JHOVE2 – Next-Generation Framework and Application for Format-Aware Characterization

Version: 2.1.0
Issued: 2013-02-14
Status: Final

Release Notes

Version 2.1.0

Version 2.1.0 of JHOVE2 includes 3 new format modules, 1 new Identifier module, 1 new displayer module, and several bug fixes and enhancements from the Issues page on the JHOVE2 wiki (https://bitbucket.org/jhove2/main/issues).

New format modules included in this release:
    ARC
    GZIP
    WARC

This release includes a new identifier module, based on the Unix "file" utility. The downloadable release is configured to run the DROID identifier that was released in version 2.0.0.

For information on how to install the "file" utility on Windows, MAC, and Unix machines, and for information on how to update the JHOVE2 Spring configuration files to employ the new Identifier module, please see the "Specification and Installation/Configuration Guide" for the File Identifier Module on the JHOVE2 wiki modules page (https://bytebucket.org/jhove2/main/wiki/documents/JHOVE2-File-module-spec-2.1.0RC2.pdf).

The new XSLDisplayer module (which extends XMLDisplayer) can do XSLT transformations on the XML output before displaying it.

Resolved issues included in this release:
- #56: Review Laurent Bihanic's Gzip code
- #125: opensp tests fail on ubuntu
- #126: 0 tag IFD error message
- #128: jargs jar has moved to a different Maven Repository -- pom.xml must be updated
- #130: Have BerkeleyDB je persistence database use user home directory by default
- #132: Tool to confirm that all Messages are represented in jhove2_messages.properties file
- #134: duplication of the Formatmodule output takes place when using the in-Memory

Persistence Manager.

- #136: Windows driver script doesn't work outside of home directory
- #140: Incorrect "PostScript" name for "PDF" in "OtherFormats-config.xml"
- #143: Error message for org.jhove2.module.format.tiff.IFDEntry.InvalidCountValueMessage is missing in jhove2_messages.properties file
- #146: Typo in droid signaturefile
- #147: WARC Droid Signature definition
- #148: Bug in InMemorySourceAccessor/InMemoryBaseModuleAccessor/...
- #153: Tiff Module never reports Validity.True
- #155: Problems with spaces and hyphens in file paths
- #156: Create GZip format module
- #157: Create ARC format module
- #158: Create WARC format module
- #160: org.jhove2.module.format.wave.bwf.LinkChunk missing zero-arg constructor
- #161: org.jhove2.config.spring.SpringConfigInfo must make CLASSPATH for Spring context configurable
- #162: Message org.jhove2.module.format.sgml.OpenSpWrapper.IOExceptionForSGMLStdErrFile2 in Java code is not in messages properties files
- #163: spring-test-2.5.6.jar is not included in the download zip file
- #165: TiffTagTest and ICCModuleTestBase need setUpBeforeClass() overrides
- #166: Update MessagesChecker tool to read in more than one .properties file
- #167: Wrong URL for OPenSp windows binary download in User Guide
- #168: Need documentation for new GZIP module
- #169: Need documentation for new ARC module
- #170: Need documentation for new WARC module
- #171: Document new File identifier
- #172: New BSD File -based identifier
- #173: create displayer properties file for Arc module
- #174: Create displayer properties file for gzip module
- #175: Create displayer properties file for WARC module
- #176: Update user's guide to refer to configuration info for File-based identifier

For information about issues resolved in this release, known bugs, open issues, and enhancement requests, please refer to

- ***JHOVE2 Issues page***

  https://bitbucket.org/jhove2/main/issues?sort=version

For detailed installation and configuration instructions please refer to:

- *JHOVE2 User's Guide*
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2-Users-Guide_20110222.pdf.

For detailed guidance on developing additional format modules please refer to:

- *JHOVE2 Architectural Overview*
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2-Architecture-v2-0-0.pdf

- *JHOVE2 Programmer's Guide*
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2Programmer2-0-0.pdf

Questions concerning the use of JHOVE2 and module development should be addressed to
JHOVE2-TechTalk-l@listserv.ucop.edu.

Specific errors or suggestions may be reported to the JHOVE2 issue tracker at
https://bitbucket.org/jhove2/main/issues?sort=id.

CALIFORNIA DIGITAL LIBRARY

Stephen Abrams
Patricia Cruse
John Kunze
Marisa Strong
Perry Willett

PORTICO

John Meyer
Sheila Morrissey

STANFORD UNIVERSITY

Richard Anderson
Tom Cramer
Hannah Frost

BIBLIOTHÈQUE NATIONALE DE FRANCE

Laurent Bihanic

NETARKIVET.DK

Nicholas Clarke

Version 2.0.0

JHOVE2 is a next-generation framework and application for format-aware characterization. Characterization is the process of deriving *representation information* about a formatted digital object that is indicative of its significant nature and is useful for purposes of classification, analysis, and use. Effective and efficient means of characterization is a key component of any digital preservation program.

JHOVE2 supports four specific aspects of characterization:

- *Identification*. The process of determining the *presumptive* format of a digital object on the basis of suggestive extrinsic hints and intrinsic signatures, both internal (e.g. magic number) and external (e.g. file extension).

- *Validation*. The process of determining the level of *conformance* to the normative syntactic and semantic rules defined by the authoritative specification of the object's format.

- *Feature extraction*. The process of reporting the *intrinsic properties* of a digital object significant for purposes of classification, analysis, and use.

- *Assessment*. The process of determining the level of *acceptability* of a digital object for a specific purpose on the basis of locally-defined policy rules.

The object of JHOVE2 characterization can be a file, a subset of a file, or an aggregation of an arbitrary number of files that collectively represent a single coherent digital object. JHOVE2 can automatically process objects that are arbitrarily nested in containers, such as file system directories or Zip files.

The JHOVE2 project seeks to build on the success of the original JHOVE characterization tool (http://hul.harvard.edu/jhove) by addressing known limitations and offering significant new functions. These enhancements include:

- Streamlined APIs incorporating increased modularization and uniform design patterns.

- Object-focused, rather than file-focused, characterization, with support for arbitrarily-nested container formats and formats instantiated across multiple files.

- Signature-based identification using DROID (http://sourceforge.net/projects/droid).

- Rules-based assessment to support determinations of object *acceptability* in addition to validation of format *conformity*.

- Extensive user configuration of modules, characterization strategies, localized messages, and formatted results.

- Performance improvements using Java buffered I/O (java.nio).

- Persistence manager to support the characterization of an arbitrary number of objects with a fixed memory footprint.

The JHOVE2 project is a collaborative undertaking of the University of California Curation Center at the California Digital Library, Portico, and Stanford University, with generous funding from the Library of Congress as part of its National Digital Information Infrastructure and Preservation Program (NDIIPP).

JHOVE2 is made freely available under the terms of the BSD open source license for all project-developed code; some third-party libraries may be covered by other open source licenses.

http://jhove2.org/
JHOVE2-Announce-l@listserv.ucop.edu
JHOVE2-Techtalk-l@listserv.ucop.edu

Version 2.0.0 of JHOVE2 supports all the major technical objectives of the project, including a more sophisticated modular architecture; signature-based file identification; policy-based assessment of objects; recursive characterization of objects comprising aggregate files and files arbitrarily-nested in containers; and extensive configuration and reporting options. It provides a stable interface against which developers can create additional format modules.

Format modules, and profiles, included in this release are:

- ICC color profile
- SGML
- Shapefile   *Main, Index, dBASE*
- TIFF   *4 – 6, Class B, G, R, P, Y, TIFF/IT, TIFF/EP, Exif, GeoTIFF, DNG*
- UTF-8   *ASCII*
- WAVE   *Broadcast Wave Format*
- XML
- Zip

Please note that the Zip module comprises a non-validating partial module, which accomplishes recursive JHOVE2 descent on the contents of the Zip file, but does not yet validate the Zip file itself against the standard.

Version 2.0.0 of JHOVE2 can be downloaded from https://bitbucket.org/jhove2/main/downloads. Download packages are available in Zip and tar.gz form.

For information about issues resolved in this release, known bugs, open issues, and enhancement requests, please refer to

- ***JHOVE2 Issues page***

  https://bitbucket.org/jhove2/main/issues?sort=version

For detailed installation and configuration instructions please refer to:

- ***JHOVE2 User's Guide***
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2-Users-Guide_20110222.pdf.

For detailed guidance on developing additional format modules please refer to:

- ***JHOVE2 Architectural Overview***
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2-Architecture-v2-0-0.pdf

- ***JHOVE2 Programmer's Guide***
  http://bitbucket.org/jhove2/main/wiki/documents/JHOVE2Programmer2-0-0.pdf

Questions concerning the use of JHOVE2 and module development should be addressed to JHOVE2-TechTalk-l@listserv.ucop.edu.

Specific errors or suggestions may be reported to the JHOVE2 issue tracker at https://bitbucket.org/jhove2/main/issues?sort=id.

Development planning

Additional JHOVE2 functionality is scheduled for inclusion in subsequent releases:

- Version 2.1.0
  - ARC and Gzip modules
    (*integration of third-party development by Bibliothèque nationale de France / Atos*)
  - Grid and NetCDF modules
    (*integration of third-party development by Wegener Institute for Polar and Marine Research*)
  - JPEG 2000 module

- Version 2.2.0
  - PDF module

JHOVE2 project team

California Digital Library

Stephen Abrams
Patricia Cruse
John Kunze
Isaac Rabinovitch
Marisa Strong
Perry Willett

Portico

John Meyer
Sheila Morrissey

Stanford University

Richard Anderson
Tom Cramer
Hannah Frost

Library of Congress

Martha Anderson
Justin Littman

With help from

Walter Henry
Nancy Hoebelheinrich
Keith Johnson
Evan Owens