

Capture, manage and provide access to Web-based content

Web Harvester



With OCLC's Web Harvester, you can easily capture and manage born-digital, Web-based content. The Web Harvester integrates this process with other digital collection building using CONTENTdm® Digital Collection Management Software.

.....

What is the Web Harvester?

Organizations that want to collect and add Web-based content to their CONTENTdm digital collections may use OCLC's Web Harvester. Web-based content includes single, print-like documents published on the Web as well as entire Web sites. This content is highly transient and at risk of being lost, yet often falls directly within your collecting scope. With the Web Harvester, Web content can be captured and managed in conjunction with your other digital materials.

Integrated workflow

The Web Harvester enables catalogers using the OCLC Connexion client (version 2.1+) to harvest content from the Web, review it and then add the harvested items to their CONTENTdm collections during the Connexion cataloging process.

By integrating digital collection development and capture with standard cataloging workflows, the Web Harvester provides you with an additional option for expanding participation in growing and maintaining your digital collections of Web-based materials.

For added security, digital masters of the harvested content may also be ingested into the OCLC Digital Archive. The Digital Archive provides a secure storage environment for you to easily manage and monitor the health of your master files and digital originals.

Enhanced discovery

Harvested items added to CONTENTdm using the Web Harvester are discoverable from your CONTENTdm Web interface, as well as WorldCat.org and WorldCat Local.

Key benefits

- **Collect and manage** transient, born-digital Web-based content
- **Use the OCLC Connexion client** for easy capture of Web content and integration into your library's cataloging workflow
- **Manage and provide access** to Web-based content in concert with your other digital collections via CONTENTdm
- **Integrate** with your cataloging and digital collections building processes
- **Preserve a copy of** your harvested content by ingesting it into OCLC's Digital Archive
- **Provide global discovery** via WorldCat

Each harvested item added to CONTENTdm using the Web Harvester is associated with its WorldCat record via a persistent URL based on the OCLC number of the WorldCat record.

Added Benefits

Web Harvester users also are able to use the Connexion digital import feature, allowing you to submit your digitized materials to your CONTENTdm collections.

And, since the Web Harvester uses CONTENTdm, you have full access to its complete set of digital content management features.

Capture, manage and provide access to Web-based content
Web Harvester

How the Web Harvester works

1. Using the Connexion client, create a new WorldCat record, or locate an existing record, describing the Web-based content you want to capture.
2. Enter the parameters for the harvest, as well as Issue, Title and Date (if you are harvesting an issue of a serial).
3. Submit the harvest.
4. Review the harvest to make sure the content is correct.
5. Submit the content to a CONTENTdm collection and, optionally, to the Digital Archive. Upon submitting the content, the 856 link in your bibliographic record is automatically updated and the bibliographic record is mapped to Dublin Core for inclusion in CONTENTdm.
6. End users can now access your content via WorldCat.org, WorldCat Local, OCLC FirstSearch and your CONTENTdm Web interface.

The Web Harvester requires CONTENTdm 4.3 or later, an OCLC Cataloging subscription and the use of CONTENTdm Hosting Services.

Preservation

With an additional subscription to the OCLC Digital Archive, your master files will be automatically placed in the Archive's secure, managed storage system.

The Digital Archive is an option that provides a secure environment for you to easily manage and monitor the health of your digital content. It provides a foundation for digital preservation of all your digital collections.

To learn more about the Digital Archive, please visit: www.oclc.org/digitalarchive/

For more information

To learn more about the digital archive, please visit: www.oclc.org/webharvester/.

If you have questions, please contact your OCLC representative or e-mail digitalcollections@oclc.org.

