

# Digital Curation Tools: Metadata Enhancement with Selenium IDE

Daniel Gelaw Alemneh & Andrew Weidner  
University of North Texas Libraries: Digital Libraries Division  
Poster # 411

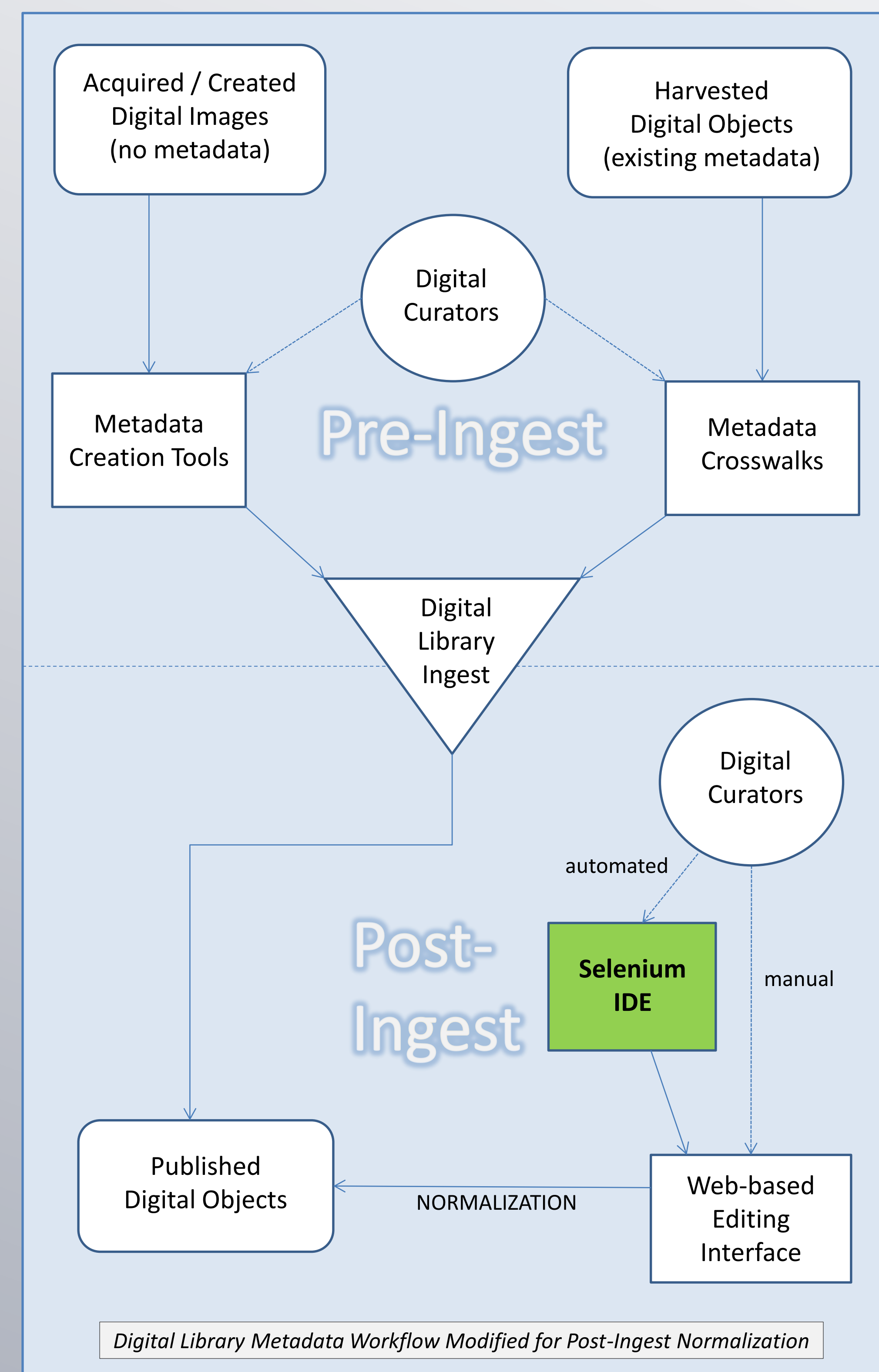


## Introduction

Digital lifecycle management starts when an item is created (born-digital) or selected for digitization (analog) and continues through image post-processing, metadata capture, derivative creation, and preservation for long-term access. Quality metadata is crucial to implementing reliable, usable, and sustainable digital libraries. Recognizing the role of standardized metadata in digital resource lifecycle management, the University of North Texas (UNT) Libraries actively promote metadata-based digital resource management.

The UNT Digital Libraries Division utilizes various tools to ensure metadata consistency and precision across all digital resources and facilitate digital curation activities. This poster illustrates a workflow that uses Selenium IDE to edit large sets of published metadata records quickly and accurately with minimal human intervention.

## Selenium IDE for Digital Libraries

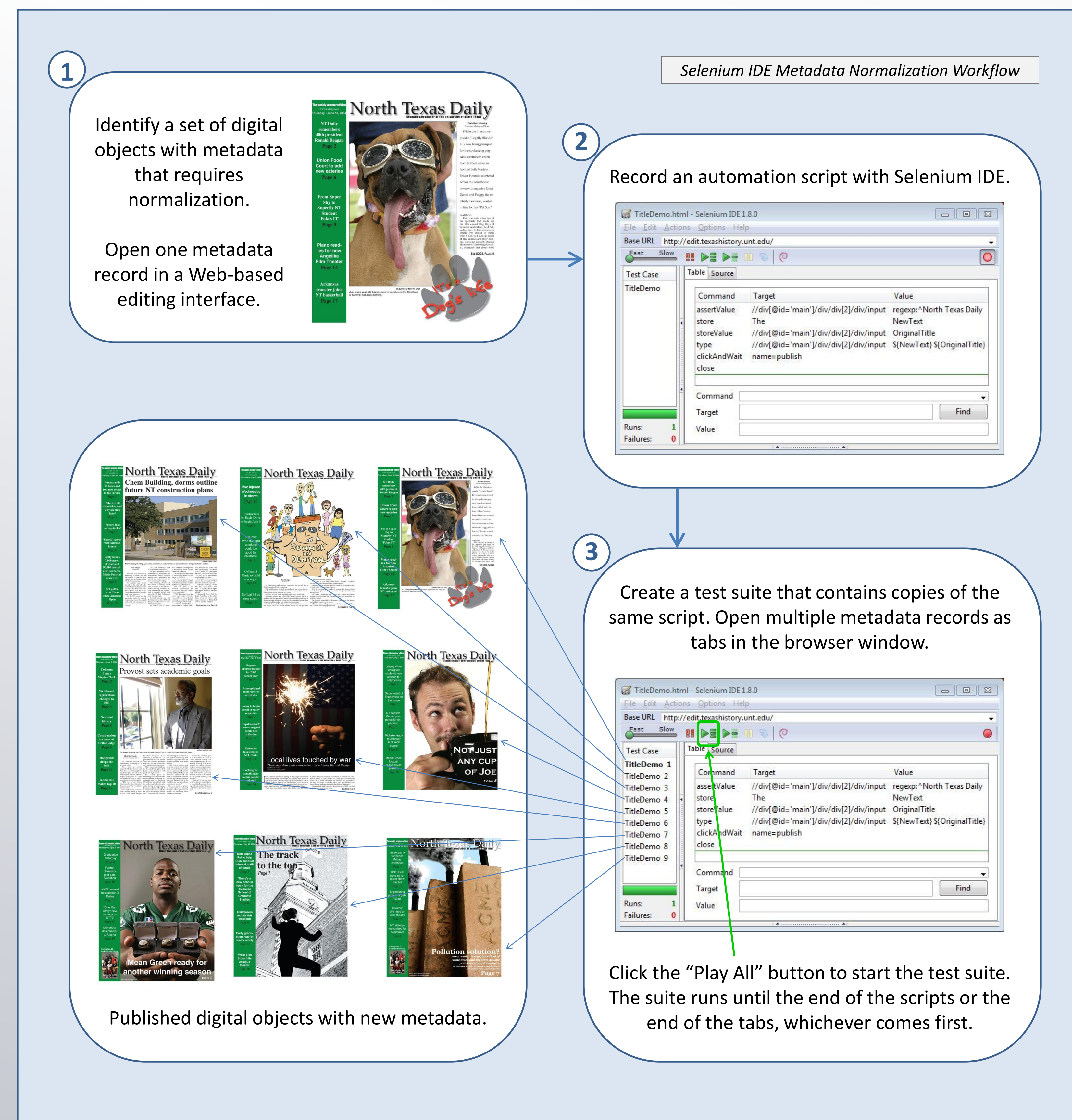


Selenium IDE is a free and open source add-on for the Firefox Web browser. It provides an integrated development environment in which to create, debug and run custom scripts that automate actions in the browser window.

A typical Selenium IDE workflow begins by identifying a set of digital objects that require normalization. If the set is large and the required changes are uniform, a digital curator creates a Selenium IDE script that performs the changes, publishes the new metadata, and closes the browser tab. After testing and debugging to ensure that the script performs correctly, the curator creates a test suite. Using the content management system's search interface, the curator opens multiple object records as Web browser tabs. Finally, the curator runs the test suite. Each script in the test suite works on a tab in the browser window until there are no scripts left in the suite or no tabs left in the browser. The curator repeats the process until the entire record set is normalized.

## Selenium IDE Commands for Metadata Normalization

Command	Function	Usage
<i>assertValue</i>	Tests an element's value.	Verify that a Web page element, such as a drop-down menu or text input field, contains a specific value. If the element contains the specified value, the script continues. If not, the script stops. Use regular expressions to match dynamic text content.
<i>click</i>	Performs a mouse click.	Click links, check boxes, select radio buttons, etc.
<i>clickAndWait</i>	Performs a mouse click and waits for a new page to load.	Use this command when a clicked link loads a new page. The script waits for the new page to load before executing the next command.
<i>close</i>	Closes the tab or browser window.	Use this command to enable fully automated batch editing with a suite of identical scripts. After opening a set of records in multiple tabs, the <i>close</i> command at the end of each script allows the next script in the suite to work on the next tab in succession until there are either no more scripts in the suite or no more tabs to edit.
<i>keyPress</i>	Sends a specified key press.	Selectively delete or add characters in a text input field. Unlike the <i>type</i> command, <i>keyPress</i> allows users to modify existing text rather than replace it entirely.
<i>select</i>	Selects a value for an element.	Select a choice from a drop-down menu.
<i>setCursor-Position</i>	Places the cursor at a specified point in a text string.	Use this command with <i>keyPress</i> to modify existing text at the appropriate point in a text string.
<i>store</i>	Stores a new text string in a variable.	Use this command to store new text in a user-defined variable. The stored text can then be pasted in another field with the <i>type</i> command.
<i>storeValue</i>	Stores an existing text string in a variable.	Use this command to copy existing text from an input field. The stored text, contained in a user-defined variable, can then be pasted in another field with the <i>type</i> command.
<i>type</i>	Enters text in an input field.	Populate an input field with the specified text, including stored variables. This command overwrites any existing text.



## Summary

Large digital collections present challenges when producing descriptive metadata. Naming schemes and element definitions can vary widely, requiring substantial rework to meet local repository standards. Successful metadata enhancement strategies involve mechanisms for both pre- and post-ingest metadata normalization. Automated metadata normalization with Selenium IDE improves efficiency and accuracy during the time- and labor-intensive post-ingest data entry process.

From institutional repository platforms (e.g., DSpace) to commercial content hosting sites (e.g., Flickr), Selenium IDE can be used to edit metadata in any content management system that has a Web-based editing interface. Selenium IDE is a highly recommended addition to the metadata enhancement toolkit for any institution that serves digital objects in a content management system with a Web-based administrative interface.