



Technologies For Appraising and Managing Electronic Records

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 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, or the U.S. government.
- Contributions by: Peter Bajcsy, Kenton McHenry, Rob Kooper, Michal Ondrejcek, Jason Kastner, William McFadden, and Sang-Chul Lee

Outline

- **Introduction**
- **A discovery of relationships among digital file collections (file2learn)**
- **A comprehensive comparison of contemporary documents (doc2learn)**
- **Automated file format conversions and conversion quality assessment (Polyglot)**
- **Summary**

Introduction

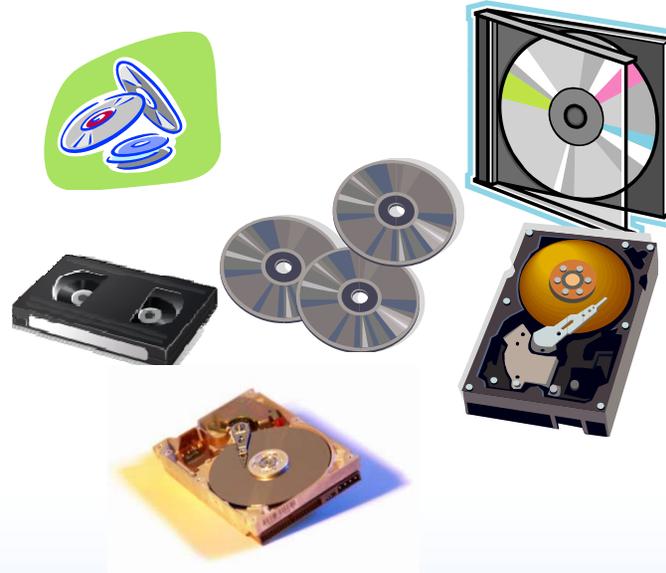
Supporting NARA's Strategic Plan

- According to *The Strategic Plan of The National Archives and Records Administration 2006–2016*. “Preserving the Past to Protect the Future”
 - **“Strategic Goal:** We will preserve and process records to ensure access by the public as soon as legally possible”
 - **“D. We will improve the efficiency with which we manage our holdings from the time they are scheduled through accessioning, processing, storage, preservation, and public use.”**

To Be Preserved!



Digital representation of information & knowledge



Preservation



AGENCY



ARCHIVES

Do We Know the Answers?

Questions During Appraisal of Electronic Records Series

- (1) Given M full DVDs with files, **which files are related?**
- (2) Given N versions of the 'same' file, **which file version(s) should be preserved?**

Do We Know the Answers?

- (3) Given P file formats, **which file format to use and which conversion software to use** so files would be possible to view in a long run?
 - **How much information is lost during file format conversion?**
- (4) **What is the granularity of information** that one should preserve about a decision process in order to reconstruct it?

Goal: Design Technologies for Appraising and Managing Electronic Records

- Technologies should address the following problems:
 - (1) a discovery of relationships among digital file collections (file2learn)
 - (2) a comprehensive comparison of contemporary documents (doc2learn)
 - (3) automated file format conversions and conversion quality assessment (Polyglot)

A Discovery of Relationships Among Digital File Collections

Discovering Relationships Among Files

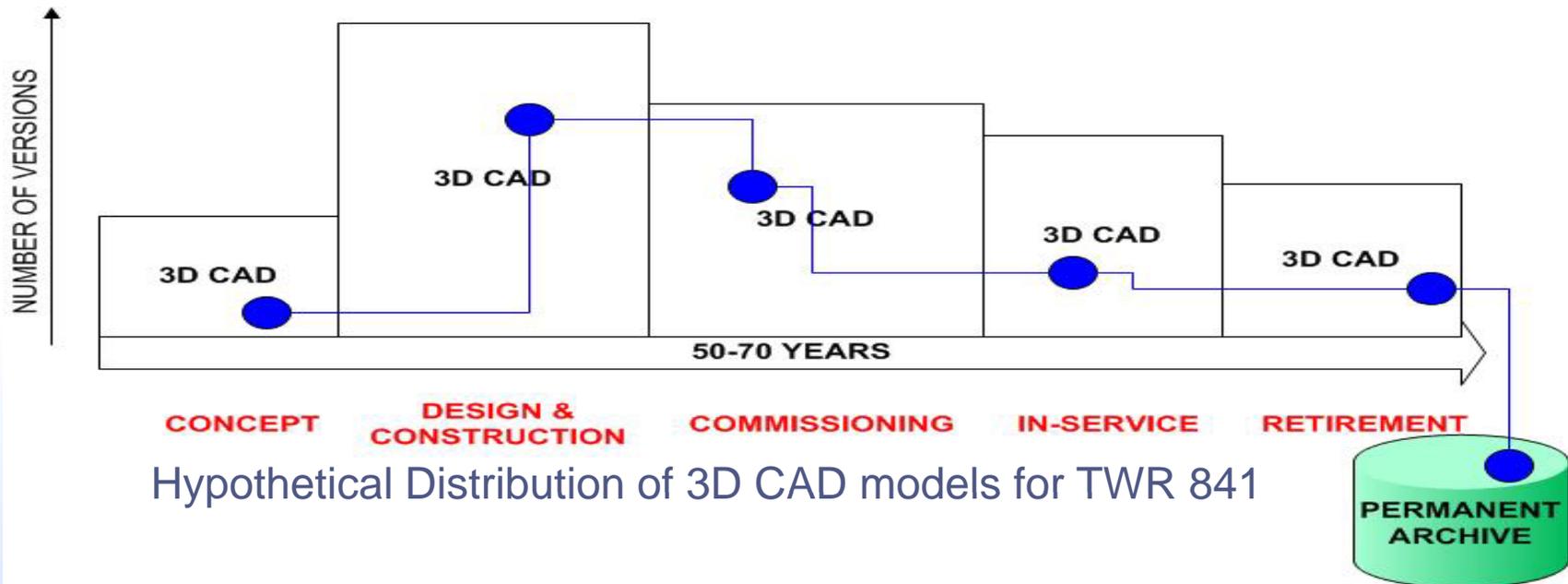
- How should one establish **relationships among electronic records** coming
 - From disparate sources or
 - From the same source at multiple time instances?
- **Need to Understand the Complexity of the Problem**

Discovering Relationships Among Files: Components

- Metadata describing electronic records
 - **How to extract** metadata?
 - **How to automate** metadata extraction from multiple data types, e.g., 2D drawings and 3D CAD models?
- Storage of metadata
 - **What ontology** to use to represent the extracted metadata?
 - **How to represent and store** data and metadata?
- Exploratory and Search Capabilities
 - **How to automate discovery** of relationships?
 - How to support discovery of relationships between electronic records corresponding to the same physical objects but **different multidimensional observations**?

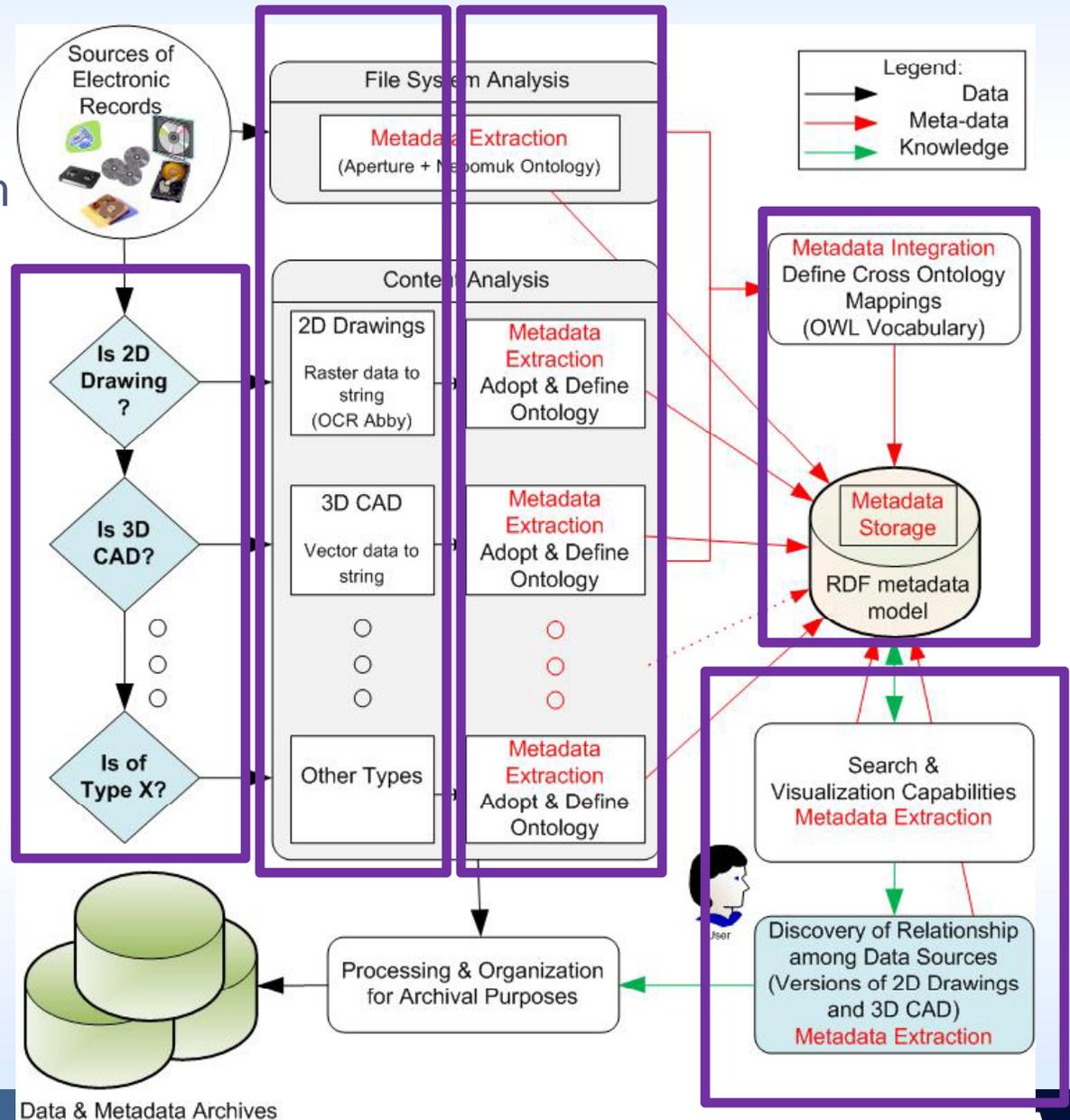
Relationships Among Multiple Data Types

- **Example Data:** Torpedo Weapon Retriever 841
 - 784 existing 2D image drawings and $N > 22$ 3D CAD models
- How to establish relationships among the 3D CAD models and 2D image drawings during a product lifecycle?



Methodology

- File Identification
- Information Extraction from
 - File System
 - File Content
- Information Organization
 - Taxonomy (classification)
 - Ontology (relationships)
- Information Representation, Integration and Storage
 - XML
 - RDF
- Relationship Discovery



File Identification and File System Analyses

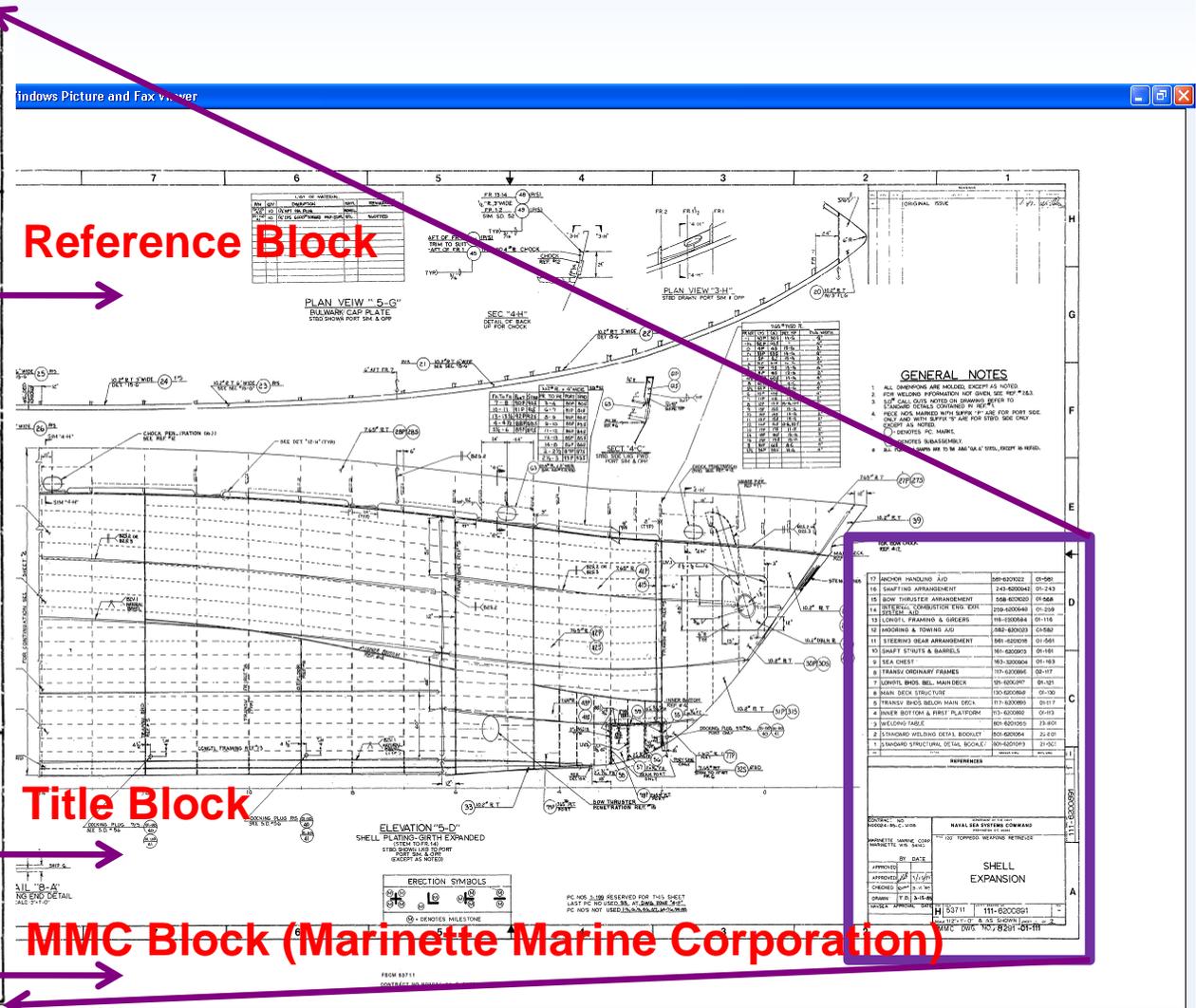
- **File Identification**
 - What is the file format?
 - Is the file format well formed?
- **Approach: Used DROID built on top of the PRONOM File Registry with additional NCSA support of 3D file identification**
- **Metadata extraction about a file system**
 - Where is the file located?
 - What is the file size, time stamp, etc.?
- **Approach: Use any file system information extraction software, such as Aperture (cross platform, open source, active development), Google desktop, OS specific solutions (e.g., Apple Spotlight, Linux, MS Search)**

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

16	SHAFTING ARRANGEMENT	243-6200942	01-243
15	BOW THRUSTER ARRANGEMENT	568-6201020	01-568
14	INTERNAL COMBUSTION ENG. EXH. SYSTEM A/D	259-6200940	01-259
13	LONGTL FRAMING & GIRDERS	116-6200994	01-116
12	MOORING & TOWING A/D	582-6201023	01-582
11	STEERING GEAR ARRANGEMENT	561-6201018	01-561
10	SHAFT STRUTS & BARRELS	161-6200903	01-161
9	SEA CHEST	163-6200904	01-163
8	TRANSV. ORDINARY FRAMES	117-6200896	02-117
7	LONGTL BHD. BEL. MAIN DECK	121-6200897	01-121
6	MAIN DECK STRUCTURE	130-6200898	01-130
5	TRANSV BHD. BELOW MAIN DECK	117-6200895	01-117
4	INNER BOTTOM & FIRST PLATFORM	113-6200892	01-113
3	WELDING TABLE	801-6201065	23-801
2	STANDARD WELDING DETAIL BOOKLET	801-6201064	22-801
1	STANDARD STRUCTURAL DETAIL BOOKLET	801-6201063	21-801
NO	TITLE	NAVSEA DWG	MMC DWG

REFERENCES

CONTRACT NO N00024-85-C-210B		DEPARTMENT OF THE NAVY NAVAL SEA SYSTEMS COMMAND WASHINGTON D C 20382	
MARINETTE MARINE CORP MARINETTE WIS 54143		TITLE 120' TOPPED WEAPONS RETRIEVER	
BY DATE			
APPROVED			
APPROVED	JLB	3/15/85	
CHECKED	RAM	3-15-85	
DRAWN	T.B.	3-15-85	
NAVSEA APPROVAL DATE			
SHEET NO	53711	DRAWING NO	111-6200891
SCALE	1/2" = 1'-0" & AS SHOWN	SHELL EXPANSION	
MMC DWG. NO. 78291-01-111			



Reference Block

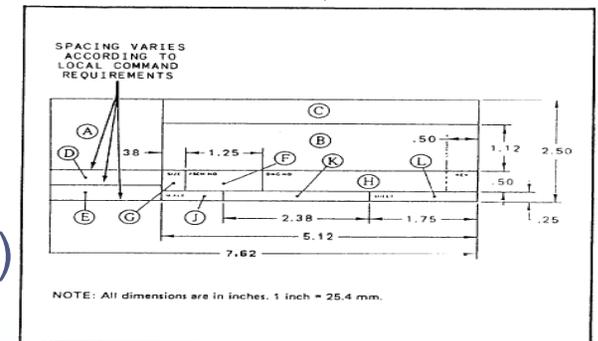
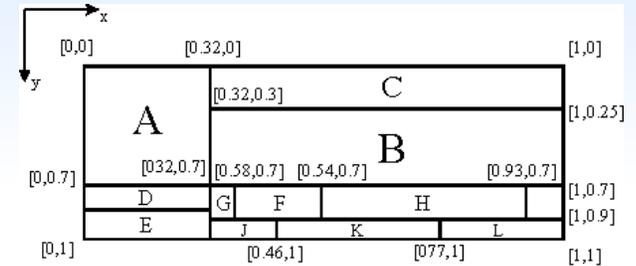
Title Block

MMC Block (Marinette Marine Corporation)

Title Block: Ontology and Metadata Representation

Ontology for sub-fields:

- A – Record of preparation (<tdrw:recordOfPreparation>),
- B – Drawing title (<tdrw:drawingTitle>),
- C – Preparing Activity <tdrw:preparingActivity>,
- F – Code identification number (<tdrw:FSCMNumber>),
- G – Drawing size (<tdrw:drawingSize>),
- H – Drawing number (<tdrw:drawingNumber>),
- J – Scale (<tdrw:drawingScale>),
- K – Specification number (<tdrw:drawingNumber>)
- L – Sheet number (<tdrw:sheetNumber>).



Resource Description Framework (RDF):

- Metadata representation: subject – predicate - object

MMC and Reference Blocks: Organization

- MMC Blocks

MMC DWG. NO. 8291-02-506

H 53711 506-6200985 B
SCALE 1/2" = 1'-0" & AS SHOWN SHEET 2 OF 7

H 53711 506-6200985 B
SCALE 1/2" = 1'-0" & AS SHOWN SHEET 2 OF 7

CONTRACT NO. N00024-85-C-2108	DEPARTMENT OF THE NAVY NAVAL SEA SYSTEMS COMMAND WASHINGTON, D.C. 20382	437-02
MARINETTE MARINE CORP. MARINETTE WIS. 54143	TITLE 120' TORPEDO WEAPONS RETRIEVER	
BY DATE APPROVED: <i>D. HARTMAN</i> 4-24-85	MISC. CONTROL & INDICATING SYSTEMS WIRING DIAGRAM	A
APPROVED: <i>A. ROY</i>		
CHECKED: <i>D. FROHM</i> 4-25-85		
DRAWN: <i>R.H.I.</i> 4-25-85		
NAVSEA APPROVAL DATE	SHEET 1 OF 4	C
H 53711 437-6200979		
SCALE NONE		

MMC DWG. NO. 8291-03-437

Inconsistencies

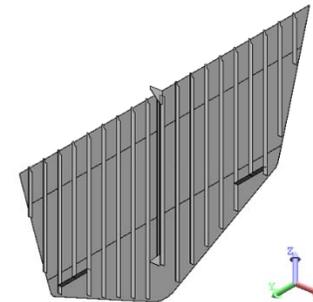
LIST OF MATERIAL									
SYMBOL	QTY.	DESCRIPTION	MATERIAL OR MANUFACTURER	SPEC. GRADE OR MODEL	UNIT	REMARKS	REVISIONS		
F1	7	1 1/2" IPS. WRENCH	WRENCH	WRENCH			1	ORIGINAL LIST	
F2	1	2" IPS 90° S.R. ELL. SCH. 40 B.W.	BLK STL	ASTM A234			2	ADDED P/E WRENCH & BRACKET	
F3	6	2" IPS 90° L.R. ELL. SCH. 40 B.W.	BLK STL	ASTM A234			3	REPLACED FILED WITH T2151 AND T2415 ON SMOOTHER TUBES PER SCH 40 TO NAME USE OF STD WRENCH IN 201	
F4	2	2 1/2" IPS 90° L.R. ELL. SCH. 40 B.W.	BLK STL	ASTM A234			4		
F5	14	1 1/2" IPS DECK SLEEVE	BLK STL	COM.L.		MMC MAKE			
F6	4	1 1/2" IPS 90° ELL. 3000" S.W.	BLK STL	ASTM A105					
F7	0	1 1/2" IPS COMPRESSION SLEEVE	VARIOUS	COM.L.					
F8	4	3/4" IPS COUPLING 3000" S.W.	BLK STL	ASTM A105					
F9	3	2" IPS 90° L.R. ELL. SCH. 40 B.W.	BLK STL	ASTM A234					
F10	6	1/2" IPS BOURLETT 3000" S.W.	BLK STL	ASTM A105					
F11	16	3/4" IPS 90° ELL. 3000" S.W.	BLK STL	ASTM A105					
F12	5	3/4" IPS UNION 3000" S.W.	BLK STL	ASTM A105					
F13	1	3/4" IPS TEE 3000" S.W.	BLK STL	ASTM A105					
F14	8	3/4" IPS DECK SLEEVE	BLK STL	COM.L.		MMC MAKE			
F15	2	WRENCH				TATE TEMCO #5870			
F16	0	1 1/2" IPS CAP 3000" S.W.	BLK STL	ASTM A105					
F17	8	STOWER PLATE 2 1/4" X 2 1/4" X 1/4" ANGLE X 2 1/2" LG.	BLK STL	COM.L.		MMC MAKE			
F18	4	1 1/2" IPS UNION 3000" S.W.	MI GALV	A-197					
F19	4	1 1/2" IPS COUPLING 3000" S.W.	BLK STL	ASTM A105					
F20	4	2 1/2" IPS 45° ELL. SCH. 40 B.W.	BLK STL	ASTM A234					
F21	2	3" IPS 45° ELL. SCH. 40 B.W.	BLK STL	ASTM A234					
F22	1	3" IPS 90° ELL. S.R. SCH. 80 B.W.	BLK STL	ASTM A234					
F23	5	1 1/2" COMP. FPT. 3000"	STEEL	ASTM A193					
F24	5	1 1/2" PIPE PLUG. NPT. COUNTER SINK	STEEL			STOCKHAM FIG. 185			
J1	9	3" IPS FLANGE ISO 50.0 W.R.	BLK STL	ASTM A181					
J2	9	3" IPS FLANGE ISO 50.0 W.R. PIP. SCH. 80	BLK STL	ASTM A181					
J3	44	3/8" - 11 UNF 2		2-40-080048	01-2-40				
J4	44	3/8" - 11 UNF 1		508-620020	01-568				
J5	6	3" IPS FLANGE		209-620040	01-209				
J6	2	2 1/2" IPS FLANGE		116-620094	01-116				
J7	2	2 1/2" IPS FLANGE		582-620023	01-582				
J8	2	3" IPS FLANGE		561-620018	01-561				
J9	2	3" IPS FLANGE		161-620093	01-161				
J10	2	2 1/2" IPS FLANGE		163-620094	01-163				
V1	2	2 1/2" IPS VALVE		117-620086	02-117				
V2	9	5" IPS VALVE		131-620097	01-121				
V3	2	2 1/2" IPS VALVE		130-620088	01-130				
V4	2	2 1/2" IPS VALVE		117-620085	01-117				
V5	2	2 1/2" IPS VALVE		113-620082	01-113				
V6	2	2 1/2" IPS VALVE		801-620105	23-801				
V7	2	2 1/2" IPS VALVE		801-620105	23-801				
V8	2	2 1/2" IPS VALVE		801-620105	23-801				
V9	2	2 1/2" IPS VALVE		801-620105	23-801				
V10	2	2 1/2" IPS VALVE		801-620105	23-801				
V11	2	2 1/2" IPS VALVE		801-620105	23-801				
V12	2	2 1/2" IPS VALVE		801-620105	23-801				
V13	2	2 1/2" IPS VALVE		801-620105	23-801				
V14	2	2 1/2" IPS VALVE		801-620105	23-801				
V15	2	2 1/2" IPS VALVE		801-620105	23-801				
V16	2	2 1/2" IPS VALVE		801-620105	23-801				
V17	2	2 1/2" IPS VALVE		801-620105	23-801				
V18	2	2 1/2" IPS VALVE		801-620105	23-801				
V19	2	2 1/2" IPS VALVE		801-620105	23-801				
V20	2	2 1/2" IPS VALVE		801-620105	23-801				
V21	2	2 1/2" IPS VALVE		801-620105	23-801				
V22	2	2 1/2" IPS VALVE		801-620105	23-801				
V23	2	2 1/2" IPS VALVE		801-620105	23-801				
V24	2	2 1/2" IPS VALVE		801-620105	23-801				
V25	2	2 1/2" IPS VALVE		801-620105	23-801				
V26	2	2 1/2" IPS VALVE		801-620105	23-801				
V27	2	2 1/2" IPS VALVE		801-620105	23-801				
V28	2	2 1/2" IPS VALVE		801-620105	23-801				
V29	2	2 1/2" IPS VALVE		801-620105	23-801				
V30	2	2 1/2" IPS VALVE		801-620105	23-801				
V31	2	2 1/2" IPS VALVE		801-620105	23-801				
V32	2	2 1/2" IPS VALVE		801-620105	23-801				
V33	2	2 1/2" IPS VALVE		801-620105	23-801				
V34	2	2 1/2" IPS VALVE		801-620105	23-801				
V35	2	2 1/2" IPS VALVE		801-620105	23-801				
V36	2	2 1/2" IPS VALVE		801-620105	23-801				
V37	2	2 1/2" IPS VALVE		801-620105	23-801				
V38	2	2 1/2" IPS VALVE		801-620105	23-801				
V39	2	2 1/2" IPS VALVE		801-620105	23-801				
V40	2	2 1/2" IPS VALVE		801-620105	23-801				
V41	2	2 1/2" IPS VALVE		801-620105	23-801				
V42	2	2 1/2" IPS VALVE		801-620105	23-801				
V43	2	2 1/2" IPS VALVE		801-620105	23-801				
V44	2	2 1/2" IPS VALVE		801-620105	23-801				
V45	2	2 1/2" IPS VALVE		801-620105	23-801				
V46	2	2 1/2" IPS VALVE		801-620105	23-801				
V47	2	2 1/2" IPS VALVE		801-620105	23-801				
V48	2	2 1/2" IPS VALVE		801-620105	23-801				
V49	2	2 1/2" IPS VALVE		801-620105	23-801				
V50	2	2 1/2" IPS VALVE		801-620105	23-801				
V51	2	2 1/2" IPS VALVE		801-620105	23-801				
V52	2	2 1/2" IPS VALVE		801-620105	23-801				
V53	2	2 1/2" IPS VALVE		801-620105	23-801				
V54	2	2 1/2" IPS VALVE		801-620105	23-801				
V55	2	2 1/2" IPS VALVE		801-620105	23-801				
V56	2	2 1/2" IPS VALVE		801-620105	23-801				
V57	2	2 1/2" IPS VALVE		801-620105	23-801				
V58	2	2 1/2" IPS VALVE		801-620105	23-801				
V59	2	2 1/2" IPS VALVE		801-620105	23-801				
V60	2	2 1/2" IPS VALVE		801-620105	23-801				
V61	2	2 1/2" IPS VALVE		801-620105	23-801				
V62	2	2 1/2" IPS VALVE		801-620105	23-801				
V63	2	2 1/2" IPS VALVE		801-620105	23-801				
V64	2	2 1/2" IPS VALVE		801-620105	23-801				
V65	2	2 1/2" IPS VALVE		801-620105	23-801				
V66	2	2 1/2" IPS VALVE		801-620105	23-801				
V67	2	2 1/2" IPS VALVE		801-620105	23-801				
V68	2	2 1/2" IPS VALVE		801-620105	23-801				
V69	2	2 1/2" IPS VALVE		801-620105	23-801				
V70	2	2 1/2" IPS VALVE		801-620105	23-801				
V71	2	2 1/2" IPS VALVE		801-620105	23-801				
V72	2	2 1/2" IPS VALVE		801-620105	23-801				
V73	2	2 1/2" IPS VALVE		801-620105	23-801				
V74	2	2 1/2" IPS VALVE		801-620105	23-801				
V75	2	2 1/2" IPS VALVE		801-620105	23-801				
V76	2	2 1/2" IPS VALVE		801-620105	23-801				
V77	2	2 1/2" IPS VALVE		801-620105	23-801				
V78	2	2 1/2" IPS VALVE		801-620105	23-801				
V79	2	2 1/2" IPS VALVE		801-620105	23-801				
V80	2	2 1/2" IPS VALVE		801-620105	23-801				
V81	2	2 1/2" IPS VALVE		801-620105	23-801				
V82	2	2 1/2" IPS VALVE		801-620105	23-801				
V83	2	2 1/2" IPS VALVE		801-620105	23-801				
V84	2	2 1/2" IPS VALVE		801-620105	23-801				
V85	2	2 1/2" IPS VALVE		801-620105	23-801				
V86	2	2 1/2" IPS VALVE		801-620105	23-801				
V87	2	2 1/2" IPS VALVE		801-620105	23-801				
V88	2	2 1/2" IPS VALVE		801-620105	23-801				
V89	2	2 1/2" IPS VALVE		801-620105	23-801				
V90	2	2 1/2" IPS VALVE		801-620105	23-801				
V91	2	2 1/2" IPS VALVE		801-620105	23-801				
V92	2	2 1/2" IPS VALVE		801-6201					

Summary of OCR Based Analyses

- **Manually encoded block coordinates for 784 files** in PNG (converted from originally LZW compressed TIFF files)
- **Automated OCR and executed OCR on**
 - 700 title blocks,
 - 150 reference blocks,
 - dozen of revision and list of material
 - about 200 additional areas with the drawing numbers (MMC DWG. NO.).
- **Performance benchmarks:**
 - Full OCR of TB, MMC and RF for about 50 image files (105 blocks) took about 6 hours on a quad core machine

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
<pre>FILE_DESCRIPTION(/* description */ (""), /* implementation_level */ '2;1'); FILE_NAME(/* name */ "", /* time_stamp */ "", /* author */ ("), /* organization */ ("), /* preprocessor_version */ '', /* originating_system */ "", /* authorization */ ' ');</pre>	<pre>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', '');</pre>	<pre>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', '');</pre>

Exploratory Framework – User Interface Overview

The screenshot displays the File2Learn application interface, which is divided into several functional areas:

- Filter for Files:** Two red-bordered boxes on the left and right sides of the interface contain lists of files. The left list includes files like `Rudder_Arms_dwg.jpg`, `Rudder_Assembly.stp`, and `Tiller_Weldment.stp`. The right list includes `Rudder_Arms_02.jpg`, `Rudder_Assembly.stp`, and `lockfile.lck`. The word "Files" is written in red next to each list.
- Graph of Relationships Between Selected Files:** A central purple-bordered area displays a graph showing relationships between files. Nodes include file URIs such as `file/5f3f99df-043a-4a3e-b013-37e587a3b0da/Tiller_Weldment.stp` and `project.org,2006:/2.0/files/File`. Relationships are labeled with terms like `droid:/IdFile/hasIdentQuality`, `http://isda.ncsa.uiuc.edu/2009/isda/hasIsdaClassification`, and `rdf:type`. A "Not identified" node is also present.
- Preview of Selected Data:** Two purple-bordered boxes at the bottom corners show 3D CAD models of mechanical parts. The left preview shows a grey metal component, and the right preview shows a similar component with a different view or configuration. The text "Preview of Selected Data" is written in purple above each model.
- Data Viewer:** A window at the bottom left shows a 3D model of a mechanical part, and another at the bottom right shows a 3D model of a similar part.
- Context View:** A window at the top center shows the URI and a "Link Resources" button.
- File List:** A window at the top left shows a list of files with expandable folders.

The interface also includes a menu bar (File, Window, Help) and a status bar at the bottom right with the text "NaraDemo".

Exploratory Framework – User Interface Overview

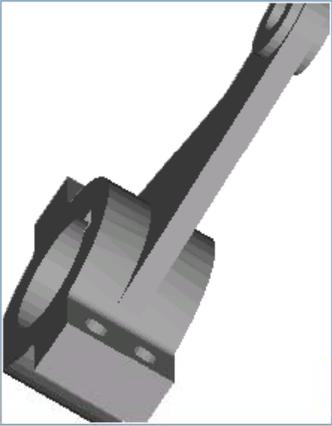
File2Learn
File Window Help

Additional Import/Export and Preference Options

type filter text

- Rudder_Arms_dwg.jpg
- Rudder_Assembly.stp
- Rudder_Plate.jpg
- Rudder_Plate.stp
- Rudder_Plate_02.jpg
- Rudder_plate_dwg.jpg
- Rudders_w_hull.jpg
- Starboard_Rudder_Stock.stp
- TWR_aft_port.jpg
- TWR_aft_stbd.jpg
- Tiller_Weldment.stp
- Tiller_Weldments.jpg
- Tiller_weldment_dwg.jpg
- Upper_Rudder_Arm.stp
- lockfile.lck

Data Viewer



Context view

URI: << Link Resources >>

Predicate	Left Resource	Right Resource
http://purl.org/creator	tag://isda.ncsa.uiuc.edu,2...	
http://isda.ncsa		
hasFileName	Tiller_Weldment.stp	Tiller_Weldments.jpg
hasFilePath	C:/NARADemo/RG 428 - G...	C:/NARADemo/RG 428 - Ge
hasIdentQ	Not Identified	Positive
tag		
URI	tag://isda.ncsa.uiuc.edu,2...	tag://isda.ncsa.uiuc.edu,20
http://isda.ncsa		
schema	AUTOMOTIVE_DESIGN { 1...	
http://isda.ncsa		
implemental	2;1	
http://isda.ncsa		
extensionU	Standard for the Exchang...	
hasExtensio	isda/154	
hasIsdaClas	ISDA_3D	ISDA_2D
hasIsdaId	http://kastner-wxp.ncsa...	
http://isda.ncsa		
name	C:\\Documents and Settin...	
originatingS	Autodesk Inventor 11	
preprocess	Autodesk Inventor 11	
timestamp	2007-08-15T07:31:08	
http://isda.ncsa		
hasFormatN		JPEG File Interchange Form
hasFormatV		1.02
hasMimeTyp		image/jpeg
hasPronomI		http://www.nationalarchiv
hasStatus		Positive (Specific Format)
http://www.w3		
22-rdf-synt	tag:tupeloproject.org,200...	tag:tupeloproject.org,2006

type filter text

- Rudder_Arms_02.jpg
- Rudder_Arms_dwg.jpg
- Rudder_Assembly.stp
- Rudder_Plate.jpg
- Rudder_Plate.stp
- Rudder_Plate_02.jpg
- Rudder_plate_dwg.jpg
- Rudders_w_hull.jpg
- Starboard_Rudder_Stock.stp
- TWR_aft_port.jpg
- TWR_aft_stbd.jpg
- Tiller_Weldment.stp
- Tiller_Weldments.jpg
- Tiller_weldment_dwg.jpg
- Upper_Rudder_Arm.stp
- lockfile.lck

Data Viewer

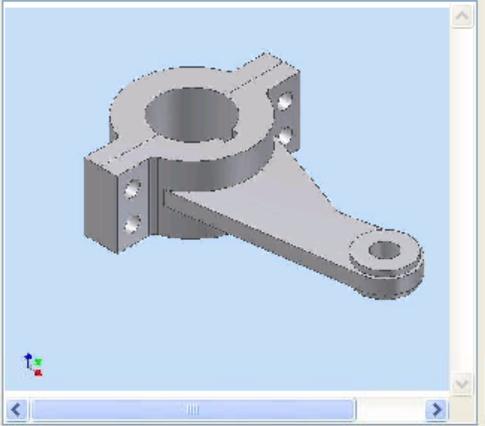


Table of Relationships Between Selected Files

NaraDemo

Exploratory Framework: Modes of Operations

- **Detection of discrepancies/anomalies in file descriptors**
 - OCR results
 - View 2D drawings and OCR results, and then edit OCR descriptors
 - 3D Model
 - View 3D model and content based extraction, and then edit descriptors
- **Comparison of pairs of files**
 - Pairs of 2D drawings
 - Pairs of 3D models
 - Pairs of (2D drawing, 3D model)
- **Establish file relationships**
 - Insert logical links to relate a pair of files

Detection of Anomalies in OCR Results

The screenshot displays the File2Learn application interface, which is used for analyzing and visualizing data from technical drawings. The interface is divided into several panes:

- Types:** A tree view on the left showing a hierarchy of files and folders, including '6200891-01.png' and 'isda/MMCBlock'. A red box highlights the '6200891-01.png' file.
- Context View:** A central pane showing RDF predicates and their associated resources. A red box highlights the URI 'http://isda.ncsa.uiuc.edu/NARA/tdrw/74'. Another red box highlights the predicate '22-rdf-syntax-ns#type' and its value 'http://isda.ncsa.uiuc.edu'. A third red box highlights the MMCMumber 'MMC DWG. NO. 291-01-111'.
- Data Viewer:** A pane at the bottom showing the OCR'd image of a technical drawing. A red box highlights the text 'NO. 7-8291 -01-111' extracted from the drawing. A purple box highlights a table in the drawing's title block.

Comparison of Files

Color encoding:

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

Predicate	Left Resource	Right Resource
hasFilePath	C:/OCR/6201019-02/6201019-02-tbc.png	C:/OCR/6201019-01/6201019-01-tb.png
tag	http://isda.ncsa.uiuc.edu/NARA/tdrw/97ec3247-eb1f-461a-...	http://isda.ncsa.uiuc.edu/NARA/tdrw/777fc5bd-fffe-4321-9928-be47afc78129/title
URI	http://isda.ncsa.uiuc.edu/2009/isda	tag://isda.ncsa.uiuc.edu,2009:file/4b2ab164-44a0-4f8d-b8d5-34cc301a89b4/Upper
LogicalLink	http://www.w3.org/1999/02-22-rdf-syntax-ns#type	http://isda.ncsa.uiuc.edu/NARA/tdrw/TitleBlock
FSCMNumber	53711	53711
drawingNumber	562-6201019	562-6201019
drawingScale	SCALE 1/2"=1'-0" & AS SHOWN	SCALE 1/2"=1'-0" & AS SHOWN
drawingSize	H	H
drawingTitle		DEPARTMENT OF THE NAVY □ NAVAL SEA SYSTEMS COMMAND □ WASHINGTON OC
fieldInfo		NAVSEA APPROVAL DATE
miniBox	http://isda.ncsa.uiuc.edu/NARA/tdrw/97ec3247-eb1f-461a-...	http://isda.ncsa.uiuc.edu/NARA/tdrw/777fc5bd-fffe-4321-9928-be47afc78129/title
miniBox	http://isda.ncsa.uiuc.edu/NARA/tdrw/97ec3247-eb1f-461a-...	http://isda.ncsa.uiuc.edu/NARA/tdrw/777fc5bd-fffe-4321-9928-be47afc78129/title
miniBox		http://isda.ncsa.uiuc.edu/NARA/tdrw/777fc5bd-fffe-4321-9928-be47afc78129/title
miniBox		http://isda.ncsa.uiuc.edu/NARA/tdrw/777fc5bd-fffe-4321-9928-be47afc78129/title
preparingActivity		120' TORPEDO WEAPONS RETRIEVER □ RUDDER & STEERING GEAR DETAILS
recordOfPreparation		CONTRACT NO N00024 85-C-2108 □ MARINETTE MARINE CORP MARINETTE WIS 5
sheetNumber	SHEET 2 OF 2	SHEET 1 OF 2

Establish File Relationships

File2Learn

File Window Help

Types

type filter text

- Rudder_Arms_dwg.jpg
- Rudder_Assembly.stp
- Rudder_Plate.jpg
- Rudder_Plate.stp
- Rudder_Plate_02.jpg
- Rudder_plate_dwg.jpg
- Rudders_w_hull.jpg
- Starboard_Rudder_Stock.stp
- TWR_aft_port.jpg
- TWR_aft_stbd.jpg
- Tiller_Weldment.stp
- Tiller_Weldments.jpg
- Upper_Rudder_Arm.stp
- lockfile.lck

Context View

<< Link Resources >>

Predicate	Left Resource	Right Resource
http://isda.ncsa.uiuc.edu/hasFileName	Upper_Rudder_Ar...	
http://isda.ncsa.uiuc.edu/hasFilePath	C:/NARADemo/RG...	C:/OCR/6201019-01/6201019-...
http://isda.ncsa.uiuc.edu/hasIdentQuality	Not identified	
http://purl.org/dc/element/creator	tag://isda.ncsa.uiu...	
tag/URI	tag://isda.ncsa.uiu...	http://isda.ncsa.uiuc.edu/NARA...
http://isda.ncsa.uiuc.edu/schema	AUTOMOTIVE_DES...	
http://isda.ncsa.uiuc.edu/implementationLevel	2;1	
http://isda.ncsa.uiuc.edu/extensionUsedBy	Standard for the E...	
http://isda.ncsa.uiuc.edu/hasExtension	isda/154	
http://isda.ncsa.uiuc.edu/hasIsdaClassification	ISDA_3D	
http://isda.ncsa.uiuc.edu/hasIsdaId	http://jkastner-wx...	
http://isda.ncsa.uiuc.edu/name	C:\Documents an...	
http://isda.ncsa.uiuc.edu/originatingSystem	Autodesk Inventor...	
http://isda.ncsa.uiuc.edu/preprocessorVersion	Autodesk Inventor...	
http://isda.ncsa.uiuc.edu/timestamp	2007-08-15T07:29...	
http://www.w3.org/1999/02-22-rdf-syntax-ns#type	tag:tupeloproject...	http://isda.ncsa.uiuc.edu/NARA...
http://isda.ncsa.uiuc.edu/FSCMNumber		53711
http://isda.ncsa.uiuc.edu/drawingNumber		562-6201019
http://isda.ncsa.uiuc.edu/drawingScale		SCALE 1/2"=1'-0" & AS SHOWN
http://isda.ncsa.uiuc.edu/drawingSize		H
http://isda.ncsa.uiuc.edu/drawingTitle		DEPARTMENT OF THE NAVY □...
http://isda.ncsa.uiuc.edu/fieldInfo		NAVSEA APPROVAL DATE
http://isda.ncsa.uiuc.edu/miniBox		http://isda.ncsa.uiuc.edu/NARA...
http://isda.ncsa.uiuc.edu/miniBox		http://isda.ncsa.uiuc.edu/NARA...

Data Viewer



Data Viewer

isda//TitleBlock

- 6200892-02.png
- 6200895-01.png
- 6200895-02.png
- 6201018-01.png
- 6201018-02.png
- 6201019-01.png
- isda//MMCBLOCK
- isda//ReferenceBlock
 - isda/reference/00
 - isda/reference/01
- isda//TitleBlock
 - isda/actionBox/01
 - isda/actionBox/02
 - isda/actionBox/03
 - isda/actionBox/04
- 6201019-02.png

NAVAL SEA SYSTEMS COMMAND
WASHINGTON D.C. 20380

100 TORPEDO WEAPONS RETRIEVER

RUDDER & STEERING GEAR DETAILS

DATE: 53711 562-6201019

SCALE 1/2"=1'-0" & AS SHOWN

NAVSEA APPROVAL DATE

NaraDemo

Establish File Relationships: Logical Link

The screenshot displays the File2Learn interface with several panes:

- Types (Left):** Lists various file types such as `Rudder_Arms_dwg.jpg`, `Rudder_Assembly.stp`, `Rudder_Plate.jpg`, `Rudder_Plate.stp`, `Rudder_Plate_02.jpg`, `Rudder_plate_dwg.jpg`, `Rudders_w_hull.jpg`, `Starboard_Rudder_Stock.stp`, `TWR_aft_port.jpg`, `TWR_aft_stbd.jpg`, `Tiller_Weldment.stp`, `Tiller_Weldments.jpg`, `Tiller_weldment_dwg.jpg`, `Upper_Rudder_Arm.stp`, and `lockfile.lck`.
- Context View (Center):** Shows a graph of relationships. A central node `tdrw:777fc5bd-fffe-4321-9928-be47afc78129/titleblock` is highlighted with a red box. It is connected to `C:/OCR/6201019-01/6201019-01-tb.png` via the property `droid:/IdFile/hasFilePath` and to `tdrw:TitleBlock` via `rdf:type`. Below the red box, a logical link is shown: `http://isda.ncsa.uluc.edu/2009/isda/LogicalLink`. This link connects to a file node `.../2009/.../file/4b2ab164-44a0-4f8d-b8d5-34cc301a89b4/Upper_Rudder_Arm.stp`, which is also connected to `t.org,2006:/2.0/files/File` via `droid:/IdFile/hasFilePath` and to the logical link via `rdf:type`.
- Types (Right):** Lists a hierarchy of types including `isda//TitleBlock`, `6200892-02.png`, `6200895-01.png`, `6200895-02.png`, `6201018-01.png`, `6201018-02.png`, `6201019-01.png`, `isda//MMCBlock`, `isda//ReferenceBlock`, `isda/reference/00`, `isda/reference/01`, `isda//TitleBlock`, `isda/actionBox/01`, `isda/actionBox/02`, `isda/actionBox/03`, `isda/actionBox/04`, and `6201019-02.png`.
- Data Viewer (Bottom Right):** Displays a technical drawing titled "RUDDER & STEERING GEAR DETAILS" from the "DEPARTMENT OF THE NAVY NAVAL SEA SYSTEMS COMMAND". It includes a table with the following data:

COMP	4243	DATE	53711	562-6201019
DATE	4/18/05	DATE	53711	562-6201019

A Comprehensive Comparison of Contemporary Documents

Support of Appraisals by Enabling Comparisons

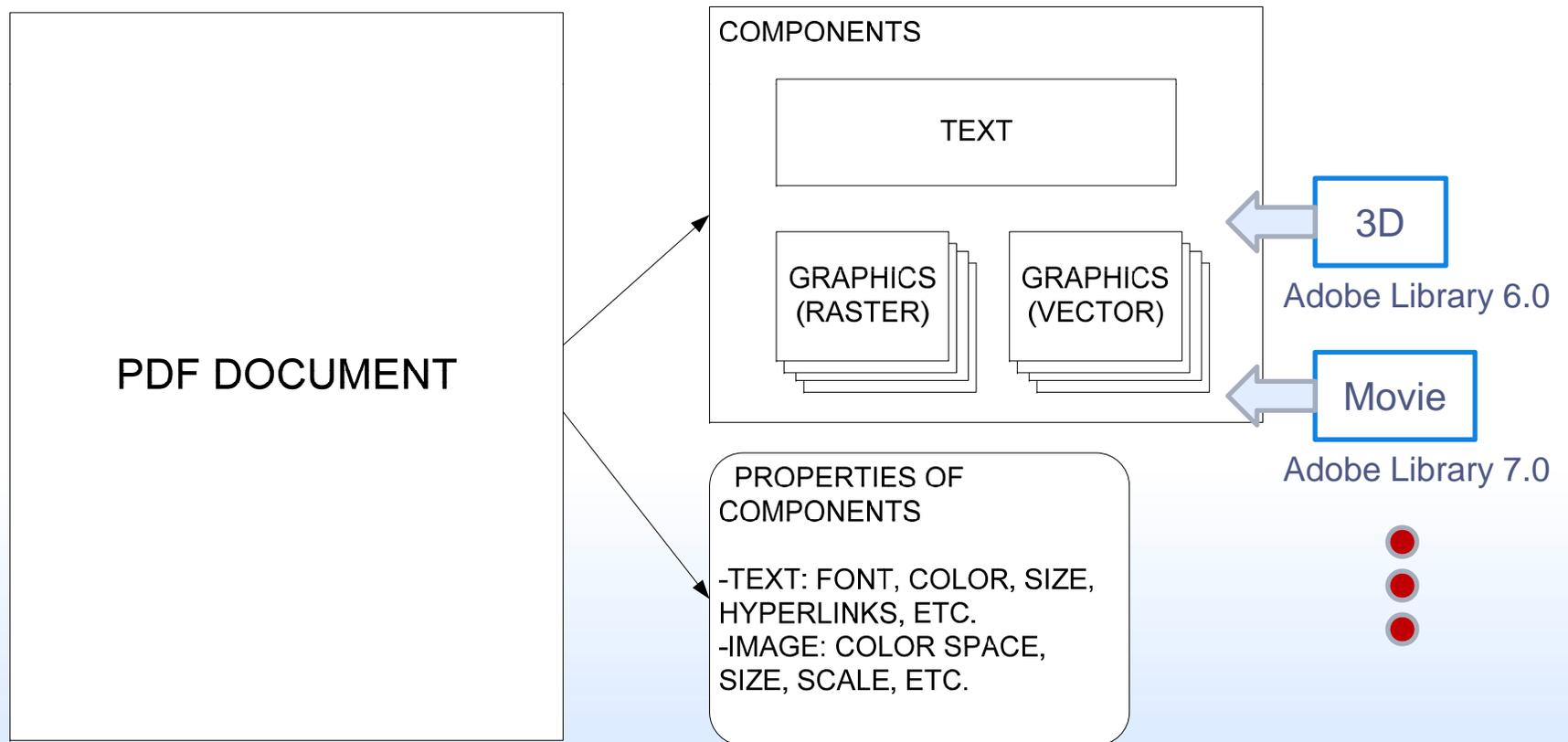
- **How to compare containers with heterogeneous information (text, images, vector graphics, animation, 3D, etc.)?**
 - Methodology
 - Metrics
 - Weighting factors for fusion
- **How to quantify similarities between the same type of information?**
 - Encodings and Representations
 - Metrics
 - Local versus global differences



One of the most important attribute is acc
attributes to be ~~two~~-twenty degrees.

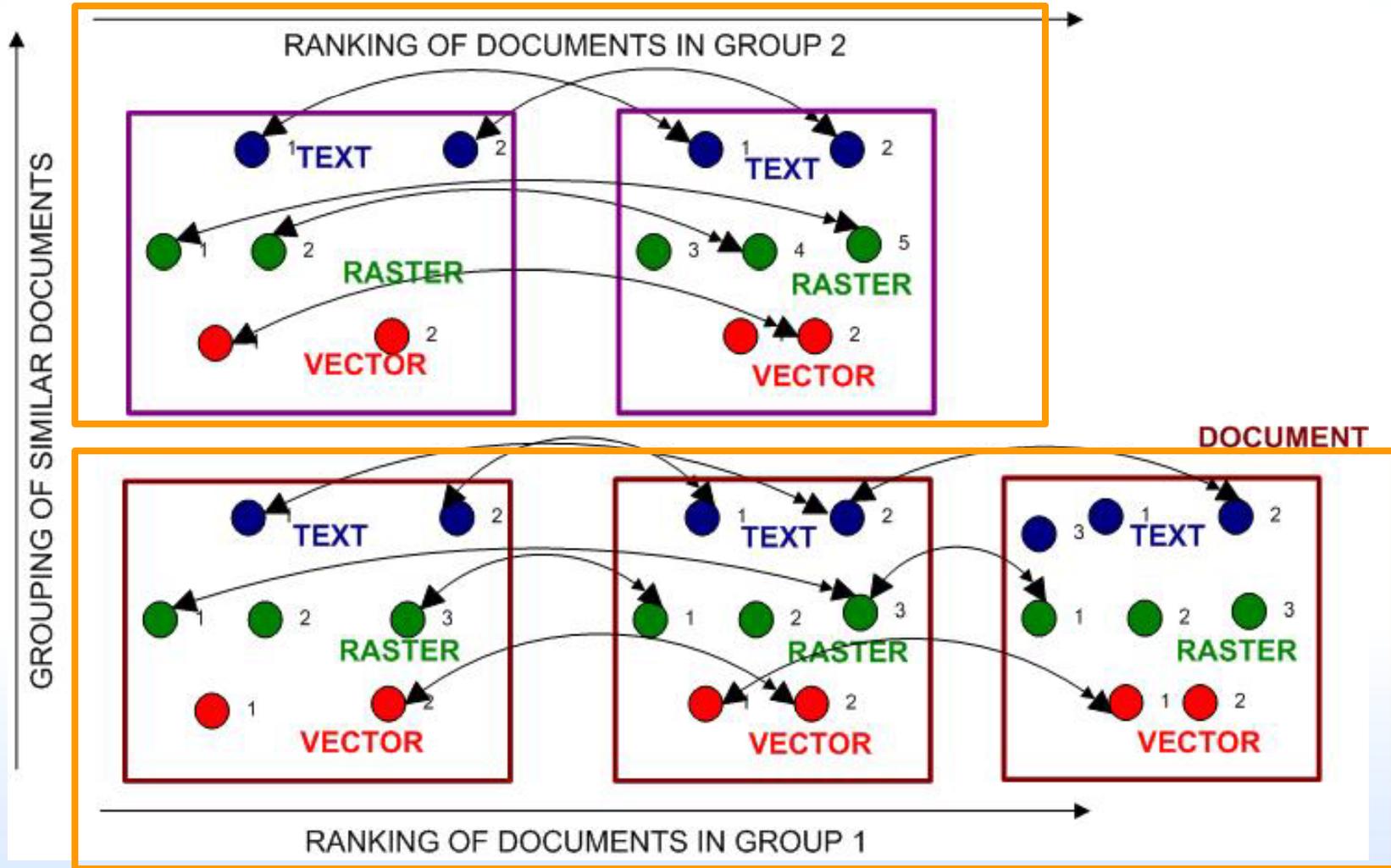
Example: Adobe Portable Document Format (PDF)

- **Why PDF? - PDF is just an example of a container**
 - Office environment (Adobe PDF, PS, MS Word, HTML, ...)
 - Satellite measurements (HDF, netCDF, ...)



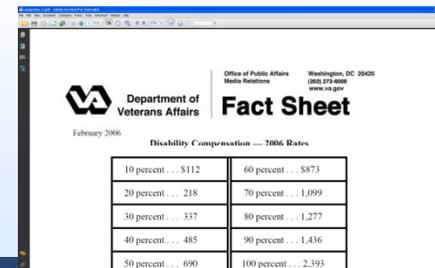
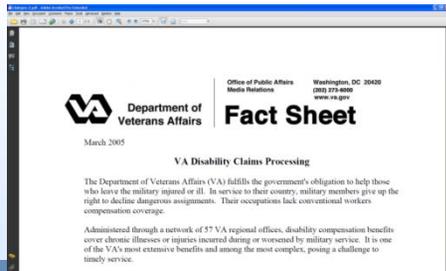
Comparisons

WANTED: GROUPING, RANKING, AND INTEGRITY VERIFICATION



Example: Compare Veterans Affairs Fact Sheets in PDF and MS Word file formats

- Test data: 108 files from RG 015 - Records of the Department of Veterans Affairs/Fact Sheets/www1.va.gov/opa/fact/docs.
 - These files are Veterans Affairs Fact Sheets and are stored in both PDF and MS Word file formats (54 MS word and 54 PDF files).
- **Which files have identical content?**
- **Demo: 6 files**
 - **amwars-2.pdf, amwars.pdf**
 - **claimpro-2.pdf, claimpro.pdf**
 - **combrates-2.pdf, combrates.pdf**

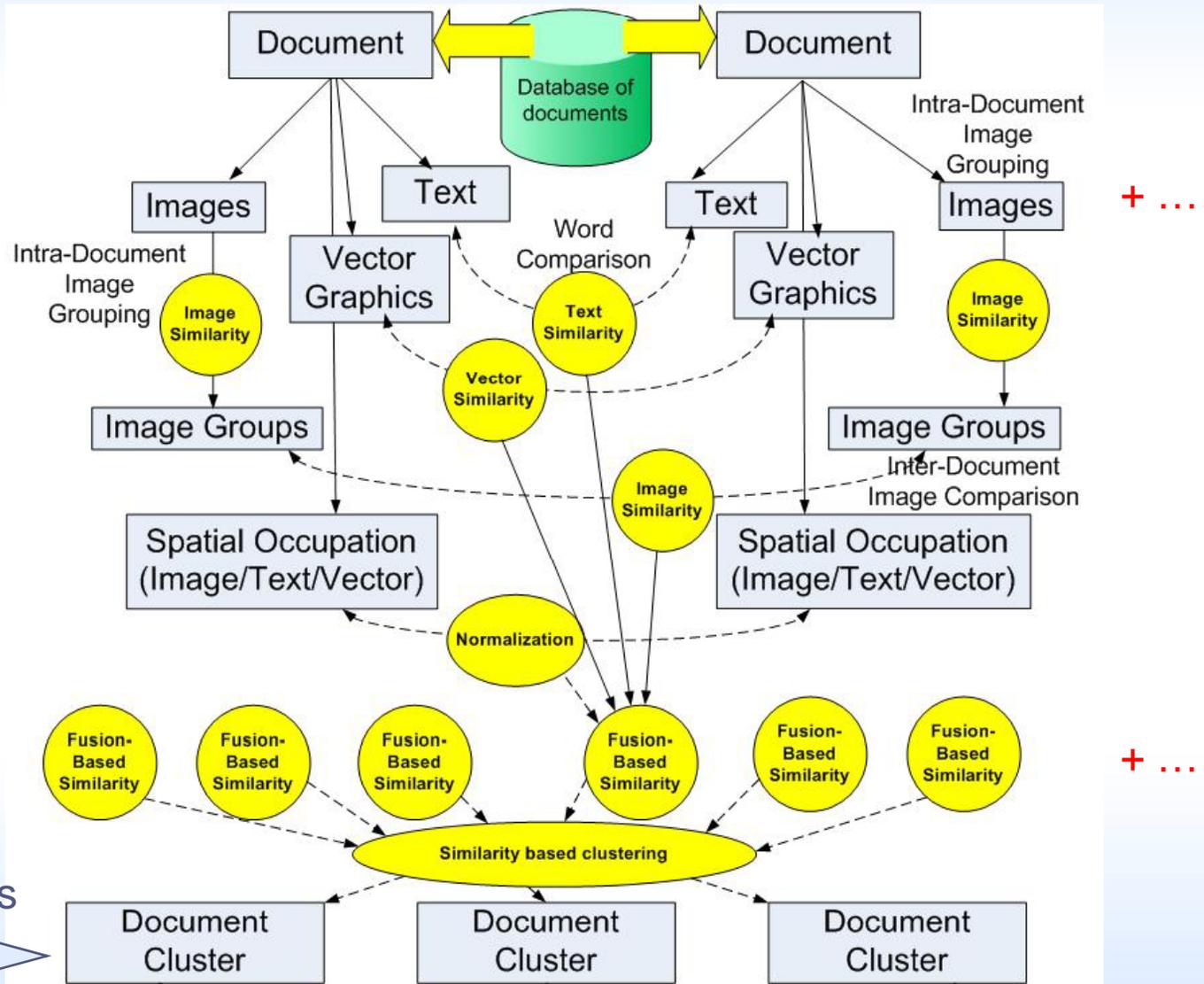


Methodology

Pair-wise comparison of the same digital objects

Comparison of multiple and heterogeneous digital objects

Relationship to Permanent Records



Exploration of Text Components

Department of Veterans Affairs
Office of Public Affairs
Washington, DC 20306
www.va.gov

February 2006

Fact Sheet

Disability Compensation — 2006 Rates

10 percent	\$112	60 percent	\$873
20 percent	218	70 percent	1,099
30 percent	337	80 percent	1,277
40 percent	485	90 percent	1,456
50 percent	690	100 percent	2,393

(Rates in 2006)

Additional Payments for Dependents:
Veterans whose service-connected disabilities are rated at 30 percent or more are entitled to additional allowances for dependents. Dependents upon the disability rating of the veteran, monthly allowances for a spouse range from \$60 to \$153 and for a dependent child, \$2 to \$91. Additional amounts are provided for each additional child and there is a higher rate for children as child after age 18.

Other Major Factors Affecting Payment Levels:
Adjustments to rates are based on a number of factors in addition to dependents. Among factors that can have a significant effect on amounts are:

- * Veterans with service-connected disabilities may receive compensation at a basic rate as high as \$6,875 per month. Veterans' special monthly compensation rates apply when a veteran experiences loss or loss of use or severe loss, loss or more of the senses of sight, hearing or speech, or a representative loss of a reproductive organ or the loss, or loss of benefit from a health veteran.
- * Adjustments may be made for veterans receiving aids, such as including individuals who need assistance with travel, feeding or certain other activities of daily living. This adjustment is referred to as "aid and attendance."

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	Num Images
1	lamwars-2.pdf		1
2	lamwars.pdf		1
3	claimpro-2.pdf	57KB Fri May 22 08:49:59 CDT 2009	3
4	claimpro.pdf	57KB Fri May 22 08:49:59 CDT 2009	3
5	combrates-2.pdf	50KB Fri May 22 08:49:59 CDT 2009	3
6	combrates.pdf	50KB Fri May 22 08:49:59 CDT 2009	3

LOADED FILES

TEXT

Occurrence of words

Occurrence of numbers

"Ignore" words

Page List

Page No	Number of Words
All	393
1	148
2	202
3	162

All Frequency

No	Word	Frequency
0	need	1
1	reopened	1
2	total	1
3	discussion	1
4	capacity	1
5	starting	1
6	annually	1
7	100	2
8	2393	1
9	signed	1
10	90	1
11	evaluated	1
12	point	1
13	degree	1
14	practicably	1
15	Fall	1
16	calculates	1
17	http/www....	1
18	based	2
19	but	2
20	workers	1
21	together	1
22	though	1
23	5845	1
24	were	2
25	hearing	1
26	injuries	1
27	regulation	1
28	other	3
29	have	4
30	specific	1
31	one	4
32	state	1
33	beneficial	1
34	91	1
35	veteran's	1
36	with	6

Minimum frequency: 0 Set

Text Frequency

No	Word	Frequency
0	need	1
1	total	1
2	discussion	1
3	benefit	1
4	maintain	1
5	scar	1
6	higher	3
7	received	2
8	its	1
9	accounting	1
10	over	2
11	veteran	11
12	levels	1
13	spouse	1
14	laws	1
15	year	8
16	bureau	1
17	knee	1
18	these	1
19	turn	1
20	#	3
21	recommends	1
22	arthritis	1
23	under	1
24	number	3
25	residual	1
26	available	1
27	meet	65
28	29	1
29	30	4
30	breast	1
31	musculoskel...	1
32	price	2
33	individually	2
34	February	1
35	levels	1

Minimum frequency: 0 Set

Integer Frequency

No	Word	Frequency
0	1436	1
1	12006	4
2	22005	1
3	18	1
4	4485	1
5	5218	1
6	6337	1
7	1099	1
8	820	1
9	14000	1
10	1040	3
11	1180	2
12	1260	2
13	13100	2
14	142393	1
15	1590	1
16	161	2
17	1710	4
18	180	2
19	1965000	1
20	2030	2
21	21775854	1
22	221277	1
23	2370	2
24	24788000	1
25	25690	1
26	2650	1

Minimum frequency: 0 Set

Float Frequency

No	Word	Frequency
0	2.6	2
1	12.3	1

Minimum frequency: 0 Set

73|where
74|which
75|will
76|with
77|would

New Remove Load Save

Compare List

No	Word	Info
5	are	3:3:3:3:1:1:1:1
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	2:2:10:10:17:17
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	16:16:20:20:14:14
16	is	4:4:6:6:16:16
17	last	14:14:1:1:1:1
18	new	2:2:1:1:3:3
19	not	4:4:3:3:6:6
20	number	3:3:8:8:3:3
21	of	15:15:42:42:32:32
22	on	6:6:7:7:9:9
23	other	14:14:3:3:3:3
24	public	1:1:1:1:1:1
25	requiring	1:1:1:1:2:2
26	rolls	1:1:1:1:2:2
27	the	7:7:54:54:46:46
28	to	3:3:34:34:15:15
29	total	13:13:1:1:1:1
30	upon	1:1:1:1:1:1
31	va	4:4:19:19:6:6
32	veterans	14:14:7:7:16:16
33	which	1:1:2:2:2:2
34	who	2:2:2:2:1:1

Save

Exploration of Vector Graphics Components

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	1anwars-2.pdf		88KB/Fri May 22 08:49:59 CDT 2009	2	1
2	2anwars.pdf			1	1
3	3claimpro-2.pdf			1	1
4	4claimpro.pdf			1	1
5	5combrates-2.pdf			1	1
6	6combrates.pdf			1	1

LOADED FILES

Preview

Word Frequency Image Color Frequency Vector Graphics Frequency

Page List

Page Number	Number of Paths
All	1
0	1
1	1
2	1

Graphics List

Graphics Type	count
1,0,0,0,2,0,0,0,1,56	

Preview

Zoom In Zoom Out

Occurrence of v/h lines

Compare List

No	Word	Info
5	are	3:3:3:3:1:1:1:1
6	as	2:2:19:9:12:12
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	2:2:10:10:17:17
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	16:16:20:20:14:14
16	is	4:4:6:6:16:16
17	last	14:14:1:1:1:1
18	new	2:2:1:1:3:3
19	not	4:4:3:3:6:6
20	number	3:3:8:8:3:3
21	of	15:15:42:42:32:32
22	on	6:6:7:7:9:9
23	other	14:14:3:3:3:3
24	public	1:1:1:1:1:1
25	requiring	1:1:1:1:2:2
26	rolls	1:1:1:1:2:2
27	the	7:7:54:54:46:46
28	to	3:3:34:34:15:15
29	total	13:13:1:1:1:1
30	upon	1:1:1:1:1:1
31	va	4:4:19:19:6:6
32	veterans	14:14:7:7:16:16
33	which	1:1:2:2:2:2
34	who	2:2:2:2:1:1

Save

Office of Public Affairs Washington, DC 20304
www.va.gov

Department of Veterans Affairs **Fact Sheet**
February 2006

Disability Compensation — 2006 Rates

30 percent	\$112	60 percent	\$871
30 percent	218	70 percent	1,999
30 percent	337	80 percent	1,277
40 percent	485	90 percent	1,456
50 percent	690	100 percent	2,392

Effective 1/1/07

Additional Payments for Dependents

Veterans whose service-connected disabilities are rated at 30 percent or more are entitled to additional allowances for dependents. Depending upon the disability rating of the veteran, monthly allowances for a spouse range from \$60 to \$120 and for a dependent child, \$2 to \$91. Additional amounts are provided for each additional child and there is a higher scale for children as shown after age 18.

Other Major Factors Affecting Payment Levels

Adjustments to rates are based on a number of factors in addition to dependents. Among factors that can have a significant effect on increases are:

- Veterans with unremediated disabilities may receive compensation at a basic rate as high as 50.8% per month.
- Some special monthly compensation rates apply when a veteran experiences loss or loss of use of one or more limbs, loses use or care of the senses of sight, hearing or speech, or experiences loss of a reproductive organ or its use, or loss of breast tissue by a female veteran.
- Adjustments may be made for veterans requiring aids, such as bedridden individuals who need assistance with eating, bathing or other activities of daily living. This adjustment is referred to as "aid and attendance."

VECTORS

Comprehensive Pair-Wise Comparison of Documents

Grouping and Visualization Control

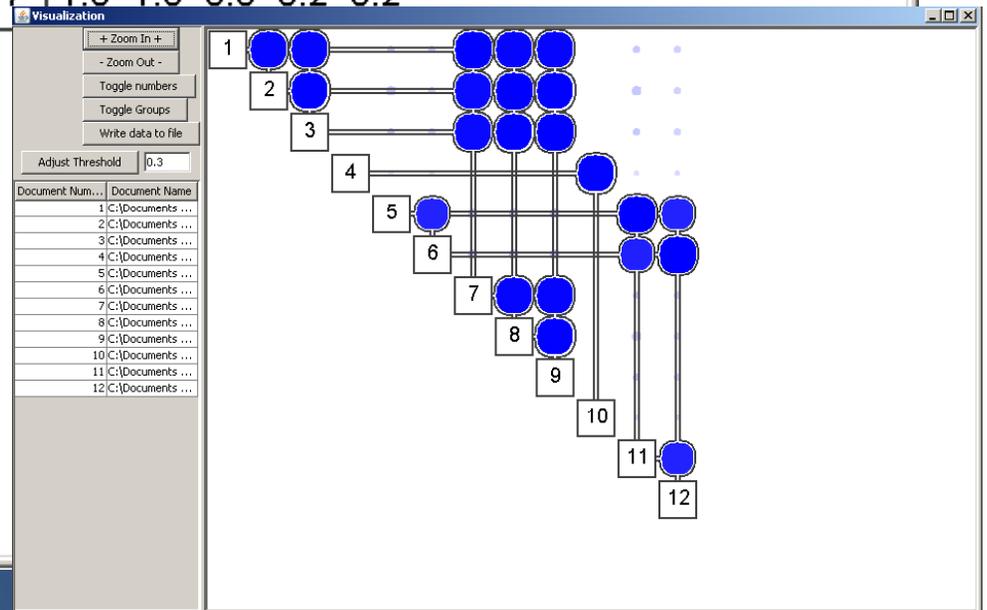
Toggle numbers
Toggle Groups
Write data to file
Adjust Threshold 0.3

Document Num...	Document Name
1	C:\Documents ...
2	C:\Documents ...
3	C:\Documents ...
4	C:\Documents ...
5	C:\Documents ...
6	C:\Documents ...
7	C:\Documents ...
8	C:\Documents ...
9	C:\Documents ...
10	C:\Documents ...
11	C:\Documents ...

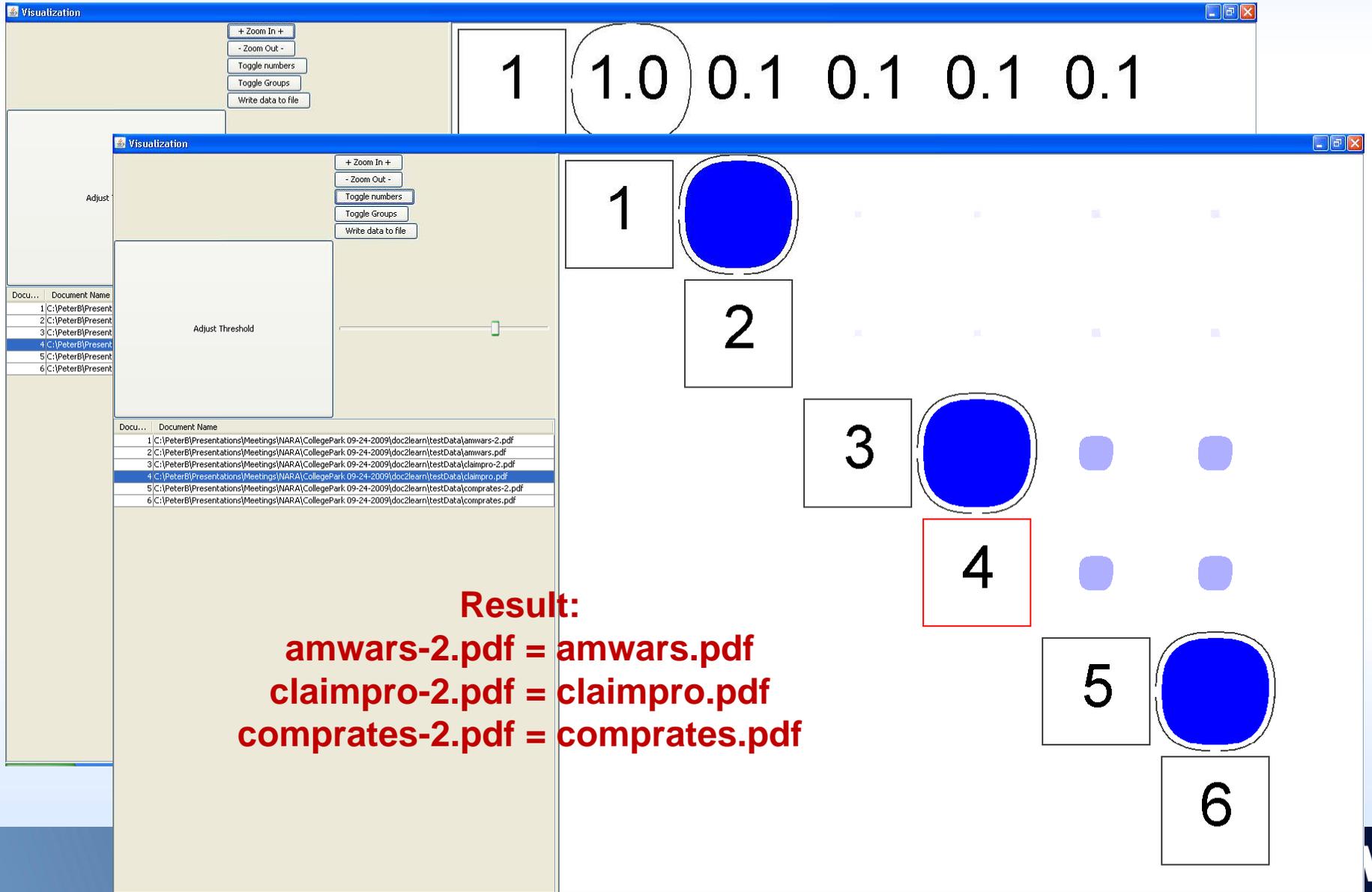
Document ID

2	1.0	0.0	0.2	0.2	1.0	1.0	1.0	0.0	0.2	0.2
3	0.0	0.2	0.2	1.0	1.0	1.0	0.0	0.2	0.2	0.2
4	0.1	0.2	0.0	0.0	0.0	1.0	0.1	0.2	0.2	0.2
5	0.9	0.3	0.3	0.3	0.1	1.0	0.9	0.2	0.2	0.2
6	0.2	0.2	0.2	0.2	0.9	1.0	0.2	0.2	0.2	0.2
7	1.0	1.0	0.0	0.2	0.2	0.2	1.0	0.2	0.2	0.2

Similarity Values

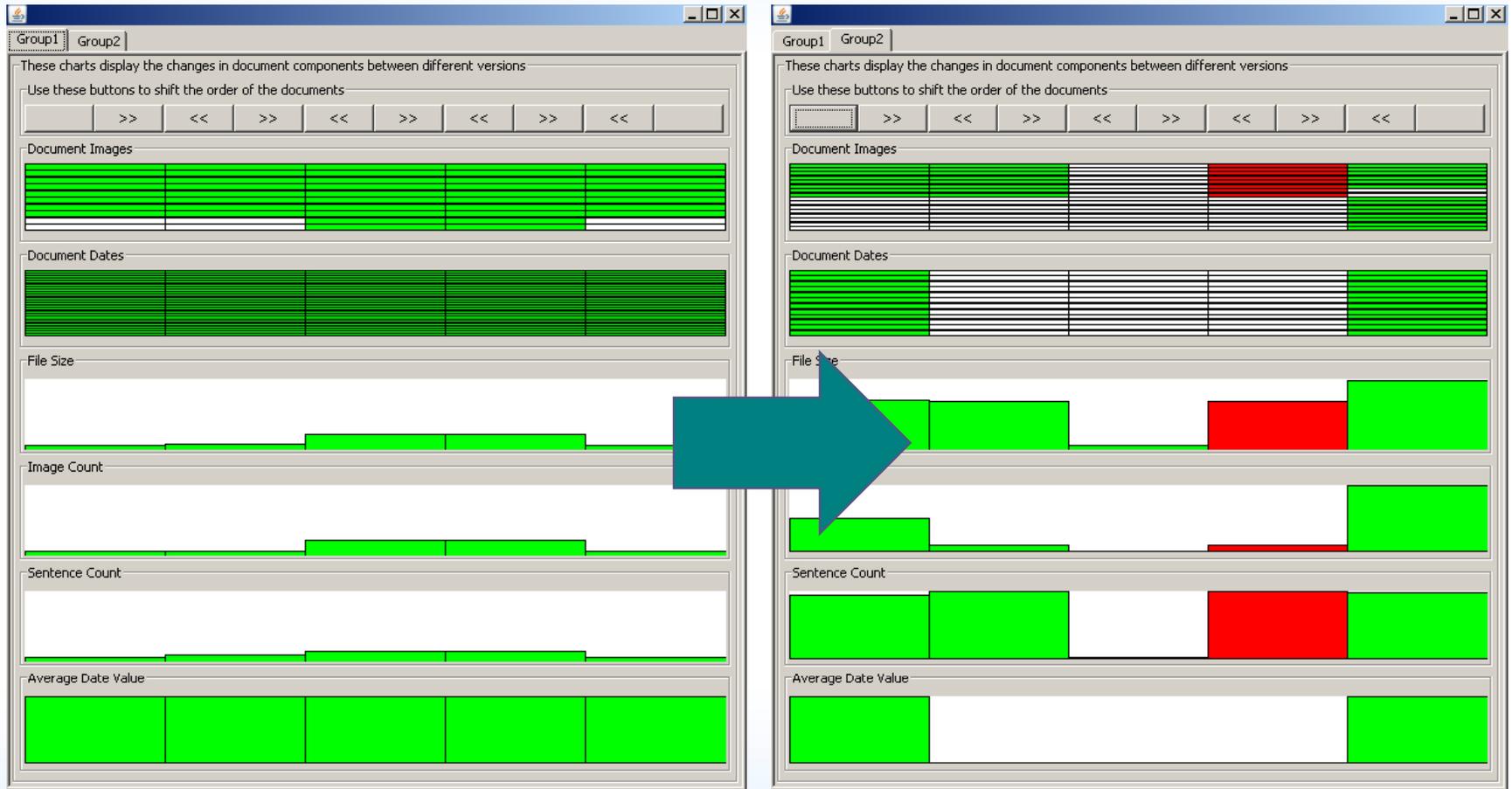


Visual Comparison for 6 Test Files



Work in progress: Group and Validate Documents

Attributes of documents



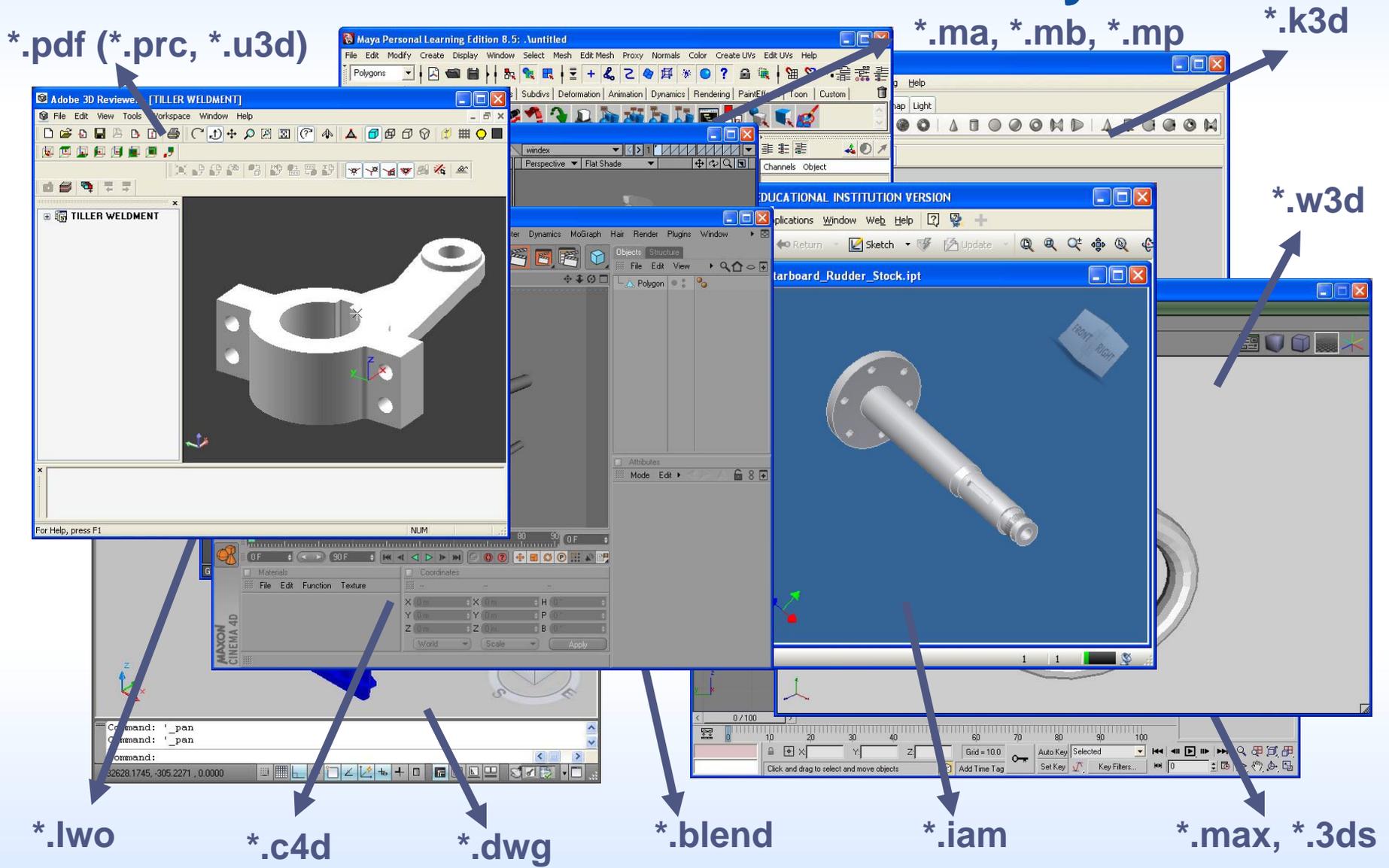
Order of documents

Automated File Format Conversions and Conversion Quality Assessment

Conversions of Electronic Records

- **Conversions of electronic records are needed** because
 - Visual exploration depends on various software packages
 - Many formats are retired (deprecated) over time
- **How to measure the degree of information preservation when files are converted from format A to format B?**
 - During conversions, information could be lost, added or modified
 - What is the importance of each byte, object, etc. ?
- **How to design a test bed for analyzing the quality of conversion and visualization software?**

Illustration of 3D File Format Reality



Our Survey about 3D Content

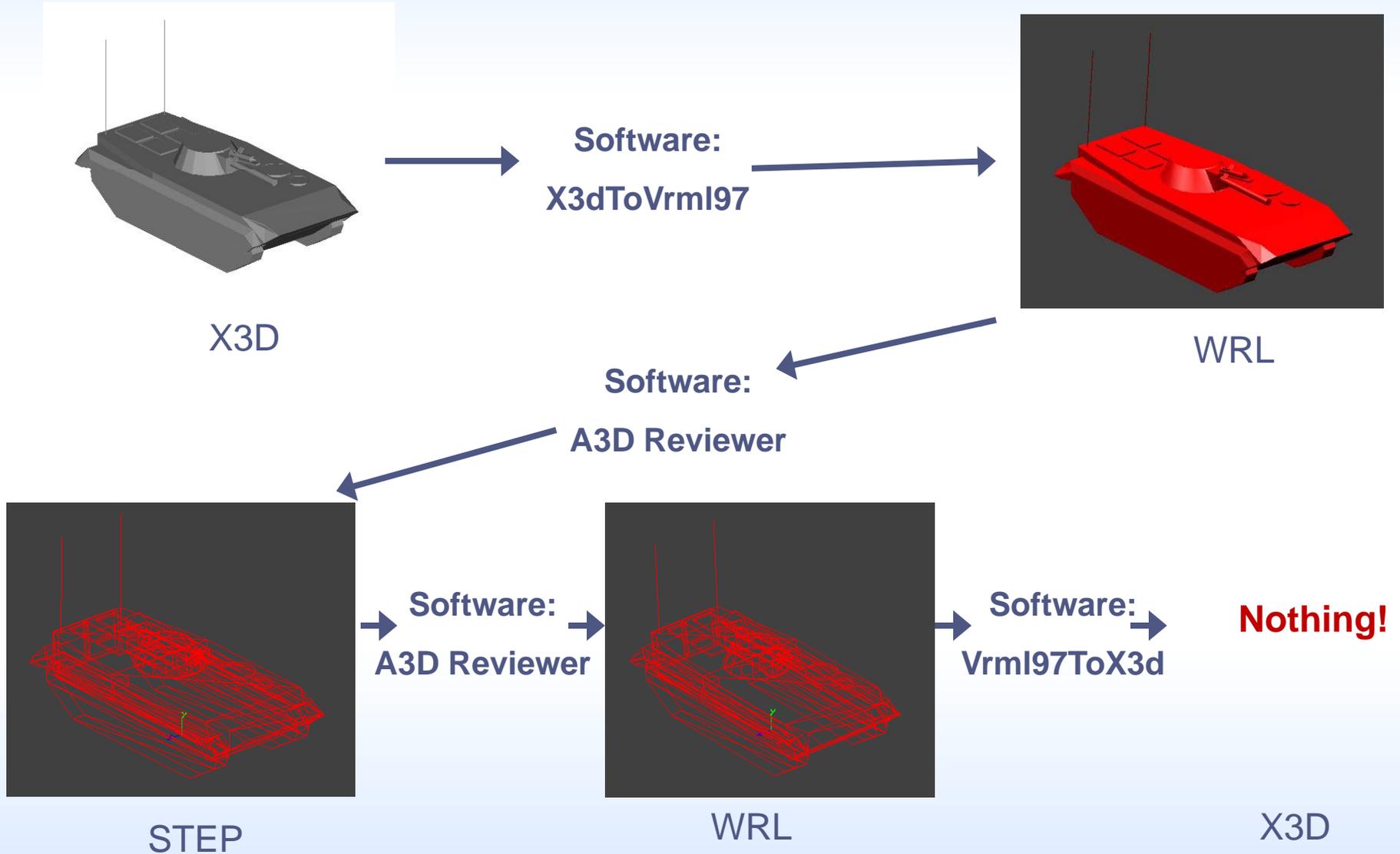
- **Q: How Many 3D File Formats Exist?**
- A: We have found more than 140 3D file formats. Many are proprietary file formats. Many are extremely complex (1,200 and more pages of specifications).
- **Q: How Many Software Packages Support 3D File Format Import, Export and Display?**
- A: We have documented about 16 software packages. There are many more. Most of them are proprietary/closed source code. Many contain incomplete support of file specifications.

Examples of 3D Formats and Stored Content

Format	Geometry				Appearance				Scene				Animation
	Faceted	Parametric	CSG	B-Rep	Color	Material	Texture	Bump	Lights	Views	Trans.	Groups	
3ds	✓	✓			✓	✓	✓	✓	✓	✓	✓		
igs	✓	✓	✓	✓	✓						✓	✓	
lwo	✓	✓			✓	✓	✓	✓					
obj	✓	✓			✓	✓	✓	✓				✓	
ply	✓				✓	✓	✓	✓					
stp	✓	✓	✓	✓	✓							✓	
wrl	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓
u3d	✓				✓		✓	✓	✓	✓	✓	✓	✓
x3d	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓

- Some content may be more important than others
 - The relative importance is situation dependent

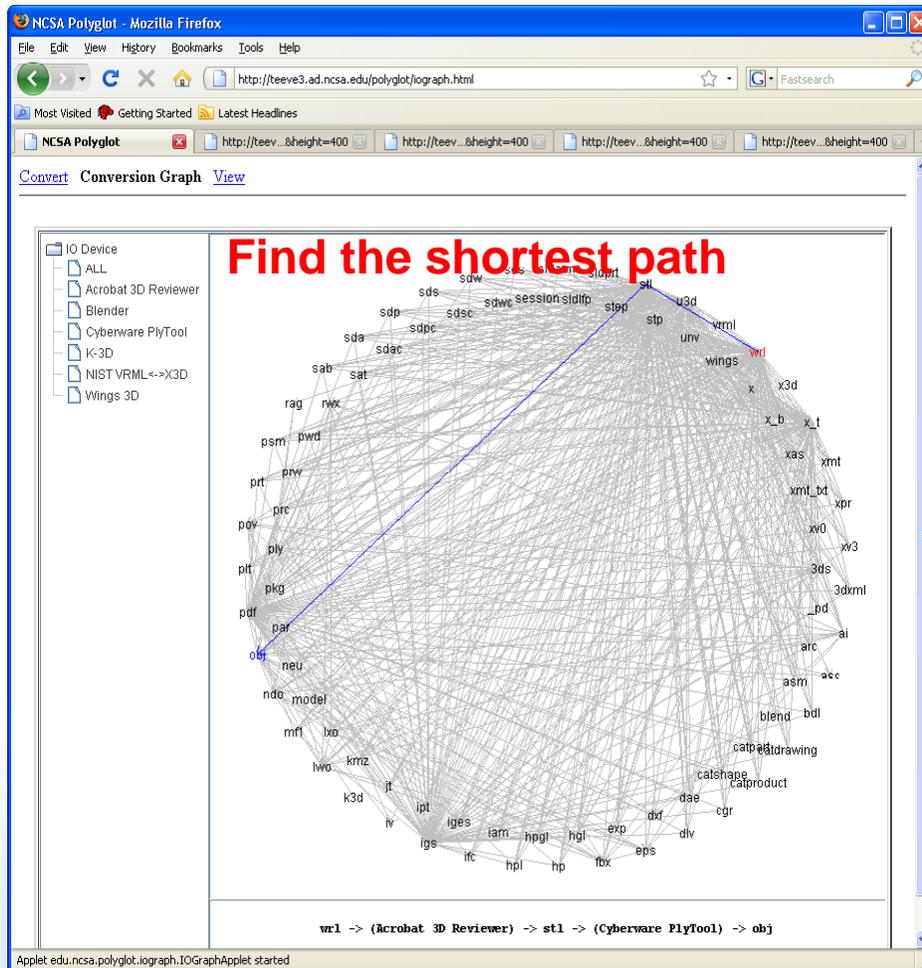
Example: Conversion of X3D to STEP to X3D



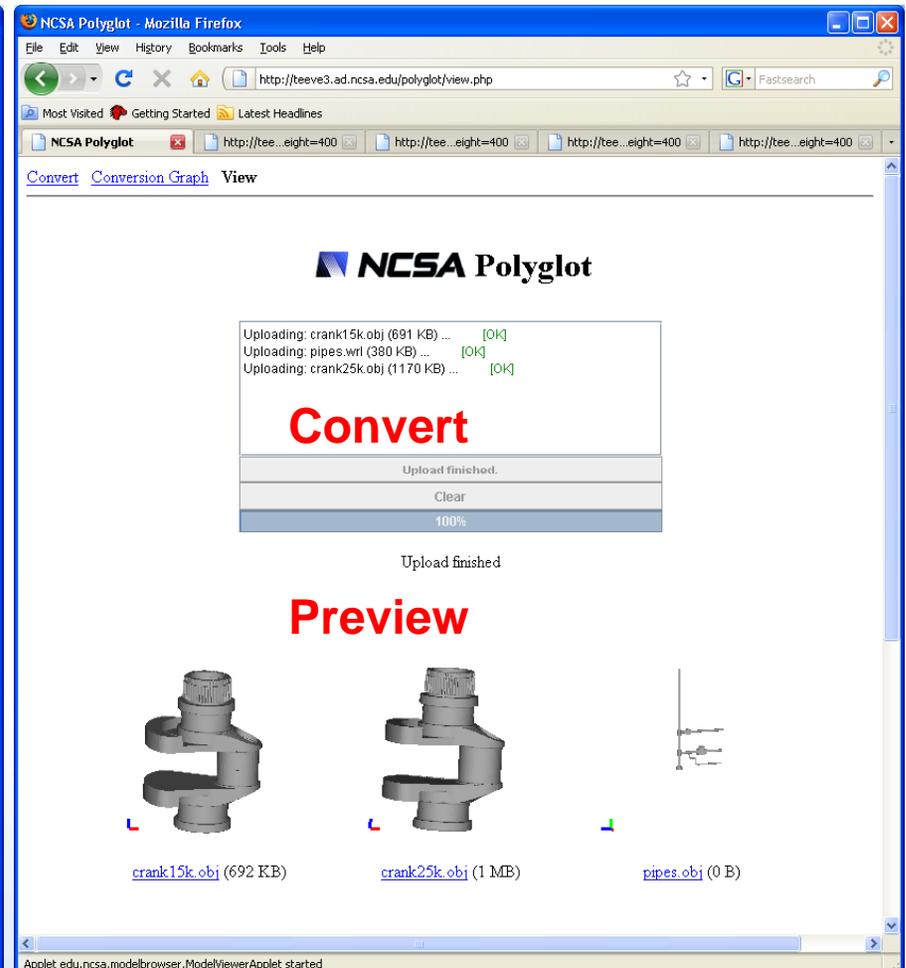
Towards a Universal Converter

- Use what is available in 3rd party software to perform conversions
 - Document **what formats can be opened/imported** by each application
 - Document **what formats can be saved/exported** by each application
 - **Automate the use of each application** and combine their abilities to perform conversions over larger set of formats

Automation of 3D File Format Mapping



The screenshot shows the NCSA Polyglot web application in Mozilla Firefox. The browser address bar displays `http://teeve3.ad.ncsa.edu/polyglot/iograph.html`. The page title is "NCSA Polyglot" and the URL is `http://teeve3.ad.ncsa.edu/polyglot/iograph.html`. The interface includes a navigation menu with "Convert", "Conversion Graph", and "View". A sidebar on the left lists supported file formats: IO Device, ALL, Acrobat 3D Reviewer, Blender, Cyberware PlyTool, K-3D, NIST VRML-<->X3D, and Wings 3D. The main area displays a complex network graph with nodes representing file formats and edges representing conversion paths. A red text overlay reads "Find the shortest path". A blue path is highlighted from the 'obj' node to the 'stl' node. At the bottom, a path is shown: `vrml -> (Acrobat 3D Reviewer) -> stl -> (Cyberware PlyTool) -> obj`. The status bar at the bottom indicates "Applet edu.ncsa.polyglot.iograph.IOGraphApplet started".



The screenshot shows the NCSA Polyglot web application in Mozilla Firefox. The browser address bar displays `http://teeve3.ad.ncsa.edu/polyglot/view.php`. The page title is "NCSA Polyglot" and the URL is `http://teeve3.ad.ncsa.edu/polyglot/view.php`. The interface includes a navigation menu with "Convert", "Conversion Graph", and "View". The main area displays the NCSA Polyglot logo and a progress bar for uploading files. The progress bar shows "Upload finished." and "100%". Below the progress bar, the text "Upload finished" is displayed. The "Convert" button is highlighted in red. Below the "Convert" button, the text "Preview" is displayed in red. Three 3D models are shown: a crankshaft, a crankshaft, and a pipe. The models are labeled with their respective file names and sizes: `crank15k.obj (692 KB)`, `crank25k.obj (1 MB)`, and `pipes.obj (0 B)`. The status bar at the bottom indicates "Applet edu.ncsa.modelbrowser.ModelViewerApplet started".

Automation of 3D File Format Conversion

- The I/O-Graph stores the information needed to convert between the formats represented in the graph.
- **In order to perform the conversion we must execute the conversion path found.**
 - Many high end graphics programs are found on the windows platform
 - Those on other platforms, such as Linux, tend to have windows ports
 - Some are command line driven (usually small converter applications).
 - Many have only GUI interfaces
 - AutoHotKey: a scripting language for the Windows GUI.

Methodology

Cloud Computing

COMPUTATIONAL
SCALABILITY

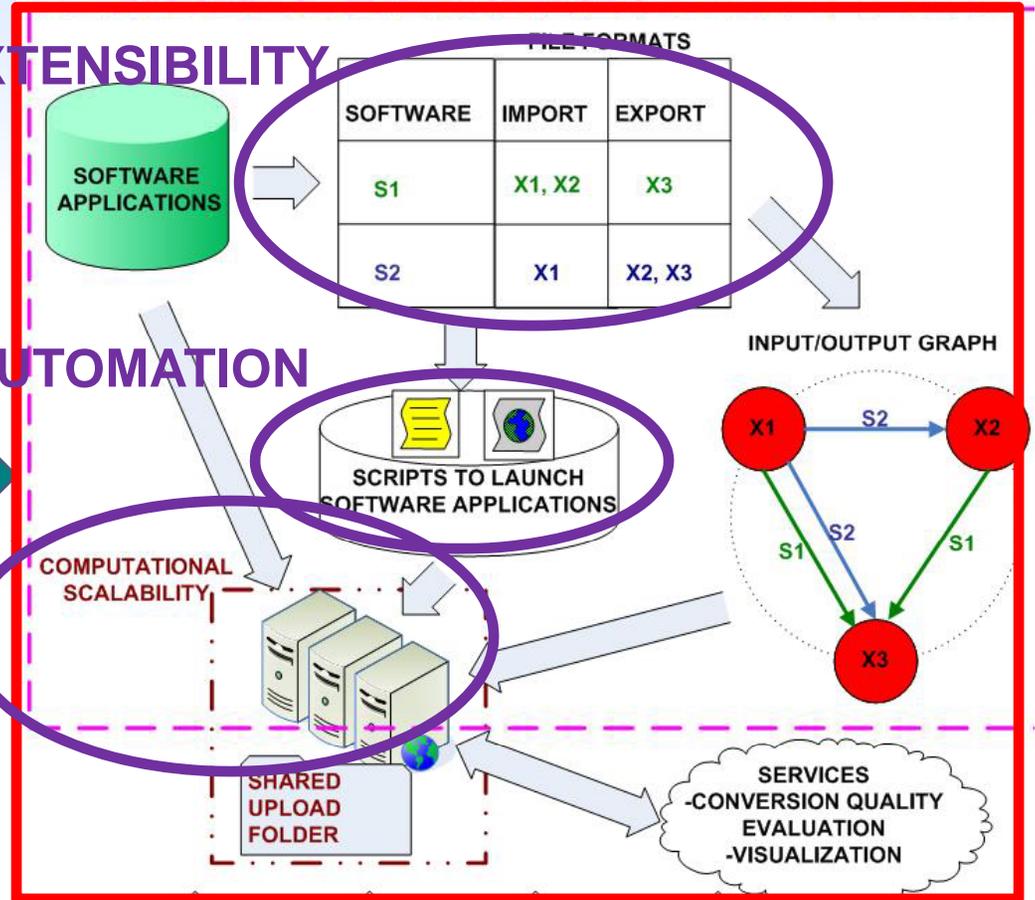
Services to Archivists

EXTENSIBILITY

AUTOMATION

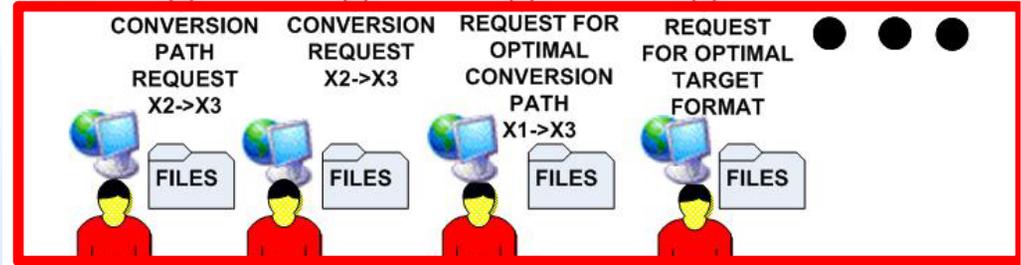
COMPUTATIONAL
SCALABILITY

EXTENSIBILITY OF FILE FORMAT CONVERSIONS



SERVER

CLIENTS



NCSA Polyglot – Conversion Services

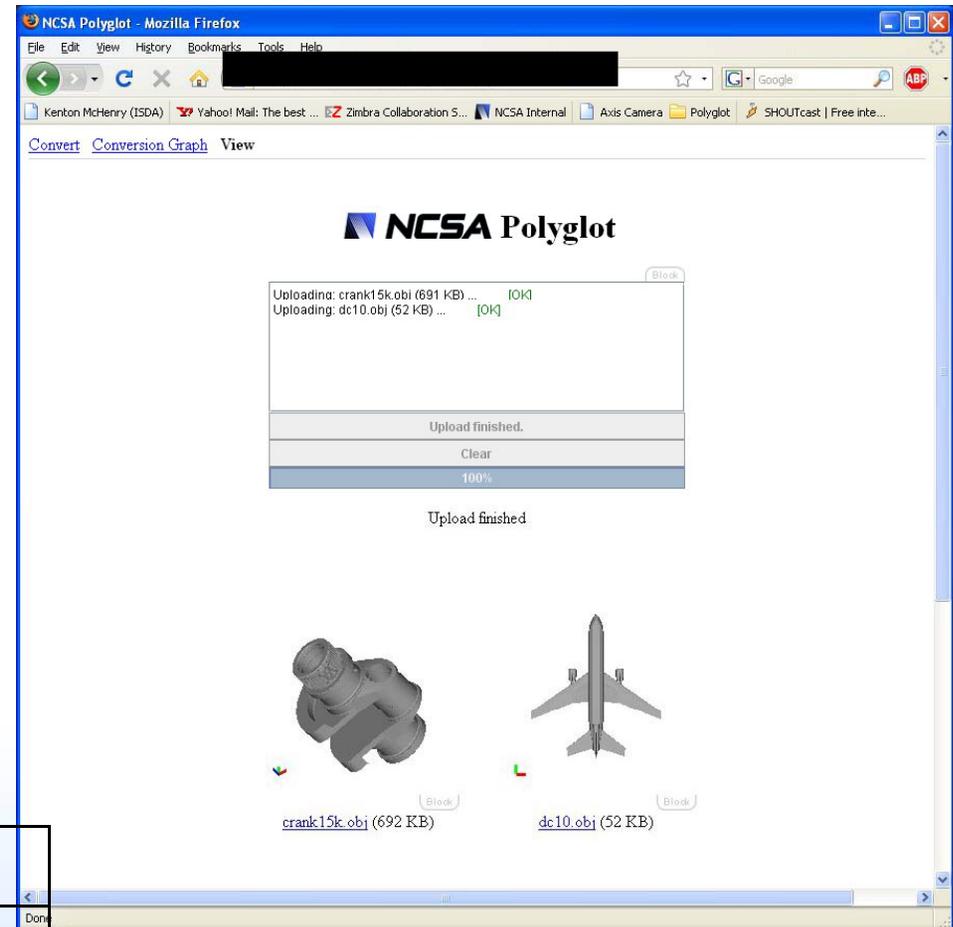
- Web interface: user can drag and drop files into upload area for conversion

- Java interface:

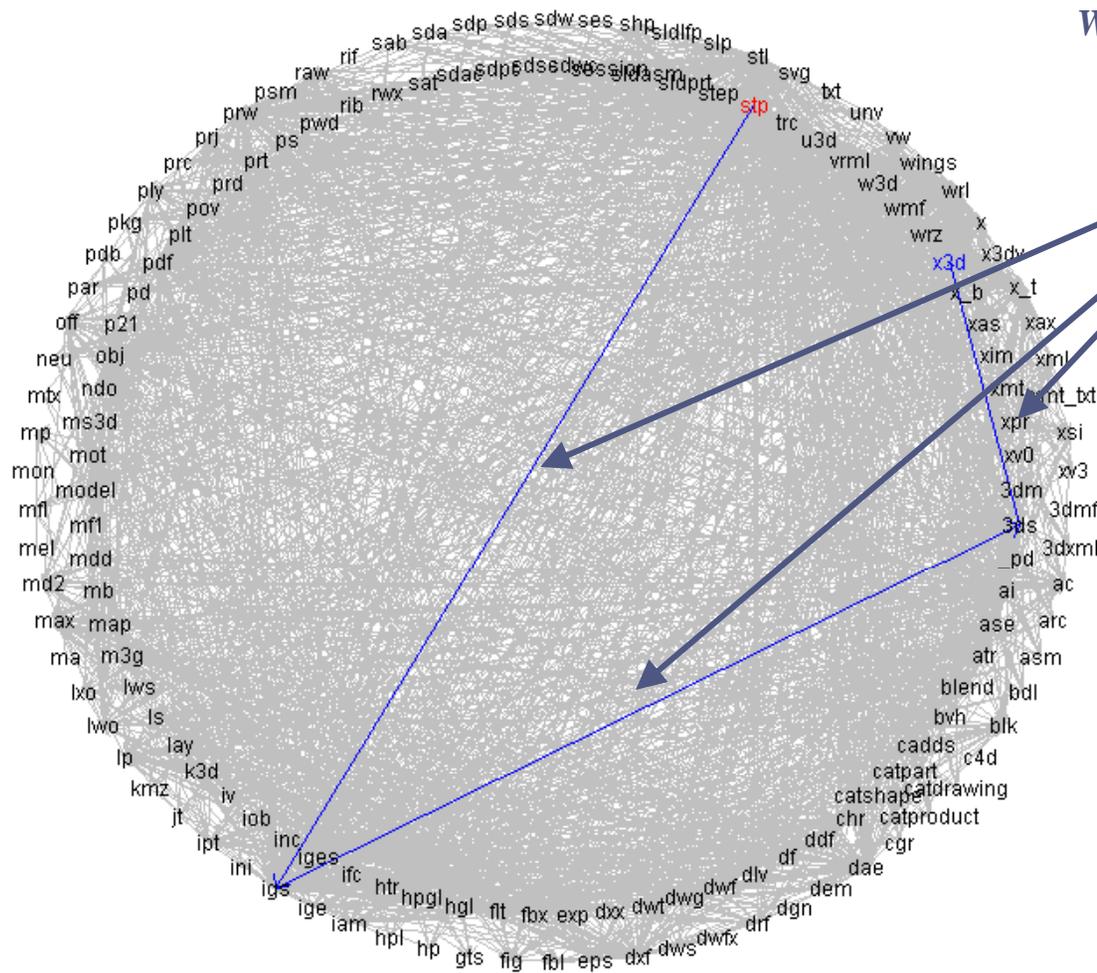
```
PolyglotRequest pgr;  
pgr = new PolyglotRequest("http://???", "obj");  
pgr.convertFile("file.wrl", ".");
```

- Scalability Test

Number of PCs	One PC	Two PCs
Processing Time	33 minutes 6 seconds	16 minutes 40 seconds



NCSA Polyplot – Data Loss Measurement Services

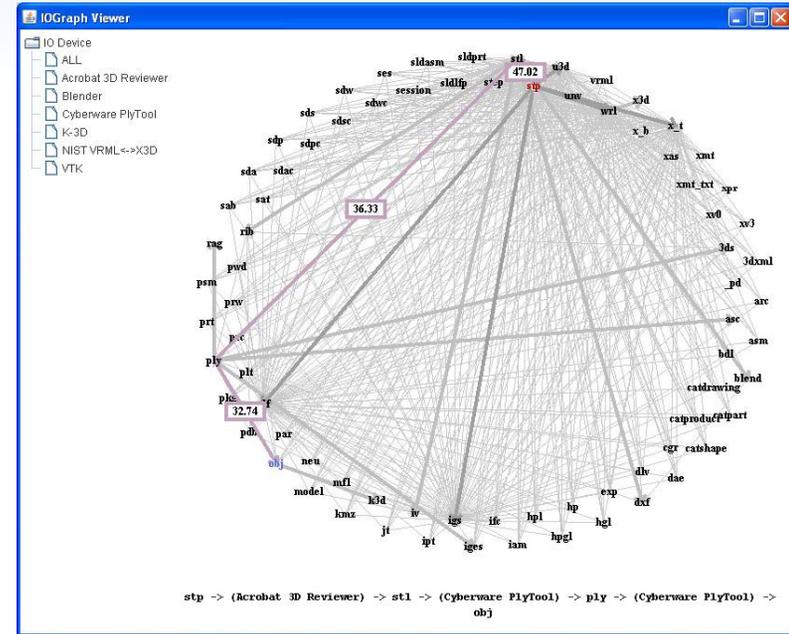


We would like to assign a value to each conversion edge ...

stp -> (Adobe 3D Reviewer) -> igs -> (3ds Max) -> 3ds -> (Blender) -> x3d

Geometry Based Content Retention

- Several metrics
- Data driven assignment
- **Example results**



Metric\Result	Single Optimal Conversion				'Best' File Format	
	Software	From	To	Information Retention	Format	Information Retention
Light Fields	Adobe 3D Reviewer	.pdf	.stp	61.67	.stp	40.73
Spin Images	Adobe 3D Reviewer	.obj	.pdf	59.07	.stl	34.89

Summary

- Technologies for appraisal of electronic records should assist archivists
 - They are designed to support decisions and data explorations by automating appraisal tasks
 - The software for doc2learn and Polyglot is available for downloading at <http://isda.ncsa.uiuc.edu/download/>
 - File2learn software – the work is still in progress
 - Feedback is very welcome
-
- Questions: Peter Bajcsy – pbajcsy@ncsa.uiuc.edu

Demo exercise

- Step 1: Check the path exists between wrl and pdf
- Step 2: drag and drop heart.wrl; select target to be pdf, click upload
- Step 3: download to desktop and open in Adobe PDF Viewer

