

From Humanities Data to Digital Humanities Data : a new opportunity for technical services

Haiqing Lin

C.V. Starr East Asian Library
University of California, Berkeley

CEAL, Denver
2019

Humanities Data vs. Digital Humanities

Data: a Case Study



Self-Portrait with Straw Hat,
 1887
 Metropolitan Museum of Art
 (F365v)

Colour Analysing: a Digital Humanities Approach











- 35 self-portraits are clustered into 6 groups by periods



Colour Objects

Cluster colors, sized by number of pixels:

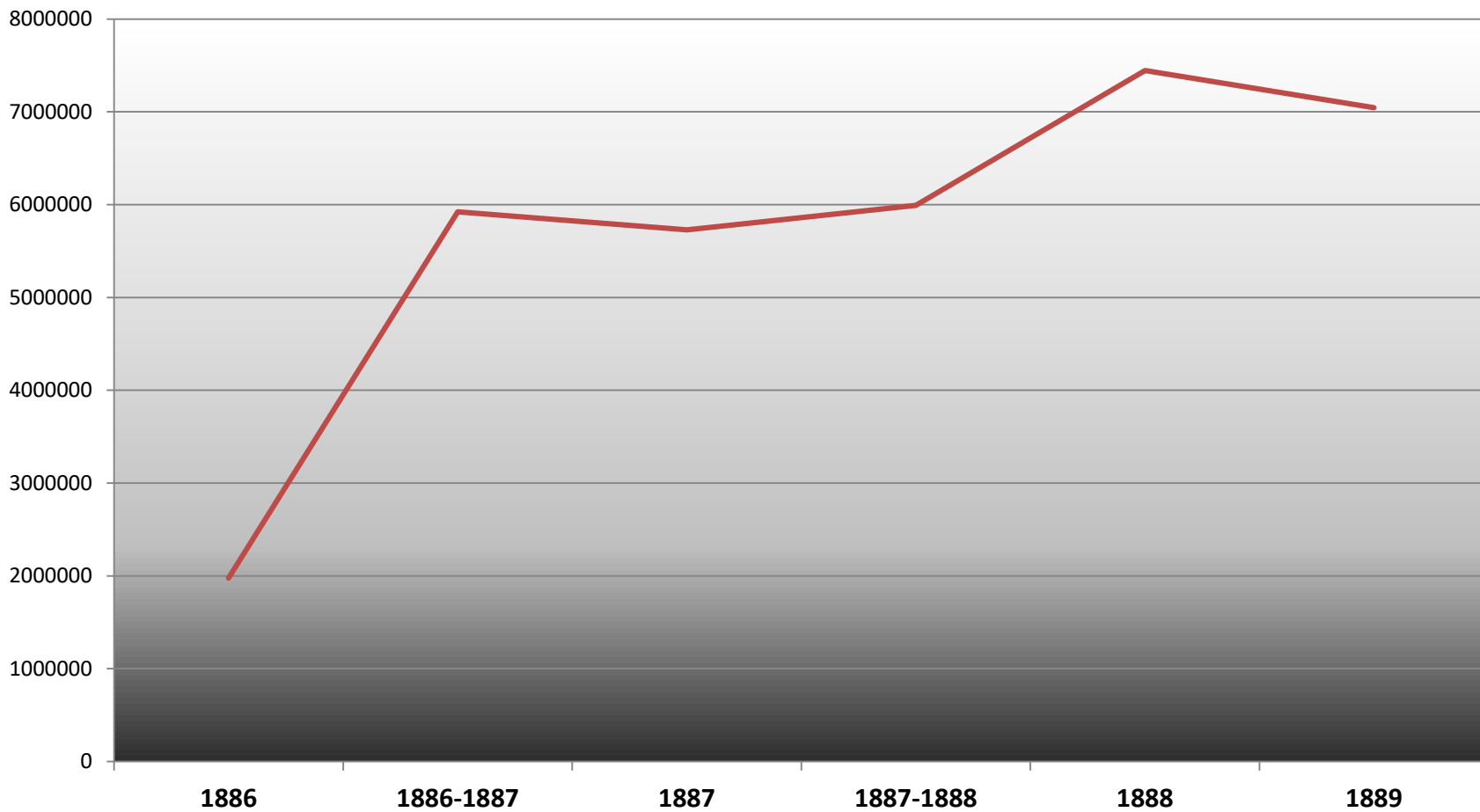


cluster	pixels	name	HEX	RGB	HSV	LCH	Lab	tags
	39.05%	 159, 159, 95 khaki $\Delta E=2.6$	#9A9C61	154 156 97	62 38 61	63 32 108	63 -10 31	khaki
	21.05%	 166, 187, 179 ashanti $\Delta E=2.1$	#A7B7AE	167 183 174	145 9 72	73 8 158	73 -7 3	air ash ashanti bel conch emerge hermitage periglacial rivendell skeptic tower blue grey
	20.29%	 101, 171, 124 tea $\Delta E=4.1$	#67B482	103 180 130	141 43 70	67 39 153	67 -35 18	aqua forest shamrock shiny tea green
	10.19%	 196, 170, 77 sundance $\Delta E=1.5$	#BEA74B	190 167 75	48 60 75	69 50 93	69 -3 50	brass roti sundance
	9.43%	 69, 64, 43 woodrush $\Delta E=1.7$	#43412D	67 65 45	53 32 26	27 12 102	27 -3 12	deep ash grasslands kelp khaki onion scrub suburban woodrush brown green grey

Colour Analysing

	avg	med	min	max	Avg Color	Med Color	Min Color	Max Color
1886								
RGB:R	49	44	15	161				
RGB:G	39	38	11	139				
RGB:B	30	29	10	114				
1886-1887								
RGB:R	100	104	23	235				
RGB:G	98	102	29	208				
RGB:B	90	85	29	184				
1887								
RGB:R	111	102.5	22	238				
RGB:G	102	96	28	212				
RGB:B	87	87	15	176				
1887-1888								
RGB:R	114	115.5	36	213				
RGB:G	106	101	26	203				
RGB:B	91	88	38	172				
1888								
RGB:R	141	144	36	228				
RGB:G	159	167	35	229				
RGB:B	113	104	26	220				
1889								
RGB:R	124	132	18	227				
RGB:G	123	119.5	33	206				
RGB:B	107	105	30	173				

Result : A Colour Curve of Portraits of Vincent van Gogh



Other Examples: Text

欽定四庫全書 周易注疏 魏 王... GitHub Catalog More... Title Search Text Search Contact

【四庫全書·文淵閣】

[007-1a] 欽定四庫全書 周易注疏 述人 唐國子博士兼太子中允贈齊州刺史吳縣開國男陸德明

宓犧氏之王天下仰則觀於天文俯則察於地理 觀鳥獸之文與地之宜近取諸身遠取諸物始畫 八卦或云因河圖而畫八卦因而重之為六十四文王拘於 羑里作卦辭周公作爻辭孔子作彖辭象辭 文言 繫辭說卦序卦雜卦謂之十翼班固曰孔子晚而

[007-1b] 好易讀之韋編三絕而為之□□即十翼也先儒說重 卦及爻辭為十翼不同解見余所撰釋文自魯商瞿子木受易於孔子 以授魯橋庇子庸子庸授江東馯 旦反徐廣音寒瞽子 弓子弓授燕周醜子家子家授東武孫虞子乘子 乘授齊田何子莊高士云字莊漢書儒林云臨淄人及秦燔書 易為卜筮之書獨不禁故授者不絕漢興田何 以齊田徒杜陵號杜田生授東武王同子中及洛 陽周王孫梁人丁寬字子襄事田何復從周王孫受古義作易說三萬 言訓故

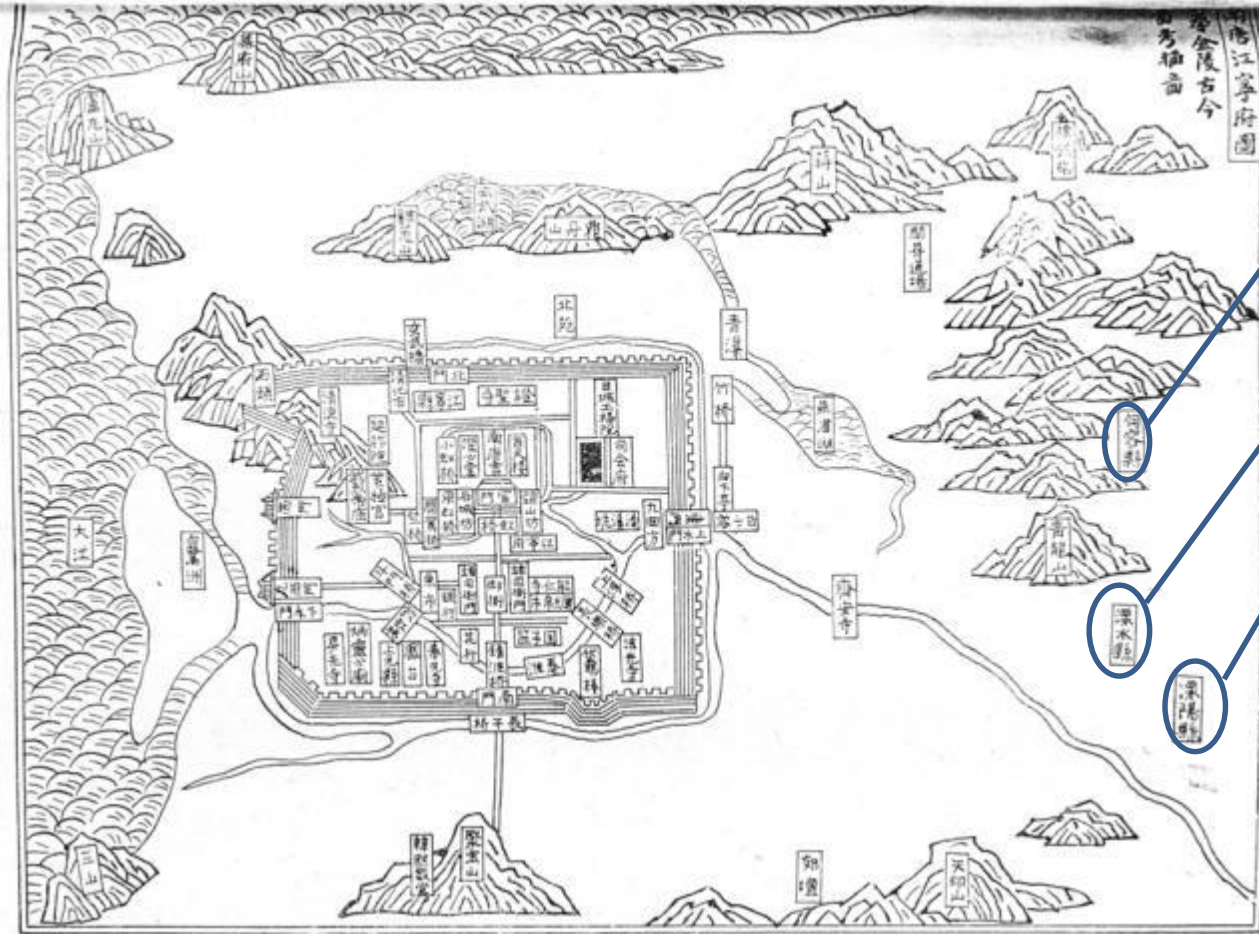
[007-2a] 舉大詁而已藝文志云易說八篇為梁孝王將軍齊服生劉向別錄云 齊人號服光皆 著易漢初言易者本之田牛同授淄川楊何字叔

Object.person: 陸德明

Object.official: 刺史

Object.word: 文言

Other Examples: Map



Object.place: 句容縣

Object.place: 溧水縣

Object.place: 溧陽縣

Humanities Data

- Definition: “Humanities data might be about the record of events – what we commonly call facts – or about texts created about them, whether these texts are written, oral, visual, aural or spatial.” (Gardiner, E., Musto, R. G., *The digital humanities: a primer for students and scholars*, New York: Cambridge University Press, 2015, page 31)
- Humanities data include text, image, audio, video, object, space and so on.

Digital Humanities Data

- Definition of Digital Humanities Data

Digital humanities data is a set of objects which are extracted from the digitalized humanities data. This set of objects consist of people, organizations, places, events and all other real-world objects. Each object has its own attributes and properties which are usually described by metadata to enable human understanding, as well as to be explicit to machine understanding. With machine understanding, digital humanities data is able to be processed by algorithms.

- From Humanities Data to Digital Humanities Data



Linking Digital Humanities Data to the Library Data: Person

激 周易註疏魏-王... GitHub Catalog More...
Title Search
Text Search
Contact

[007-1a] 欽定四庫全書
周易注解述人
唐國子博士兼太子中允贈齊州刺史吳縣開國男陸德明
宓犧氏之王天下仰則觀於天文俯則察於地理
觀鳥獸之文與地之宜近取諸身遠取諸物始畫
八卦或云因河圖而畫八卦因而重之為六十四文王拘於
羑里作卦辭周公作爻辭孔子作彖辭象辭文言
繫辭說卦序卦雜卦謂之十翼班固曰孔子晚而

[007-1b] 好易讀之韋編三絕而為之□□即十翼也先儒說重
卦及爻辭為十翼不同解見余所撰釋文自魯商瞿子木受易於孔子
以授魯橋庇子庸子庸授江東馮戶旦反徐廣音寒臂子
弓子弓授燕周醜子家子家授東武孫虞子乘子
乘授齊田何子莊高士云字莊漢書儒林云臨淄人及秦燔書
易為卜筮之書獨不禁故授者不絕漢興田何
以齊田徙陵號杜田生授東武王同子中及洛
陽周王孫梁人丁寬字子襄事田何復從周王孫受古義作易說三萬
言訓故

[007-2a] 舉大誼而已藝文志云易說八篇為梁孝王將軍齊服生劉向別錄云/
齊人號服光皆
著易漢初言易者本之田生同授淄川楊何字叔

Object.person: 陸德明

From [Library of Congress Name Authority File](#)

Lu, Deming, 1956-

URI(s)
<http://id.loc.gov/authorities/names/nr00022511>

Instance Of
[MADS/RDF PersonalName](#)
[MADS/RDF Authority](#)
[SKOS Concept](#)

Scheme Membership(s)
[Library of Congress Name Authority File](#)

Collection Membership(s)
[Names Collection - Authorized Headings](#)
[LC Names Collection - General Collection](#)

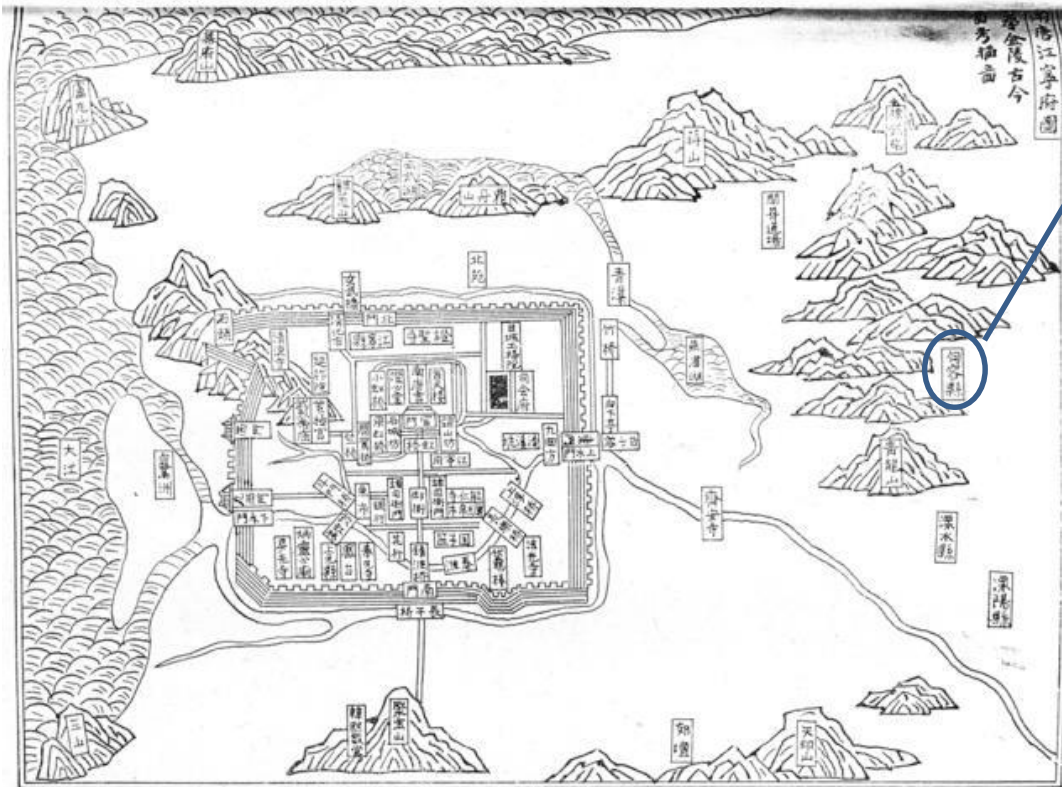
Variants
[陸德明, 1956-](#)
[陸德明, 1956-](#)

Additional Information
<http://id.loc.gov/wo/agents/nr00022511>

Exact Matching Concepts from Other Schemes
<http://viaf.org/viaf/sourceID/LC%7Cnr+00022511#skos:Concept>

Earlier Established Forms

Linking Digital Humanities Data to the Library Data: Place



Object.place: 句容县



From [Library of Congress Name Authority File](#)

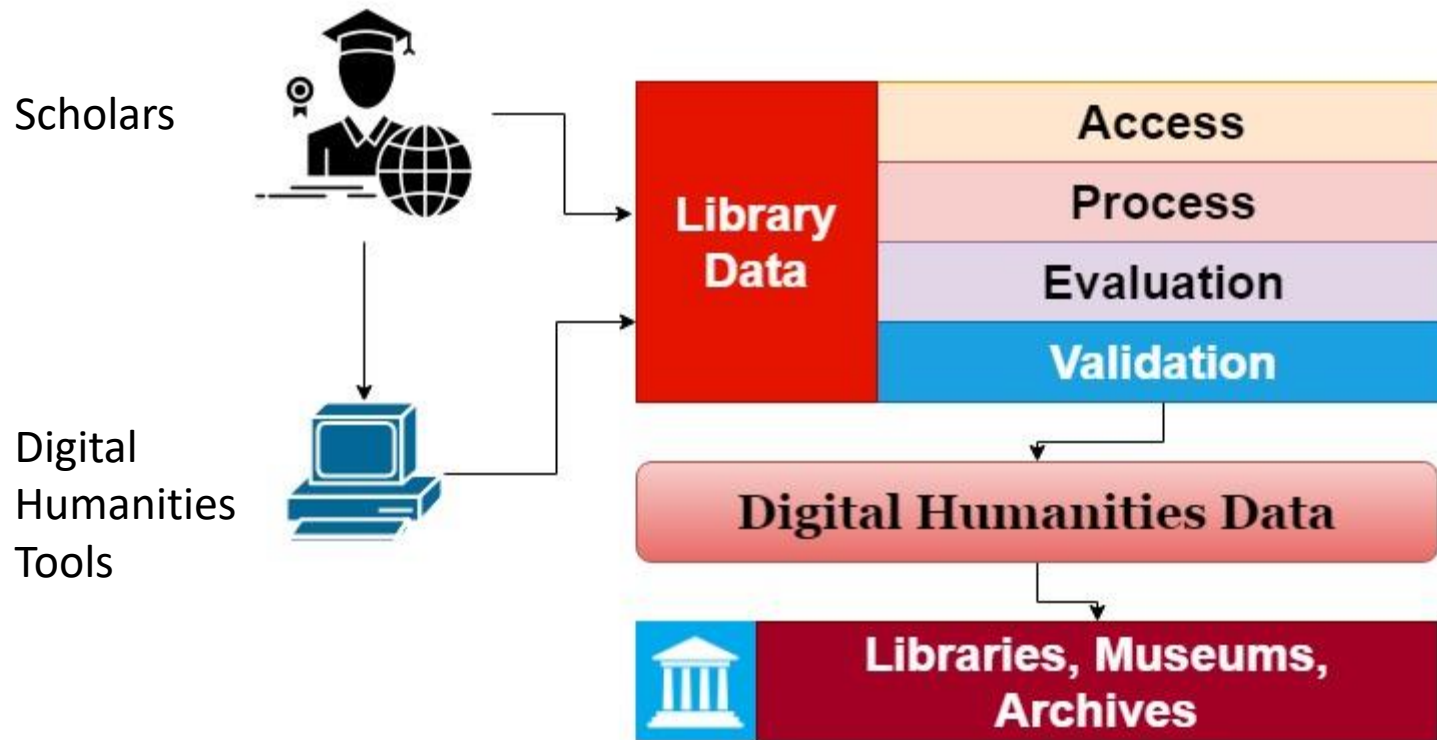
Jurong Xian (China)

- URI(s)
 - > <http://id.loc.gov/authorities/names/n82044609>
- Instance Of
 - > [MADS/RDF Geographic](#)
 - > [MADS/RDF Authority](#)
 - > [SKOS Concept](#)
- Scheme Membership(s)
 - > [Library of Congress Name Authority File](#)
- Collection Membership(s)
 - > [Names Collection - Authorized Headings](#)
 - > [LC Names Collection - General Collection](#)
- Variants
 - > Chü-yung (China)
 - > Juyong (China)
 - > KuyG-gen (China)
 - > KuyG-ken (China)
 - > Küyung (China)
- Additional Information
 - > <http://id.loc.gov/rwo/agents/n82044609>
- Exact Matching Concepts from Other Schemes
 - > <http://viaf.org/viaf/sourceID/LC%7Cn+82044609#skos:Concept>

Why Library Data

- **Reliable**
- **Evidence based**
- **Rich contents**
- **Consistency**
- **Systematic**
- **Linking to library collection**

Library data as a digital humanities infrastructure



New Role of Technical Services

- **Modeling**
FRBR, BIBFRAME, Library Reference Model (LRM)
- **Describing**
AACR2, RDF
- **Encoding**
MARC21, RDF/XML, MARCXML
- **Evaluating**

Towards Data-oriented technical services: next generation

- By the data
- Of the data
- For the data

Thank you!