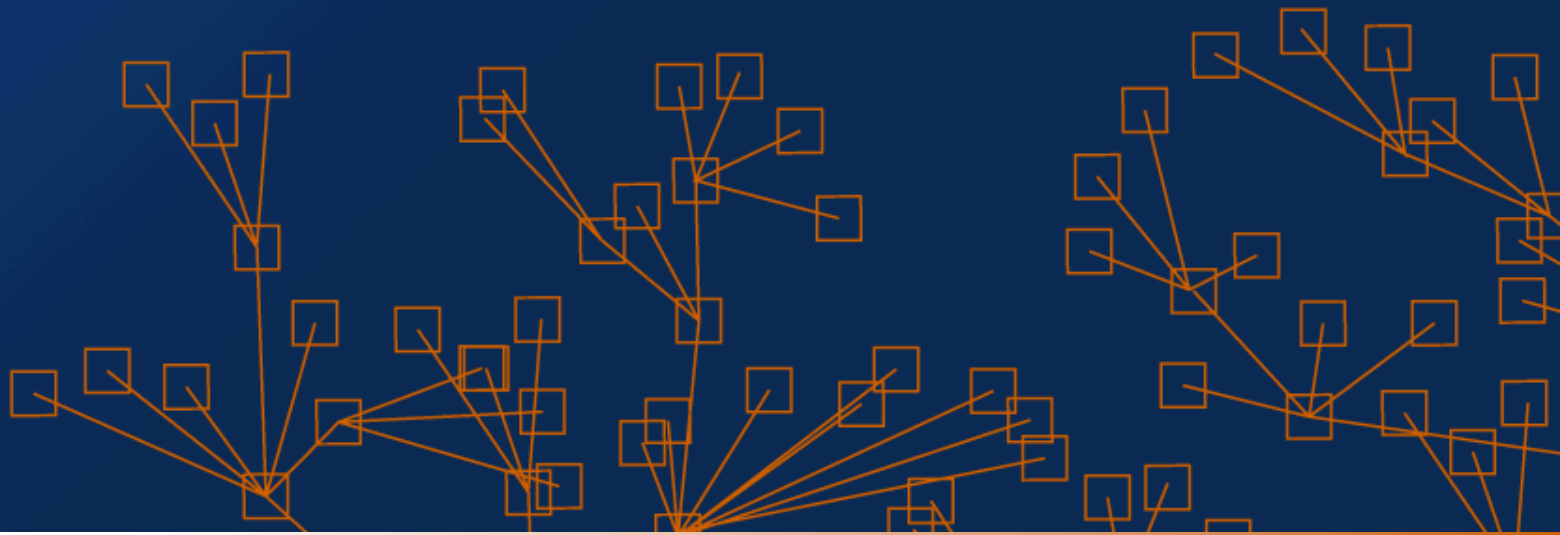




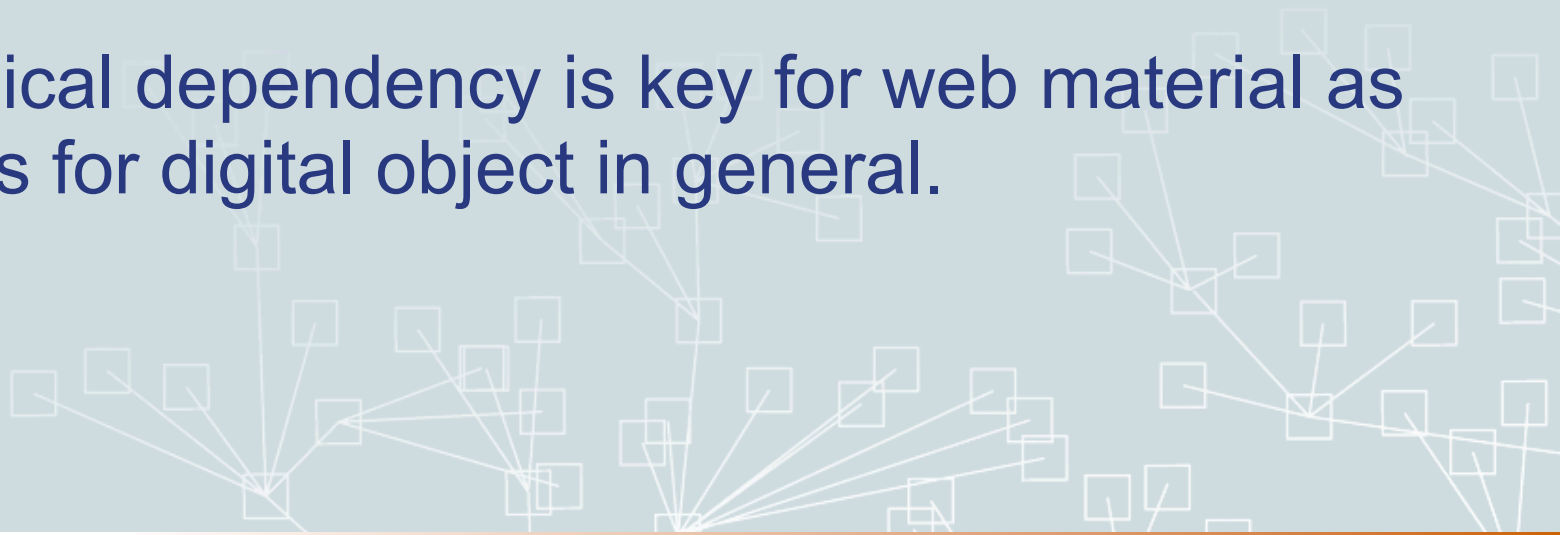
IIPC Web Archiving Metadata Set





What information is important for Web material?

- Non-isolated documents form together a network of content and description of content (docuverse)
 - less need for descriptive information
- Technical dependency is key for web material as well as for digital object in general.





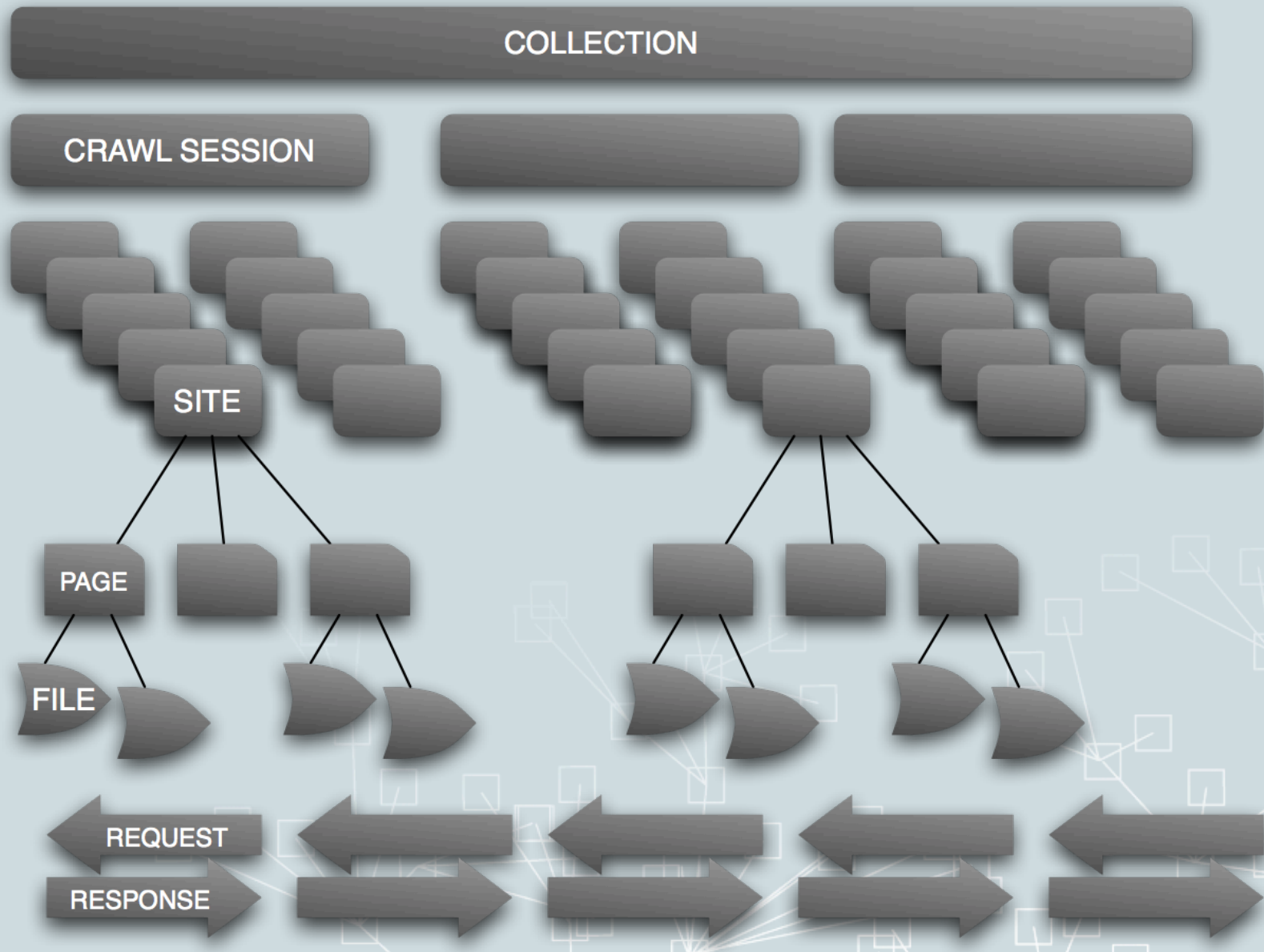
What information is important?

- Importance of selection (space/time) in WA
 - need to document it and how it has been embedded in tools (crawling tools & policy)
- each interaction with a web server can potentially generate a unique customized response
 - document the context of this interaction, or pseudo-transaction.



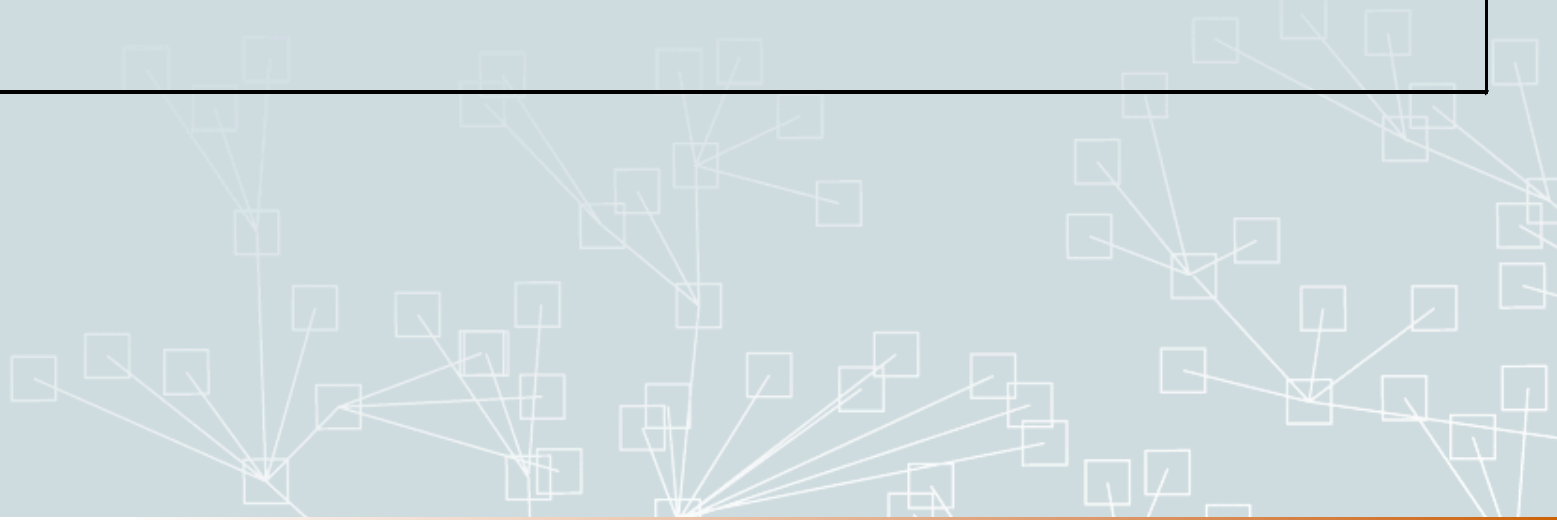


Metadata application level





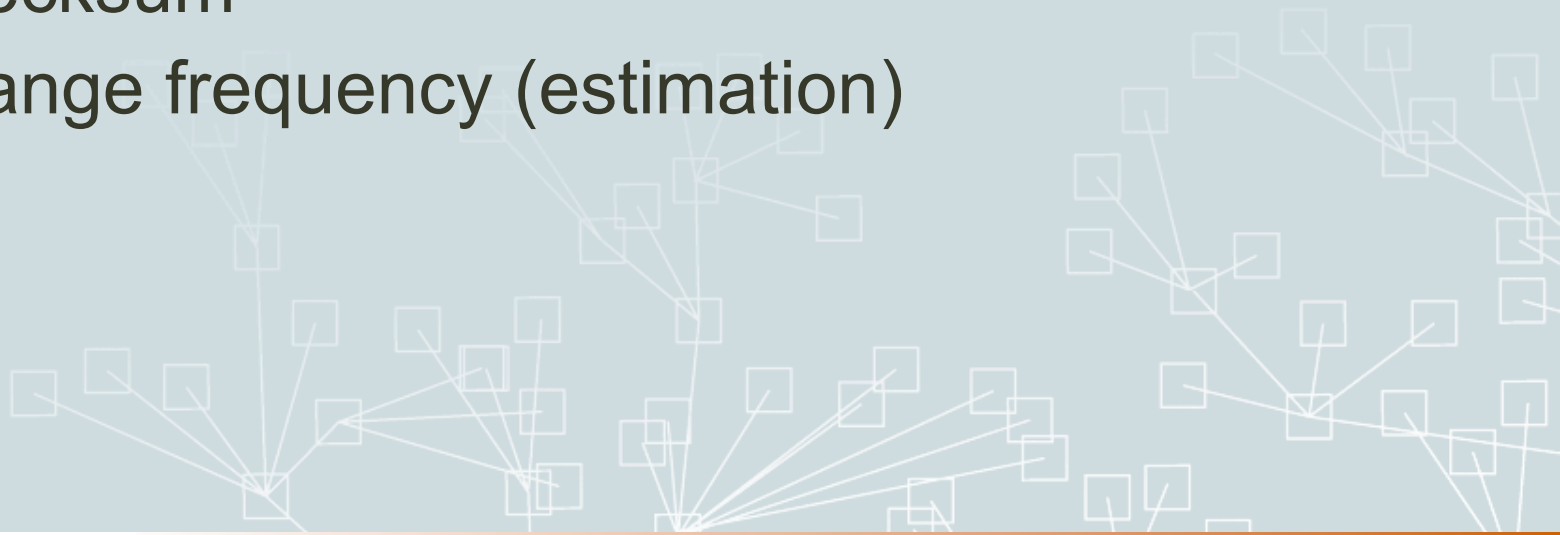
1.1					
Type	Example	Cardinality	Obligation	OAIS	
URI	URI	http://www.bnf.fr/index.htm	1	Y	PDI Reference





Metadata related to the Document

- Original location
- Size
- Format (name, version, verification tool)
- Checksum
- Change frequency (estimation)





Transaction actors

- **Crawler (type, set-up, IP)**
- **Server (IP, DNS information)**



Crawl

- Crawling session
- Client request
- Server response
- Robot exclusion status
- Credential for access (SSL)





Selection process

- Entry Point List
- Path to document
- Hop from the entry point
- Selection policy documentation
- Characterization/Evaluation (value, tool)





Change history

- Metadata
- New value
- Tool





Public version for comments: Nov. 2005

standards@netpreserve.org

IIPC site: www.netpreserve.org

julien@netpreserve.org

