Supporting Model-Based Enterprise by File Migration and Comparison Services

Presented by: Peter Bajcsy

-Research Scientist at NCSA

-Associate Director of I-CHASS, I3 Institute

-Adjunct Assistant Professor, CS & ECE UIUC

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

Acknowledgement

- This research was partially supported by a National Archive and Records Administration (NARA) supplement to NSF PACI cooperative agreement CA #SCI-9619019 and SCI 05-04064 COOP, ONR TRECC, and NCSA Industrial Partners.
- The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Archive and Records Administration, the National Science Foundation, or the U.S. government.
- Contributions by: Peter Bajcsy, Kenton McHenry, Rob Kooper, Michal Ondrejcek, Jason Kastner and Luigi Marini

Outline

- Introduction: Supporting Model-Based Enterprise
- File Format Conversion Services
- File Comparison Services
- A Discovery of Relationships Among 2D and 3D Digital File Collections
- Summary

Introduction

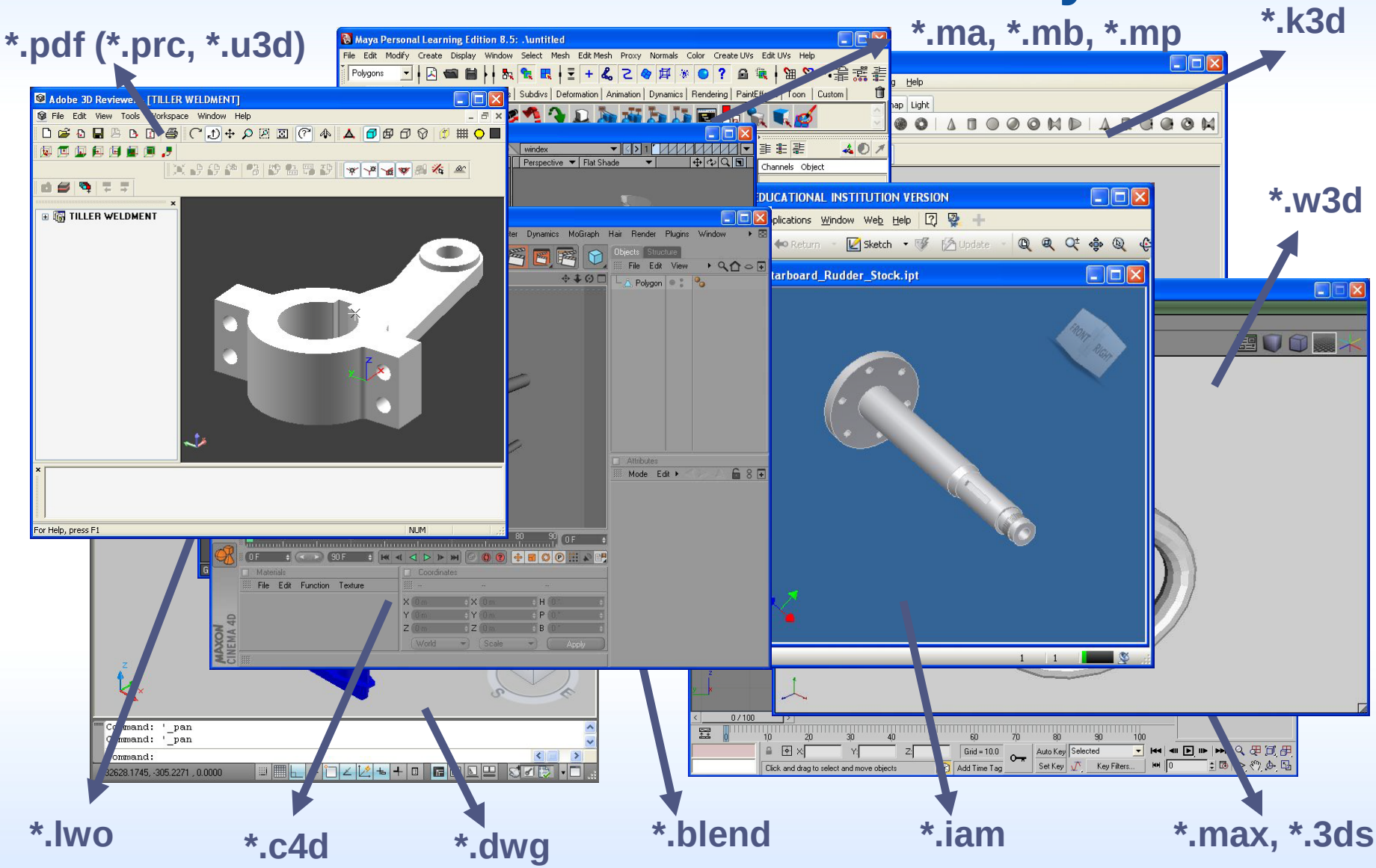
Definitions and Interpretations

- **Definition:** Model Based Enterprise (MBE) is an integrated and collaborative environment, founded on 3D product definition (Model Based Definition) shared across the enterprise, enabling rapid, seamless, and affordable deployment of products from concept to disposal.
  - http://model-based-enterprise.org/default.aspx
- **Interpretation:** MBE is an environment with many distributed files and computational services supporting file sharing, visualization, file searching, tagging, commenting, licensing, source acknowledgment, file ranking, …
  - Files ~ Products = Technical Data Packages (TDP) plus models

Introduction: File Format Conversions

- **Observation:** File content is in many file formats and hence file format conversions are hard to avoid
- **Motivation behind file format conversion services:**
  - Visual exploration depends on various software packages
  - Many formats are retired (deprecated) over time
- **Applications of file format conversions:**
  - Manage content (keep you files current)
  - Manage required processing to convert file formats
  - Preserve content for a long term

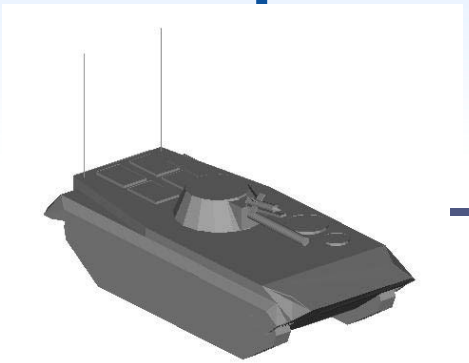
Illustration of 3D File Format Reality



Introduction: Content-Based File Comparison

- **Observation:** File content is modified when managed and updated over time
- **Motivation behind file comparison services:**
  - Content in files before and after file format conversions might not be the same
  - File content changes over time and space
  - Replicas of files might not have provenance information
- **Applications of file comparisons:**
  - Select software for file format conversion
  - Relate new files to existing files
  - Reduce cost of preservation due to redundancy

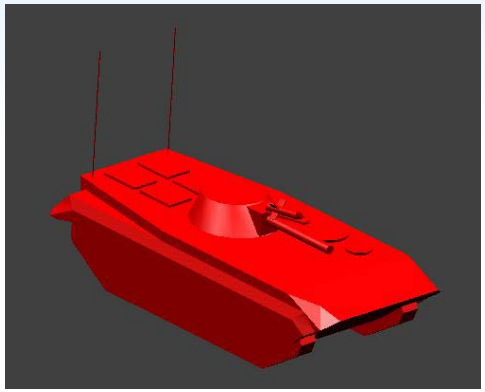
Example: Conversion of X3D to STEP to X3D



X3D

Software:

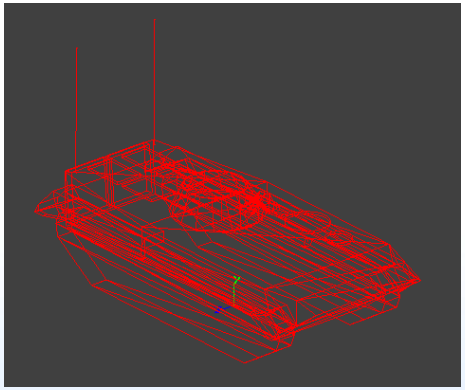
X3dToVrml97



WRL

Software:

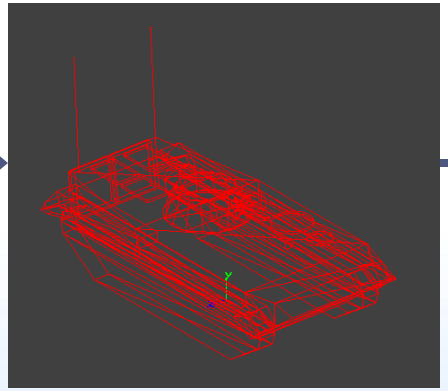
A3D Reviewer



STEP

Software:

A3D Reviewer



WRL

Software:

Vrml97ToX3d

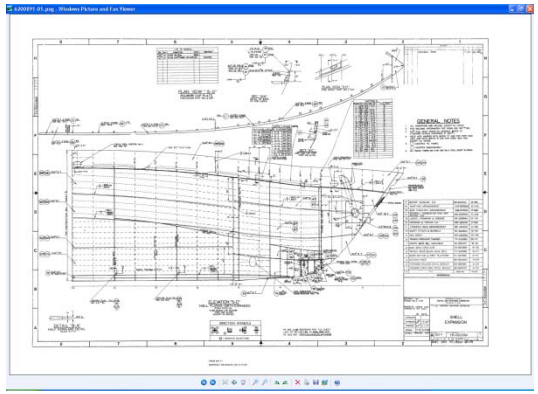
Nothing!

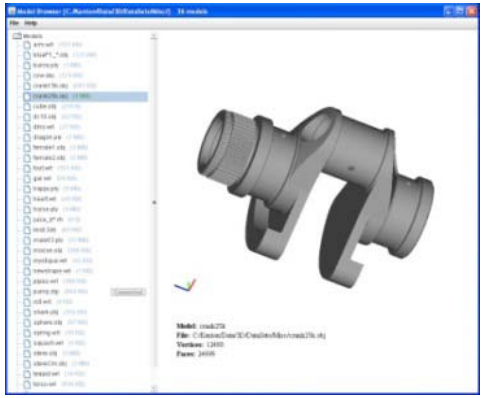
X3D

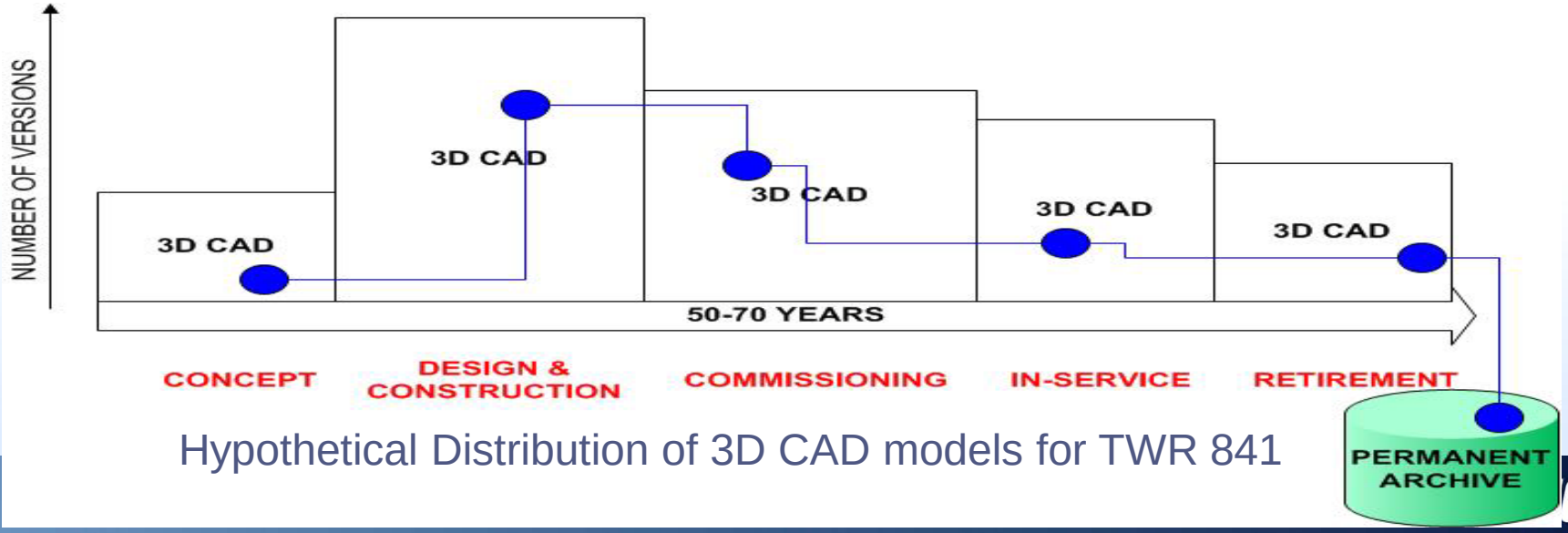
How do we provide a content-based file-to-file comparison service?

Example: Torpedo Weapon Retriever 841

- How to establish relationships among the 3D CAD models and 2D engineering drawings during a product lifecycle?

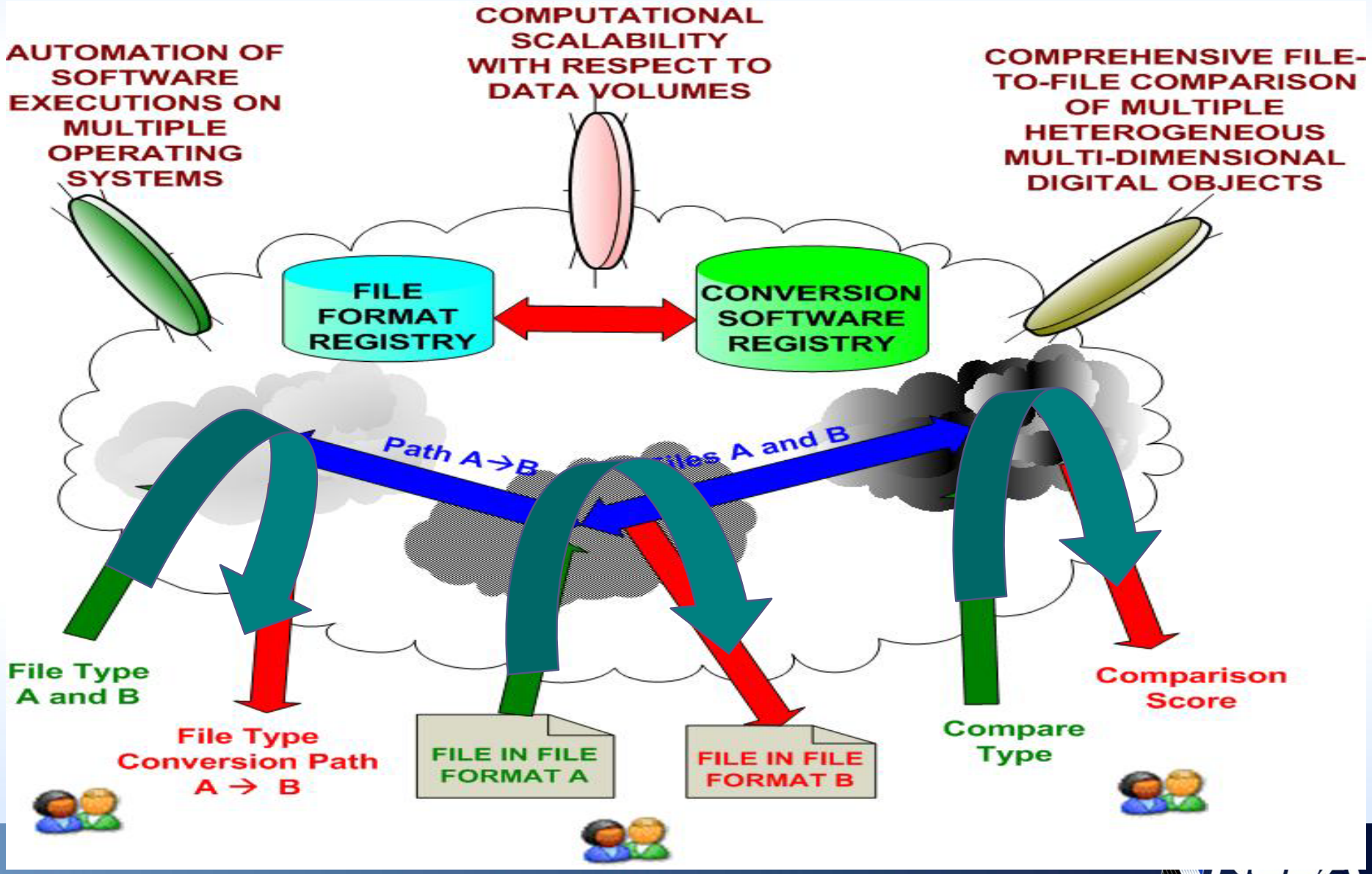






Hypothetical Distribution of 3D CAD models for TWR 841

Services for Consumers, Content Contributors, and Service Integrators



A Stack of Prototype Solutions

ANALYZE
LEARN

Image Understanding (Im2Learn)

Geospatial Image Understanding (GeoLearn, SP2Learn)

COMPARE

Comparison (Versus)

CHARACTERIZE

3D-2D Relationship Discovery (File2Learn)

Document Appraisal (Doc2Learn)

3D File Comparison (Model Browser)

CURATE

Content Management (Medici)

Image Provenance Evaluation (IP2Learn)

SHARE

Distributed Computation (Cyberintegrator)

CONVERT

Conversion Engine (Polyglot)

Conversion Software Registry (CSR)

Software Reuse (AutoScript)

Computational Scalability Testbed

Electronic Records

Closed Source Software

Hardware

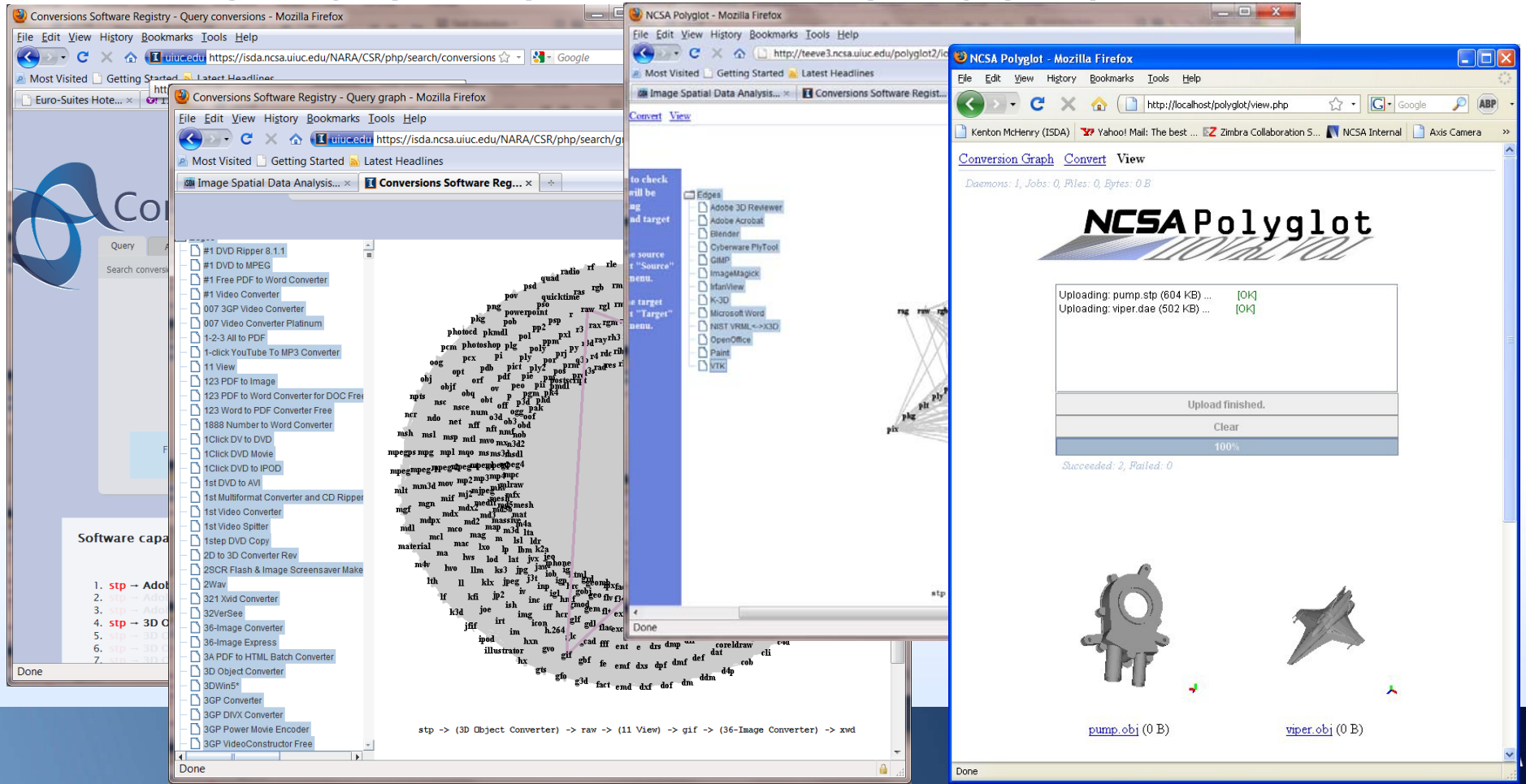
File Format Conversion Services:

+Find file format conversion software to convert from any file format to any other file format
+Execute file format conversions with any available third party software

*Open source software
*Web services are available to the general public

How to ingest content coming in many file formats and many data representations?

- Conversion Software Registry (CSR)
- Conversion Engine (Polyglot)



Conversion Software Registry (CSR)

Conversions Software Registry - Query conversions - Mozilla Firefox

File Edit View History Bookmarks Tools Help

uiuc.edu https://isda.ncsa.uiuc.edu/NARA/CSR/php/search/conversions.php

Most Visited Getting Started Latest Headlines

Conversions Software Reg... | Medici | Versus

NATIONAL ARCHIVES iSDA NCSA

E-mail:

Password: ?

Register Login

ConversionSoftwareRegistry

Query Add Edit

Search conversions | Graph conversions | All conversions | Software | Formats

Input Format:

Output Format:

Intermediaries: 0 1 2 3 4

Paths: 10 100 1000

Search

Number of conversions: 233812

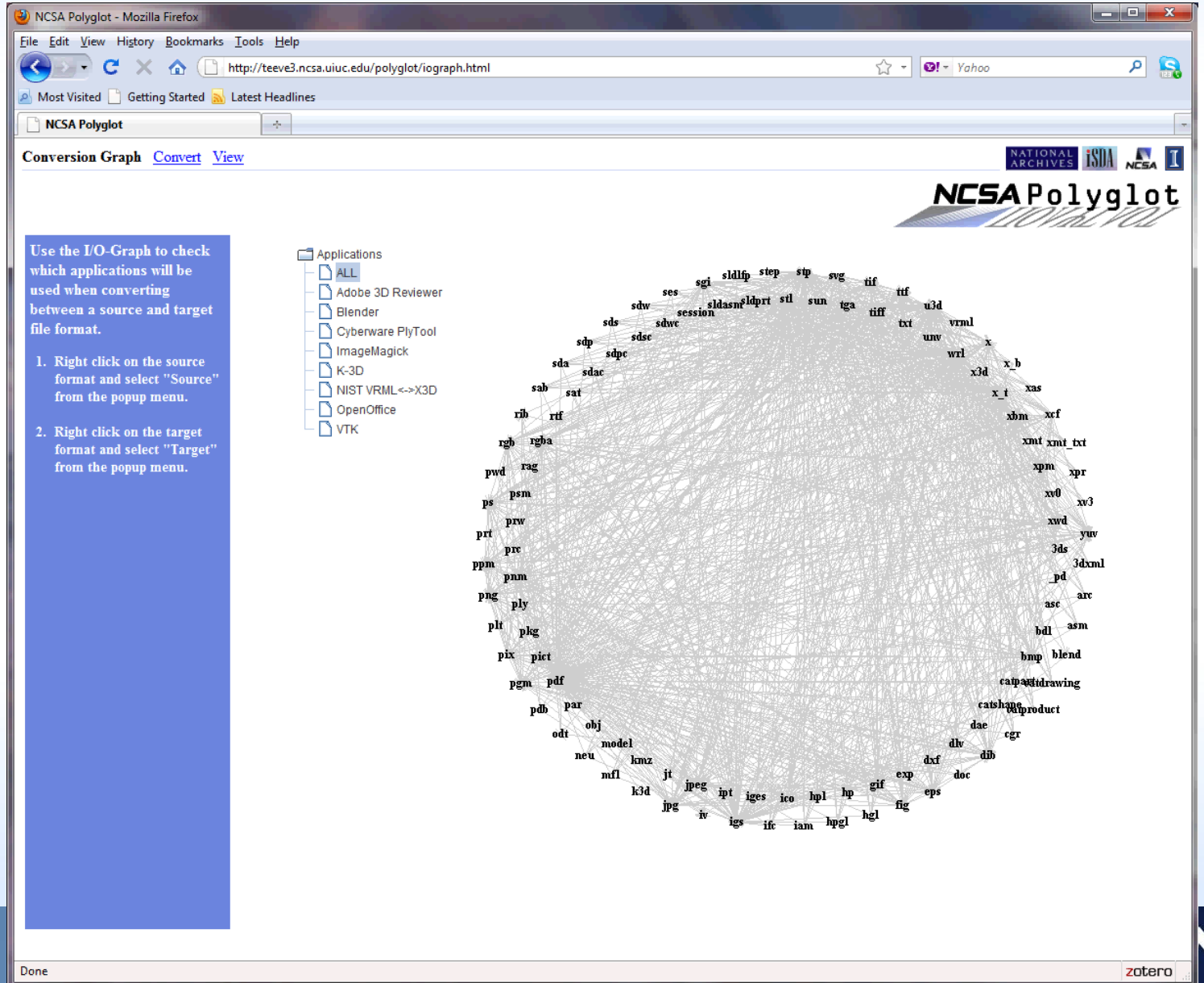
Find software capable of converting from a given source format to a desired target format by entering the corresponding file format extensions in the areas above.

Report bugs to JIRA

Done

NCSA

File Format Conversion Engine: Polyglot



File Comparison Services:

+Comparison services for any files

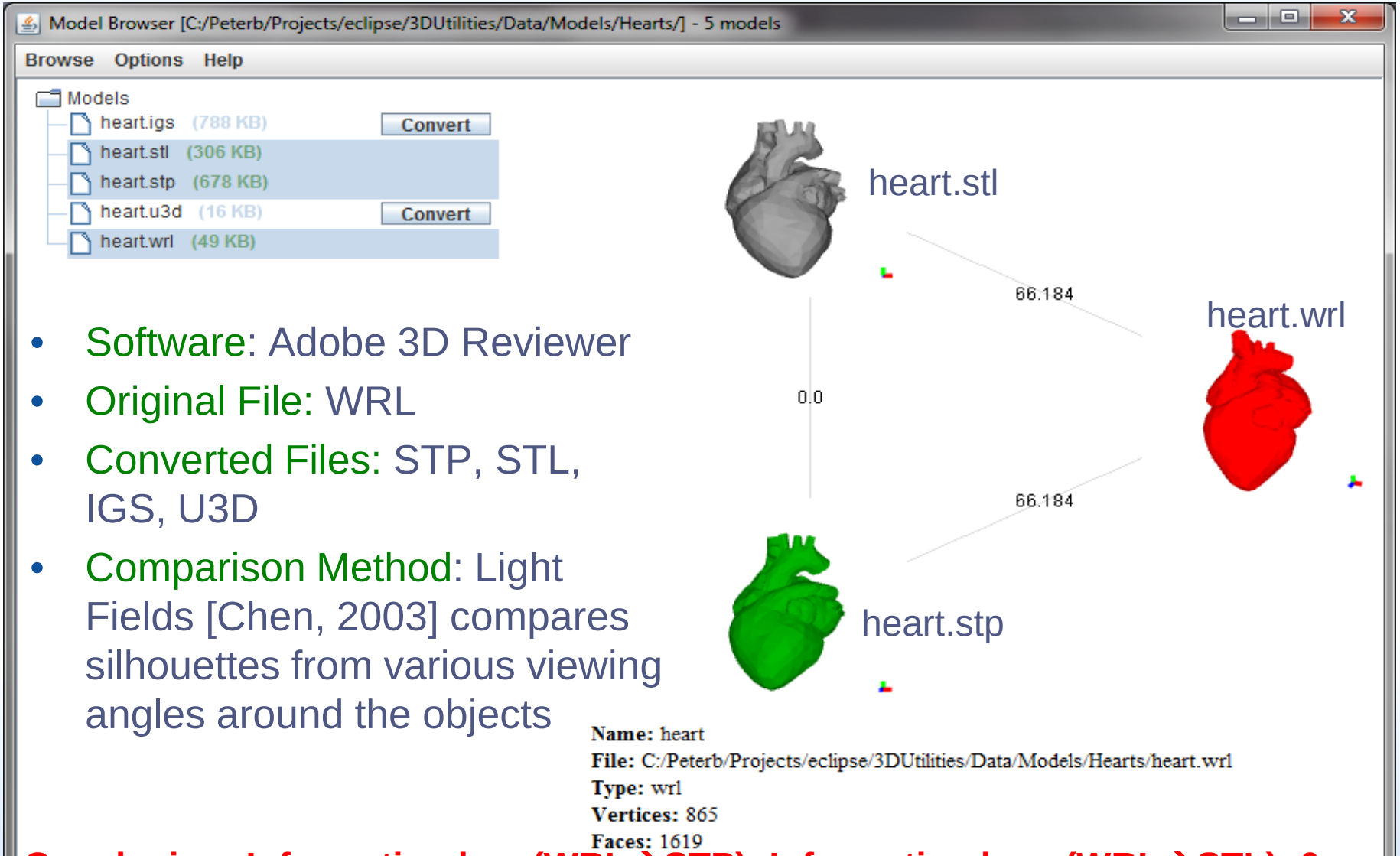
*Comparison of files containing one object type

*Comparison of file containers with multiple types of objects

*Sharing, curation and selection of files for comparison

*Comparisons to evaluate information loss due to file format conversion over a set of files

3D Comparison Example (ModelBrowser)

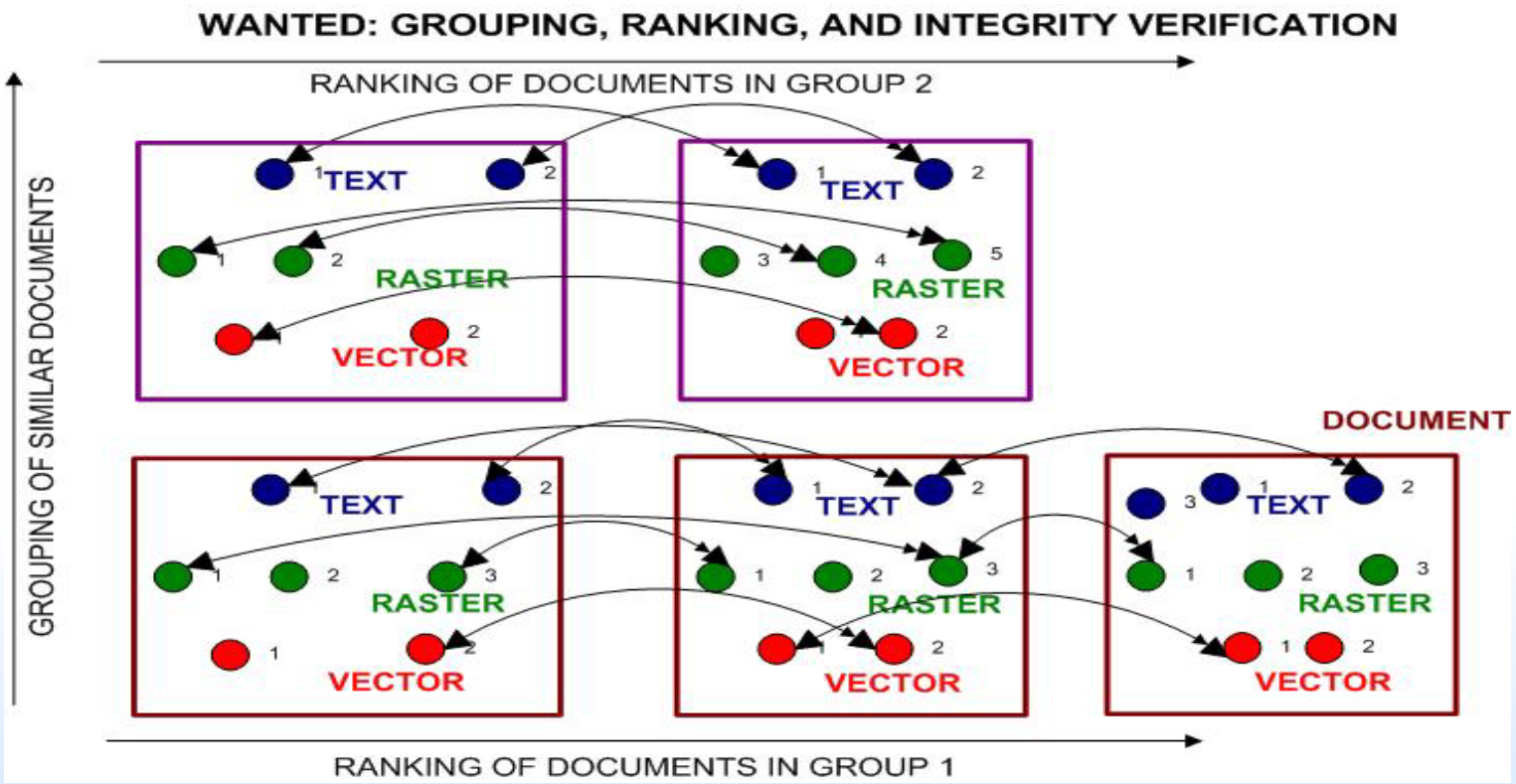


- Software: Adobe 3D Reviewer
- Original File: WRL
- Converted Files: STP, STL, IGS, U3D
- Comparison Method: Light Fields [Chen, 2003] compares silhouettes from various viewing angles around the objects

Conclusion: Information loss(WRL→STP)=Information loss (WRL→STL)>0

Comparison of File Containers (Doc2Learn)

- **Content Based Comparison of Files in the Same Format:** Given hundreds of versions of the 'same' Adobe PDF file, **which file version(s) are similar**?



Multiple Object Comparisons (Doc2Learn)

Adobe PDF documents ~ {text, images, vector graphics, ….}

Document Analyzer

Load | Launch Document | Use Ignore List | Compare | Group | Move to Ignore List | Show Comparison Visualization | Show Second Window | Done | Help

Done

Document List

No	Filename	File Size	File Date	Num Pages	Num Images
1	amwars-2.pdf	88KB	Fri May 22 08:49:59 CDT 2009	2	1
2	amwars.pdf	88KB	Fri May 22 08:49:59 CDT 2009	2	1
3	claimpro-2.pdf	57KB	Fri May 22 08:49:59 CDT 2009	3	1
4	claimpro.pdf	57KB	Fri May 22 08:49:59 CDT 2009	3	1
5	comprates-2.pdf	50KB	Fri May 22 08:49:59 CDT 2009	3	1
6	comprates.pdf	50KB	Fri May 22 08:49:59 CDT 2009	3	1

Word Frequency | Image Color Frequency | Vector Graphics Frequency

Word Frequency

Page List

Page No	Number of Words
All	383
1	240
2	187

All Frequency

No	Word	Frequency
0	cronk	1
1	total	13
2	periods	1
3	396944	1
4	4734991	1
5	2007	4
6	freedom	3
7	visit	1
8	113842	1
9	_____…	1
10	storm	1
11	11/1/55	1
12	335	1
13	3278	1
14	–1945	1
15	washington	1
16	14471	1
17	veteran	7
18	317385	1
19	projections	1
20	worldwide	6
21	10.va	1
22	202	1

Minimum frequency: 0 Set

Text Frequency

No	Word	Frequency
0	cronk	1
1	total	13
2	periods	1
3	freedom	3
4	visit	1
5	_____…	1
6	storm	1
7	11/1/55	1
8	–1945	1
9	washington	1
10	veteran	7
11	projections	1
12	worldwide	6
13	10.va	1
14	1/27/73	1
15	cook	1
16	&	1
17	death	1
18	*	2
19	number	3
20	theater	7
21	janeway	1
22	fredrak	1

Minimum frequency: 0 Set

Integer Frequency

No	Word	Frequency
0	4734991	1
1	396944	1
2	2007	4
3	113842	1
4	335	1
5	3278	1
6	14471	1
7	317385	1
8	202	1
9	15006	1
10	651022	1
11	33741	1
12	802381	1
13	600000	1
14	59297	1
15	167	1
16	3	5
17	26000	1
18	2	2
19	1	4
20	7	2
21	6	2
22	1590	1

Minimum frequency: 0 Set

Float Frequency

No	Word	Frequency

Minimum frequency: 0 Set

Ignore List

No	Word
1	
2	&
3	=
4	A
5	Data
6	For
7	In
8	Only
9	The
10	These
11	This
12	We
13	a
14	about
15	above
16	also
17	an

New | Remove | Load | Save

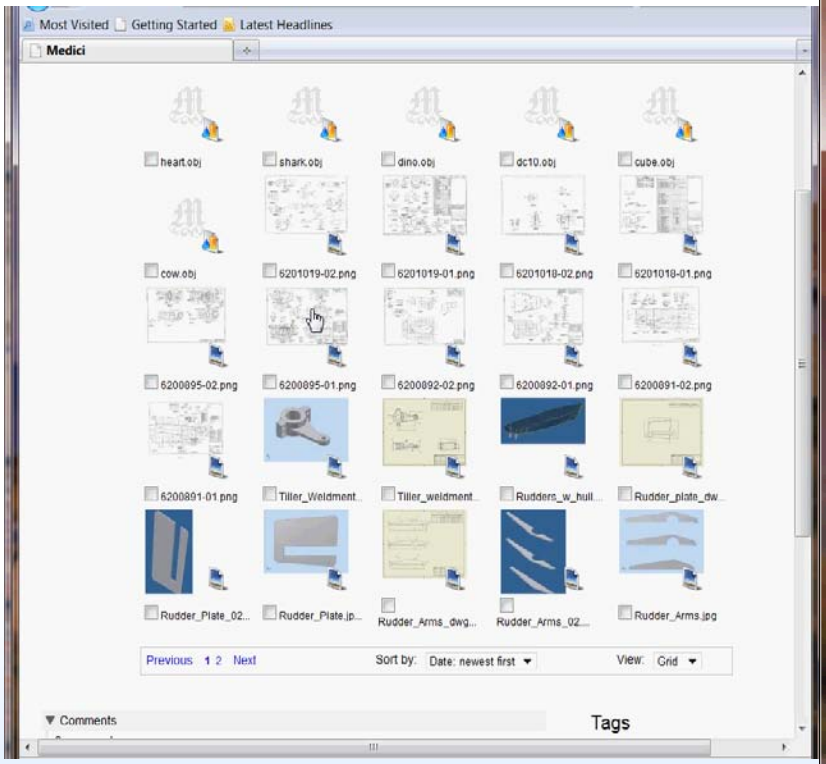
Compare List

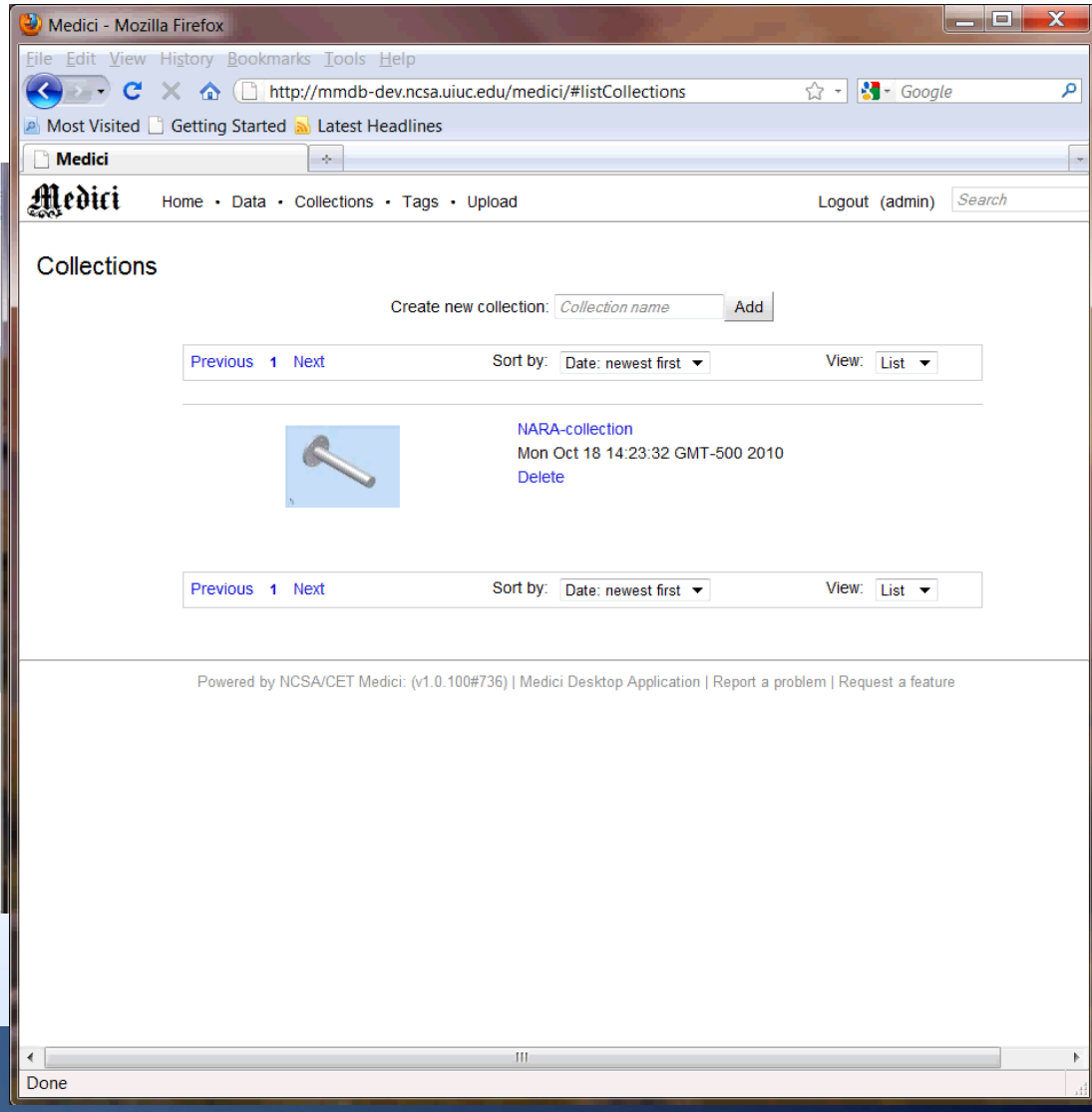
No	Word	Info
1	-	16:16:4:4:4:4
2	.	45:45:53:53:…
3	90	1:1:1:1:1:1
4	and	7:7:27:27:16…
5	are	3:3:3:3:11:11
6	as	2:2:9:9:12:12
7	based	3:3:1:1:2:2
8	be	1:1:2:2:7:7
9	benefits	1:1:4:4:1:1
10	compensation	1:1:5:5:12:12
11	for	2:2:10:10:17…
12	from	2:2:6:6:4:4
13	has	1:1:3:3:2:2
14	have	1:1:1:1:4:4
15	in	16:16:20:20:…
16	is	4:4:6:6:16:16
17	last	14:14:1:1:1:1

Save

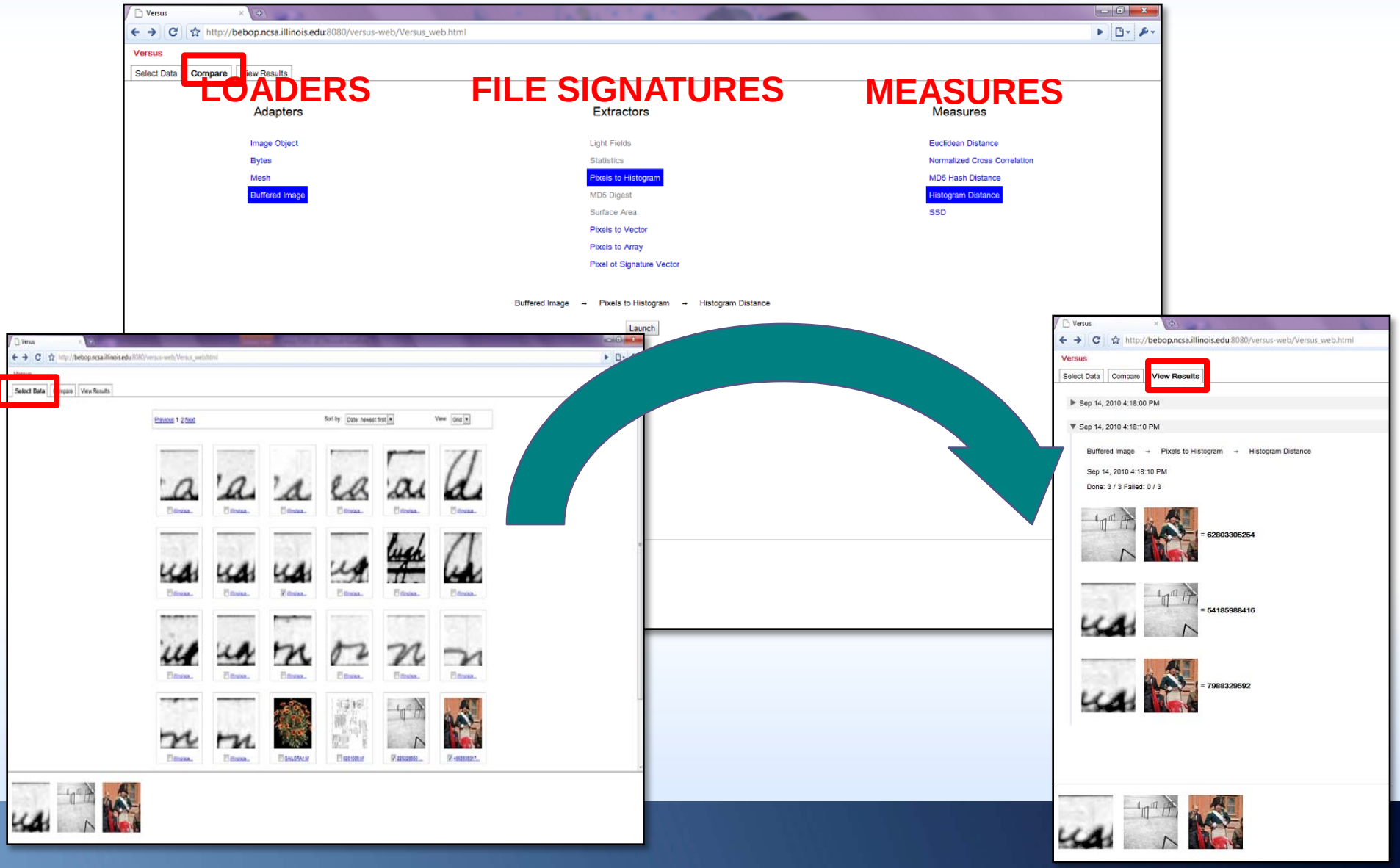
Service: Share, Curate and Select Files for Comparison (Medici)

- Multimedia content management

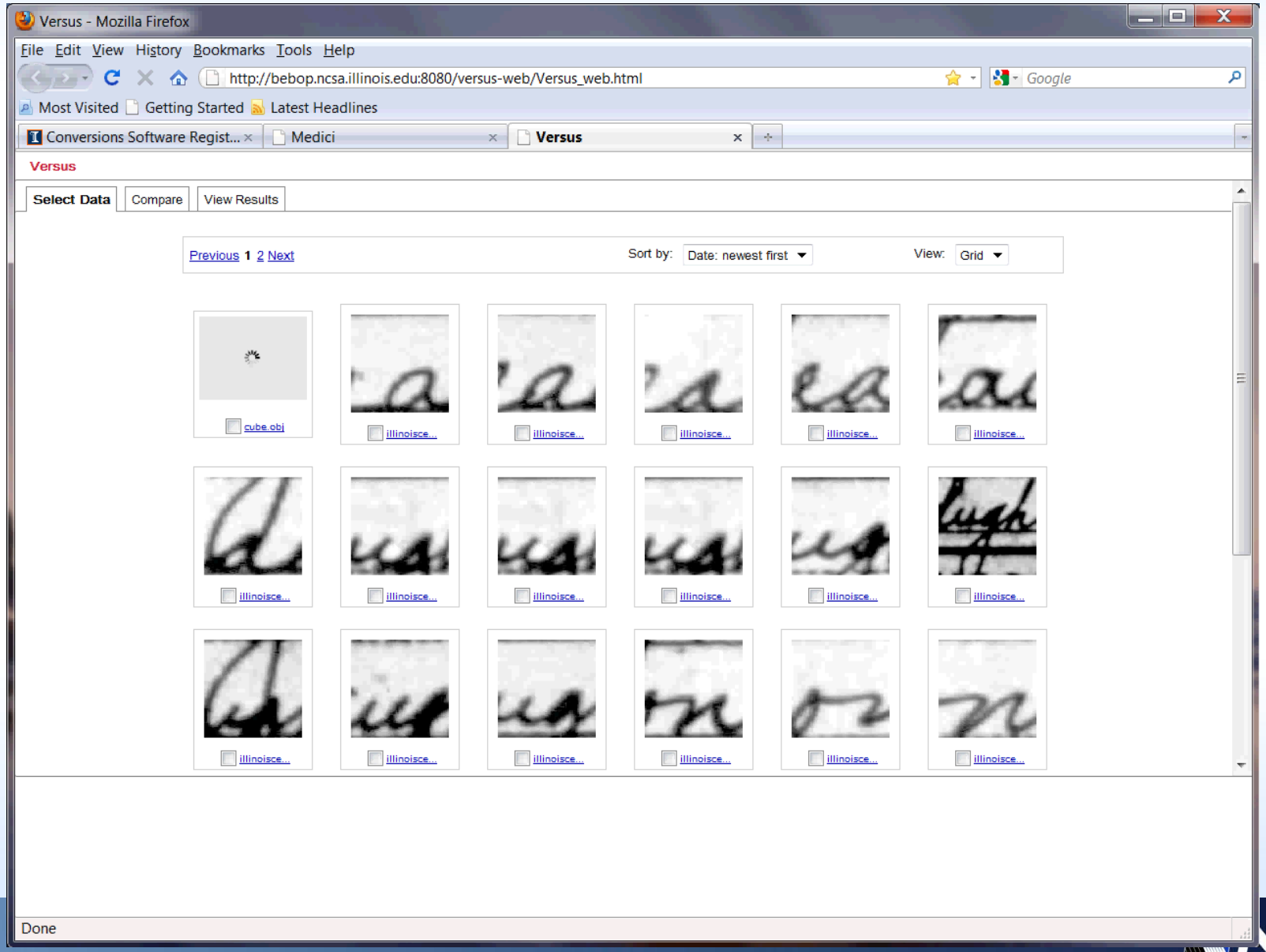




Service: Content-based File Comparison (Versus)

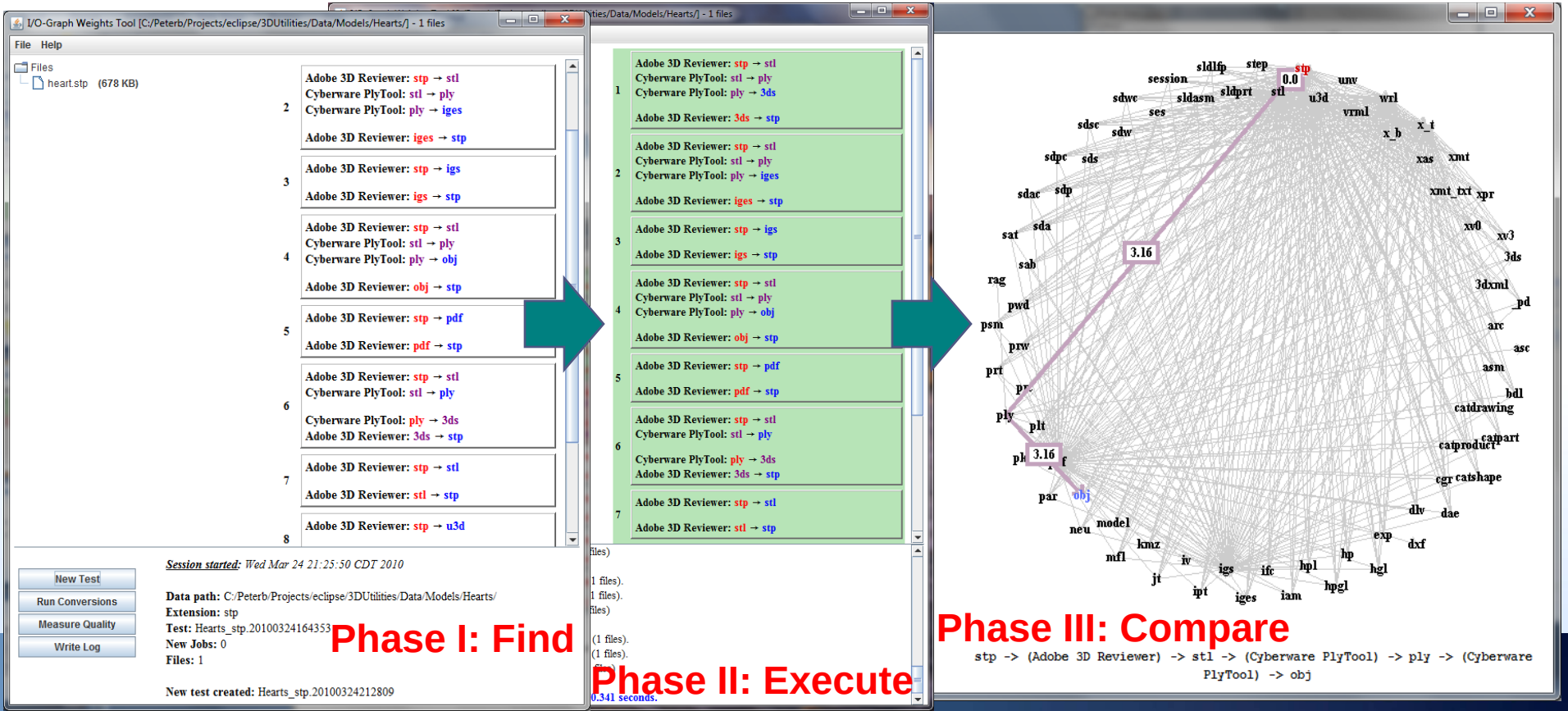


Comparison Framework: Versus



Information Loss Evaluation: Computational Requirements

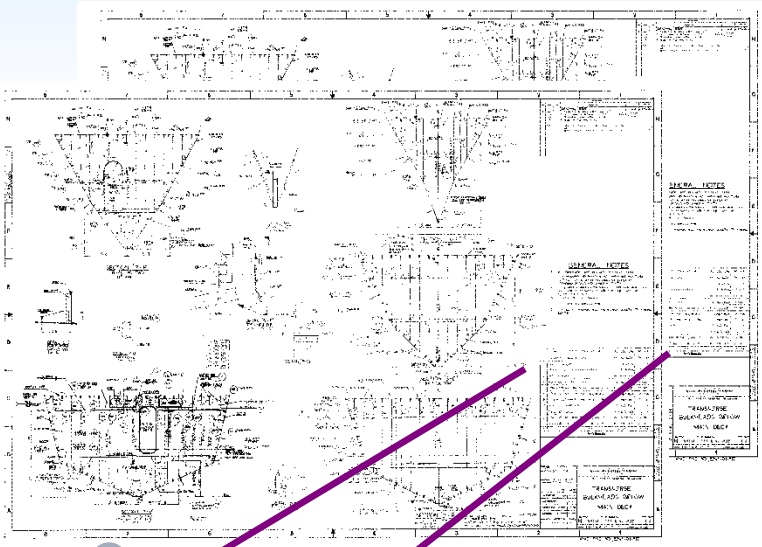
- Files: one file in STP file format
- Software: Adobe 3D Reviewer, Cyberware PlyTool
- Comparison Method: Light Fields [Chen, 2003]
- Number of closed paths: 10 (28 individual conversions)

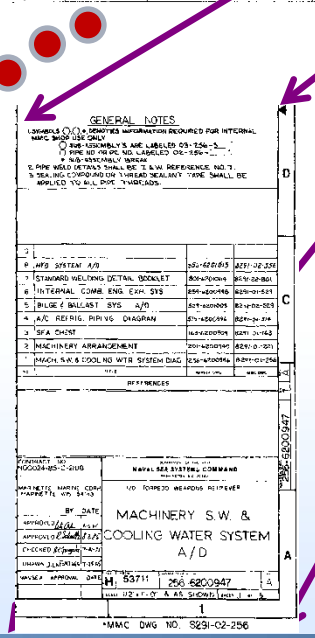


A Discovery of Relationships Among 2D and 3D Digital File Collections:

+Characterization of file content
+Automation of content signature extraction
+Semi-automated comparison and relationship discovery

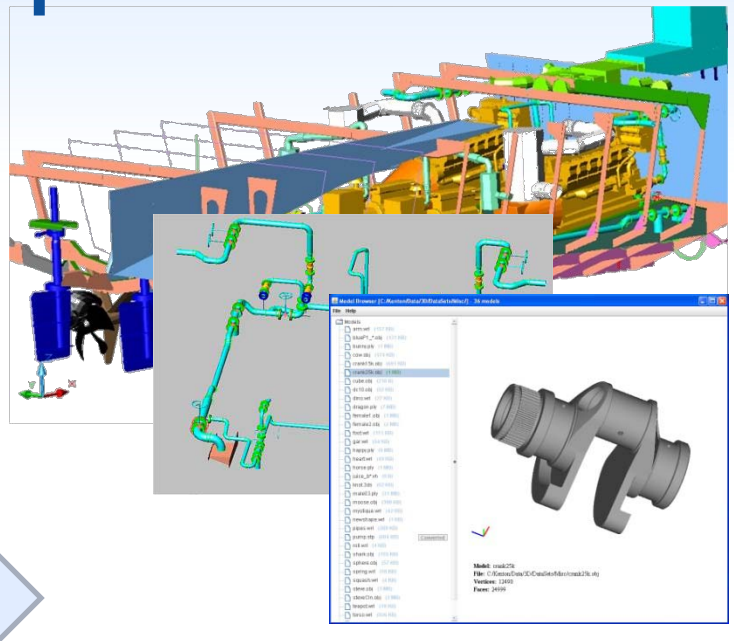
Relationship Discovery: Approach





OCR

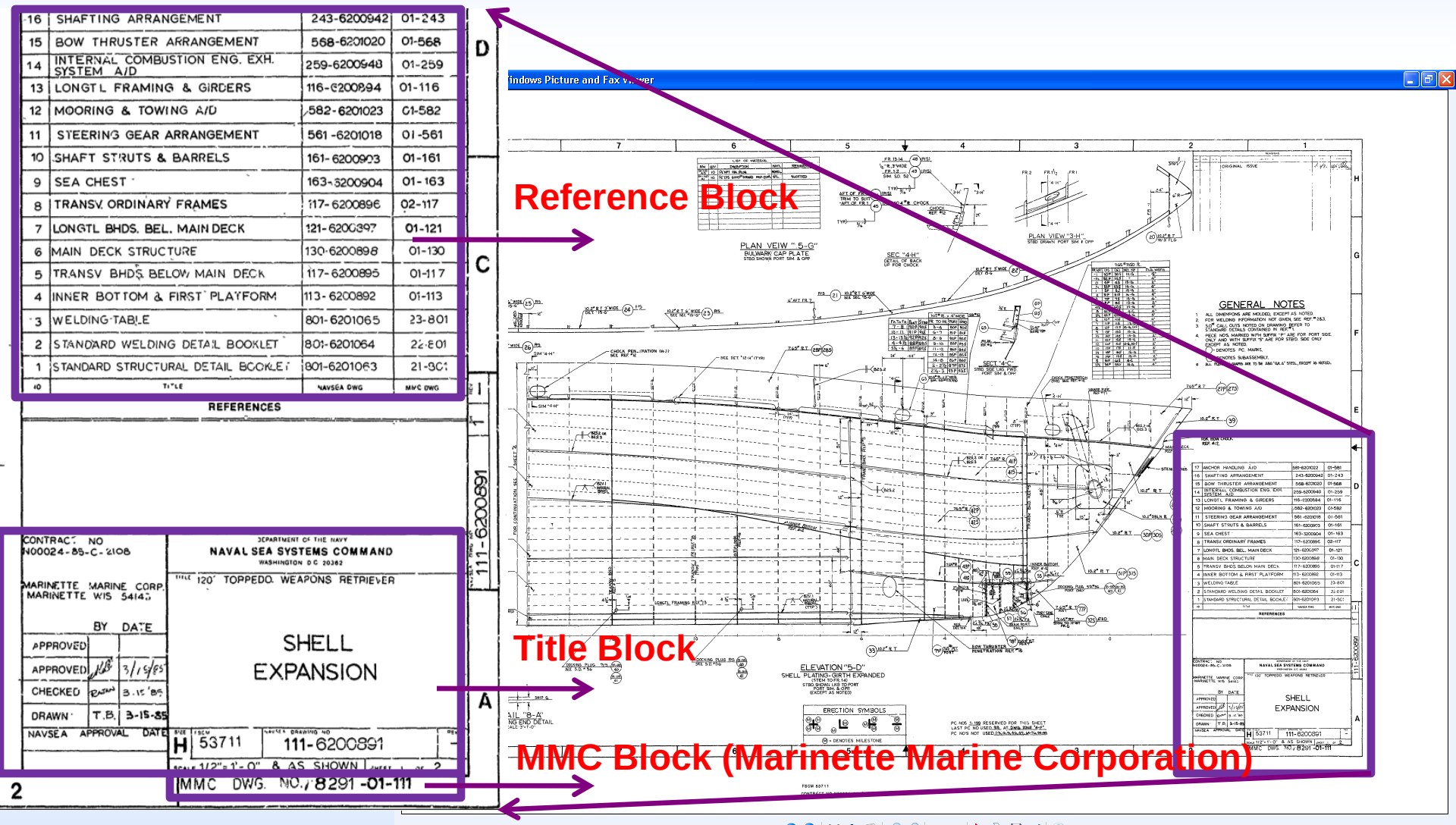
Relationship Discovery



Part name,
Author,
Software,
Date, …

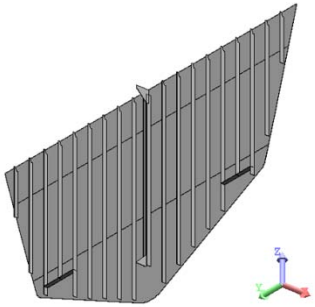
File Descriptors/Signatures

Content Analyses: Optical Character Recognition (OCR) of 2D Drawings



Content Based Extraction from STEP Files

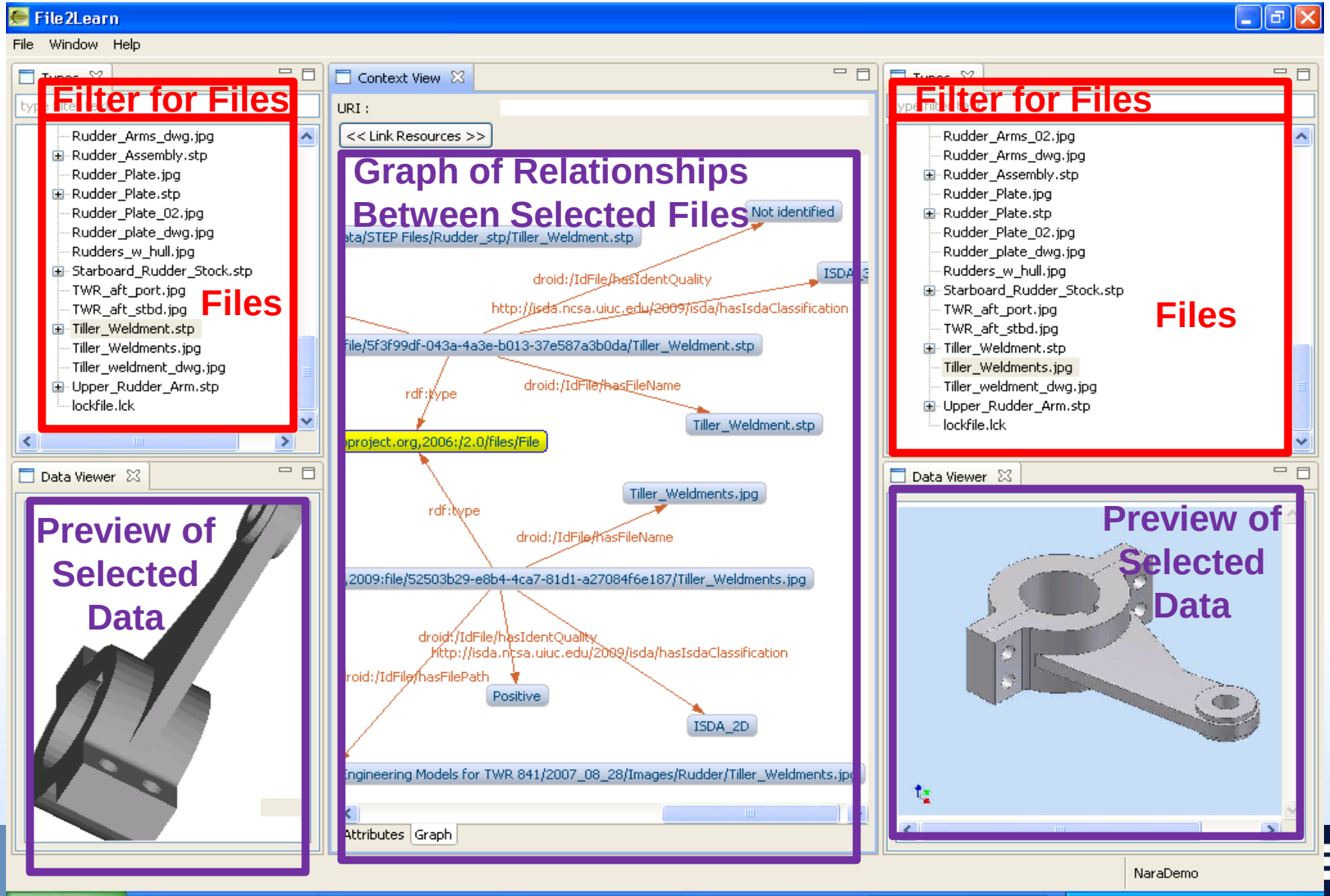
- 3D CAD models in STEP file format are searched for any ASCII strings matching English dictionary and following STEP metadata specification.



Example Metadata for TWR841 ship deck

STEP METADATA SPECIFICATION	EXPECTED STEP METADATA	PARSED STEP METADATA
FILE_DESCRIPTION(/* description */ (''), /* implementation_level */ '2;1'); FILE_NAME(/* name */ '', /* time_stamp */'', /* author */ (''), /* organization */ (''), /* preprocessor_version */ ' ', /* originating_system */ '', /* authorization */ ' ');	FILE_DESCRIPTION((''), /* implementation_level */ '2;1'); FILE_NAME('120 TORPEDO WEAPONS RETRIEVER, TRANSVERSE BULKHEADS BELOW, MAIN DECK', '04-10-86', ('LDOBSON'), ('NAVAL SEA SYSTEMS COMMAND'), ' ', 'IDA-STEP', ' ');	FILE_DESCRIPTION((''), '2;1'); FILE_NAME('D:\\NARA\\Archieve_data_samples\\BHD_FR12\\U2110_BHD12_2007_05_09.stp', '2007-05-10T13:45:37', ('rakowpj'), (''), 'Autodesk Inventor 11', 'Autodesk Inventor 11', '');

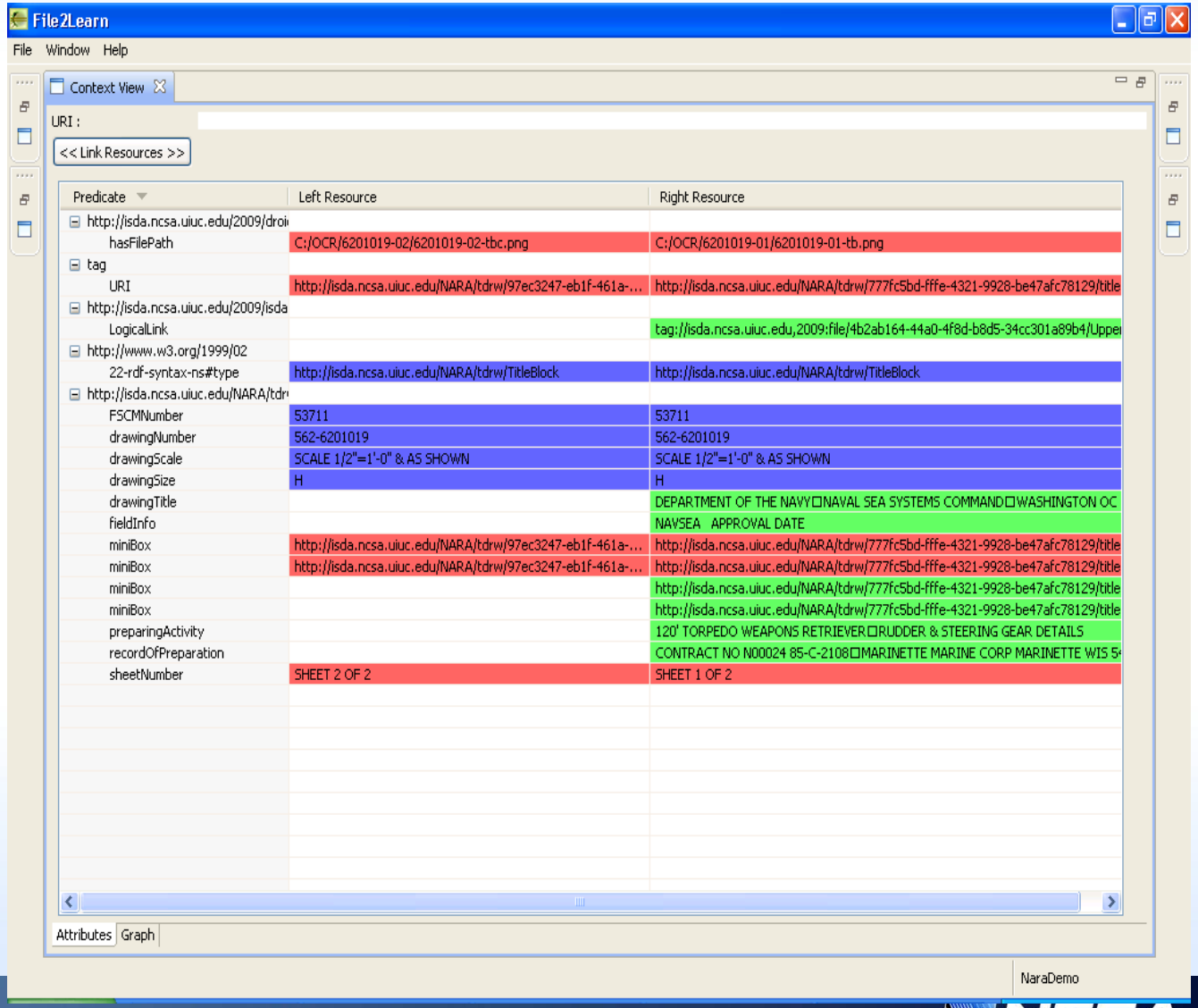
Exploratory Framework – User Interface Overview



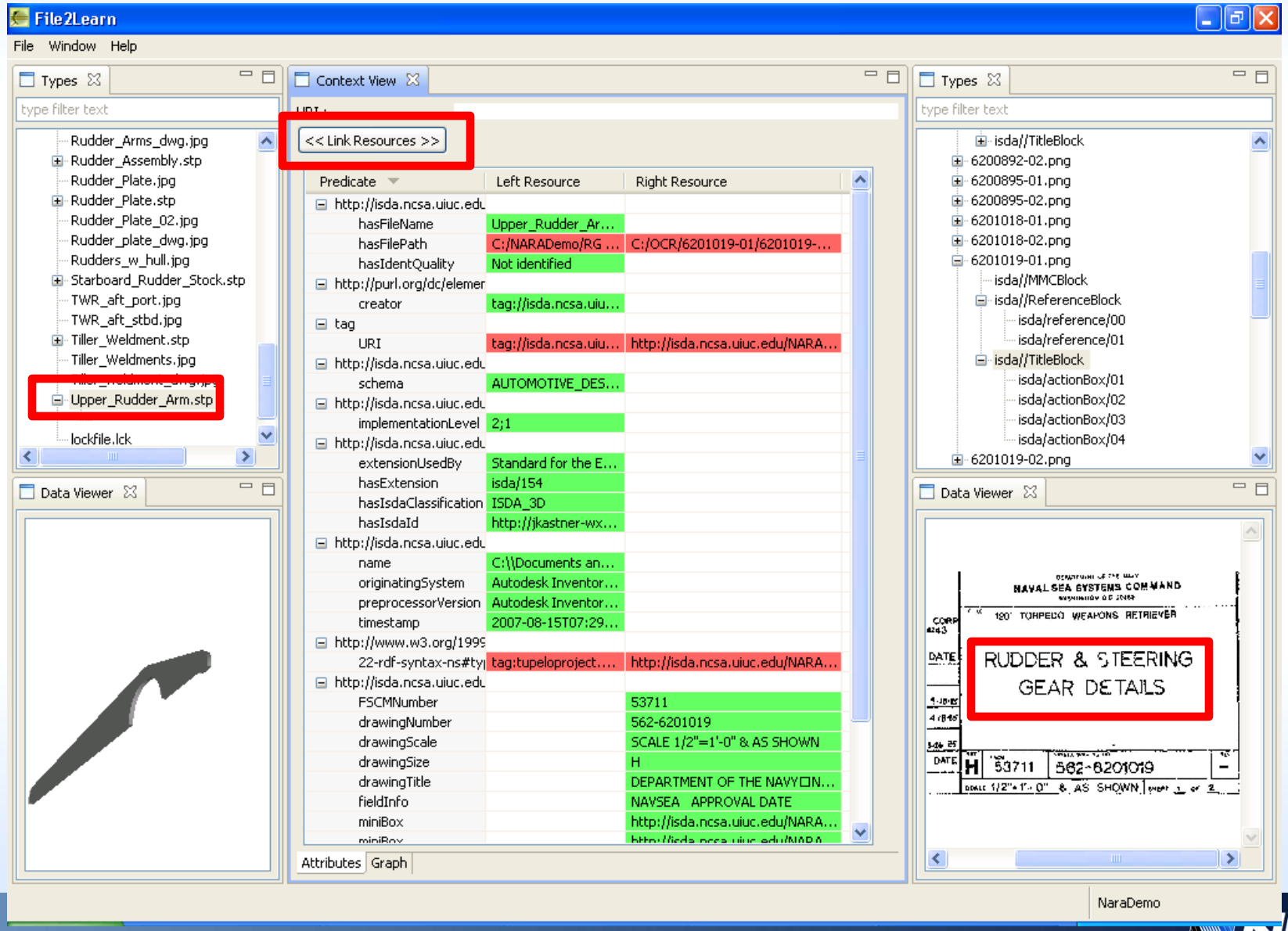
Comparison of Files

Color encoding:

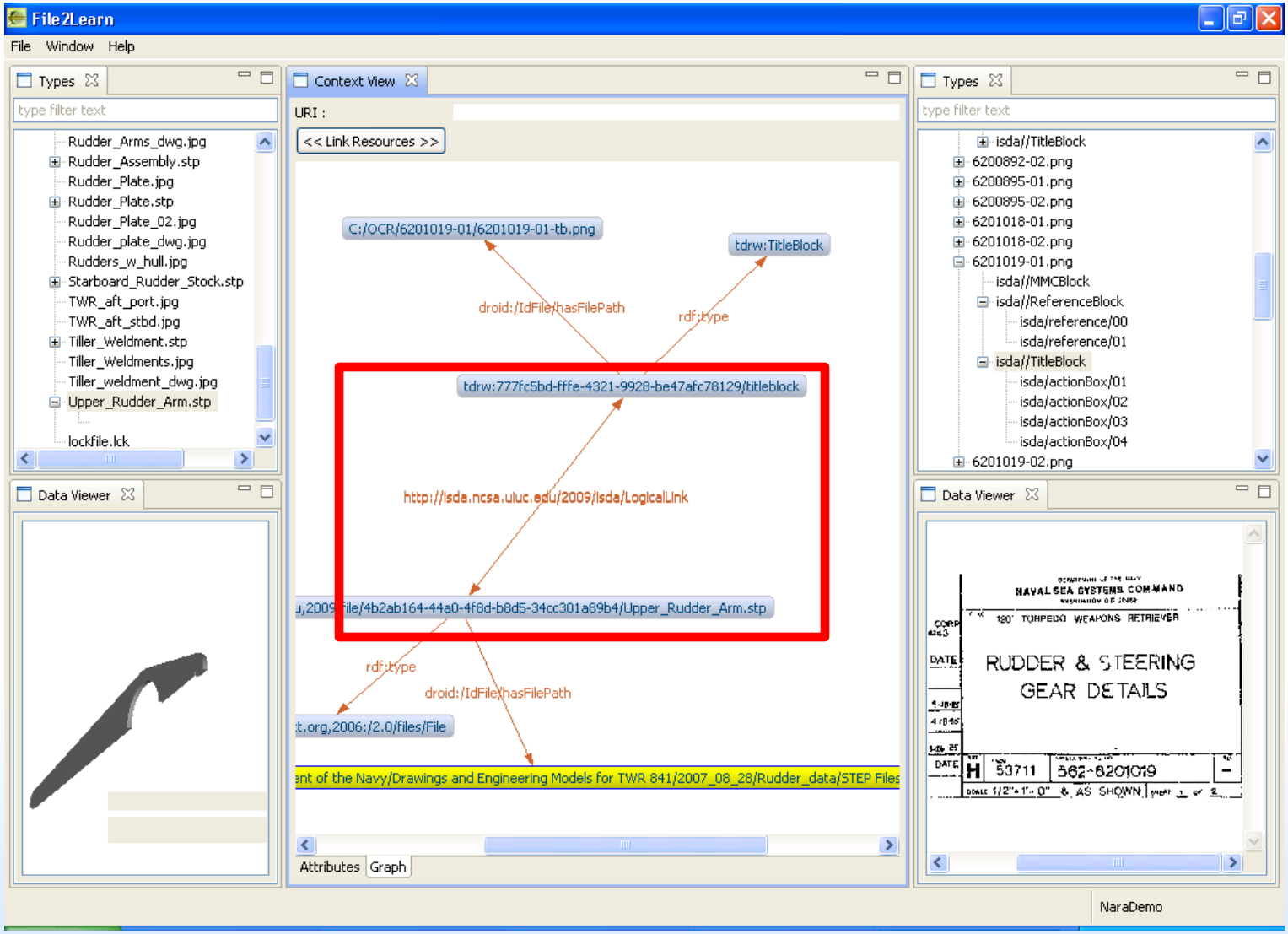
- **Predicates and values match**
- **Predicates match**
- **Predicate occurs only in one file**



Establish File Relationships



Establish File Relationships: Logical Link



Summary

- Model Based Enterprise (MBE) can benefit from distributed **file format conversion and comparison services**
- In general, finding relationships is still an open problem
- **Automation and scalability** are the keys to keeping files current, to dealing with the quantities, and to finding information
- MBE and Preservation Solutions
  - Forward looking solutions should be based on standards
  - Contemporary solutions have to be based on understanding historical electronic records

Software Summary

- The technologies are documented at http://isda.ncsa.uiuc.edu/
- Most of the software is available for downloading at http://isda.ncsa.uiuc.edu/download/
- Feedback is very welcome

- Questions: Peter Bajcsy – pbajcsy@ncsa.illinois.edu