THE LIBRARY OF CONGRESS: DIGITAL PRESERVATION VIDEO SERIES

BagIt: Transferring Content for Digital Preservation

We live in a networked world, where information is always in motion. You transfer content when you add digital music files to your collection. Or when you backup your files and store a copy elsewhere for safekeeping.

Libraries and archives need to do the same thing when they transfer content for digital preservation. They must ensure that content is moved from place to place reliably.

But content transfer can fail for a variety of reasons: human error, equipment failure, network disruption, inscrutable digital voodoo. Is there a way to move digital content so that it can be successfully preserved?

It's harder than you might think. Different organizations have different policies, different technologies, different everything. When there are no standards, transfers are hard to arrange.

The Library of Congress -- with its partners, Stanford University and the California Digital Library -- has developed a solution to this problem: "Bagit."

BagIt is a specification that supports the reliable transfer of content in self-contained packages, what we call "bags." A digital bag is a container for sending your content, just like a physical envelope is used to send content through the mail. But with digital bags, you send content from one computer to another.

You can also think of bags like the folders or directories on your computer. They can hold any kind of file you can imagine: documents, pictures, music, movies…even other folders.

When preparing to transfer content, the first thing you do is create a bag. This is just like creating a new folder on your computer, and you can give the bag any name you like. There are three things that Bagit requires to be in a bag: a bag declaration tag, a list of the content files, and the content itself. Bags can be sent over computer networks or physically moved using portable storage devices.

The receiving computer analyzes the bag. If everything matches up, the files are unchanged and the transfer was successful.

Bags are containers with built-in inventory checking, so you can be sure that your content transferred intact. Bags are designed to be flexible. They can work in many different settings, including situations where the content is located in more than one place.

To make it even easier, the Library of Congress has developed a suite of tools to help you create and transfer bags. These tools have been released as open source, and are freely available on SourceForge.

Bagit helps support digital preservation by supporting the successful transfer of files from one computer to another. To learn more about Bagit and other digital preservation activities at the Library of Congress, please visit digitalpreservation.gov.