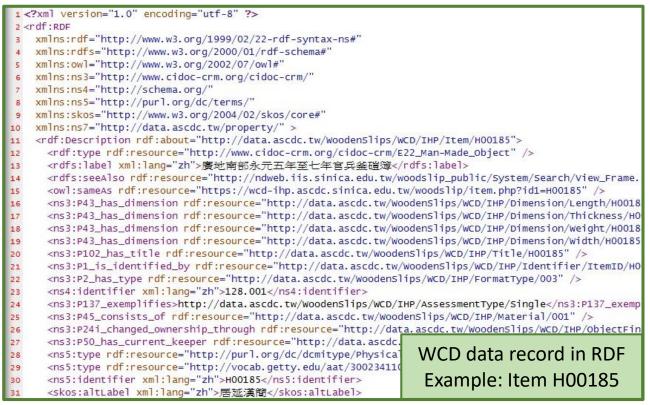
	Ontology type	Purpose for design	Model base
1	簡牘知識本體 Chinese wooden slips ontology	To describe entire metadata information of each wooden slip	CIDOC-CRM
2	漢字知識本體 ASCDC Chinese characters ontology	To describe information of each Chinese character written on the wooden slips	N/A
3	標註知識本體 Web annotation ontology	To describe annotated information on each wooden slip and characters written on it to classify different type of annotation and transform the annotated text into linked data	Web Annotation Data Model

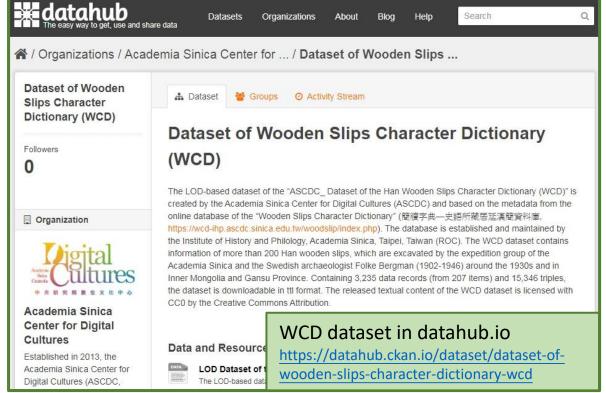
The Integrative Ontological Design of the Chinese Wooden Slips



LOD Dataset of the WCD

 Based on this integrative data model design, metadata of the wooden slips and characters on the slips are converted into LOD dataset; Released in the websites of ASCDC's LODdatasets, datahub.io and Linked Open Data Cloud





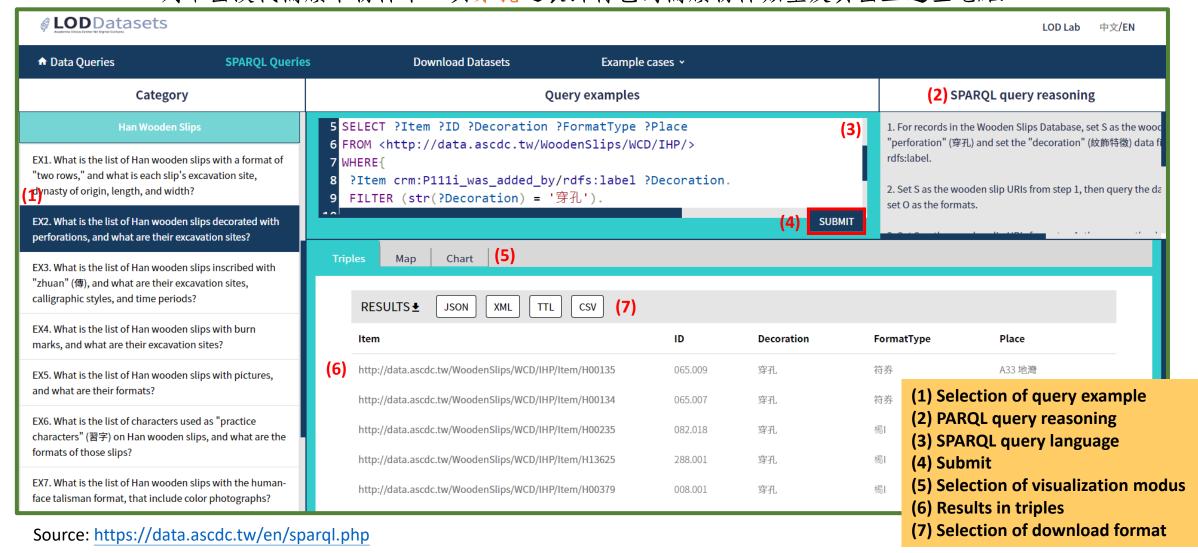
Template-based LOD SPARQL Query Examples

According to the focus of research, the template-based LOD SPARQL queries are designed in discussion with domain experts to satisfy the researcher's needs and to offer a user-friendly interface to retrieve the data content.

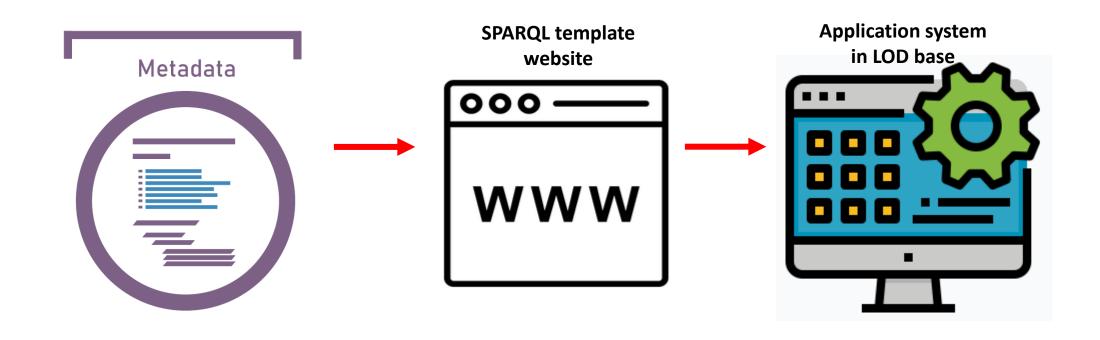
	Query examples for WCD dataset	Related research focuses
1	What is the list of Han wooden slips inscribed with "zhuan" (傳), and what are their excavation sites, calligraphic styles, and time periods?	1) Interpretation of Chinese characters 2) Comparative study on writing styles
2	What is the list of Han wooden slips decorated with perforations, and what are their excavation sites?	1) Study on the administrative History and official documents in Han Dynasty
3	What is the list of Han wooden slips with pictures, and what are their formats?	1) Comparative study on writing styles
4	What is the list of characters used as "practice characters" (習字) on Han wooden slips, and what are the formats of those slips?	1) Interpretation of Chinese characters 2) Comparative study on writing styles
5	What is the list of Han wooden slips with the human-face talisman format, that include color photographs?	1) Comparative study on writing styles
6	What is the list of Han wooden slips with a calligraphic style of "clerical script" (隸書) and a format of "two rows"?	1) Interpretation of Chinese characters 2) Manuscript restoration between slips
7	What is the list of Han wooden slips inscribed with "yue" (月), "shi" (石), "dou" (斗), or "sheng" (升), and excavated from the "A8 City Ruins" (A8破城子)?	1) Interpretation of Chinese characters 2) Manuscript restoration between slips
8	What is the list of Han wooden slips inscribed with the place of origin "lingshichong" (令史充), and the proximate time periods of wooden slips in the same numbered bundle?	Interpretation of Chinese characters Study on the career development of frontier officers in Han Dynasty
9	What is the list of Han wooden slips inscribed with "zhuan" (專), with a calligraphic style of "clerical script" (隸書)?	1) Interpretation of Chinese characters 2) Comparative study on writing styles

Template-based LOD SPARQL Query Examples

Query Example: What is the list of Han wooden slips decorated with perforations, and what are their excavation sites? 列舉出漢代簡牘中物件中,具穿孔之裝飾特色的簡牘物件類型及其出土遺址地點?

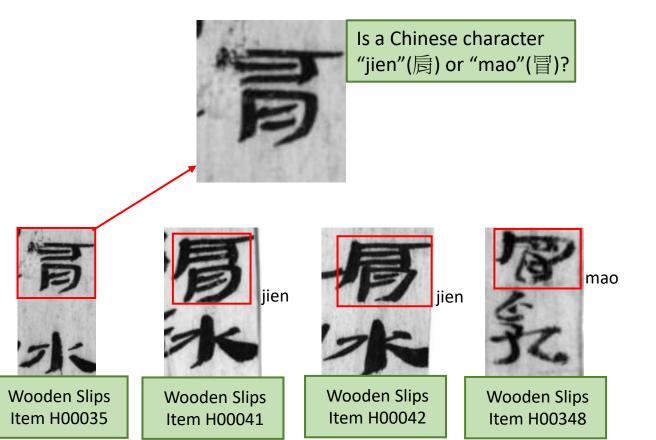


Template-based Queries to Application System



Template-based Design as Support for Study of Wooden Slips

- Interpretation of Chinese characters: to identify the written character and written style of a single character.
- Manuscript Restoration: to re-discovering the correct sequence or related group of slips between the scattered slips by finding the similar contextual meaning or written patterns.





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

WCD: Image Annotation

- Images of objects and characters can be annotated under different motivations and sub-categories 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 recording object data or interpreting the form and meaning of the characters.
- The annotated information will be recorded in the IIIF manifest and be reusable, exchangeable as linked data.
- Images can be researched by crowdsourcing.



The character "Jia" (♥) can be annotated under different motivations such as tagging, describing, commenting, or classifying (Item H04737)

WCD: Image Annotation

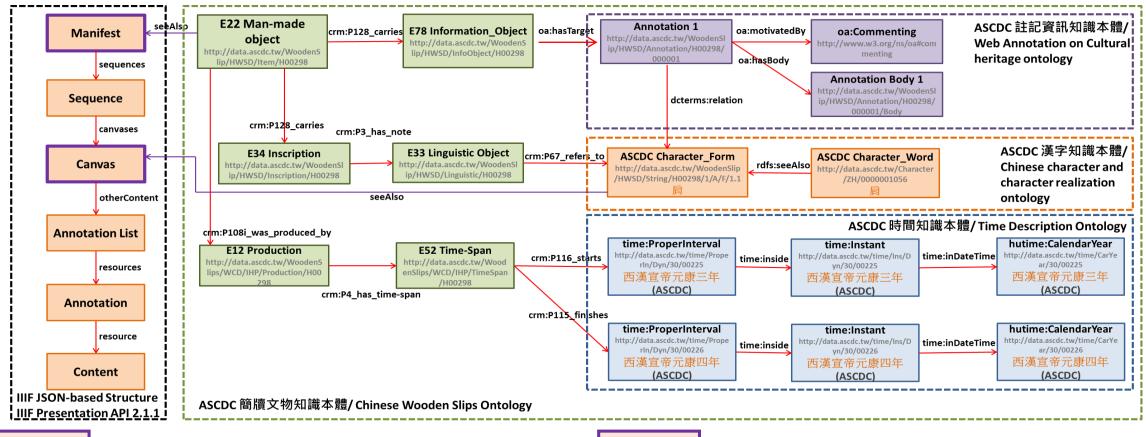
- The extension of the current IIIF annotation function is made based on W3C's Web Annotation Data Model (WADM).
- The current structure of faceted annotation contains WADM's motivations, "Tagging", "Describing", "Commenting", and "Classifying", for different annotation purposes.
- When annotating the images, the property
 oa:motivation is applied to distinguish different
 motives, while dcterms:type is reused to specify
 the sub-type of annotation within a motivation.

	Motivation types oa:motivation	Sub-types @zh	Sub-types @en dcterms:type
1	Tagging 標記	釋讀(確定)	InterpretationEnsured
		釋讀(不確定)	InterpretationUnsecured
		不區分/其他	TaggingOthers
2	Describing 描述	部首	Radical
		部件	Component
		不區分/其他	DescribingOthers
3	Commenting 評論	書寫特徵	WritingCharacteristic
		參考資料	References
		參考物件	RelatedItem
		備註	Note
		評論內容	Comment
		不區分/其他	CommentingOthers
4	Classifying 分類	書體	ScriptType
		不區分/其他	ClassifyingOthers

Structure of faceted annotation (motivations and sub-types)

Data Linkage between LOD and IIIF

 Using a linking property such as seeAlso as 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



Canvas Joint between a character's LOD and IIIF manifest

Manifest

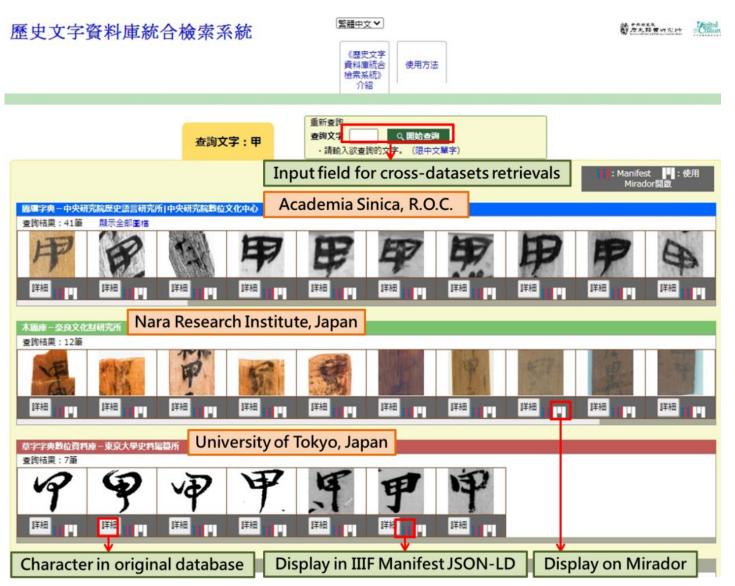
Data Linkage between LOD and IIIF: An Example

```
@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)
```



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

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

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.

```
status code: "200",
    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",
          manifest_url: "https://wcd-ihp.ascdc.sinica.edu.tw/api/GET/55/manifest",
       (12) subject: "簡牘字典一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"
木画庫-奈良文化財研究所
查詢結果:65筆
```

Closing Remarks

- Linked data is a method to make results of DH study more reusable, accessible, and enhance data retrieval across different datasets or institutes. It can also integrate with heterogeneous databases or IIIF-based image resources.
- Based on the shared linked data-structure of LOD dataset and IIIF manifest, both can be connected using semantic properties.
 Thus, a model to support DH studies, both on their descriptive component and image component, is possible.



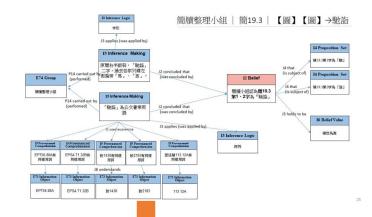
Implementing linked data-based Application system

Introducing the FAIR Principles



Future Works





Introducing the CRMinf Model for Argumentation

Integrating into the Wikidata



