

Integrating multi-dimensional information spaces

Kostas Saidis, Alex Delis
{saiko, ad}@di.uoa.gr

University of Athens

2 Oct. 2009, Corfu, Greece

2nd Workshop on Very Large Digital Libraries
(VLDL 2009)

In conjunction with ECDL 2009

Size does not matter (1)

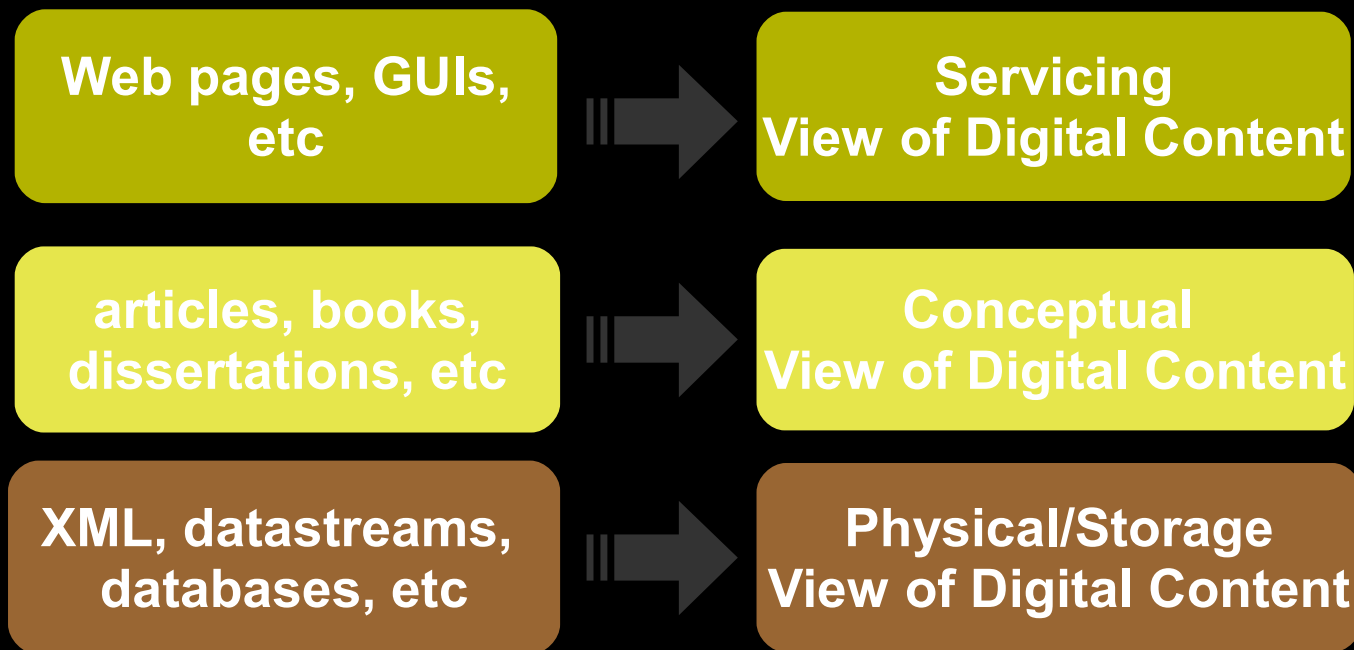
- We view Very Large DLs as systems that manage not only “large” but also “complex” information spaces
- Diverse, multi-faceted content items:
 - digitized and/or born digital intellectual works, institutional and/or personal archives, scholarly information, user-generated content
- Heterogeneous content sources:
 - databases, XML repositories
- Plethora of applications, services, use-cases

Our discussion

- Users need to share, reuse, refine and extend information in varying application contexts
- Can we supply VLDLs and related systems with a unified information space management infrastructure?
- Can this infrastructure add value by simplifying – and automating as highly as possible – the integration of diversely structured and heterogeneous information spaces?

Diverse views of information

- Different systems develop different views of digital content for different purposes.



Multi-dimensional Information Space Management

- Systems manage information in multiple dimensions, supporting diverse:
 - Information identification & discovery options
 - Information access options
 - Information conceptualization options
 - Information utilization options

Integration as a process (roughly)

1. Discovery: systems “learn about” the existence of each other
2. Identification: systems unambiguously identify their individual items
3. Access: systems access their items
4. Utilization: systems synthesize their items

Integration imposes extensions

- Realizing these steps requires dealing with a variety of information discovery, access, conceptualization and utilization options supported by involved systems
- Thus, when integrating information spaces, we practically need to extend involved systems in multiple crosscut and interdependent options
- Hard, cost-consuming, may require source-code modifications and/or system redesign

Integration requires automation

Information integration/interoperation is about “enabling information that originates in one context to be used in another in ways that are as highly automated as possible”

[The DOI Handbook, Edition 4.4.1, The International DOI Foundation]

Our point

- If we simplify the process of extending systems' multi-dimensional information management options
- We simplify the process of integrating their information spaces

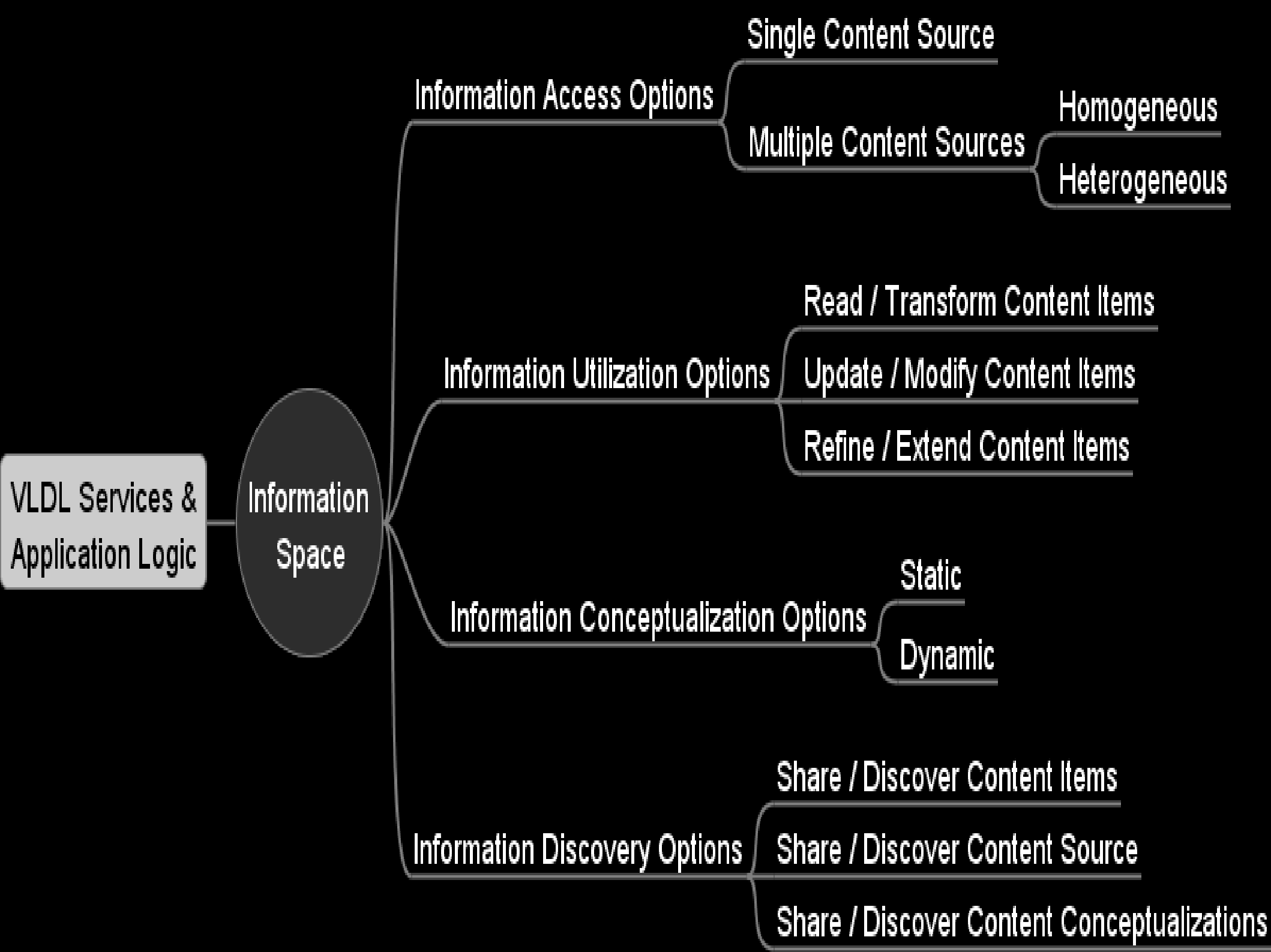
Simplify ~ automate as highly as possible

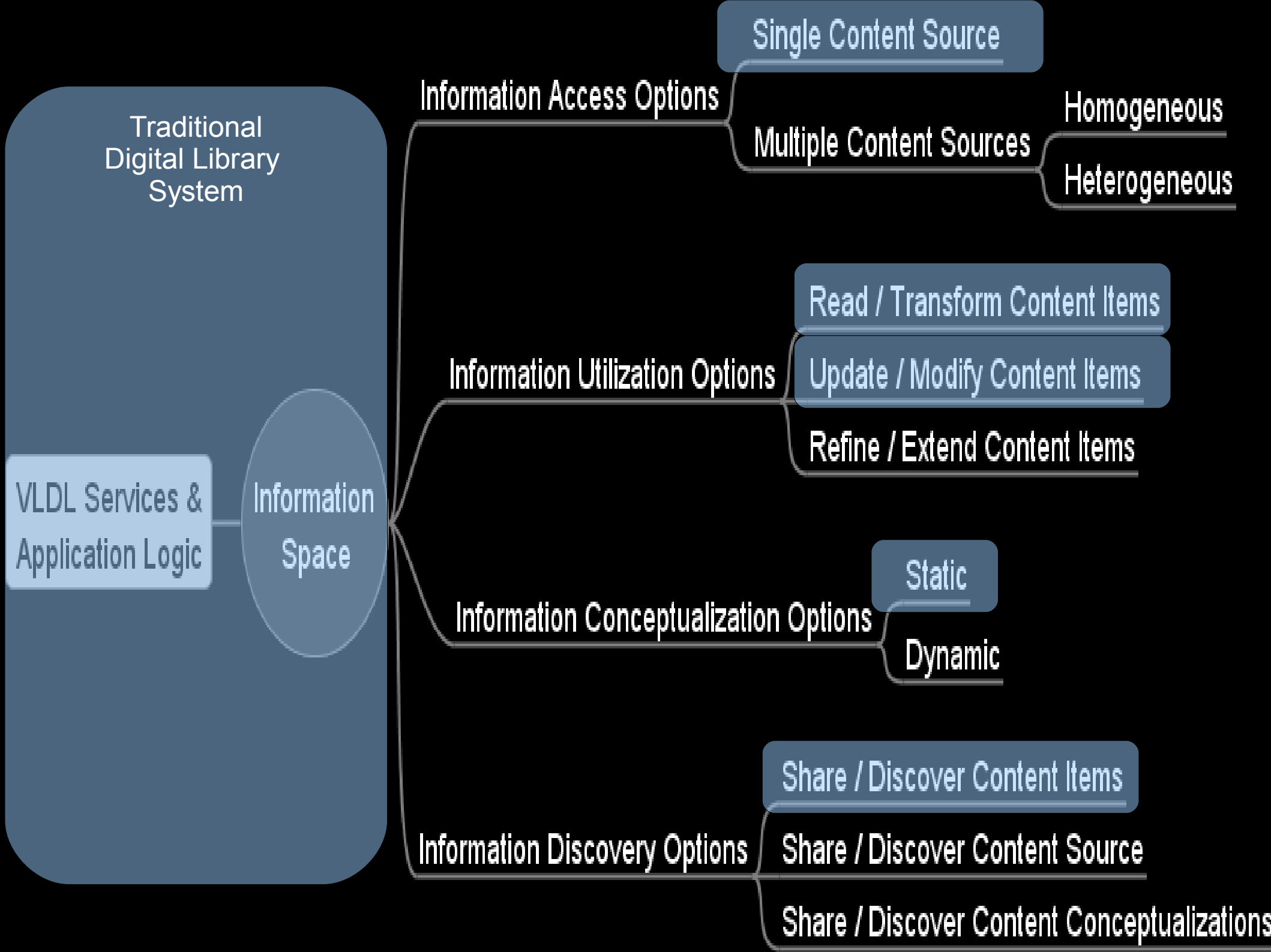
WWW: the largest interoperable information space

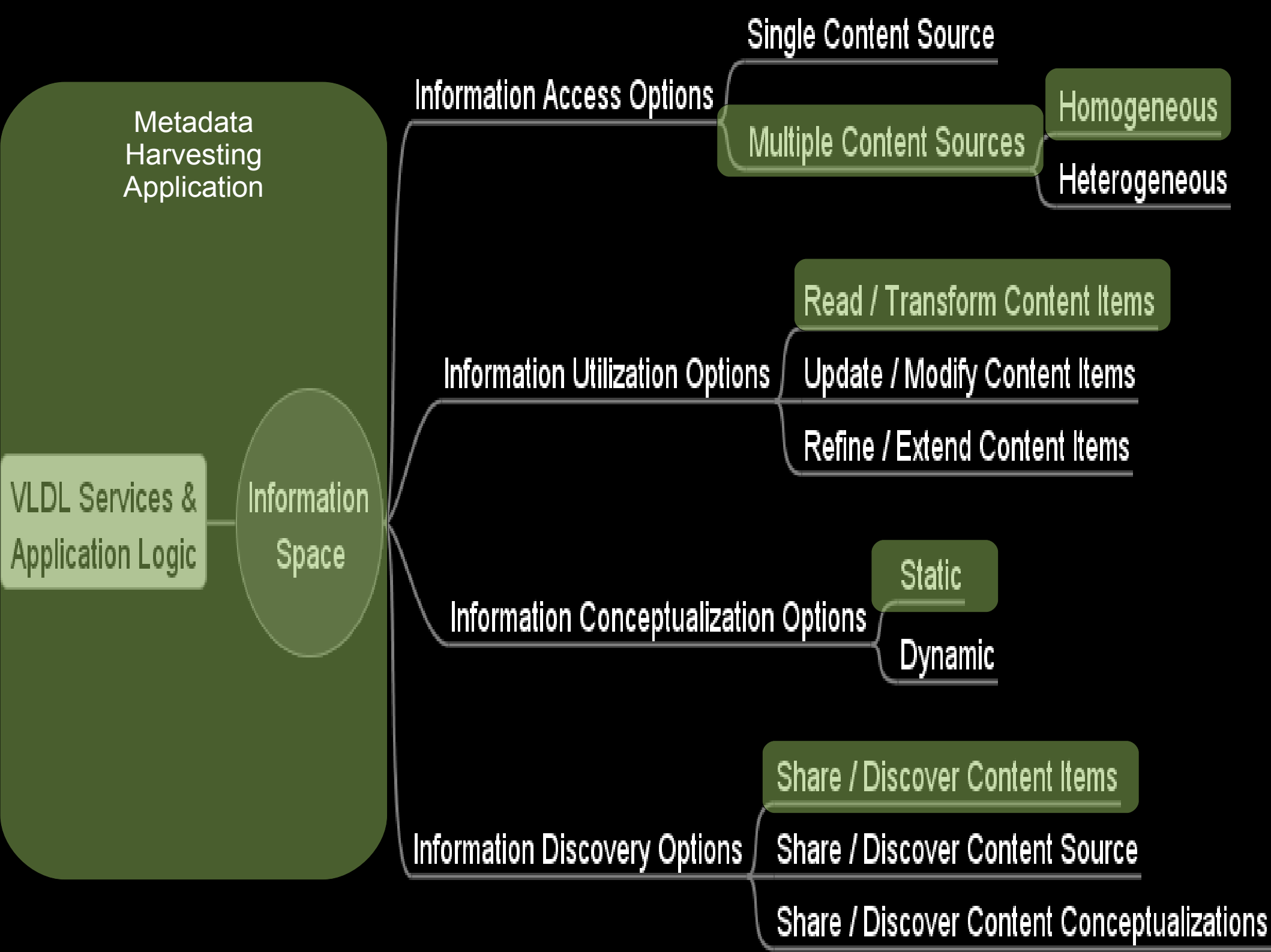
- Automates information identification and access (HTTP & URIs)
- Yet:
 - No built-in information discovery service (google)
 - A single “document-based” conceptualization
 - Information utilization follows a limited “publish/consume” paradigm
- Technologies such as Web Services & Semantic Web enhance limited information discovery, conceptualization and utilization options

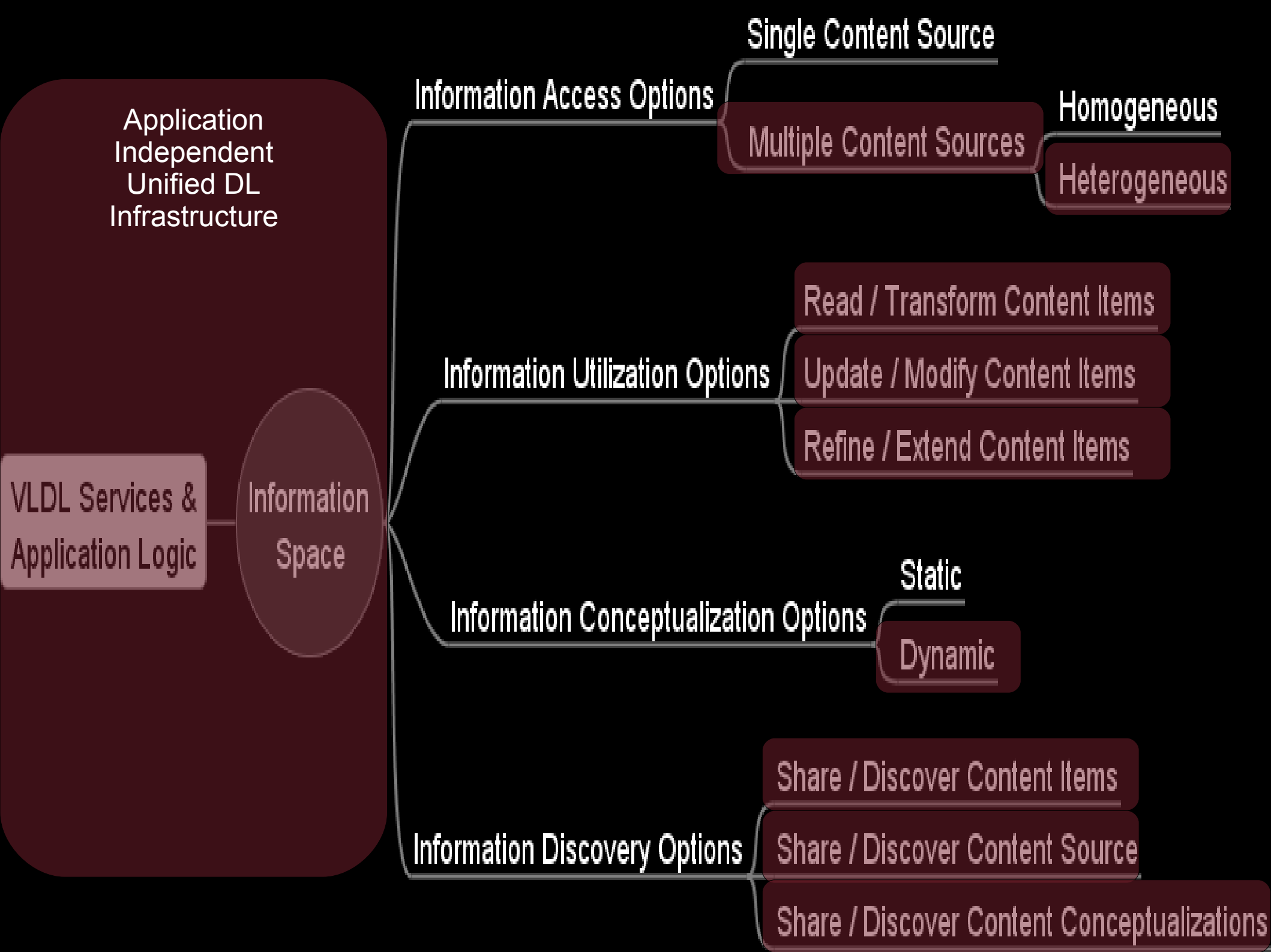
Size does not matter (2)

- Information integration/interoperation:
 - plays a crucial role in smaller-scale information spaces, too:
 - Digital libraries
 - Business & Enterprise Environments
 - Proprietary & Legacy systems
 - etc
 - is dominated by the information management options supported by involved systems









Infrastructure design (1)

- Content Source API:
 - allow systems to operate atop multiple heterogeneous sources
 - register new sources dynamically
 - use a driver-based technique
- Content Access/Update API:
 - read/modify actions that apply to any underlying content source

Infrastructure design (2)

- Content Conceptualization API:
 - support storage-independent, dynamic conceptualizations
 - Employ an inheritance mechanism to enable refinement / extension of content items
- Content Discovery API:
 - Provide 3 indexing/discovery facilities, for sharing:
 - Content items,
 - Content sources,
 - Content conceptualizations

</presentation>

