Software/Tool Comparison Worksheet

Criteria	Evaluation
Name	Salix
Website/Contact	http://nhc.asu.edu/vpherbarium/canotia/SALIX3.pdf http://nhc.asu.edu/vpherbarium/projects.html Daryl Lafferty Leslie R. Landrum Anne Barber School of Life Sciences Arizona State University P. O. Box 874501 Tempe, Arizona 85287-4501 USA Phone: (480) 965-6162 Fax: (480) 965-6899
Description	Salix is not a database management system, but a software tool to be used in cleaning, parsing, and ingesting textual data generated from specimen labels by Optical Character Recognition (OCR) software. It is designed primarily for vascular and non-vascular plant specimens, but likely has application for other collections. The software parses previously generated OCR text into selected fields and produces a comma separated variable file (.csv) that can be imported into a variety of database management systems. It should be noted that the Salix software does not perform OCR. As recommended by Salix staff, images of herbarium labels are processed via OCR software with the label data output to a Microsoft Word document (though output to text documents can also be processed). The OCR output is cleaned and edited by a technician via Word (or other text editor), highlighted and copied into the computer's copy buffer, and pasted into the Salix interface by a one-button click. Microsoft Word is recommended by Salix staff due to its robust suite of editing features. Once text is pasted into Salix, the text is parsed into several Darwin Core fields and inserted into a comma separated variable file for eventual export. Several tools are available that allow for formatting the output file, editing and adding countries, and building

	collections of key words for Salix to use in interpreting and parsing the text, which allows customization for individual herbaria.
Computer hardware Operating system	Windows PC
Additional hardware/software required	Digital camera. As typically used, Salix requires the installation of OCR software with the capability of producing a word processing or text file containing the results of the OCR process. Salix staff uses and recommends ABBYY Finereader (proprietary, but often distributed with flatbed scanners) and reports better OCR output with ABBYY than with other OCR software they have tested. Other software that might be implemented includes: Tesseract (open source), GOCR (open source), OCRopus (open source), and Adobe Acrobat Professional (proprietary). Although OCR text can be edited with Salix, word processing (Microsoft Word is recommended by Salix staff) enhances the editing process.
Features	Editable glossaries of words and phrases assist Salix in finding various portions of a label and interpreting text for parsing. Built in taxonomic authority files, with capability to search Tropicos to verify the accuracy of scientific names and to add taxon names to the local taxonomic authority file. Parses edited OCR data into a variety of common fields. Provides for converting georeferences to alternate expressions, e.g. d m s to d.ddd, converts feet to meters for elevation, and automatically checks the validity of coordinate pairs. Data can be exported in standard Darwin Core format via a comma or tab separated variable file, or extensible markup language (XML) file for import into a variety of databases. Contact information is available and Salix staff encourages feedback and suggestions as they continue to refine the software and add important features.

Market presence	Salix was funded in 2008, though initial experimentation began earlier.
Initial cost	None
Maintenance cost	None
Ease of setup/ prerequisite skills	Salix is easy to set up and to use, with a relatively gentle learning curve for most users.
Continuing IT support required?	No
Special skills required for maintenance	None
Challenges	Requires installation and use of an OCR software package. Comprehensive help documentation is not currently available, but both written and video documentation is in process. Salix staff anticipates video tutorials to be available by fall 2012 at http://vimeo.com/asuherbarium.