

BIRDS 2021 Invited

(with CHIIR 2021, online)

User Discovery and Exploration in Future Digital Libraries

by Edward A. Fox

ORCID: 0000-0003-1447-6870

- <http://fox.cs.vt.edu/talks/2021/20210319FoxBIRDS.pptx>
- fox@vt.edu <http://fox.cs.vt.edu>
- Dept. of Computer Science, Virginia Tech
- Blacksburg, VA 24061 USA

Acknowledgements

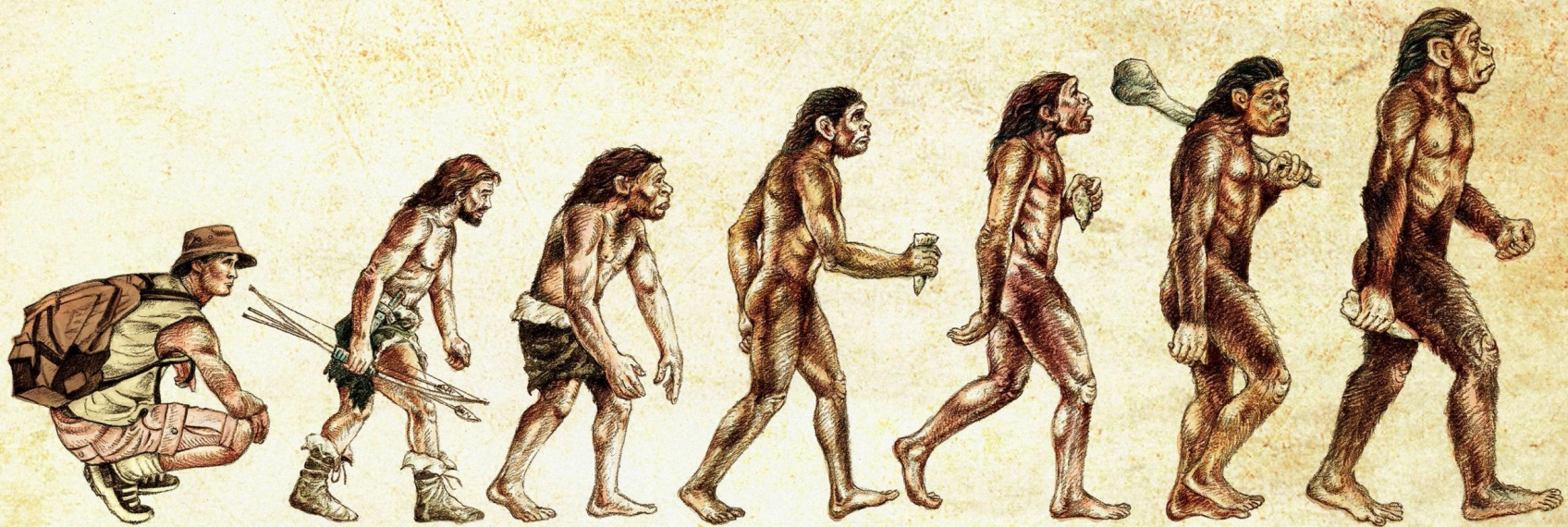
- Mentors (Licklider, Kessler, Salton)
- Virginia Tech, CS, Digital Library Research Laboratory (DLRL)
- NSF, IMLS, and many other sponsors (see <http://fox.cs.vt.edu/cv.htm>)
- Students, colleagues, co-investigators (selected): Eman Abdelrahman, Monika Akbar, Hamed Alhoori, Pranav Angara, Ashish Baghudana, Jefferson Bailey, Bipasha Banerjee, Abigail Bartolome, Warren Bickel, Matt Bock, Boots Cassel, Prashant Chandrasekar, Saurabh Chakravarty, Prashant Chandrasekar, Raja Chava, Satvik Chekuri, Yinlin Chen, Kiran Chitturi, Lois Delcambre, Noha ElSherbiny, Alexandre Falcao, Weiguo Fan, Eric Fouh, Chris Franck, Rick Furuta, Jack Geissinger, Lee Giles, Marcos André Gonçalves, Doug Gorton, Seth Guikema, Islam Harb, S.M.Shamimul Hasan, Michael Hsiao, Bill Ingram, Palakh Jude, Adheesh Juvekar, Sampanna Kahu, Tarek Kanan, Ola Karajeh, Andrea Kavanaugh, Farnaz Khaghani, Martin Klein, Nadia Kozievitch, Abhinav Kumar, Harinni Kumar, Spencer Lee, Sunshin Lee, Jonathan Leidig, Lin Tzy Li, Liuqing Li, Yi Ma, Yufeng Ma, Mohamed Magdy, Shivam Maharshi, Ashish Malpani, Madhav Marathe, Gary Marchionini, Paul Mather, Maanav Mehrotra, Pamela Murray-Tuite, **Uma Murthy**, Pranav Nakate, Michael Nelson, Sanghee Oh, Sung Hee Park, Supritha Patil, Denilson Pereira, Jeff Pomerantz, Naren Ramakrishnan, Pranavi Rambhakta, Sagnik Ray-Choudhury, Chandan Reddy, Rao Shen, Cliff Shaffer, Steve Sheetz, Don Shoemaker, Ziqian Song, Venkat Srinivasan, Hussein Suleman, Amirsina Torfi, Ricardo Torres, Adithya Upadhyaya, Saket Vishwasrao, Xinyue Wang, Kris Wernstedt, Barbara Wildemuth, Jian Wu, Zhiwu Xie, Seungwon Yang, Xiaoyan Yu, Xuan Zhang, ...

Perspectives

- Member of ACM (1967-), IEEE (2004-)
- Professor of CS (cs.vt.edu, 1983-)
- Digital libraries (dlib.vt.edu, 1991-)
- Event archiving (eventsarchive.org: web, social media, 2007-)
- CTO of startup (claimededge.co, 2018-) working to connect law and insurance firms with summarization, extensible workflows, and AI support of guidance; NSF I-Corps customer discovery

Short Version

- Exploration is a fundamental need, and so should be well supported in future DLs:
- for all Societies: Humans and/or with Computers;
- through diverse Scenarios, aided by Spaces and Structures, covering all types of Streams.
- Theory-> Design-> Implementation-> DevOps
- Integrating Sciences: computing, library, infor.; AI, Archiving, DB, DS, HCI, HPC, Hypermedia, IR, Networking, NLP, Psychology, Sociology,...

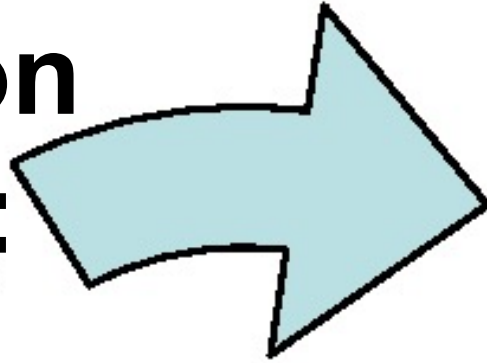


Explore the world's best prehistoric caves at Taman Negara.

Malaysia
Truly Asia

Exploration includes:

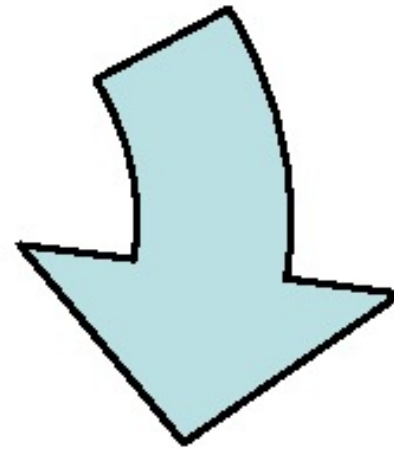
Information use



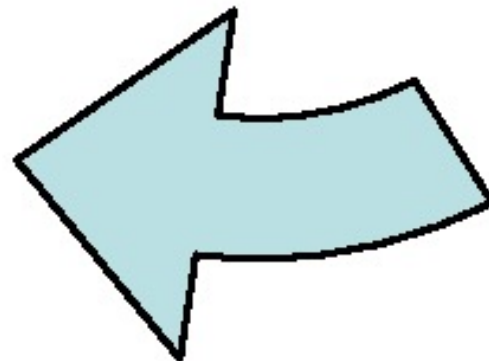
Information
needs



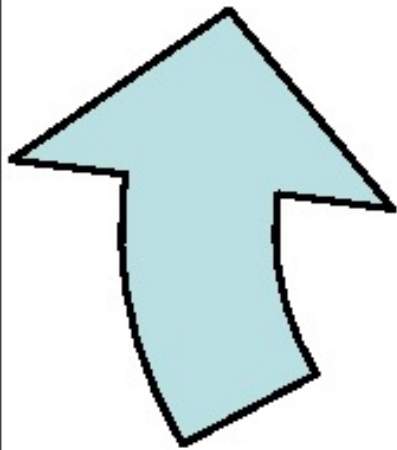
Research
processes



Information
source



Information
search

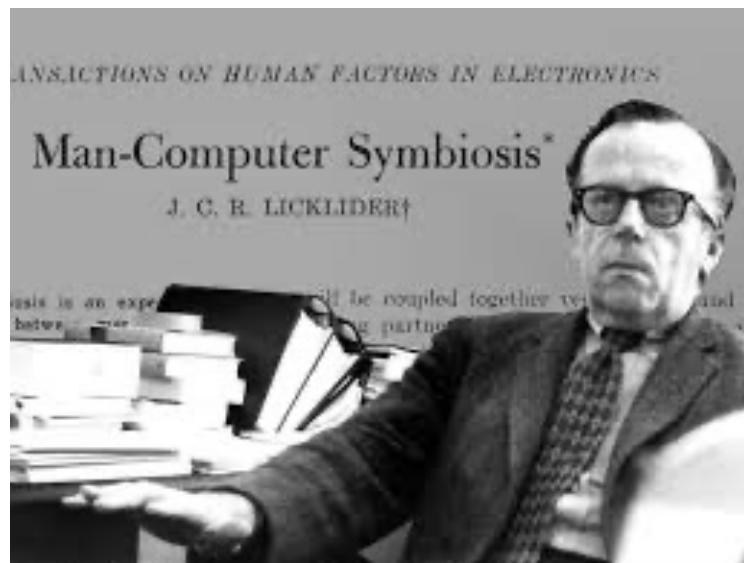
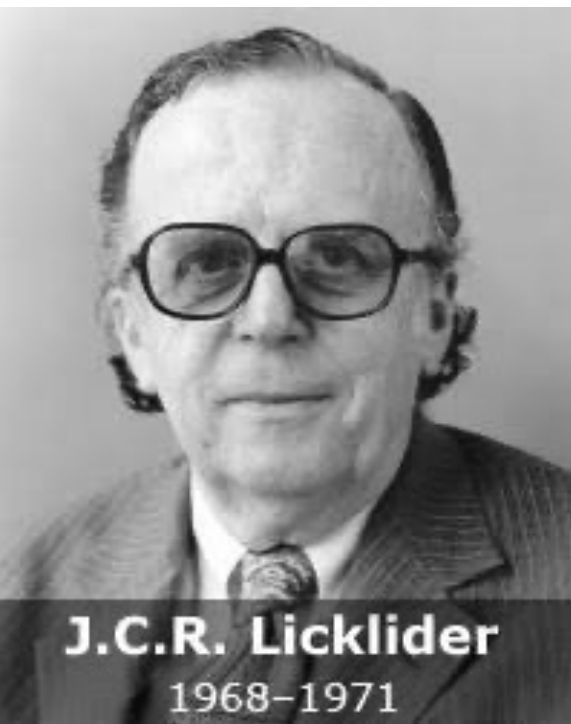


See also:
*Exploratory
Search: Beyond
the Query-
Response
Paradigm*, by
Ryen W. White,
Resa A. Roth,
Morgan &
Claypool, 2009

See also works
on: HCIR,
Information
Seeking
Behavior, Info-
Visualization,
Probing, Delving,
Investigation,
Discovery, Data
Analysis/Science

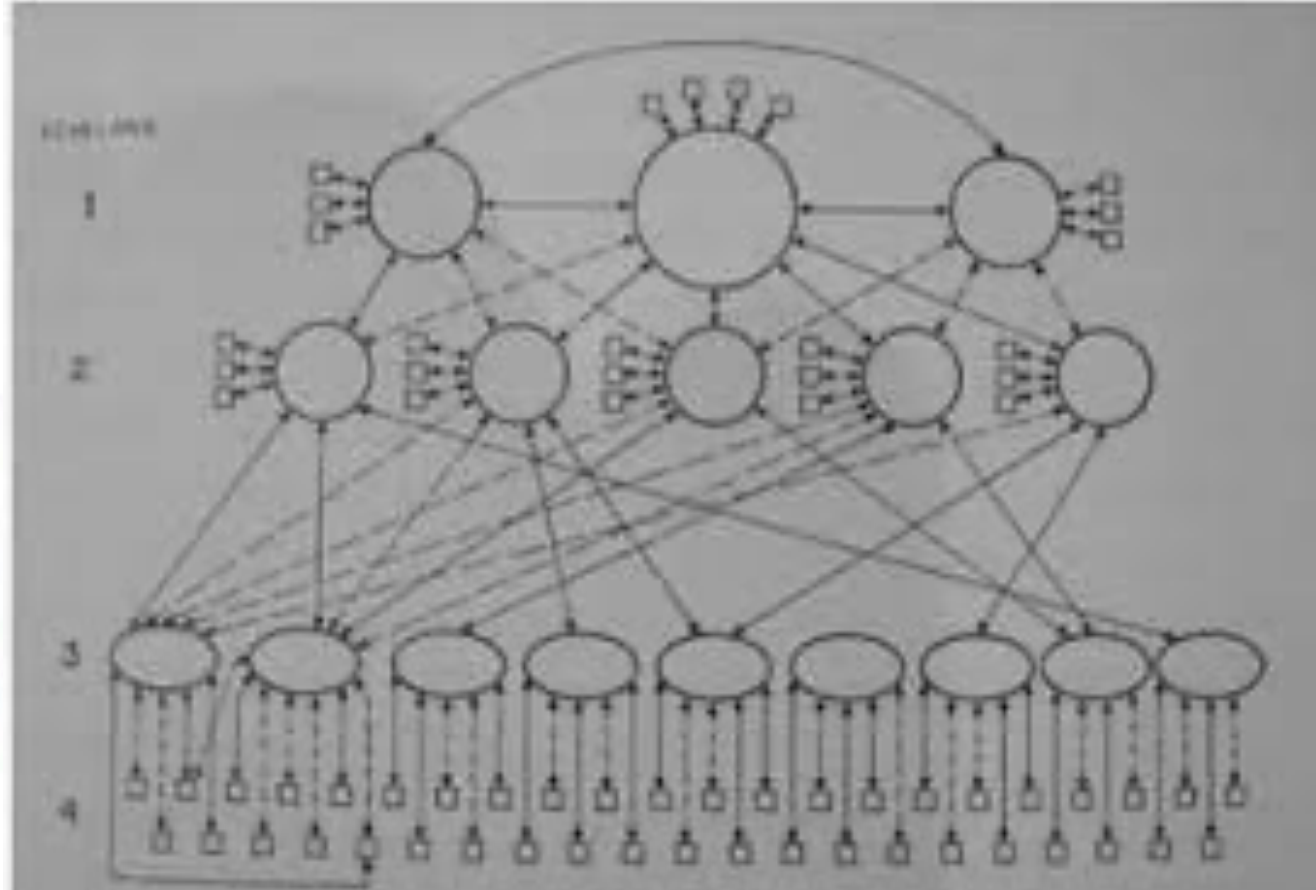
Outline

- **JCR Licklider**
- **Libraries of the Future**
- **5S**
- **Building Digital Libraries**
- **Exploring (incl. from JCDDL 2006)**
- **Future**



1962

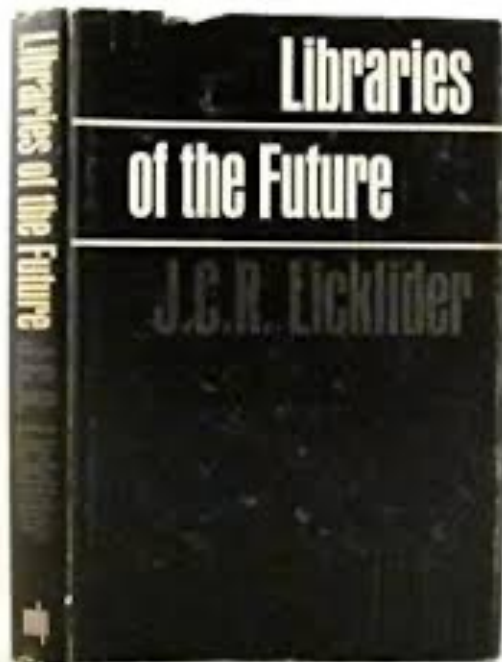
- J.C.R. Licklider of MIT discusses his "Galactic Network" concept in a series of memos
- a globally interconnected set of computers through which everyone could quickly access data and programs from any site.
- The first recorded description of the social interactions that could be enabled through networking
- Licklider later teamed up with ARPA
- the rest is history...



For: ARPA's Information Processing Techniques Office

JCR Licklider

- Grandfather of Internet
- Director, Project MAC: CTSS, Multics, AI Lab
- M. M. Waldrop's 2001 biography "The Dream Machine: J.C.R. Licklider and the Revolution That Made Computing Personal":
 - "He is almost alone in his conviction that computers can become not just superfast calculating machines, but joyful machines: tools that will serve as a new media of expression, inspirations to creativity, and gateways to a vast world of online information."



THE HOPE IS THAT, IN NOT TOO MANY YEARS, HUMAN BRAINS AND COMPUTING MACHINES WILL BE COUPLED TOGETHER VERY TIGHTLY, AND THAT THE RESULTING PARTNERSHIP WILL THINK AS NO HUMAN BRAIN HAS EVER THOUGHT AND PROCESS DATA IN A WAY NOT APPROACHED BY THE INFORMATION-HANDLING MACHINES WE KNOW TODAY.

- J. C. R. LICKLIDER -

LIAQUOTES.COM

J.C.R Licklider
Proposed the need for HCI in 1960



Image from <http://www.cpedia.com/wiki/Licklider%20J>

We need to substitute for the book a device that will make it easy to transmit information without transporting material.

- J. C. R. Licklider

WIKIQUOTE.COM

Overview

- As JCR Licklider forecast in 1965 in his “Libraries of the Future” (report for CLR)
- Highly effective human-computer symbiosis in procognitive systems that mediate online interactions with knowledge
- Knowledge: acquisition, organization, use
- Analysis: files, documents, texts, words
- Marry information retrieval, question-answering, NLP
- Adaptive self-organization, associative chaining
- Leverage: set theory, spaces, functions, relations, predicate calculus, higher-order knowledge representations

Closing Challenge

- (4) A sympathetic, cooperative, verbal, community is a fundamental essential for the development of a sophisticated verbal mechanism. To develop complex language behavior . . .
- (5) no one seems likely to design or invent a formal system capable of automating sophisticated language behavior. The best approach ... to call for a formal base plus an overlay of experience gained in interaction with the cooperative verbal community.

Outline

- JCR Licklider
- Libraries of the Future
- **5S**
- Building Digital Libraries
- Exploring (incl. from JCDDL 2006)
- Future

5S Motivation

- DLs are not benefiting from formal theories as have other CS fields: DB, IR, PL, etc.
- DL construction: difficult, ad-hoc, lacking support for tailoring/customization
- Conceptual modeling, requirements analysis, and methodological approaches are rarely supported in DL development.
- See my 9/2004 invited talk in Beijing “Digital Libraries for Education: Foundations to Case Studies”
<http://fox.cs.vt.edu/talks/2004/200409BeijingDL.ppt>



MORGAN & CLAYPOOL PUBLISHERS

Theoretical Foundations for Digital Libraries

*The 5S (Societies, Scenarios, Spaces,
Structures, Streams) Approach*

Edward A. Fox
Marcos André Gonçalves
Rao Shen

*SYNTHESIS LECTURES ON INFORMATION
CONCEPTS, RETRIEVAL, AND SERVICES*

Gary Marchionini, *Series Editor*

**Ch. 2: Exploration;
App. D2 pp. 91-100**



MORGAN & CLAYPOOL PUBLISHERS

Key Issues in Digital Libraries

Integration and Evaluation

Rao Shen
Marcos André Gonçalves
Edward A. Fox

*SYNTHESIS LECTURES ON INFORMATION
CONCEPTS, RETRIEVAL, AND SERVICES*

Gary Marchionini, *Series Editor*



MORGAN & CLAYPOOL PUBLISHERS

Digital Library Technologies

*Complex Objects, Annotation,
Ontologies, Classification,
Extraction, and Security*

Edward A. Fox
Ricardo da Silva Torres

*SYNTHESIS LECTURES ON INFORMATION
CONCEPTS, RETRIEVAL, AND SERVICES*

Gary Marchionini, *Series Editor*



MORGAN & CLAYPOOL PUBLISHERS

Digital Libraries Applications

*CBIR, Education, Social Networks,
eScience/Simulation, and GIS*

Edward A. Fox
Jonathan P. Leidig

*SYNTHESIS LECTURES ON INFORMATION
CONCEPTS, RETRIEVAL, AND SERVICES*

Gary Marchionini, *Series Editor*

Definition: Digital Libraries are complex systems that

1. help satisfy info needs of users (societies)
2. provide info services (scenarios)
3. organize info in usable ways (structures)
4. present info in usable ways (spaces)
5. communicate info with users (streams)

5S Model: Definitions

5S	Definition
Streams	Sequences of elements of an arbitrary type
Structures	Labeled directed graphs
Spatial	Sets and operations on those sets
Scenarios	<i>Sequences of events</i> that modify <i>states</i> of a computation in order to accomplish some functional requirement.
Societies	Sets of communities and relationships among them

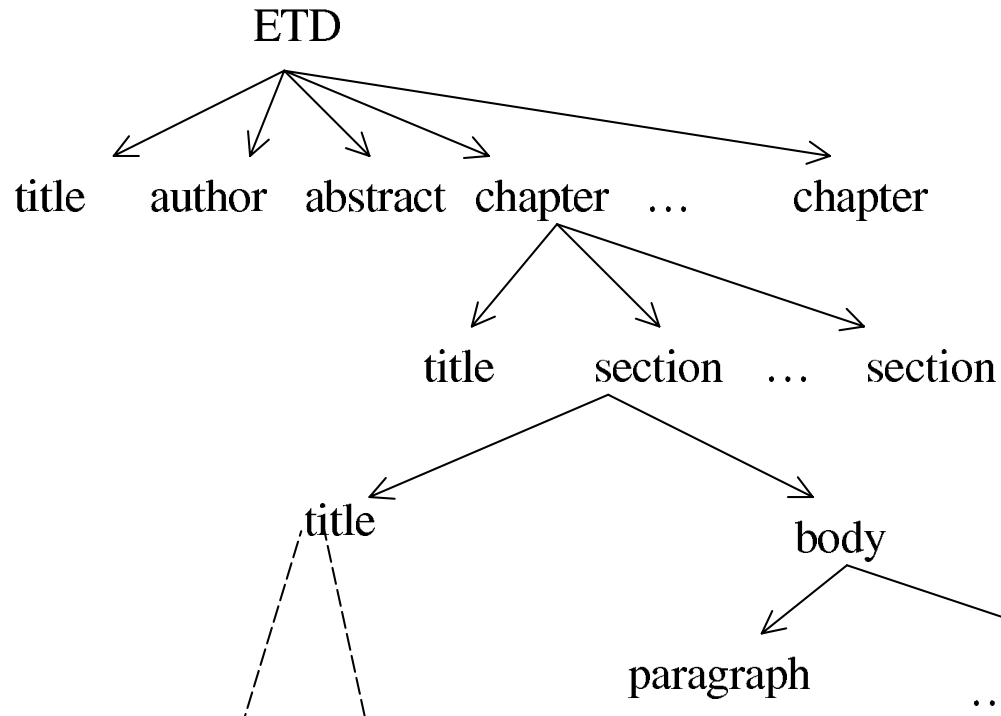
5S for DLs

Compare:
5 elements
(China)

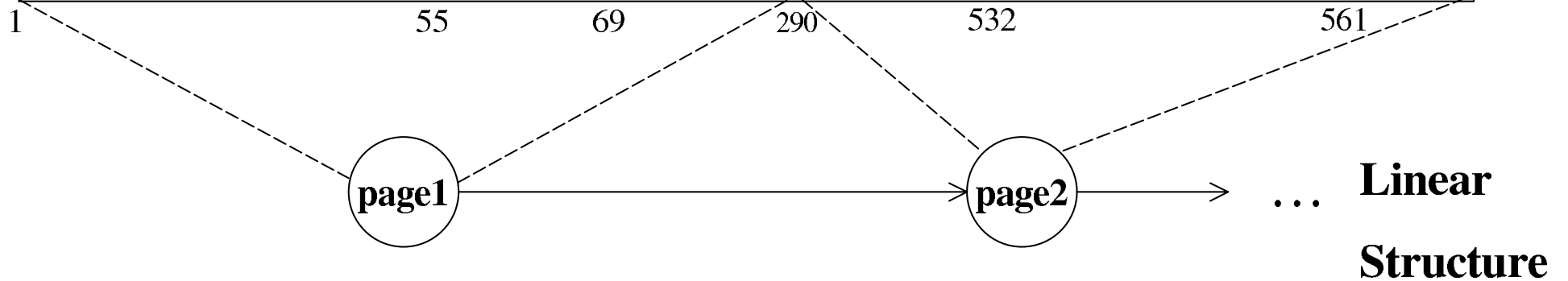
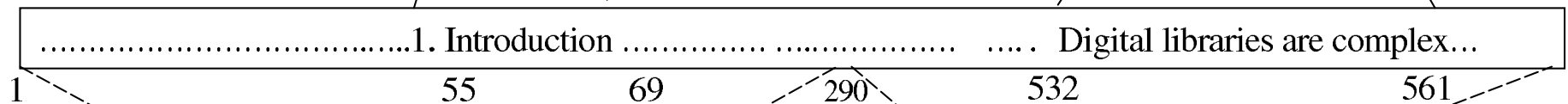


Ss	Examples	Objectives
Streams	Text; video; audio; image	Describes properties of the DL content such as encoding and language for textual material or particular forms of multimedia data
Structures	Collection; catalog; hypertext; document; metadata	Specifies organizational aspects of the DL content
Spaces	Measure; measurable, topological, vector, probabilistic	Defines logical and presentational views of several DL components
Scenarios	Searching, browsing, recommending	Details the behavior of DL services
Societies	Service managers, learners, teachers, etc.	Defines managers, responsible for running DL services; actors, that use those services; and relationships among them

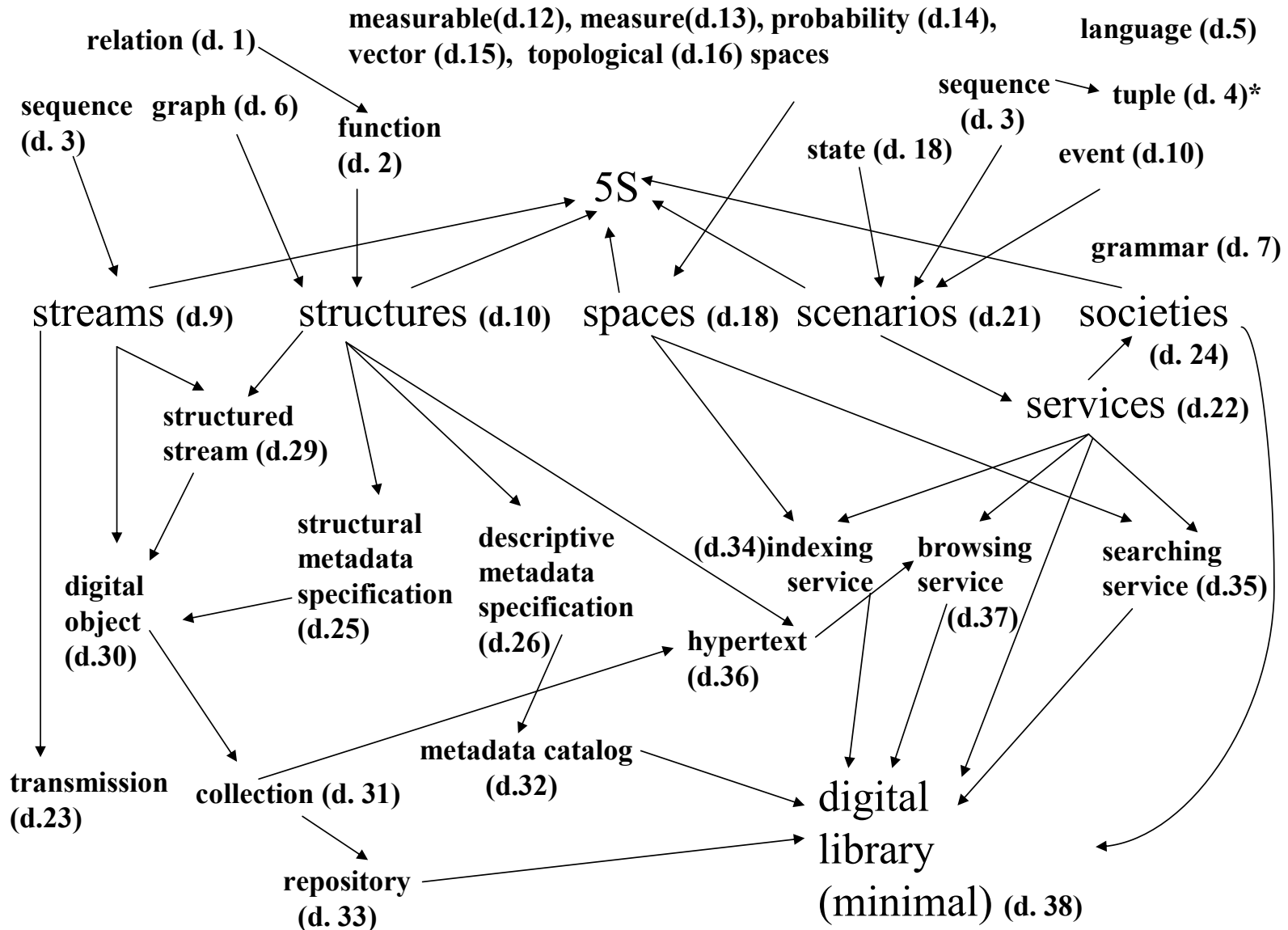
Hierarchical Structure



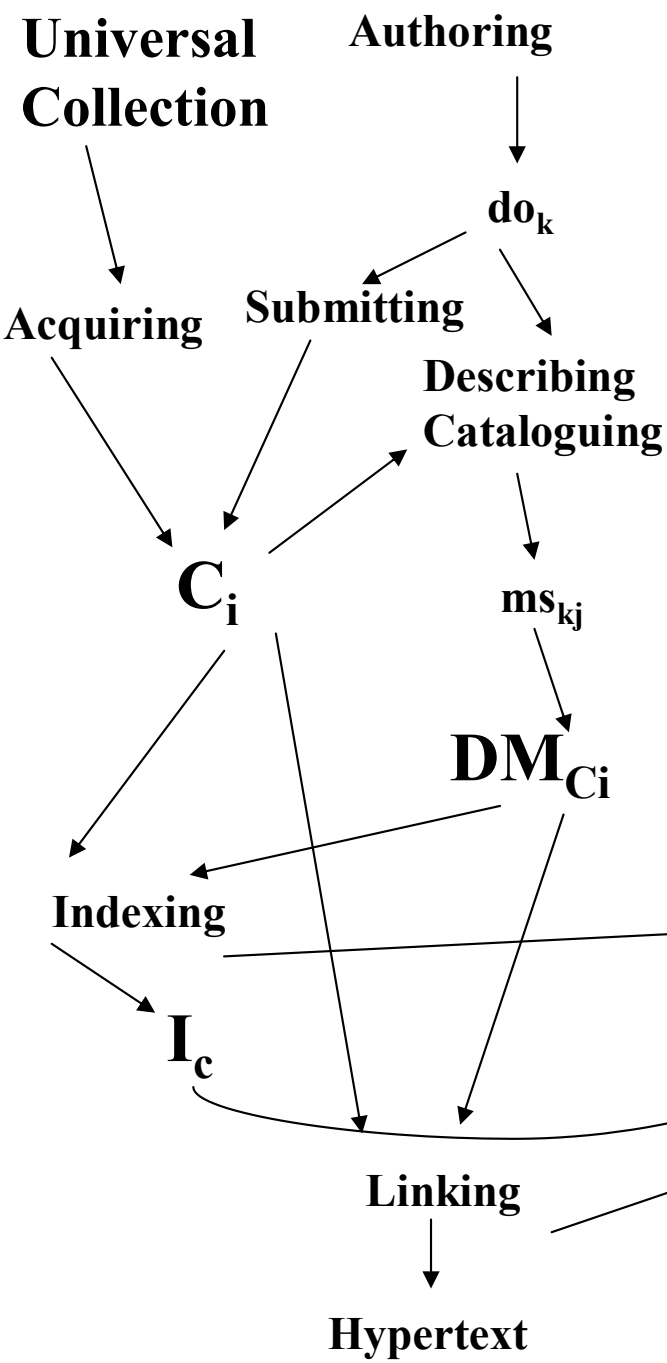
Textual Stream



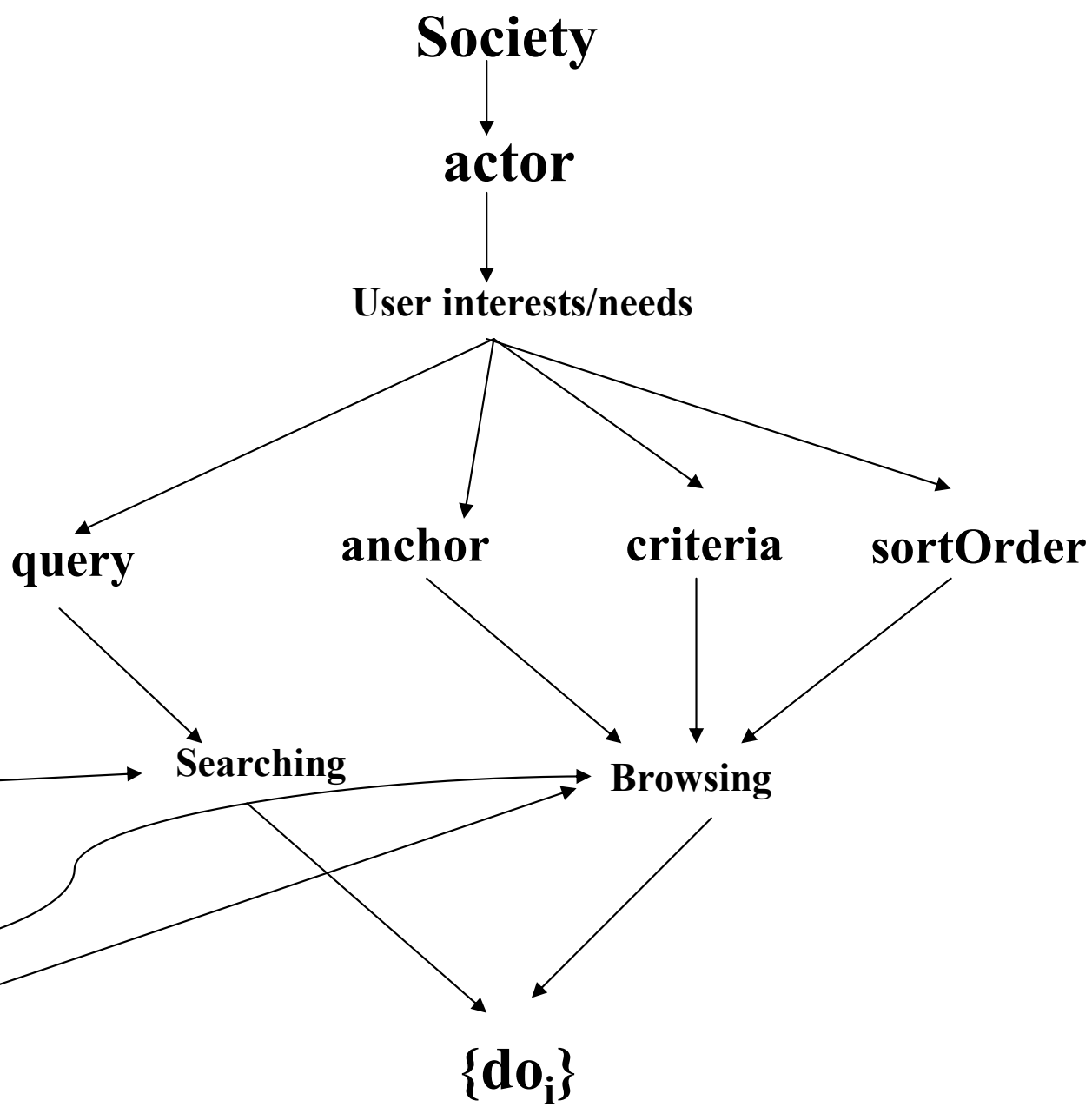
Overview of 5S and DL formal definitions and compositions

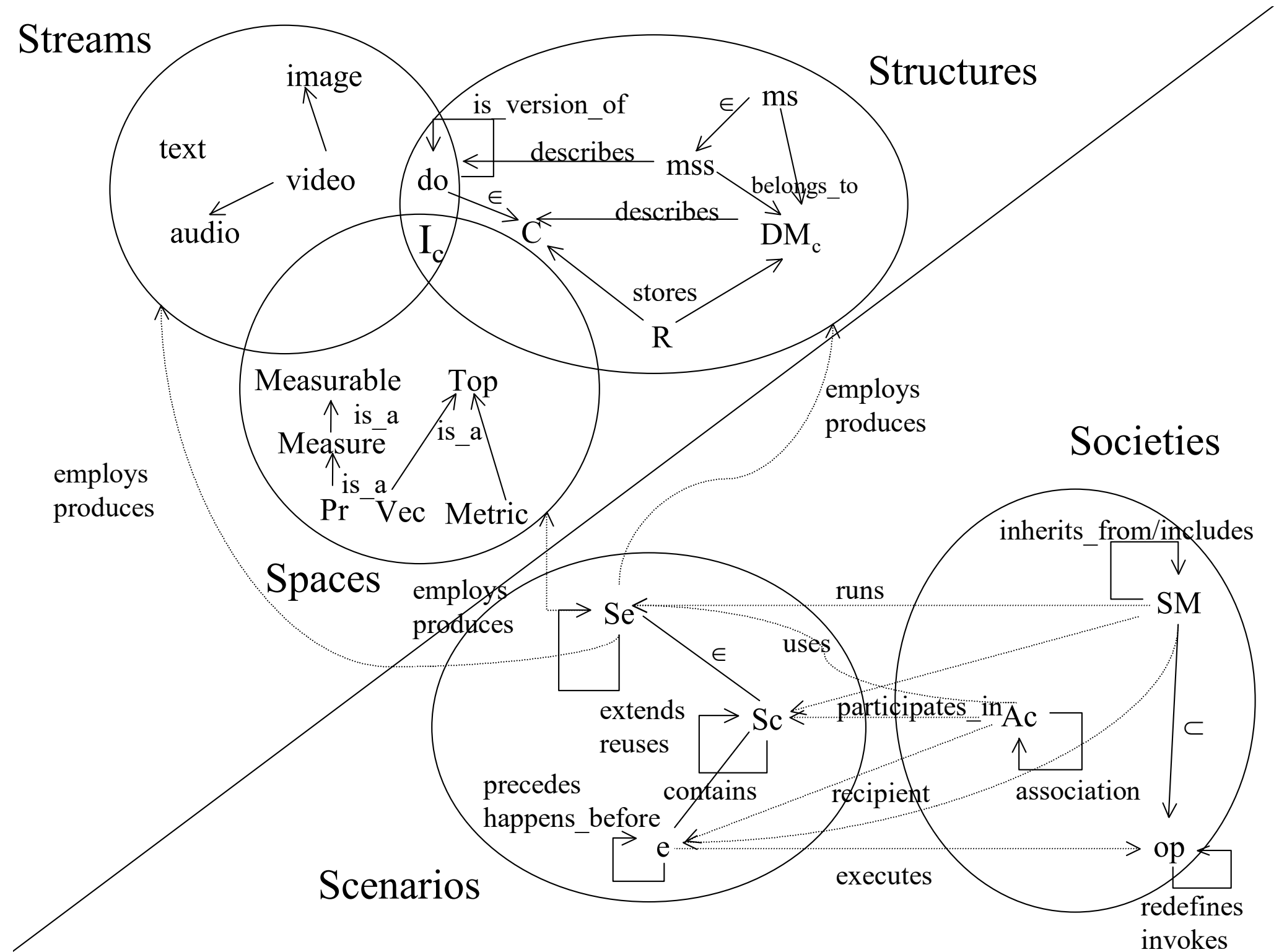


Infra-structure Services (fundamental)



Information Satisfaction Services (fundamental)





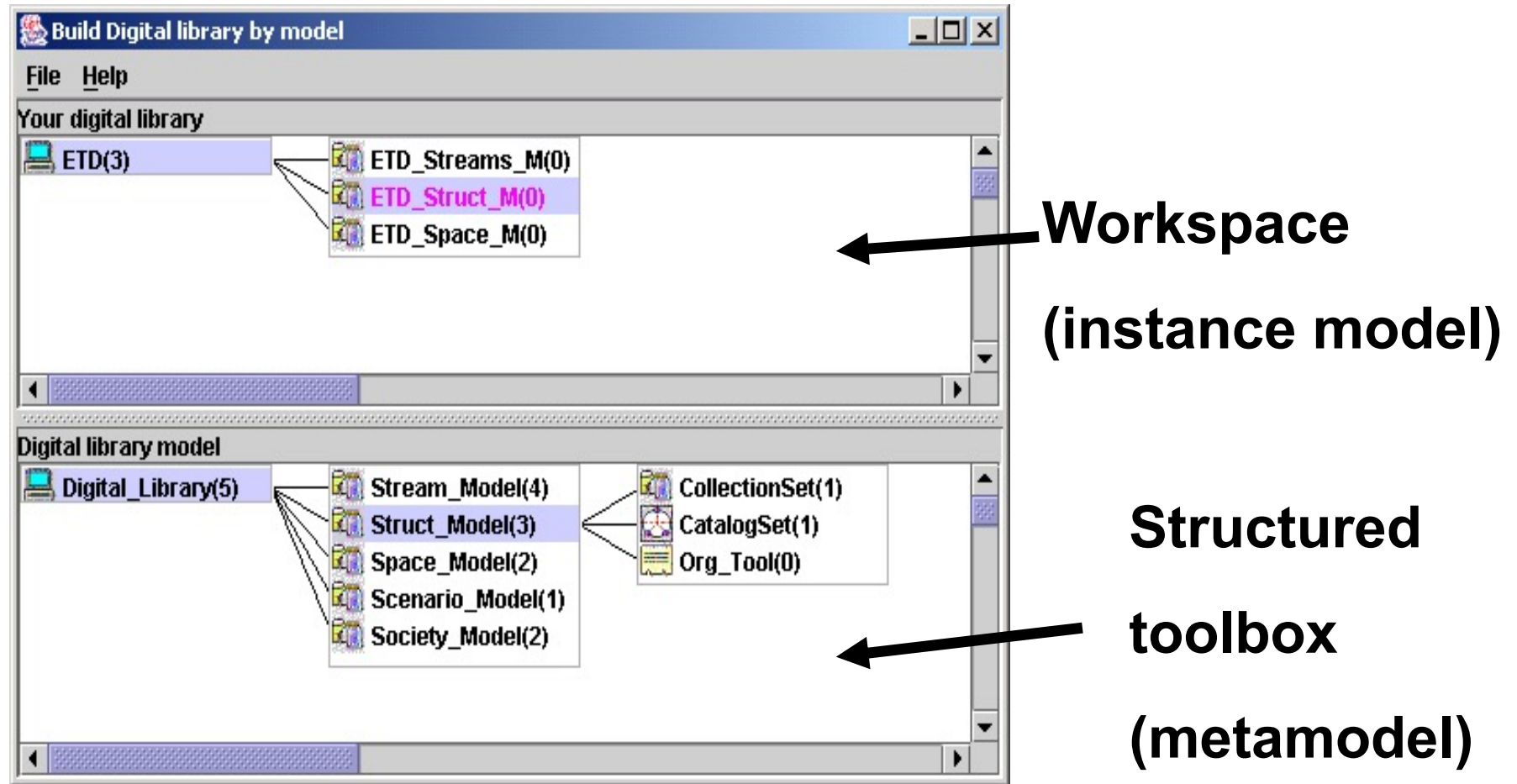
Outline

- JCR Licklider
- Libraries of the Future
- 5S
- **Building Digital Libraries**
- Exploring (incl. from JCDDL 2006)
- Future

DL Services/Activities Taxonomy

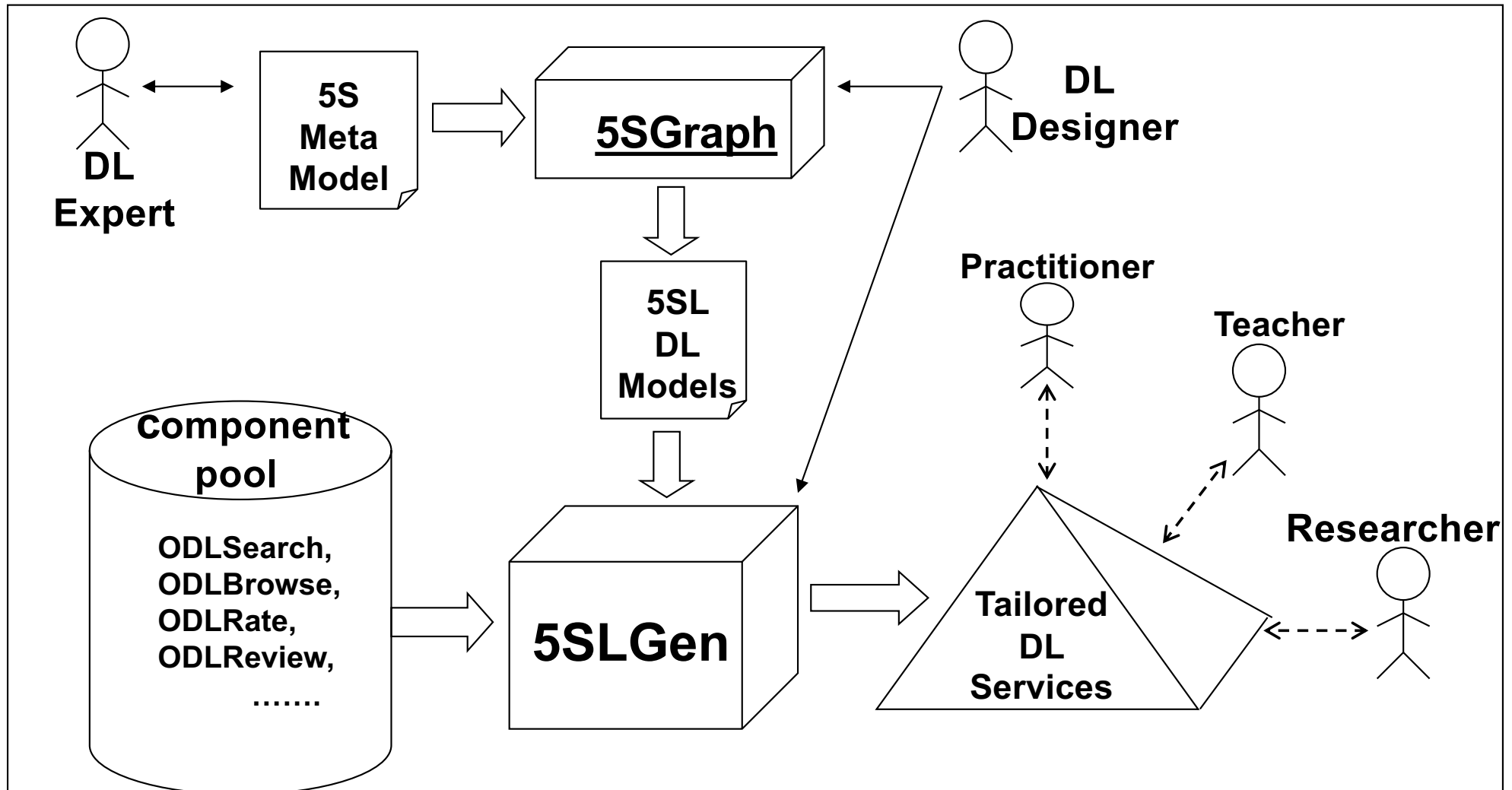
Infrastructure Services			Information Satisfaction Services
<i>Repository-Building</i>		<i>Add Value</i>	
<u>Creational</u>	<u>Preservational</u>		
Acquiring Cataloging Crawling (focused) Describing Digitizing Federating Harvesting Purchasing Submitting	Conserving Converting Copying/Replicating Emulating Renewing Translating (format)	Annotating Classifying Clustering Evaluating Extracting Indexing Measuring Publicizing Rating Reviewing (peer) Surveying Translating (language)	Browsing Collaborating Customizing Filtering Providing access Recommending Requesting Searching Visualizing

5SGraph: No Code DL



See <https://www.gartner.com/en/documents/3956079/magic-quadrant-for-enterprise-low-code-application-platf> from August 2019: "By 2024, low-code application development will be responsible for more than 65% of application development activity."

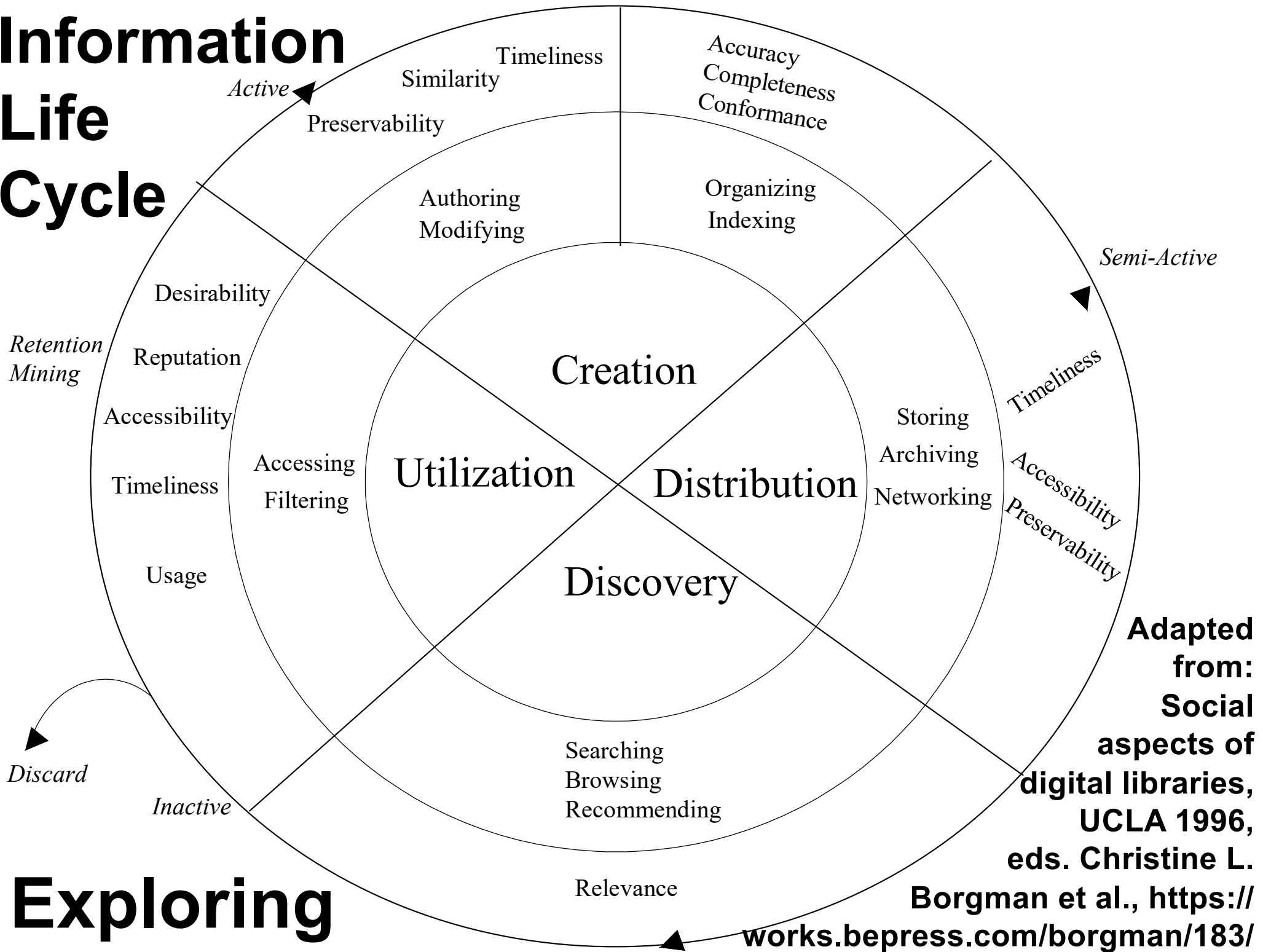
5S-based Architecture for DL Modeling and Generation



Outline

- JCR Licklider
- Libraries of the Future
- 5S
- Building Digital Libraries
- **Exploring (incl. from JC DL 2006)**
- Future

Information Life Cycle



Exploring

Exploring Digital Libraries: Integrating Browsing, Searching, and Visualization

JCDL 2006, Chapel Hill, NC, June 12, 2006

**Rao Shen, Naga Srinivas Vemuri, Weiguo Fan,
Ricardo da S. Torres, and Edward A. Fox**

DOI 10.1145/1141753.1141755

<http://fox.cs.vt.edu/talks/2006/20060612JCDL06exploring.ppt>

Introduction

- **What's exploring?**
 - ◆ **searching, browsing, investigating, studying, or analyzing**
 - ◆ **for purposes of discovery,**
 - ◆ **pursuing truth or facts about something**

Introduction (Cont.)









- **Are browsing and searching duals or can they be converted to each other when certain conditions are met?**
- **Can we generalize these DL exploring services within a formal DL framework?**
- **Can the formal generalization guide development of exploring services for domain focused DLs?**

Exploring Services Formalization



- ***Within the 5S Framework***
- ***Generalize DL exploring services such as browsing, searching, clustering, and visualization.***
- ***Prove theorems and lemmas based on the generalization***
 - ◆ ***searching \longleftrightarrow browsing***

Theorems & Lemmas related to Operations

Theorems and Lemmas	Searching Op_s	Browsing Op_b	Clustering Op_{clu}	Visualization Op_{viz}
 Theorem 1	✓			
 Theorem 2		✓		
 Theorem 3 (Op_s followed by Op_{clu})	✓		✓	
 Theorem 4 (Op_s followed by Op_{viz})	✓			✓
 Lemma 1	✓	✓		
 Lemma 2	✓	✓		
 Lemma 3 (Op_b followed by Op_s)	✓	✓		
 Lemma 4 (Op_s followed by Op_b)	✓	✓	✓	

Theory-based approach to describing DL Exploring Services

— guides us to design and implement
exploring services for ETANA-DL

- • Multi-dimensional browsing
- • Searching and browsing integration
- • Visualization
- • Usability evaluation

Evaluation of ETANA-DL services (N=28, Scale 0-5)

Browse	Search	EtanaViz	Save navigation path (SNP)	Search within browsing context (SWBC)
4.0	4.0	4.0	4.5	4.5

Conclusions

- **Approach DL exploring services based on a DL theory.**
- **Develop theorems indicating browsing and searching can be converted and switched to each other under certain conditions.**
- **Provide a systematic and functional method to design and implement DL exploring services in an integrated archaeological DL, ETANA-DL, which was used as a case study.**
- **Made contributions to aid both users and developers of DLs.**

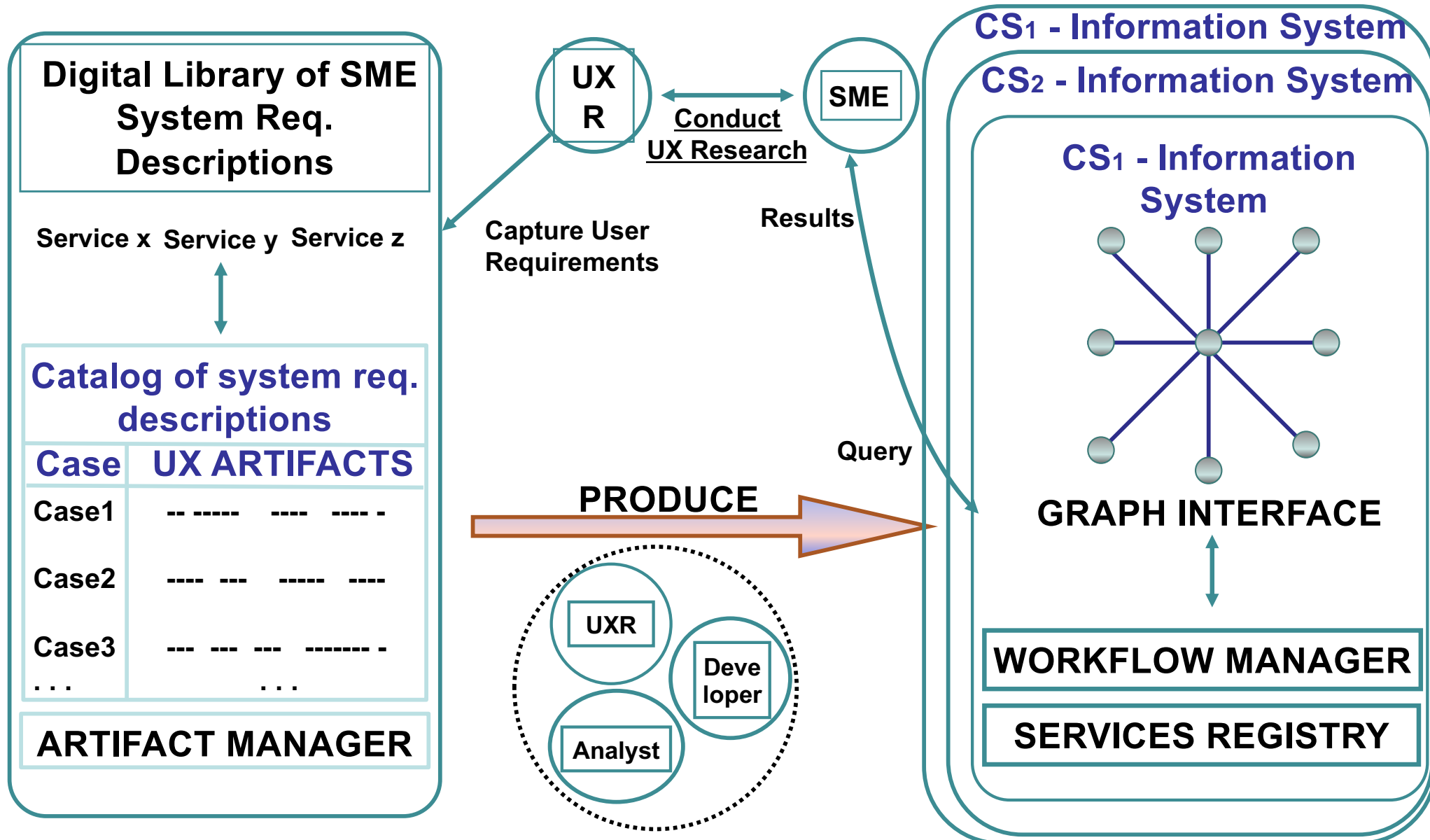
Outline

- JCR Licklider
- Libraries of the Future
- 5S
- Building Digital Libraries
- Exploring (incl. from JCDDL 2006)
- **Future**

Working toward the Future

1. Collecting information for DS
2. Curating and archiving content (IS)
3. Text and data analysis (DS)
4. Discovering personas, customers (HCI)
5. Mapping goals to tasks to services (UX, SMEs, DevOps) -> knowledge representation
6. Dynamically solving information needs (IR)
7. Semi-automatically continuously improving the information system (DS, IR, IS, HCI)

Future: Prashant Chandrasekar's Architecture of DL for SME Exploration



Future

1. Facilitate curiosity and wonder
2. Aid learning, discovery, and leveraging
3. Promote truth and understanding
4. Enable collaboration in the small and large
5. Support specialization and synthesis
6. Meet short and long term goals, with history
7. Tailor to groups, personalize to individuals
8. Cover the full information life cycle

Summary

- JCR Licklider
- Libraries of the Future
- 5S
- Building Digital Libraries
- Exploring (incl. from JCDDL 2006)
- Future

Thank You!

fox@vt.edu

Questions? Discussion?

Upcoming related publications:

- Edward A. Fox and Prashant Chandrasekar. 2021. How Should One Explore the Digital Library of the Future? Data and Information Management, Volume 5, in press, <https://sciendo.com/journal/DIM>
- Edward A. Fox. 2021. Building and Using Digital Libraries for Electronic Theses and Dissertations. Journal of Electronic Theses and Dissertations (J-ETD), Volume 1, in press, <http://j-etd.org>