Technical uplift of the Web Curator Tool

Ben O’Brien, Jeffrey van der Hoeven
IIPC WAC 2019
WCT facts & figures

- The Web Curator Tool (WCT) is an open-source workflow management application for selective web archiving
- WCT 1.0: first release in 2006
- WCT 2.0: major upgrade, available since November 2018
- Currently 17 organizations worldwide are using it or considering start using

Key features:

- Easy to use by any collection manager
- Embedded Heritrix H1 & H3 crawler engine
- Supports permissions, job scheduling, harvesting, quality review, and the collection of descriptive metadata.
Web Curator Tool
2019 Technical Uplift:
Summary

Legend

current technology
- Support for Heritrix 3
- Simplified consolidated install
- Better documentation

improved technology
- Modern software components
- Platform for future changes, including crawler abstraction and component APIs/REST APIs
- Cloud-ready

presentation layer
scheduling
application framework
persistence layer

Web Curator Tool

current version 2.x ➔ upcoming version 3.x

application container

application runtime

build technology

Struts 1.2.8/Tiles 1 ➔ Spring MVC 5.1.x/Tiles 3
Quartz scheduler 1.5.2 ➔ Quartz scheduler 2.3.1
Spring 1.2.7 ➔ Spring 5.1.x
Hibernate 3.1.3 ➔ Hibernate 5.4.2 with JPA

Tomcat 8.5/9 and Jetty 7.5 ➔ Spring Boot 2.1

Java 8 ➔ Java 11

Maven 3.5 ➔ Gradle 5.x
Check out our poster!

Visit our website at:

http://dia-nz.github.io/webcurator/