

*Digital Library Federation Fall Forum*  
Providence, November 12-14, 2008

# JHVE2

## Needs Assessment and Functional Requirements

Stephen Abrams  
*California Digital Library*

Evan Owens  
*Portico*

Tom Cramer  
*Stanford University*

# Agenda

- Project goals, deliverables, and schedule
- New terminology and concepts
- Functional requirements

## JH<sup>VE</sup>2 project

- A next-generation architecture for format-aware object characterization
  - Three-fold goals:
    - Re-factor the existing architecture to achieve higher performance, simplify system integration, and encourage third-party enhancement
    - Provide significant new function
    - Implement modules
- Collaborative project of CDL, Portico, and Stanford University
  - Funded by Library of Congress/NDIIPP
  - Open source BSD license

## New function

- Complex object data model
- Generic plug-in interface
- Common data structure passed between modules to enable stateful processing
- Identification de-coupled from validation
- Standardized handling of format profiles and error reporting
- Symbolic display of binary formats
- API-level support for editing

## Format support

- Based on project partner requirements and budgetary constraints
  - *Image*: JPEG 2000, TIFF
  - *Audio*: WAVE
  - *Text*: SGML, UTF-8, XML
  - *Document*: PDF
  - *GIS*: Shapefile
  - *Color*: ICC
  - And their well-known variants, e.g. TIFF/IT, TIFF/EP, GeoTIFF, EXIF, DNG, ...
- Unfortunately precluding some JHOVE-supported formats
  - AIFF, GIF, HTML, JPEG

## Schedule

- Months 1-6      Outreach, design, and prototyping
- Months 7-9      Core APIs and framework
- Months 10-24    Module implementation

## For more information

Wiki [confluence.ucop.edu/display/JHOVE2Info/Home](https://confluence.ucop.edu/display/JHOVE2Info/Home)

Mailing lists [JHOVE2-Announce-L](#)  
[JHOVE2-Techtalk-L](#)

*(Subscribe via the wiki)*

# Terminology

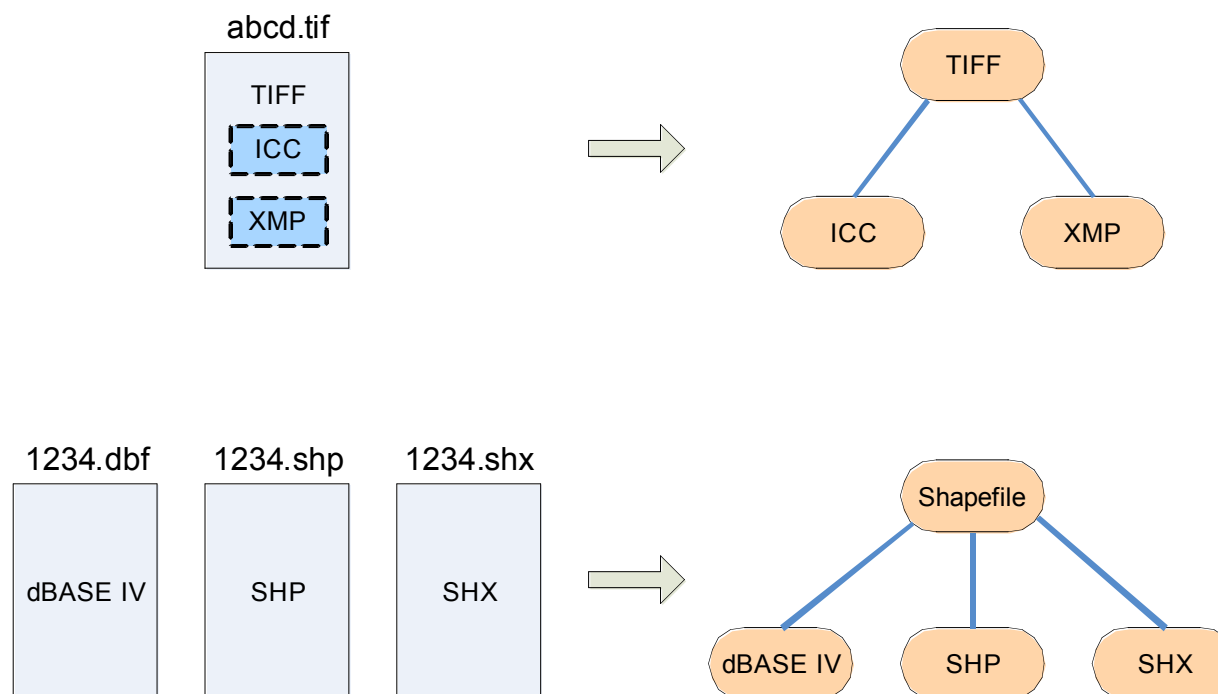
- What is it?
  - *Identification* Determining presumptive format through signature matching
- What is it, really?
  - *Validation* Determining conformance to commonly-accepted normative requirements
- What about it?
  - *Feature extraction* Reporting intrinsic properties significant to preservation planning and action
- What should you do with it?
  - *Assessment* Determining acceptability for a given purpose on the basis of locally-defined policies



# Objects, not files

*Source units*

*Reportable units*



## Functional requirements

- Given the use of new terminology to define characterization, is the conceptual basis for the project clear?
- What does "assessment" mean in the context of your existing or planned preservation workflows?
- Ambiguous statements in format specifications can be dealt with in two ways:
  - Configurable validation criteria, or
  - The application of assessment rules to the results of prescribed validation criteria

Which seems preferable?

## Functional requirements

- Given a potential trade-off between:
  - More substantive assessment capabilities
  - More supported formatswhat would your prioritization be?
- Regarding the functional requirements, what is missing?  
What is most important?
- Would you be interested in pursuing 3rd party development, or co-development, of missing functionality?

## Functional requirements

- Do you have interesting test data that you can contribute to a JHOVE2 testbed?
- Any other questions or suggestions?