

Exploring the Intellectual Structure of Digital Stewardship

Fang-Ming Hsu, Chun-Min Lin, Chu-Mei Chiu
Department of Information Management, National Dong Hwa University, Taiwan

Digital Stewardship(DS)

- ▶ Institute of Museum and Library Service(IMLS) first proposed the idea of ‘digital stewardship’ (DS) in 2007 in the annual meeting of National Archives and Records Administration (NARA).
- ▶ The definition of ‘digital stewardship’ is mainly about the preservation and usage of digital data and records.
- ▶ Records with long-term preservation value, either paper-based or digital records, governmental or private, are included in the realm of digital stewardship that deals with the management procedures, technology, usage, and promotion of digital records.



DS in this research

- ▶ Governments in every country have lots of digital records, archives, rules, and regulations that need to be permanently preserved.
- ▶ The repository and management of digital electronic governmental records have become a great concern of many governments and thus form the focus of present study.
- ▶ The present research concentrated on the long-term preservation of governmental Electronic Records (ER) and Digital Archives (DA).



Intellectual structure of DS

- ▶ To explore the intellectual structure of DS, this study used co-word analysis to investigate the deployment of terms in academic research articles in ER and DA fields.
- ▶ Digital stewardship is a relatively newer phrase and includes both ER and DA. Thus, key terms revolving ER and DA were employed in searching in databases. The analysis of research regarding digital stewardship would base on the synthesis of the search results.



ER and DA

- ▶ ER and DA represents two different aspects of issues in management field. However, their research domains overlap in some way and both have continuity.
- ▶ In this study, the method of co-word analysis in bibliometrics was employed on both ER- and DA-related research articles.
- ▶ The purpose was to ascertain that co-word analysis can be applied to the analysis of cross-topic research.



Importance of Intellectual Structure

- ▶ The analysis and understanding of the intellectual structure and the evolution of a scientific discipline can be useful for academics, students, and practitioners (Borkhovich, Bricker, & Simkins, 1994; Locke & Perera, 2001).
- ▶ For academics, its usefulness derives from the opportunity to position their investigations within the field of study, identify potential new directions and declining perspectives, and summarize the most relevant literature and the relationships between key works in the area.

Casillas, J., & Acedo, F. (2007). Evolution of the intellectual structure of family business literature: A bibliometric study of FBR. *Family Business Review*, 20(2), 141-162.

Importance of Intellectual Structure (continued)

- ▶ For students, having a map of the conceptual structure of a discipline can be of great interest in order to develop an overview of the field of study, understand the relationships between paradigms, and identify the essential works on each one of them.
- ▶ Practitioners can find a guide to understand how the concepts and perspectives within a discipline have evolved over time, and determine which are the most analyzed topics, which the conceptual bases, and which areas deserve special attention in order to implement them in business.

Casillas, J., & Acedo, F. (2007). Evolution of the intellectual structure of family business literature: A bibliometric study of FBR. *Family Business Review*, 20(2), 141-162.

Intellectual Structure of a Discipline

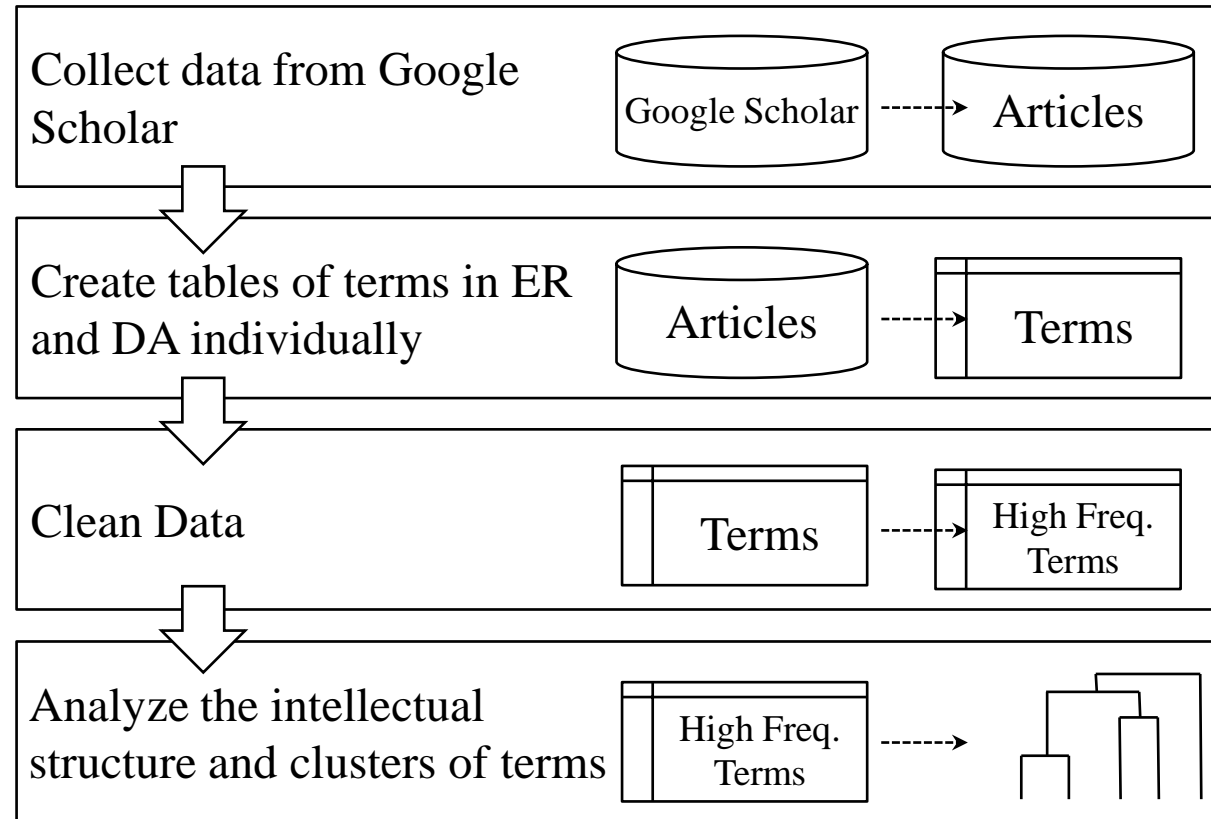
- ▶ To conclude, intellectual structure is the exploration of the classification of the trend of a research discipline and its evolvement over time.
- ▶ In order to understand the classification of a discipline, besides experts judgment, bibliometrics methods such as co-word analysis, co-citation analysis can be used. Both methods calculated the distance between and among academic research articles to know the clustering of research, and to define the classification of discipline.



Google Scholar(GS)

- ▶ In bibliometrics-related research, WoS (Web of Science) from ISI was the most used database to collect academic articles. Later on, Google Scholar (GS) developed by Google became an important database in research because of its simple interface and broad coverage of research articles.
- ▶ Amara (2012) compared the academic research articles collected by GS and WoS in business field and found GS contains more information than WoS.
- ▶ Mikki (2010) compared the academic research articles collected by GS and WoS in geography science field, and found GS contains comprehensive data which covered 85% of the data in WoS.
- ▶ Harzing (2013) analyzed Nobel-winning articles longitudinally using GS and found the stability and contents of GS were increasing rapidly. Harzing also emphasized that in the process of data collection, if comparing to WoS, GS seemed
- ▶ to include research articles without prejudice which were otherwise found in manual collection

The Flowchart of this Research



Searching Terms

	ER	DR	EA	DA	Unions
2014	707(47%)	159(11%)	320(21%)	351(23%)	1,500
2013	4,240(42%)	984(10%)	2,180(21%)	3,030(30%)	10,200
2012	4,290(37%)	976(8%)	3,470(30%)	3,360(29%)	11,700
2011	5,010(42%)	885(7%)	3,140(26%)	3,360(28%)	12,000
2010	4,970(52%)	840(9%)	1,440(15%)	2,770(29%)	9,640
2009	4,770(45%)	887(8%)	1,500(14%)	3,700(35%)	10,500
Total	23,987(43%)	4,731(9%)	12,050(22%)	16,571(30%)	55,540

The following terms were searched in Google Scholar.

"electronic records" | "digital records" | "electronic archives" | "digital archives"

- "health" - "medical" - "nursing" - "treatment"



DS-related Research Articles Being Analyzed in This Study

- ▶ Using the above mentioned searching terms, 1000 academic research articles each year, with a total of 10,000 data, were downloaded from 2004–2013 from GS. Research other than using English were deleted (4,922 articles) and 5,078 data remained. These research could be journal articles, theses, dissertations, patents, and books.



ER Top 20 Most Frequently Used Terms and Their occurrences in Each Country's National Archive Website Glossaries

No	Term	TFIDF	TF	AU	CA/US	CN	TW
1	Record	630.606	834	v	v	v	v
2	system	452.946	395	v	v		v
3	information	438.442	394	v	v	v	v
4	archive	428.569	293	v	v	v	v
5	electronic records management	413.689	240				v
6	records management	402.265	231	v	v	v	
7	management	396.769	352		v		v
8	document	385.385	227	v	v	v	
9	preservation	368.138	206	v	v	v	v
10	electronic records management system	367.596	142	v			
11	data	348.088	204	v		v	v
12	metadata	339.268	130	v	v	v	v
13	access	303.289	160	v		v	
14	standard	298.991	156	v	v	v	v
15	technology	287.810	151		v		v
16	agency	264.473	98	v			
17	archivist	250.211	108				v
18	strategy	245.818	109		v	v	
19	application	244.941	105			v	v
20	electronic document	243.490	91		v	v	

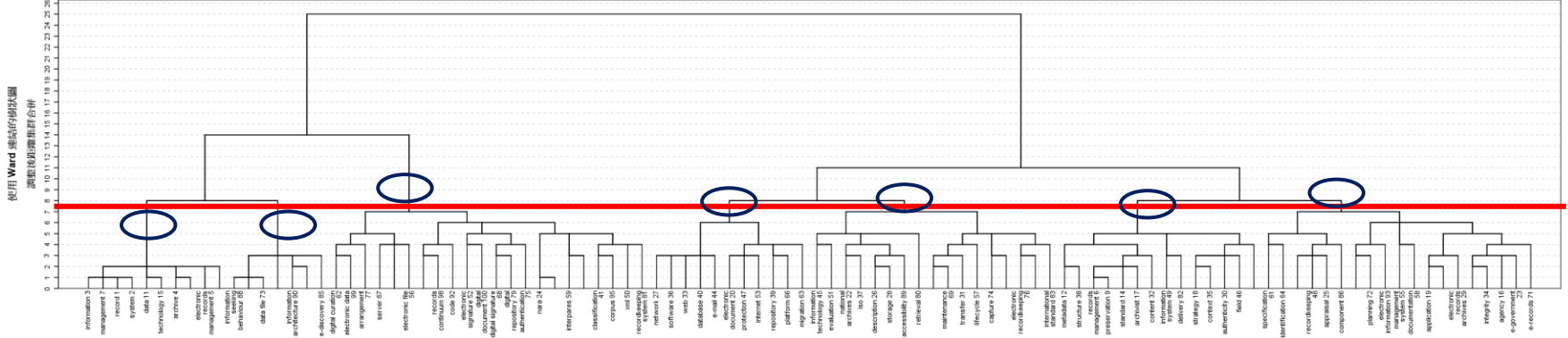


DA Top 20 Most Frequently Used Terms and Their occurrences in Each Country's National Archive Website Glossaries

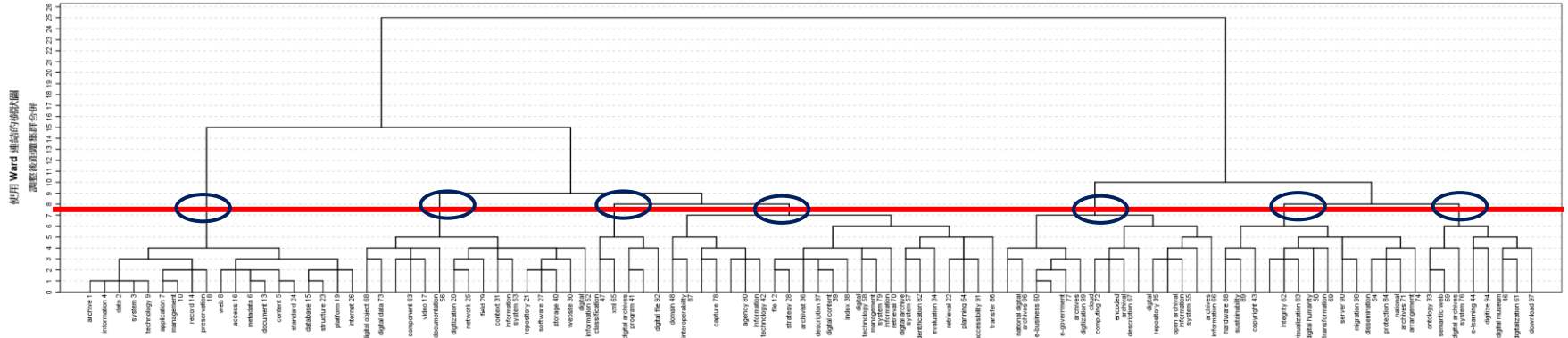
No	Term	TFIDF	TF	AU	CA/US	CN	TW
1	archive	753.652	755	v	v	v	v
2	data	646.170	441	v		v	v
3	system	594.460	452	v	v		v
4	information	593.251	540	v	v	v	v
5	content	467.756	279	v			v
6	metadata	461.894	214	v	v	v	v
7	application	442.093	260			v	v
8	web	437.841	210		v	v	v
9	technology	434.840	275		v		v
10	management	429.557	220		v		v
11	national digital archives program	407.486	149				v
12	file	373.855	135	v		v	
13	document	372.494	168	v		v	
14	record	359.106	147	v	v	v	
15	database	357.562	162	v	v		
16	access	356.374	176	v		v	
17	video	353.431	126		v	v	
18	preservation	327.361	151	v	v	v	v
19	platform	317.577	130			v	
20	digitization	285.390	105	v	v		



Clustering Analysis of ER and DA High-Frequency Terms



ER



DA

Clustering Analysis of ER High-Frequency Terms

	Term
1	access
1	archive
1	data
1	document
1	electronic records management
1	information
1	management
1	record
1	system
1	technology

	Term
2	archivist
2	authenticity
2	content
2	context
2	delivery
2	field
2	information system
2	long-term preservation
2	metadata
2	preservation
2	records management
2	standard
2	strategy
2	structure

	Term
3	agency
3	application
3	appraisal
3	component
3	digital preservation
3	documentation
3	e-government
3	electronic information
3	electronic records archives
3	electronic records management system
3	e-records
3	file
3	identification
3	information management
3	integrity
3	management system
3	planning
3	recordkeeping
3	specification

	Term
4	database
4	electronic document
4	e-mail
4	internet
4	migration
4	network
4	platform
4	protection
4	repository
4	software
4	web



Clustering Analysis of ER High-Frequency Terms (continued)

	Term
5	accessibility
5	archives management
5	capture
5	description
5	electronic recordkeeping
5	evaluation
5	hardware
5	information technology
5	international standard
5	iso
5	lifecycle
5	maintenance
5	national archives
5	records management system
5	retrieval
5	storage
5	transfer

	Term
6	arrangement
6	authentication
6	classification
6	code
6	corpus
6	digital curation
6	digital document
6	digital repository
6	digital signature
6	document management
6	electronic data
6	electronic document and records management system
6	electronic file
6	electronic signature
6	inter pares
6	long-term digital preservation
6	nara
6	recordkeeping system
6	records administration
6	records continuum
6	server
6	xml

	Term
7	data file
7	e-discovery
7	electronic laboratory notebooks
7	information architecture
7	information seeking behaviour
7	metadata encoding and transmission standard
7	ontology-based retrieval system



Clustering Analysis of DA High-Frequency Terms

Term	Term	Term	Term
1 access	2 accessibility	3 component	4 digital archives system
1 application	2 agency	3 context	4 digital museum
1 archive	2 archivist	3 digital data	4 digitalization
1 content	2 capture	3 digital information	4 digitize
1 data	2 description	3 digital object	4 download
1 database	2 digital archive system	3 digitization	4 e-learning
1 document	2 digital content	3 documentation	4 ontology
1 information	2 digital preservation	3 field	4 semantic web
1 internet	2 digital technology	3 information system	
1 management	2 digital technology	3 long-term preservation	
1 metadata	2 domain	3 network	
1 platform	2 evaluation	3 repository	
1 preservation	2 file	3 software	
1 record	2 identification	3 storage	
1 standard	2 index	3 video	
1 structure	2 information retrieval	3 website	
1 system	2 information technology		
1 technology	2 interoperability		
1 web	2 management system		
	2 national digital archives program		
	2 planning		
	2 retrieval		
	2 strategy		
	2 transfer		



Clustering Analysis of DA High-Frequency Terms (continued)

	Term
5	archives digitization
5	archives information
5	archives management
5	cloud computing
5	digital repository
5	e-business
5	e-government
5	electronic records archives
5	encoded archival description
5	file management
5	information architecture
5	metadata encoding and transmission standard
5	national digital archives
5	open archival information system
5	open archives initiative protocol for metadata harvesting

	Term
6	classification
6	digital archives program
6	digital file
6	taiwan e-learning and digital archives Program
6	xml

	Term
7	arrangement
7	copyright
7	digital humanity
7	dissemination
7	geographic information system
7	hardware
7	integrity
7	migration
7	national archives
7	protection
7	server
7	sustainability
7	transformation
7	visualization



High-frequency Terms Only Found in ER Clusters

	Terms
ER only	appraisal; authentication; data file; digital signature; document management; e-discovery; electronic data; electronic document; electronic document and records management system; electronic file; electronic information; electronic laboratory notebooks; electronic recordkeeping; electronic records management; electronic records management system; electronic signature; e-records; information seeking behaviour; inter pares; iso; lifecycle; long-term digital preservation; maintenance; nara; ontology-based retrieval system; recordkeeping; recordkeeping system; records administration; records continuum; records management; records management system;



High-frequency Terms Only Found in DA Clusters

	Terms
DA only	acquisition ; archive system; archives digitization; archives information; catalog; cloud computing ; copyright; data extraction ; data mining ; digital archive system; digital archives information; digital archives management; digital archives program; digital archives system; digital content; digital data; digital file; digital humanity; digital information; digital learning; digital museum; digital object; digital technology; digitalization; digitisation; digitization; digitize; dissemination; domain; download; e-business; e-learning; electronic portfolio; encoded archival description; file information; file management; geographic information system; index; information infrastructure; information retrieval; interoperability; national digital archives; national digital archives program; node; ontology ; open archival information system; open archives initiative protocol for metadata harvesting; record; semantic web ; stability; sustainability; taiwan e-learning and digital archives program; text mining ; transformation; verification;

High-frequency Terms Both Found in ER and DA Clusters

	Terms
ER & DA	access ; accessibility ; agency; application; archive; archives management; archivist; arrangement; authenticity; capture; classification; code; component; content; context; corpus ; data; database ; delivery; description; digital curation; digital document; digital preservation; digital repository; document; documentation; e-government ; electronic records archives; e-mail; evaluation; field; file; hardware; identification; information; information architecture; information management; information system; information technology; integrity; international standard; internet; long-term preservation; management; management system; metadata ; metadata encoding and transmission standard; migration; national archives; network; planning; platform; preservation; protection; repository; retrieval; server; software; specification; standard; storage ; strategy; structure; system; technology ; transfer; web; xml ;

Research Results and Implications

- ▶ Important terms regarding DS have been located in this study along with the relationship among and the clustering analysis of these terms.
- ▶ Based on the past research in bibliometrics, the results of co-word analysis can represent the main research fields in certain discipline.
- ▶ The research themes of these clustering found in this study will be further defined in subsequent work.
- ▶ The difference of the research between ER and DA is worth exploring. For example, while it is logical that 'appraisal' was less surveyed in DA; yet why 'digital signature' received less attention as well? Also, why 'cloud computing' and 'data mining' were seldom investigated in ER? Such phenomenon require further exploration. The results of present study may serve as a reference for future DS-related studies.

Thank You

