

COLLECTION DEVELOPMENT FOR THEMATIC WEB ARCHIVES

Pamela Graham
Columbia University

RESAW Conference
London, June 14-16, 2017

THE NATURE OF THE PROBLEM

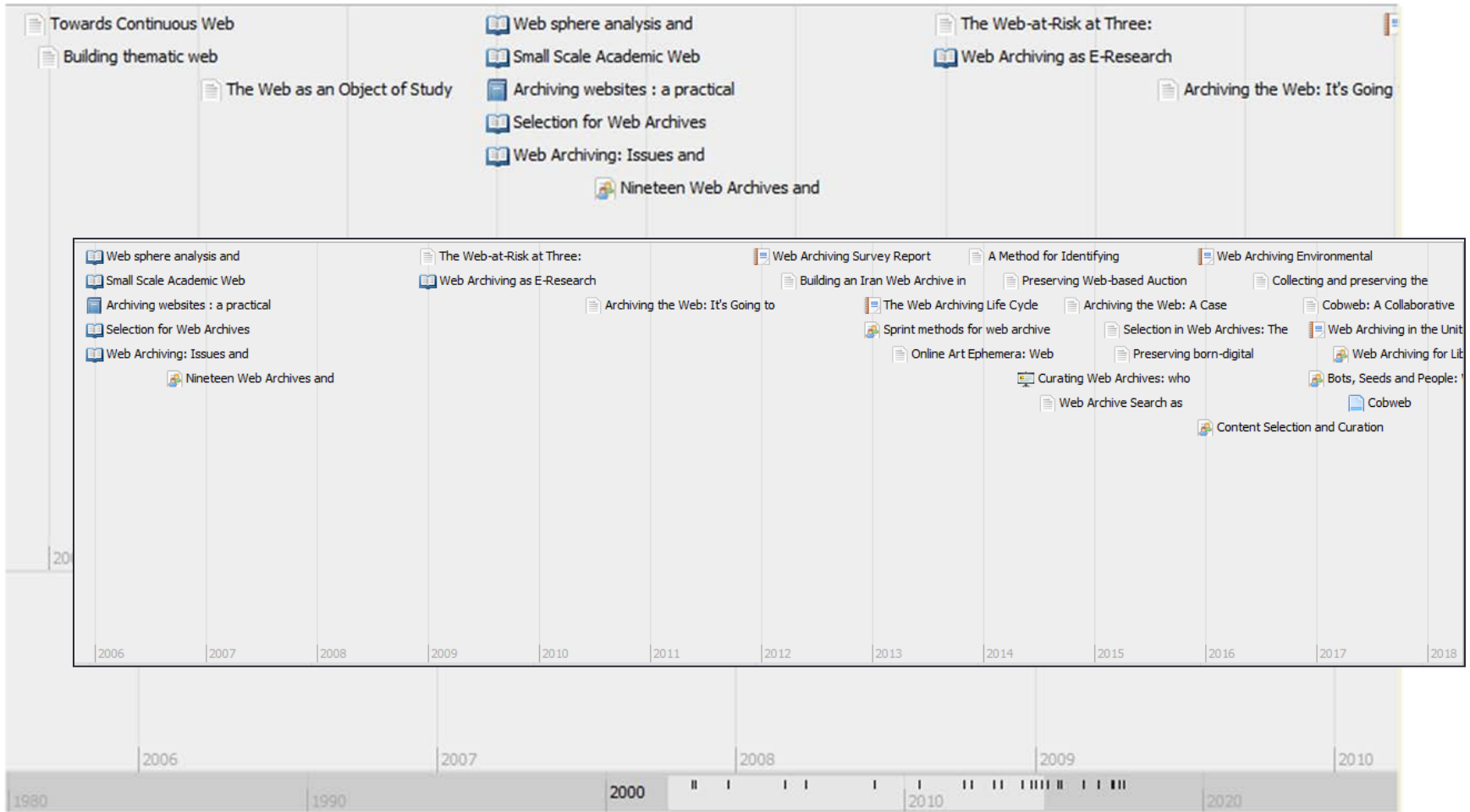
“ . . .we have not transcended the foundational preoccupation with the efficacy of our technical methods for Web capture, preservation and replay”

(Taylor 2017, p. 4)

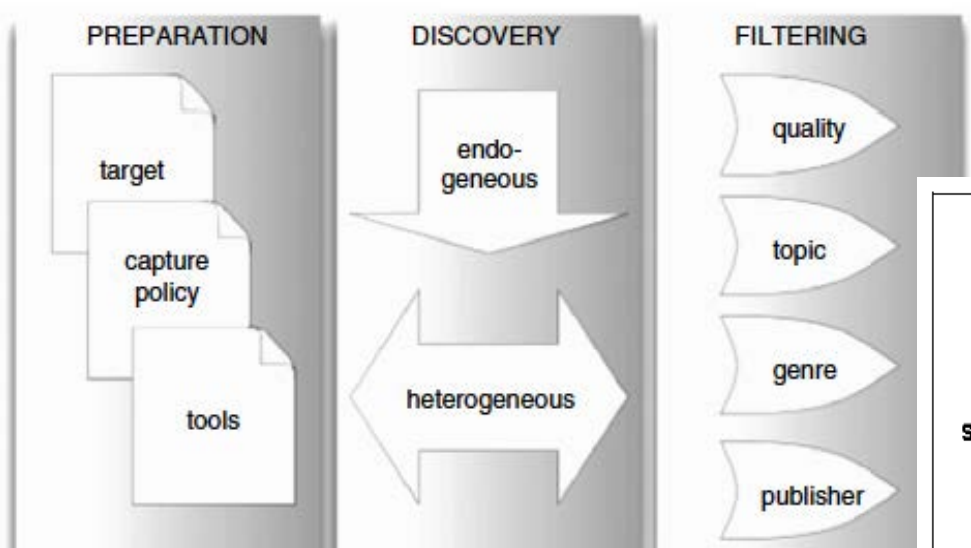
More on the problem

- Lack of communities of practice, shared knowledge and best practices
- Low profile of Collection Development
- Challenges of Selector engagement
 - “start-up” “first box”
- Underspecified in programmatic development

Sampling the literature



3.4 Selection Process



Masanès, (2006), p. 82

“Selection morphs from a human selection of discrete and stable units . . .to a more versatile and dynamic selection of paths to be followed with certain depth and time.” (Masanès, 2006, p. 74)

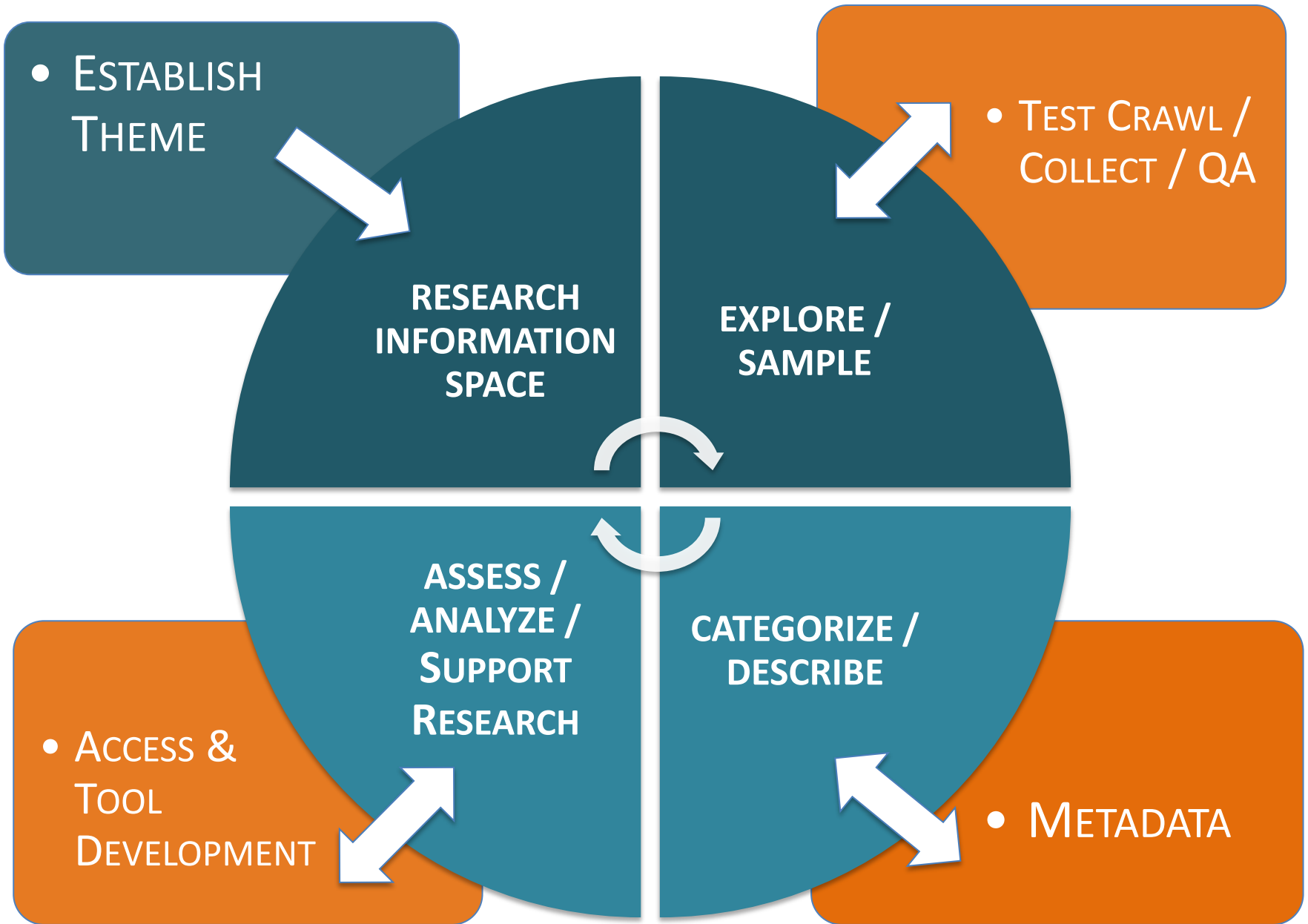
WEB CAPTURE LIFE CYCLE PROCESSES AND ELEMENTS				Who is engaged in the process		
				Service Unit	Web Capture Team	Both
PLAN	Collection Planning	Permissions Planning				
SELECT	Selection					
GET	Pre-crawl Technical Review	Crawl	Quality Review			
PREPARE/ ASSEMBLE	Store and Manage	Staff Access				
MAKE AVAILABLE	Researcher Access					
DESCRIBE	Cataloging					

Fig. 2. Putting it Together: Life Cycle Processes with Elements of Defined Web Capture Processes

Grotke, A., & Jones, G. (2006), p. 52.

Take-aways

- Phases of selection
 - Exploration, Discovery, Sense-Making
- Interconnectedness with subject and technical work
- Participation in the full life cycle
- Research, Assessment, Analysis

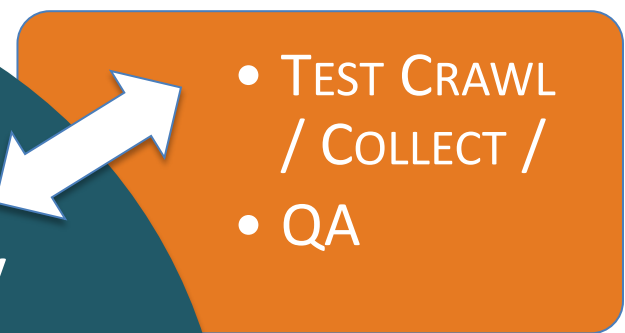


Thematic Web Archiving Selection Life Cycle

- ESTABLISH
THEME

- Nature of the web, as a primary source, platform and/or repository of knowledge in a subject area or topic
- Connection to existing collection programs and strengths?
- Applicability of existing collection development skills/knowledge
- Researcher needs and practices

Output: High-level statement documenting rationale and need for creating a collection.



- How do scholars scope and define the “web” in their studies of this field or topic
- “Exogenous sources” to guide the selection of sites
- Exclusions, what NOT to collect
- Scholarly practices and metadata schema and choices ?

- Degrees of “at-risk-ness” ?
- Scoping and frequency of capture?
- Archivability of sample websites?
- Legal and ethical issues in the collecting process; authority and trust of the web archive

Outputs: Seed lists, exogenous resources for selection and benchmarking; decisions about what NOT to collect; scoping, frequency and permissions policies.

Visualize, represent the archive in new ways; inform ongoing collecting and documentation

Support researcher use; identify collecting and access needs

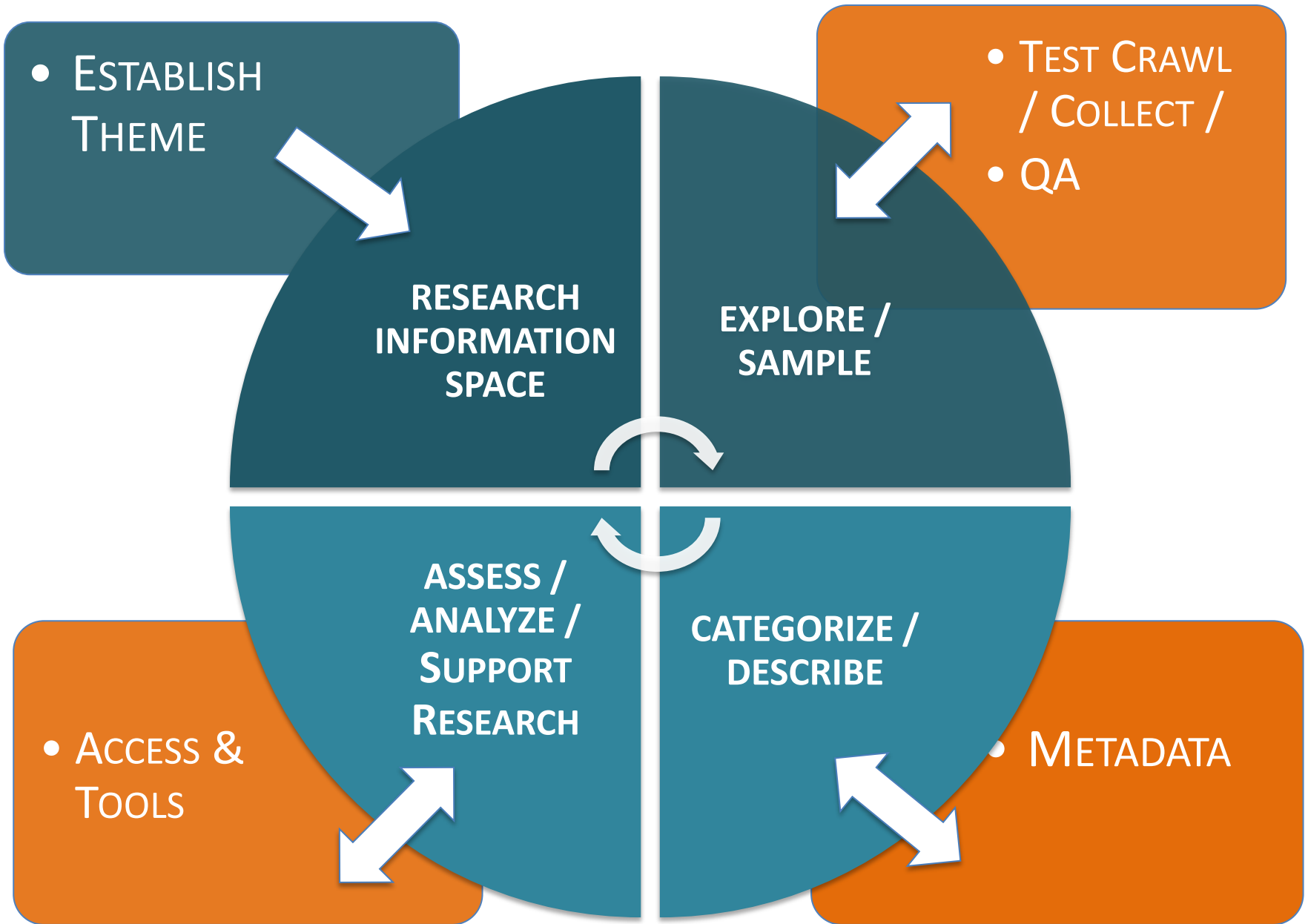
Monitor research generated through the use of web archives to inform collecting and access

Are there categories or typologies of websites emerging from the exploratory work?

How do these categories inform metadata decisions, access and ongoing collecting?



Outputs: Ongoing benchmarking and comprehensiveness metrics; Additional seeds; clearer documentation; use cases and input on access and discovery.



Thematic Web Archiving Selection Life Cycle