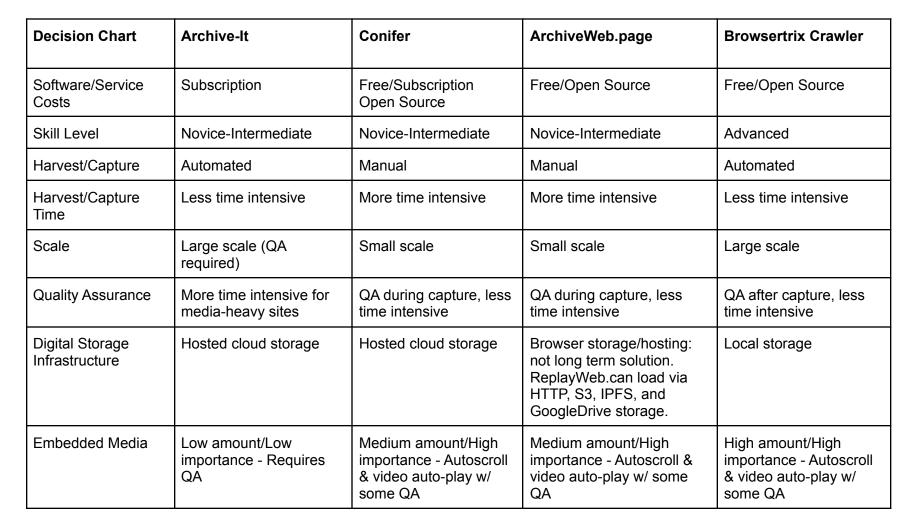
Sustainable Web Archiving at Scale:

An Introduction

June 21-22, 2021





Emulated/Remote Browsers	NA	Yes	Chrome only	Chrome only
Social Media/Privacy	Low Importance- Lower fidelity/ no login access	High importance- Higher fidelity/login credentials required/can pose access obstacles	High importance - Higher fidelity/login credentials stored separately	High importance - Higher fidelity/credentials stored separately
Cataloging	Built-in template for Dublin Core fields (standard + custom fields) OAI-PMH crosswalk to OCLC WorldCat for collection-level records ArchivesSpace integration via WASAPI. Supports full text search of metadata and content.	Unstructured. Title, list, description fields. Embeddable in web pages. Supports full text search of title and URL fields of hosted collections.	Generates JSON indexing/ supports full-text search of all content in WACZ	Generates JSON indexing/ supports full-text search of all content in WACZ
Access	Online	Online	Online or Local	Local
Export Format	WARC	WARC	WARC or WACZ	WARC or WACZ