Preservation of Engineering Artifacts

Sponsored in part by NSF IIS-0456001

William C. Regli
Department of Computer Science
College of Engineering
Drexel University
Philadelphia, PA

What is Archiving?

- *Digital Preservation:* the mitigation of the deleterious effects of technology obsolescence, media degradation, and fading human memory.

- Challenge: how to maintain interpretability of “born digital” data for 10-20… 50+ year lifespans?

Definition from Gladney

Central Question from March 2006 NIST LTKR

What, if anything, makes engineering different and/or harder than preservation in other domains?

- i.e. different from film/video, audio, scientific datasets, GIS, etc

Example: Preserving Film

Citizen Kane

- Possible Use Cases
  - View the film
  - Understand social context of the film
  - …
- Minimal info required?
  - MPEG file, production notes

The Magnificent Ambersons

If film preservation was like engineering one would preserve...

- The film
- Who made the film
  - Credits
- How the film was made
  - Where the cameras and lights were positioned for each shot
  - All shots, takes and film not in the final version
- The daily production logs
  - Who operated each piece of equipment each day
- Moods and state of the actors

What makes engineering different?

- Complexity of the data types
  - Semantics of data types must be preserved
- Diversity of the data types
  - Design, modeling, simulation, requirements, etc …
- Size of the data elements
  - Example Case Study: over 3G of data for a single machined part
- Temporal aspect
  - May need to preserve changes in an object over time
- Business process workflows within engineering organizations
  - Data must be captured and integrated from different places within the organization
- Engineering information is both descriptive and prescriptive
  - What to make and how to make it
- Lack of a well-defined stakeholder
- …

Assumptions/Realities of Preservation

- Disk space is free (near free)
  - One can literally store everything
- People do not want to do any additional work to preserve
  - One needs to be automated
- Simple formats have the greatest guarantee of being readable
  - Keep it simple
- It is hard to envision the ultimate use for the preserved data
  - Archeology, forensics, history, etc.

Vision of Memex: Digitization of the Totality of Human Knowledge!

- Soon everything can be
  recorded and indexed

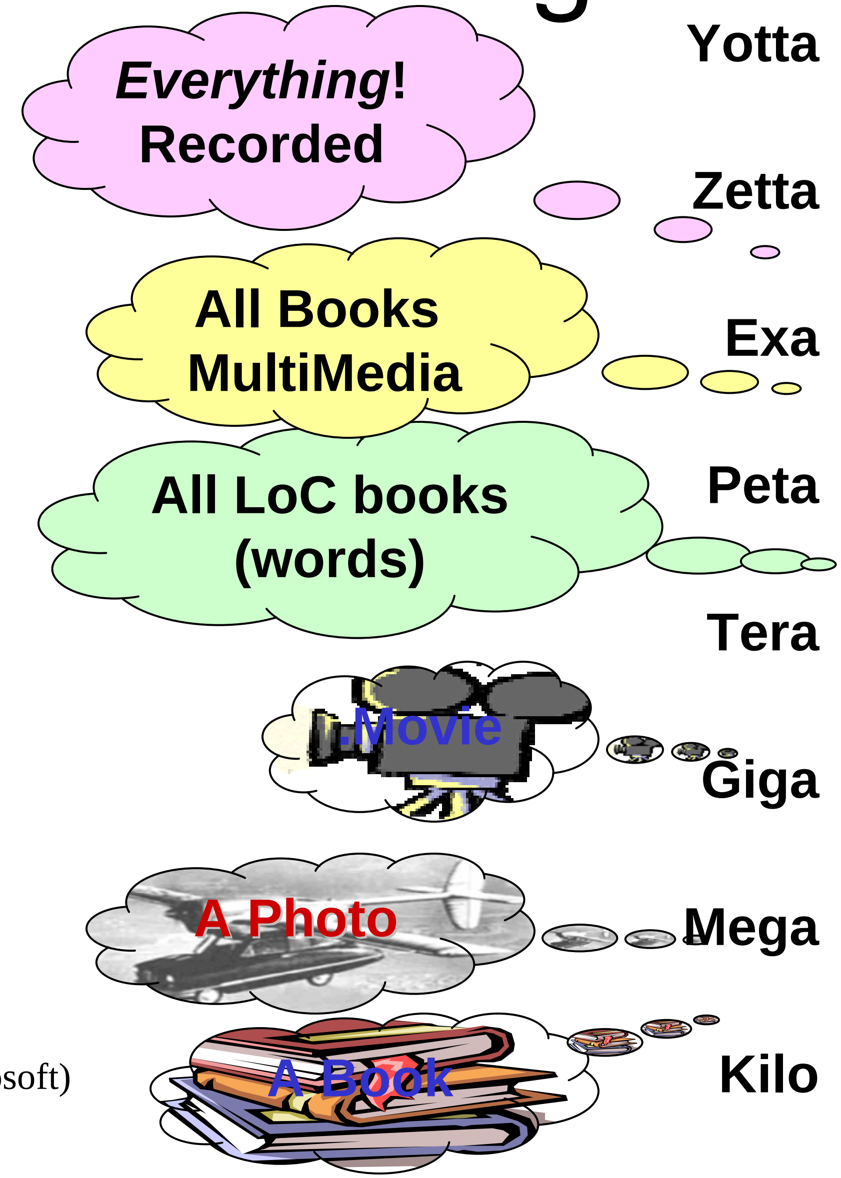
- Most data never be seen by humans

- **The Precious Resource:**
  Human Attention!
  Auto-Summarization
  Auto-Search
  is key technology.

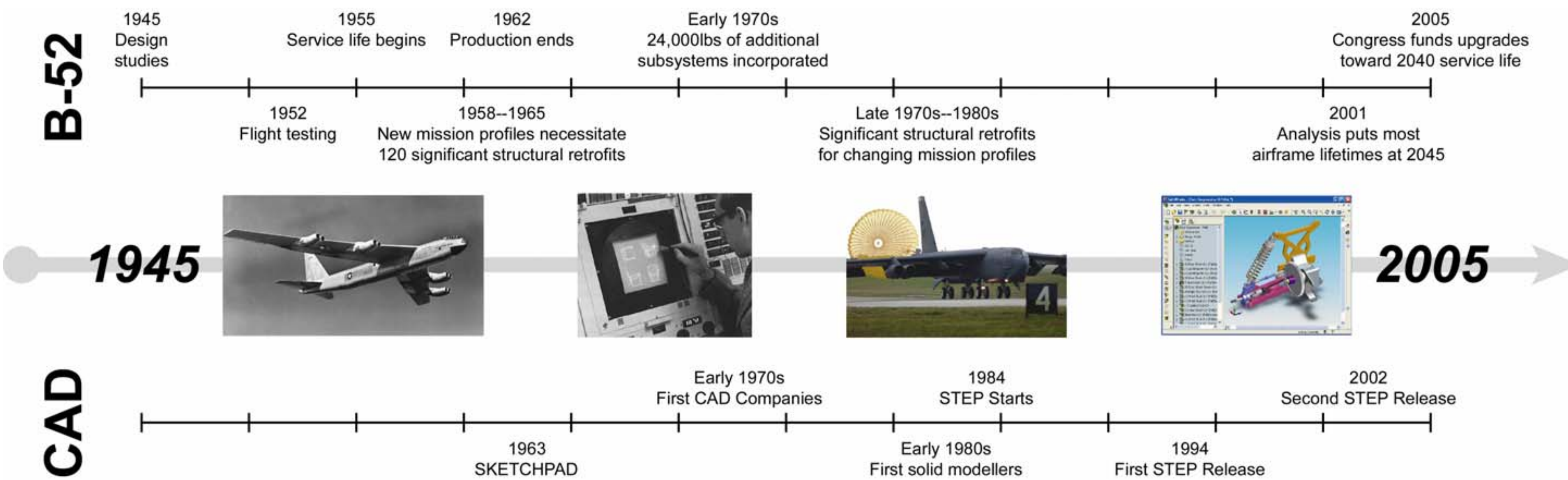
www.lesk.com/mlesk/ksg97/ksg.html

24 Yecto, 21 zepto, 18 atto, 15 femto, 12 pico, 9 nano, 6 micro, 3 milli

(From the Turning Award Lecture of Dr. James Gray, Microsoft)

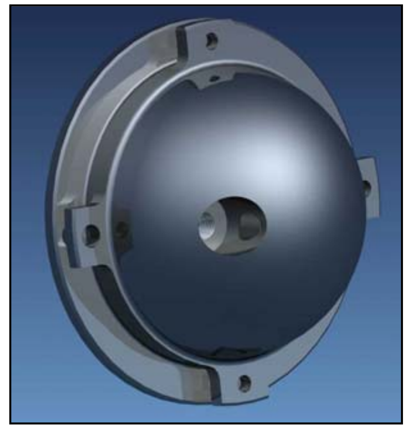


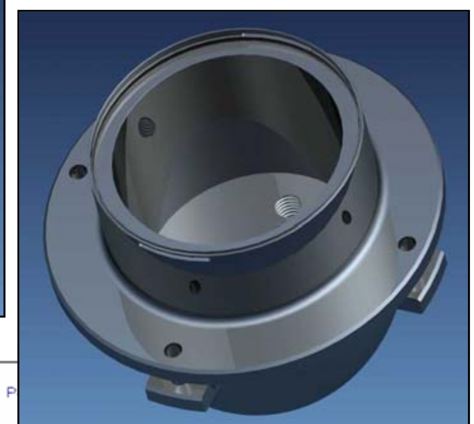
Consider the B-52

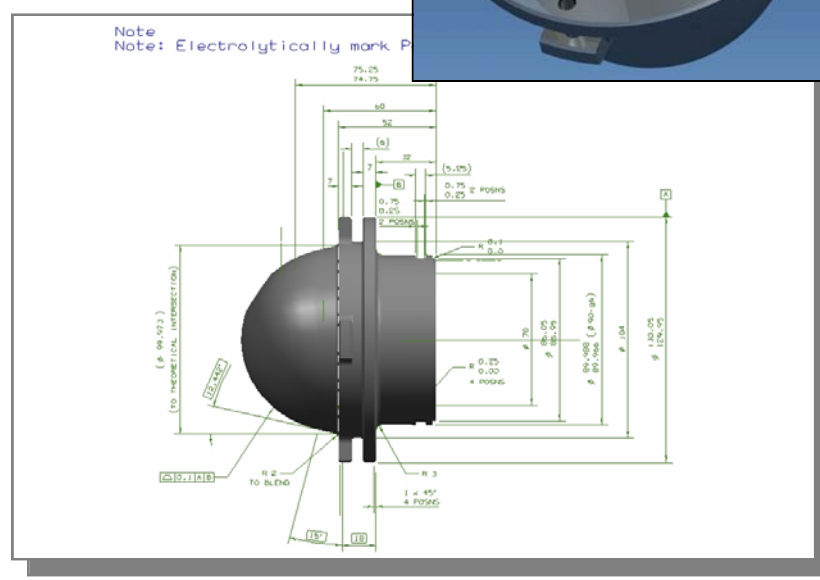


Our Challenge Problem: UK AWE Amber 2 Part

- Partner: **Kansas City Plant**
  - 50 year history
  - Primary manufacturing facility for the DoE & NNSA
  - Expertise in discrete parts, electronics, MEMS, …
- The Amber 2 part
  - High-precision machined part
  - Designed in the UK
  - Analysis at both UK AWE and KCP
  - Fabricated at KCP



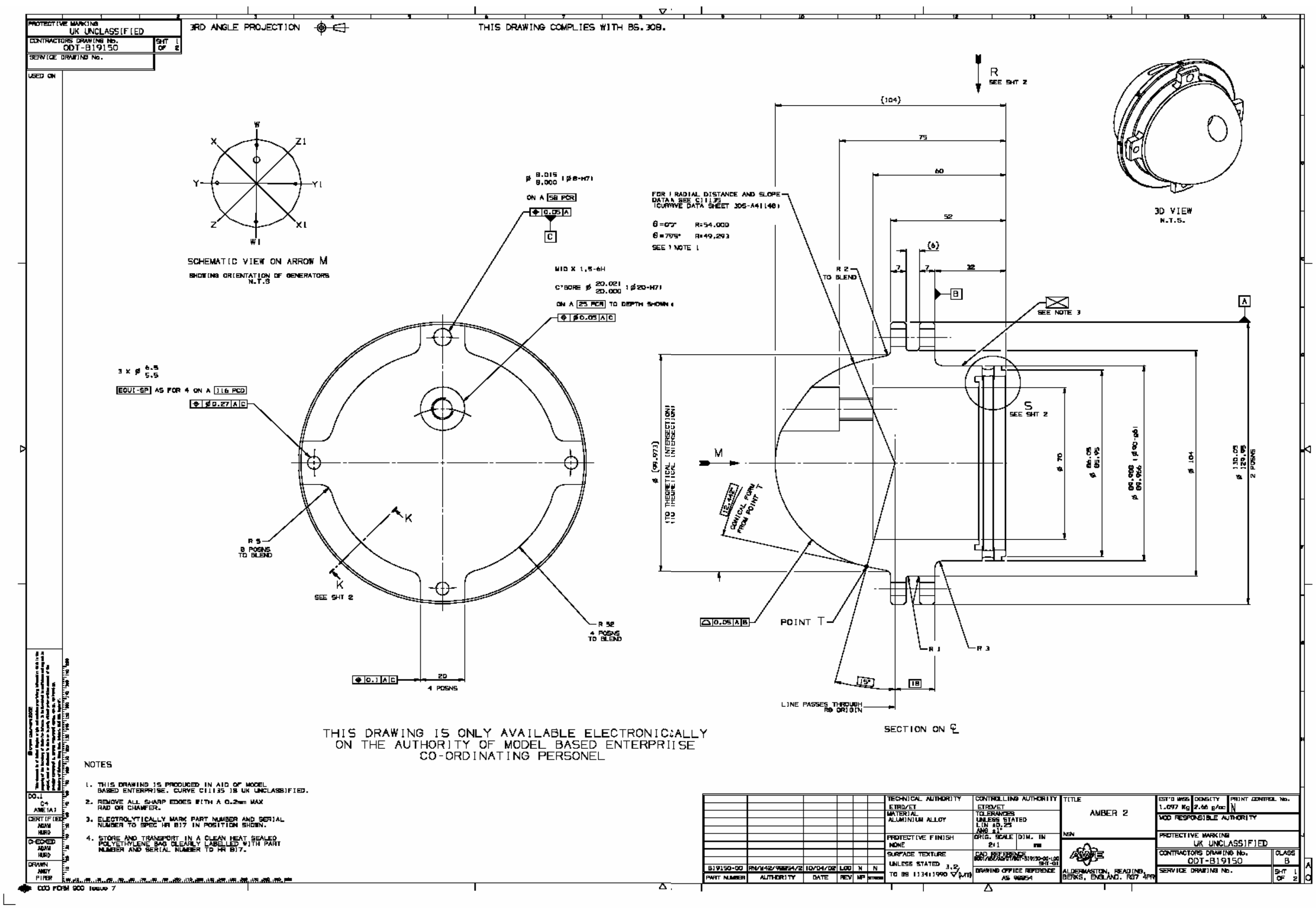




UK AWE Amber 2 Part Data

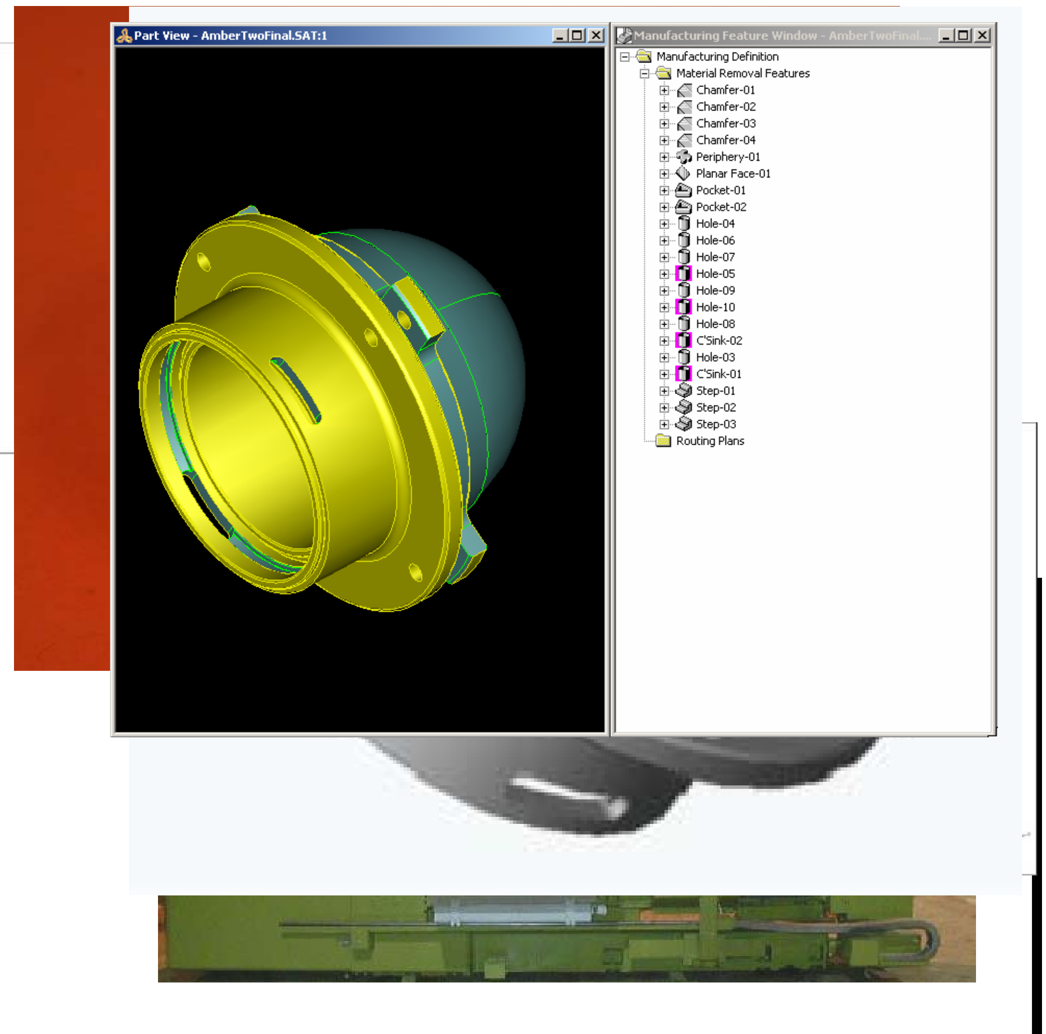
- 2D CAD Drawings
  - TIFF images
- 3D CAD data
  - Parasolid, Pro/E, STEP, ACIS, …
- Shape data
  - Mesh & point cloud
- Tolerance data
  - ASME Y14.5 tolerances and tolerance features
  - Tolerance analysis
- Analysis data
  - FEA parameters and output
- Manufacturing data
  - Features
  - Process plans
  - Manufacturing plan simulations
- Fabrication data
  - Tooling, cost, time
- Inspection data
  - Inspection plan, robotic simulation
- Documentation
  - MS Word files
  - AVIs, MPGs
  - Other files

Current Archive Format



Information Missing from Archive

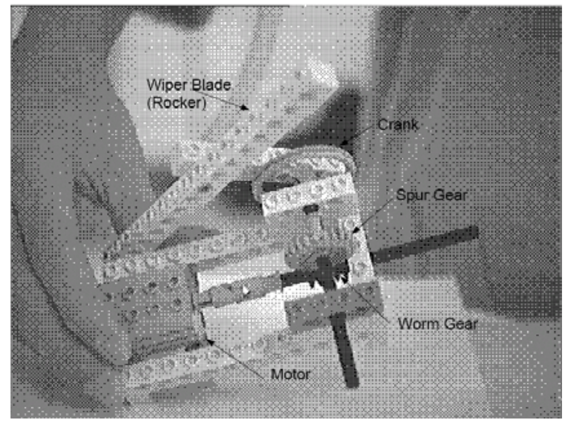
- Tolerances
- Manufacturing Planning
- Analysis
- Inspection
- Fabrication
  - Okuma LH35-N CNC lathe
- Reverse Engineering
- Provenance
- Etc.

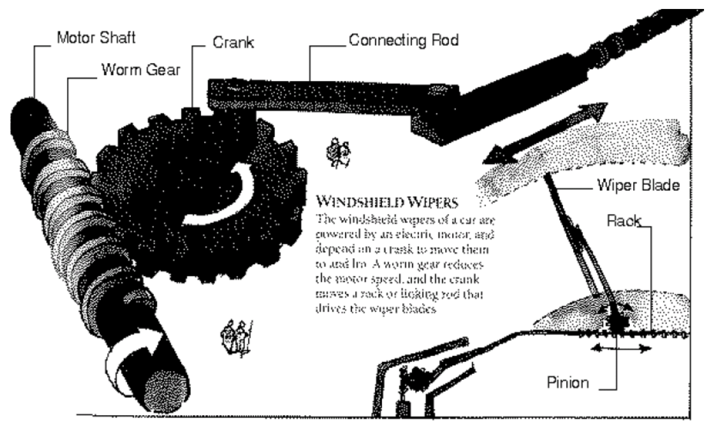


An example of a multi-disciplinary engineering model

Designing a “Windshield Wiper”

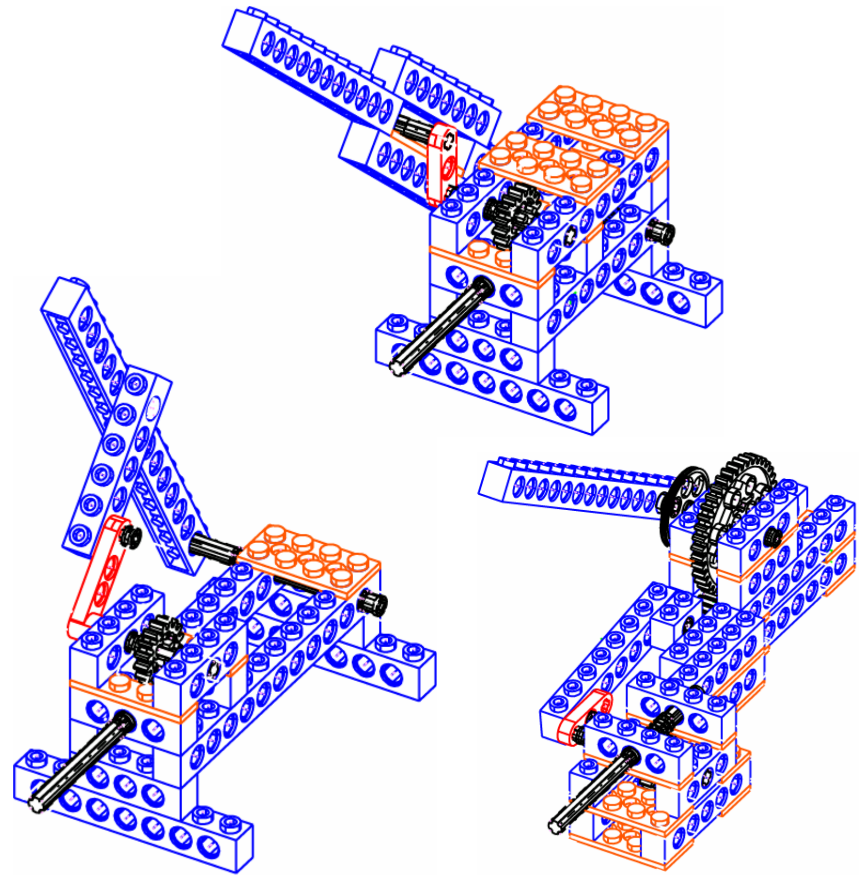
- From D. Macaulay, “How Things Work”
- What are the models relevant to engineering preservation?
  - Conceptual
  - Detailed Design (CAD)
  - Assembly
  - Functional
  - Behavioral
  - Simulation/Analysis
    - Physics
  - Manufacturing/Assembly
    - Inspection, Planning
  - Service and Maintenance
    - De-manufacturing/disposal





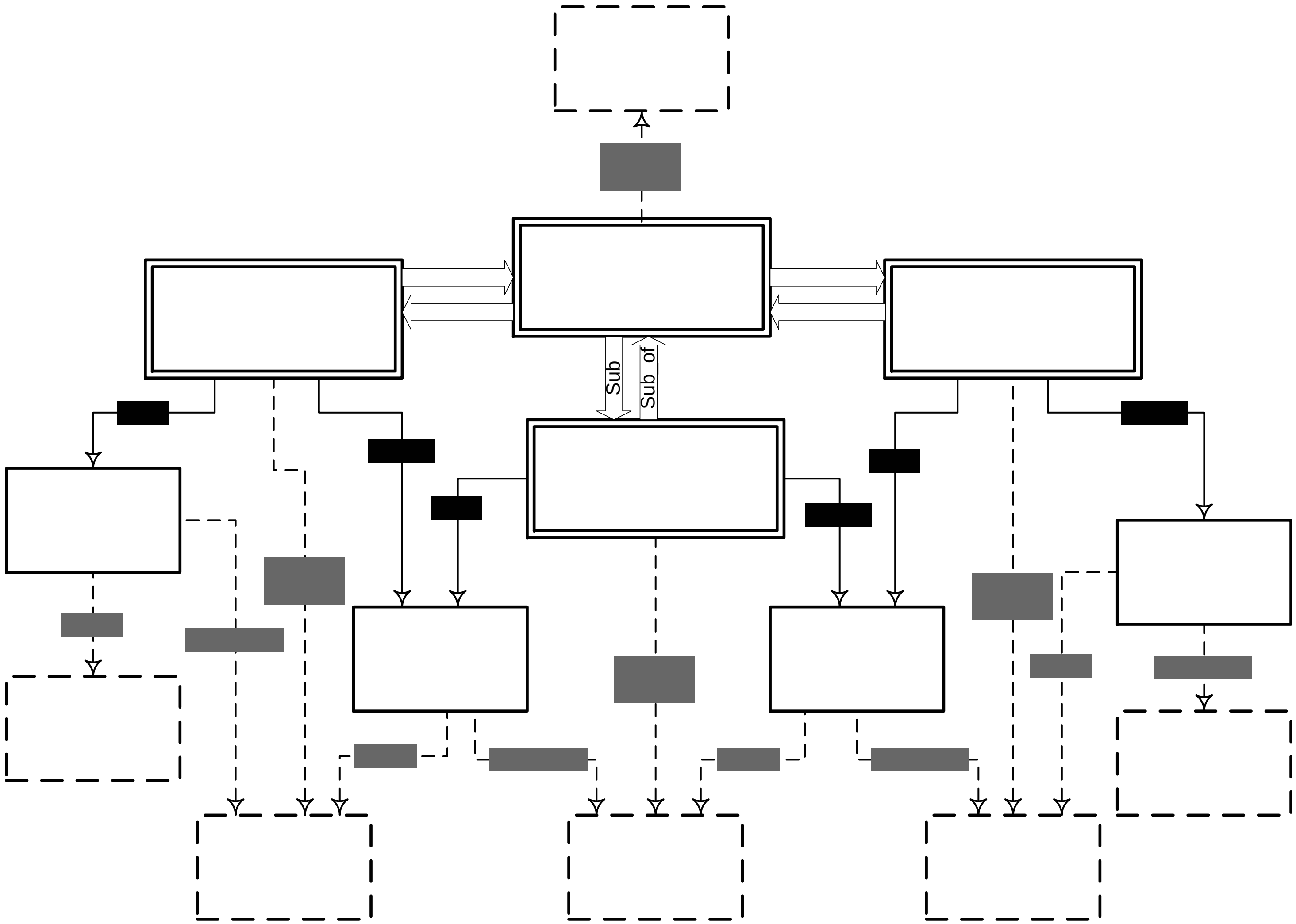
Models

- CAD Models
  - 3D models with joints and constraints
- Typically consist of
  - Part models
  - Assembly model(s)
- Formats can be 3D solid or 3D wireframe



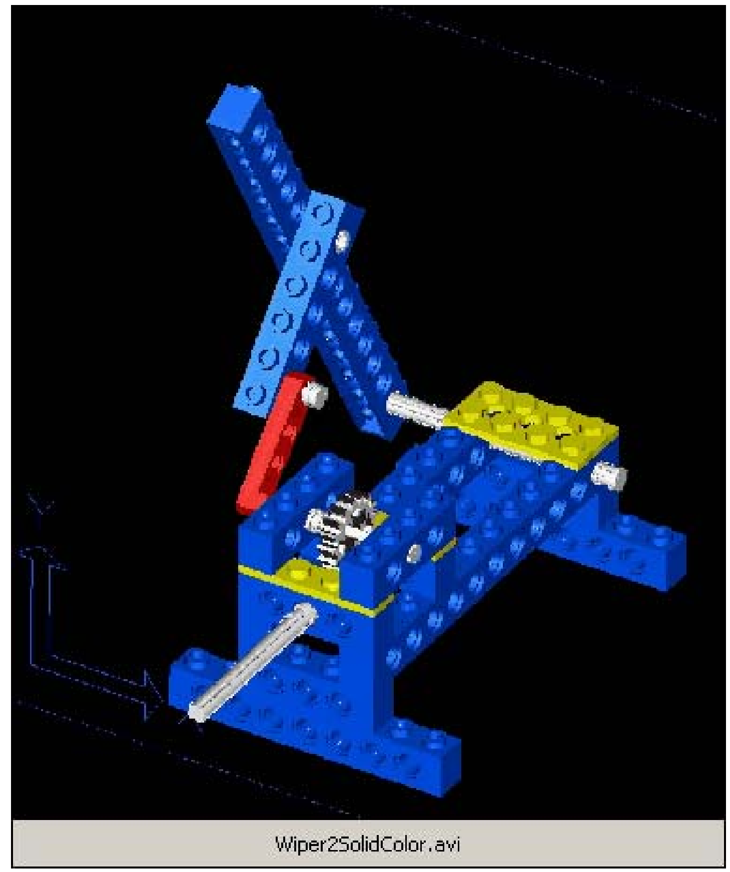
3 Lego models of a wiper assembly

Example Model: Behavior



Physics-based Models

- Kinematics (i.e. Animation)
  - Just move the parts based on joints & constraints
- Dynamics
  - Incorporate forces, motor torques, power consumption, friction, etc
- Other issues:
  - collision detection algorithms that check for intersection, calculate trajectories, impact times and impact points in a physical simulation

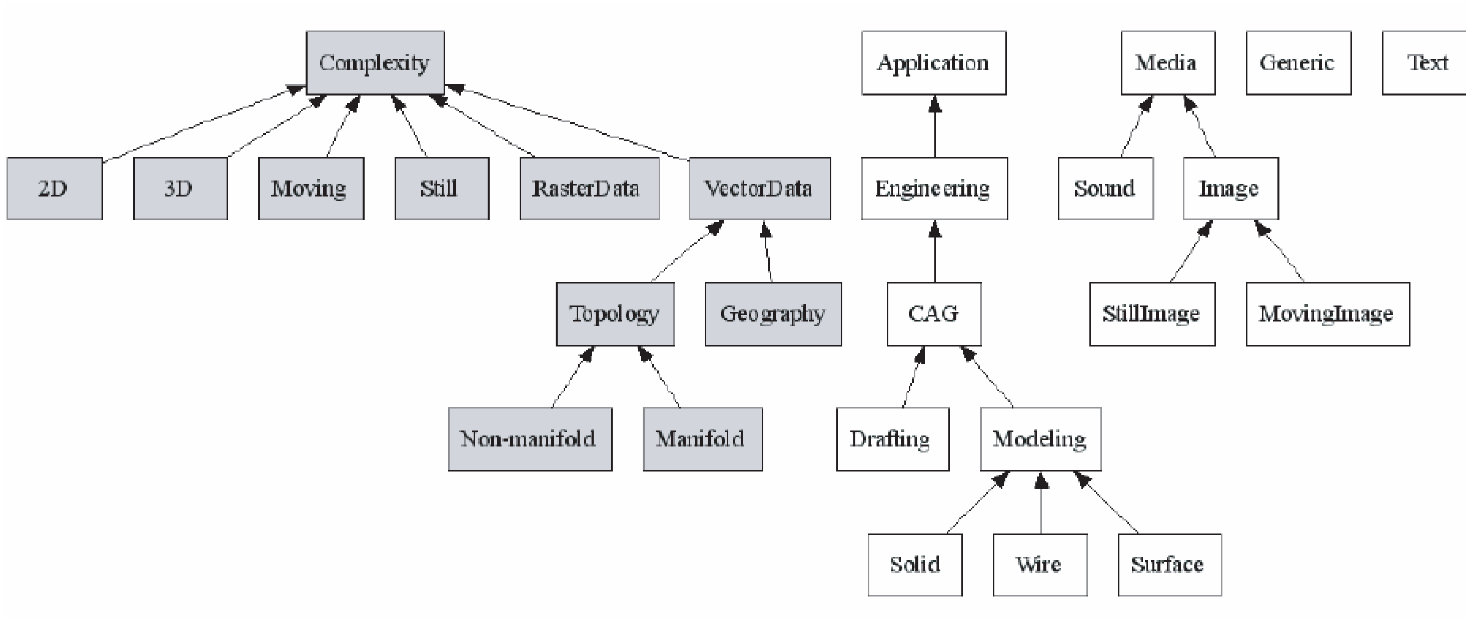


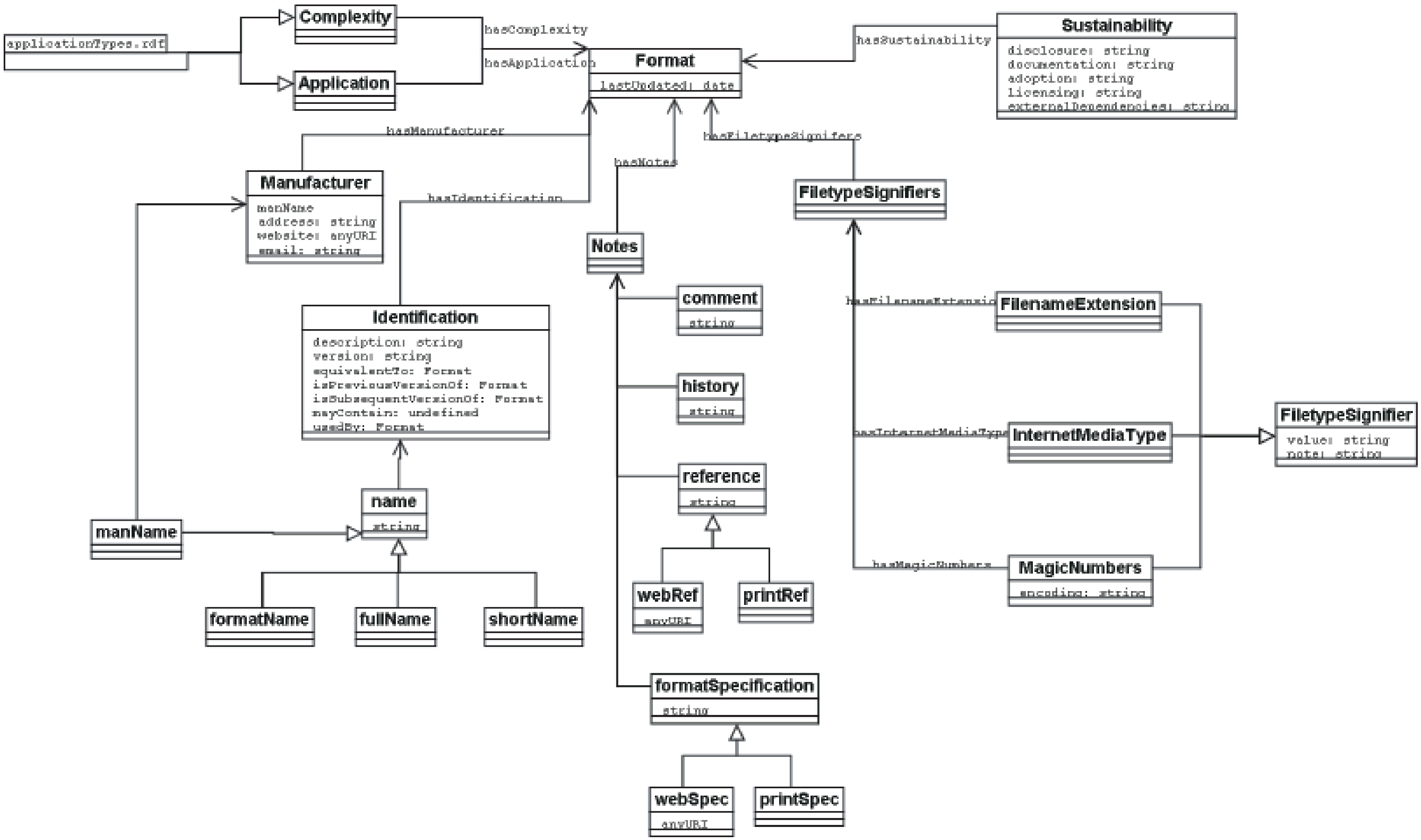
Immediate Needs
(i.e. low hanging fruit)

- Format Registries
  – CAD formats and related engineering formats
- Use cases
  – Who are the consumers of engineering archives?
- Representations for important data not currently covered by existing standards
  – STEP is not enough
- Software Tools
  – How to make preservation tangible? How to make injest transparent? How to interrogate the data once you have it?
- Open Testbeds
  – If we aren’t to have local, proprietary solutions, we need testbeds accessible to industry and academia
- Processes and best practices for preservation

Engineering Format Registry

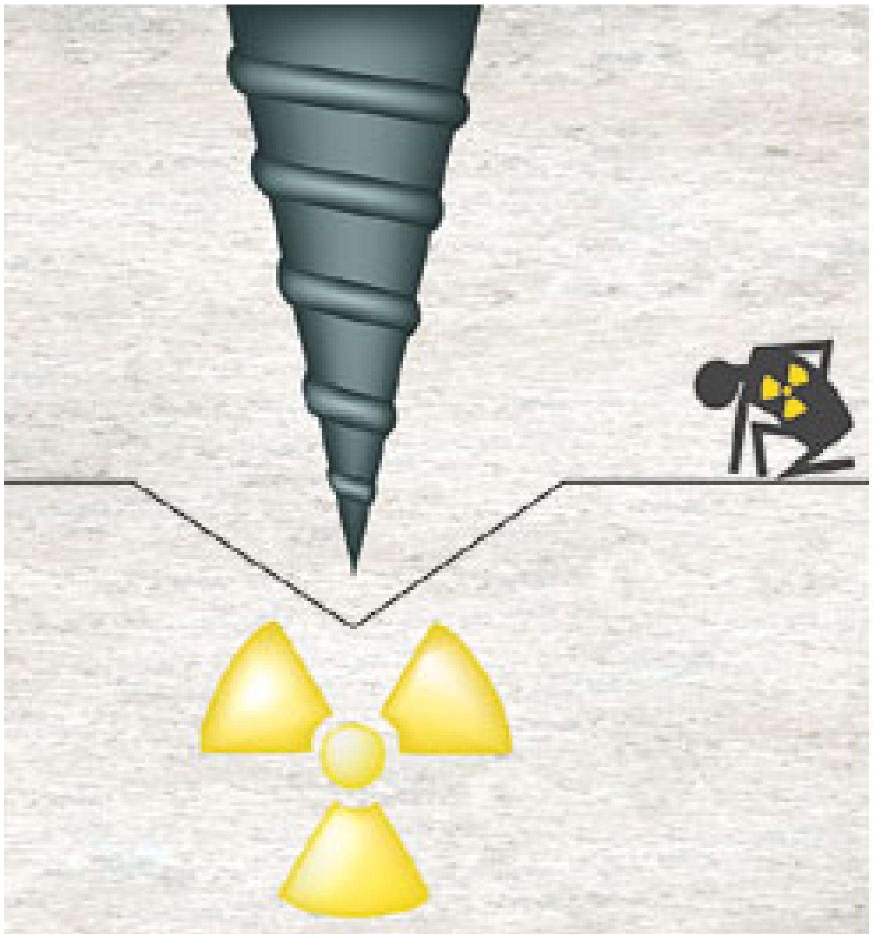
- Taxonomy of engineering data formats
- Captured with OWL with globally accessible wiki
- Technical contents
  - Format name, reference to formal specs, provenance, examples, applications that use it, etc



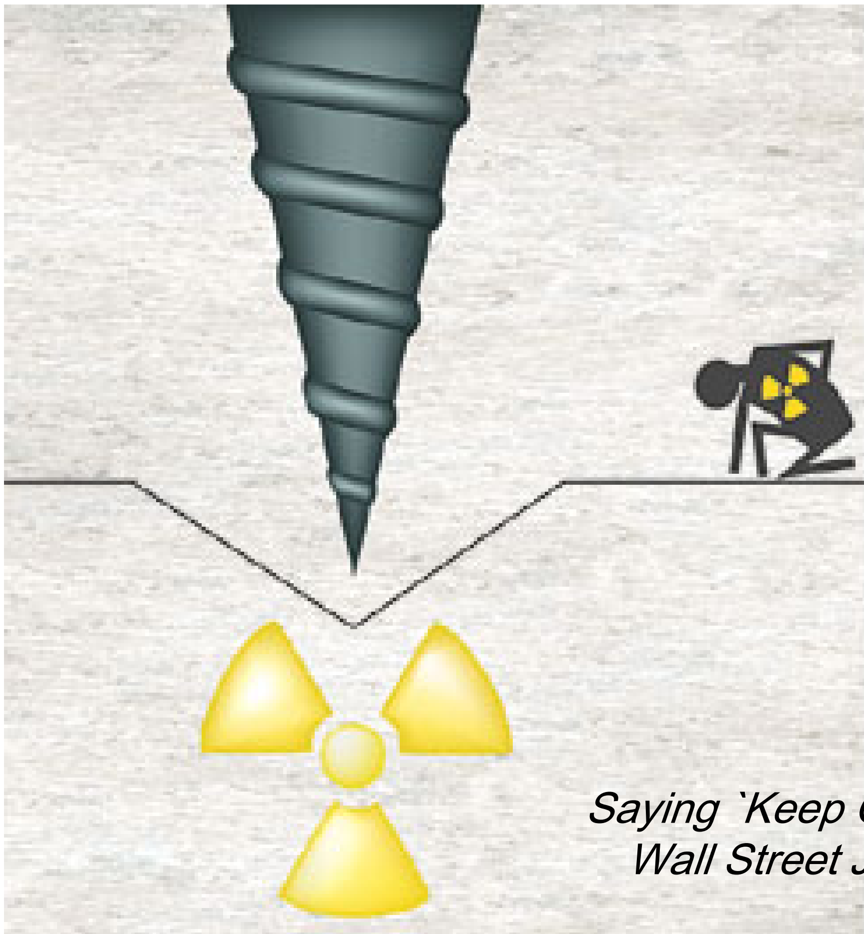


applicationTypes.rdf
Complexity
Application
hasComplexity
hasApplication
Format
lastUpdated: date
hasSustainability
Sustainability
disclosure: string
documentation: string
adoption: string
licensing: string
externalDependencies: string
hasManufacturer
hasNotes
hasFiletypeSignifers
Manufacturer
manName
address: string
website: anyURI
email: string
hasIdentification
FiletypeSignifiers
Notes
Identification
description: string
version: string
equivalentTo: Format
isPreviousVersionOf: Format
isSubsequentVersionOf: Format
mayContain: undefined
usedBy: Format
comment
string
hasFilenameExtensio
FilenameExtension
history
string
hasInternetMediaType
InternetMediaType
FiletypeSignifier
value: string
note: string
reference
string
name
string
manName
webRef
anyURI
printRef
hasMagicNumbers
MagicNumbers
encoding: string
formatName
fullName
shortName
formatSpecification
string
webSpec
anyURI
printSpec

Representations: Example



Representations: Example

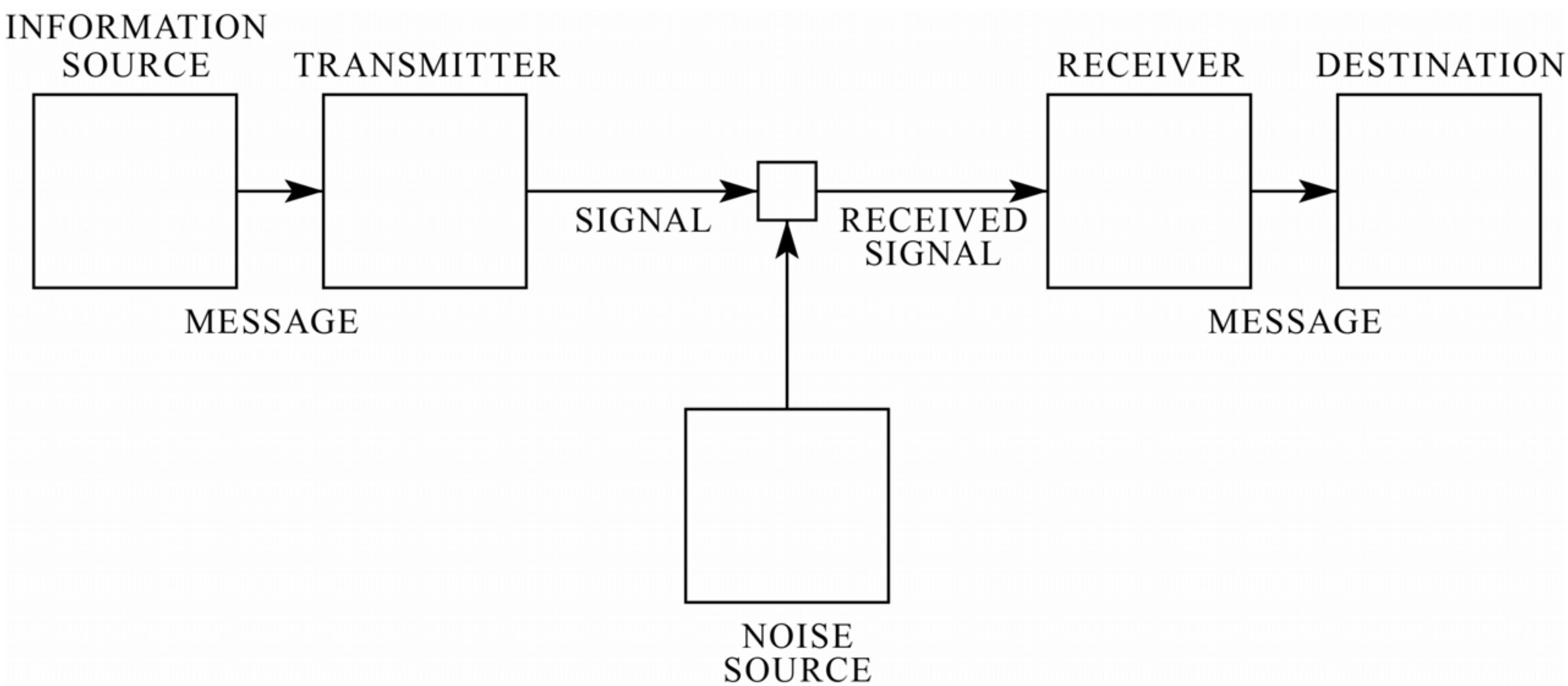


Saying `Keep Out' To Folks in 12,003
Wall Street Journal Feb 10, 2003

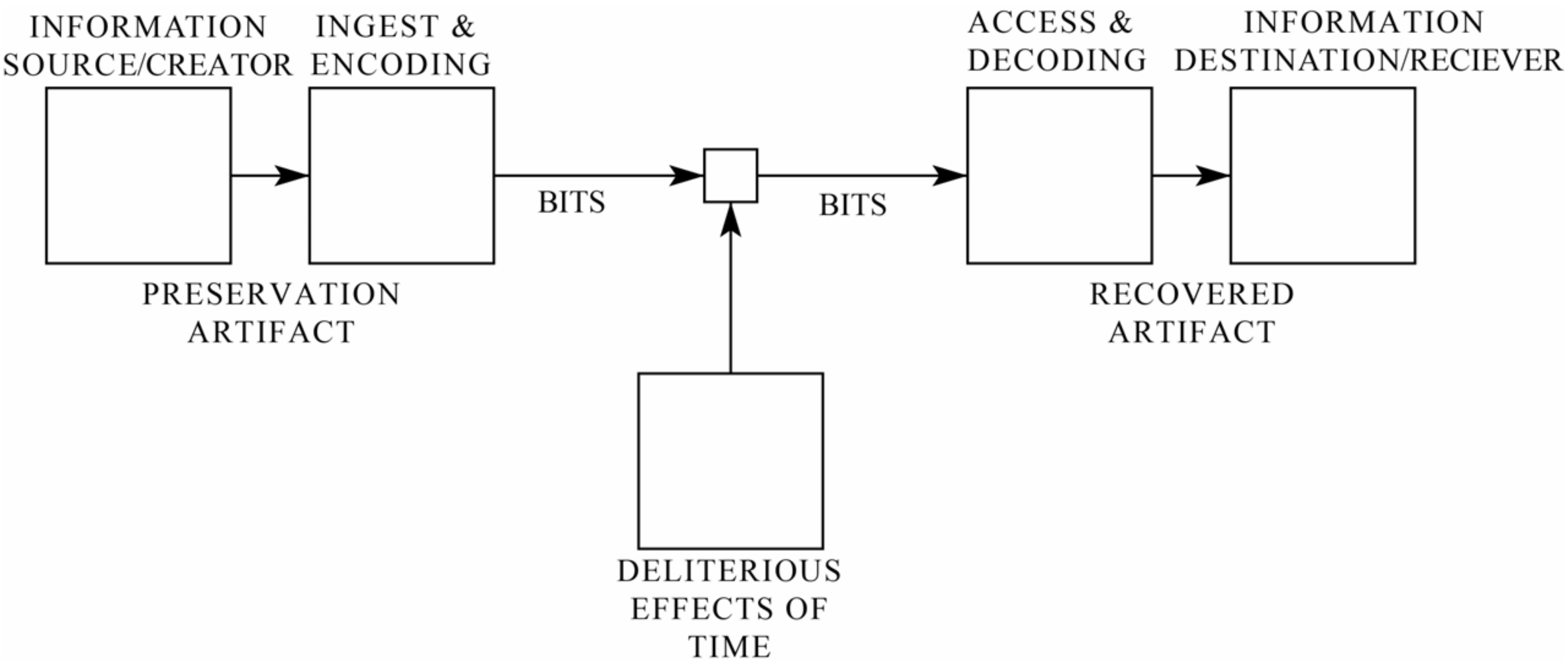
Representations

- Need to choose representations that will be accessible "in the future"
  - This rules out CAD native files
  - Redundancy and simplicity
- But such representations may
  - not capture the complexity of the engineered product
  - be hard to populate automatically

From Shannon (1948)



Shannon
(modified for digital preservation)

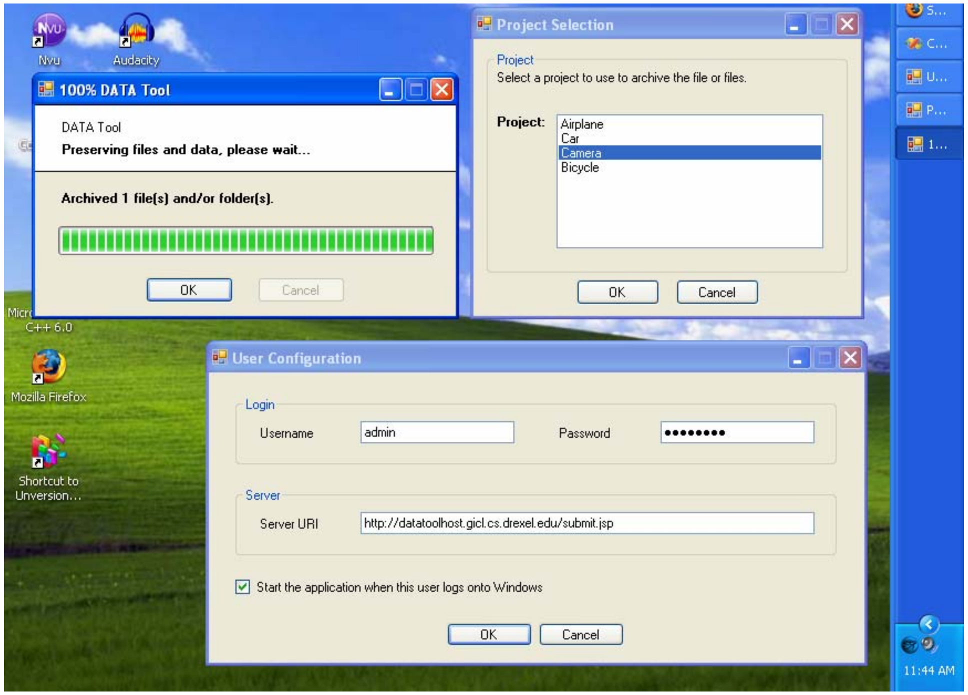


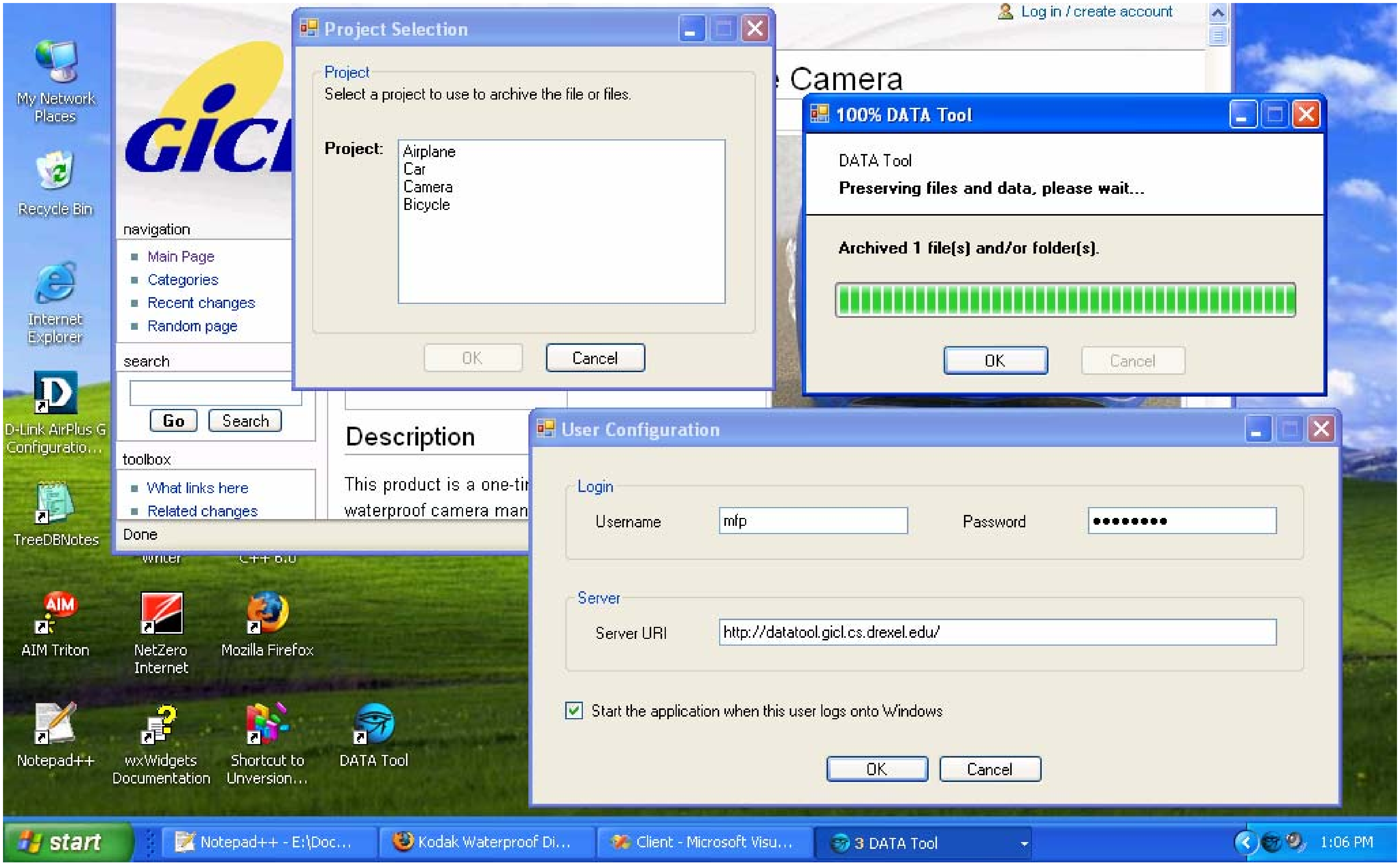
Representation of Engineering Information

- STEP may only be part of the solution
  - Other options may include
    - Low level formats (SMF, STL, VRML) that are self-documenting
    - Mathematical formats that are semantically stable
      - OWL, Prolog, FoL
    - Workflow, business practices, associations among data elements, etc
    - May include other digital media (video, audio)
- Capturing something is better than nothing
  - Need automated tools to populate representations

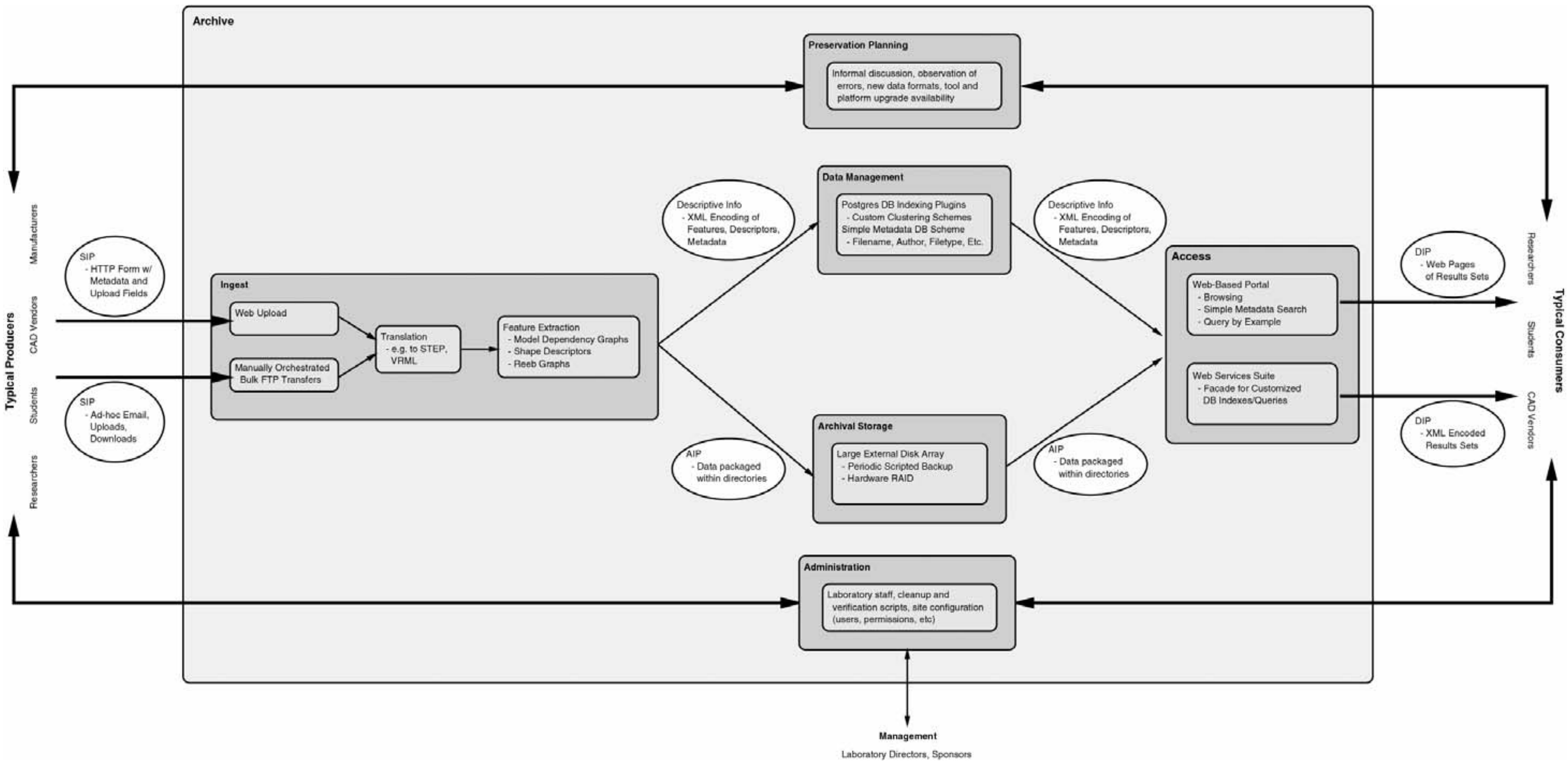
Archive Tools

- Objective: insert archiving activity seamlessly into existing workflows
- Idea: drag and drop, context-specific ‘backup’
  - Engineer ‘backs up’ files and an intelligent archive agent determines file context and related metadata

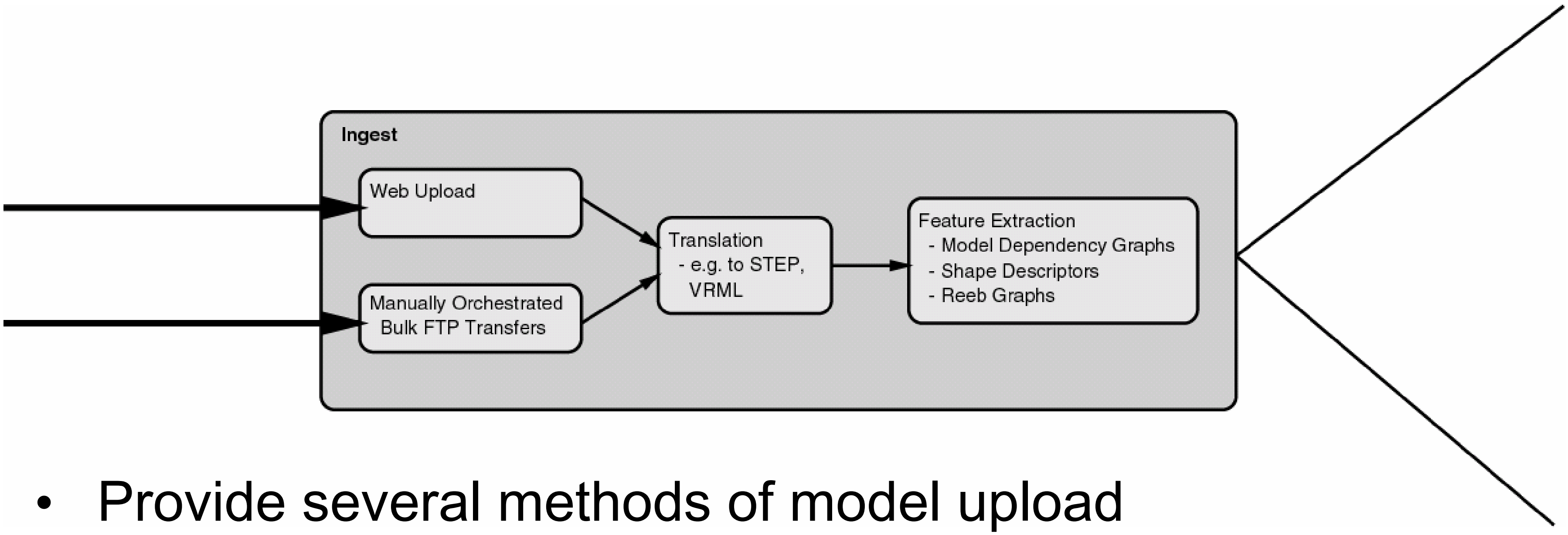



Project Selection
Project
Select a project to use to archive the file or files.
Project:
Airplane
Car
Camera
Bicycle
OK
Cancel
100% DATA Tool
DATA Tool
Preserving files and data, please wait...
Archived 1 file(s) and/or folder(s).
OK
Cancel
User Configuration
Login
Username
mfp
Password
Server
Server URI
http://datatool.gicl.cs.drexel.edu/
Start the application when this user logs onto Windows
OK
Cancel
Log in / create account
Camera
navigation
Main Page
Categories
Recent changes
Random page
search
Go
Search
toolbox
What links here
Related changes
Done
Description
This product is a one-ti
waterproof camera man
My Network Places
Recycle Bin
Internet Explorer
D-Link AirPlus G Configuratio...
TreeDBNotes
AIM Triton
NetZero Internet
Mozilla Firefox
Notepad++
wxWidgets Documentation
Shortcut to Unversion...
DATA Tool
start
Notepad++ - E:\Doc...
Kodak Waterproof Di...
Client - Microsoft Visu...
3 DATA Tool
1:06 PM

OAIS Semantic Web Service Framework Overview

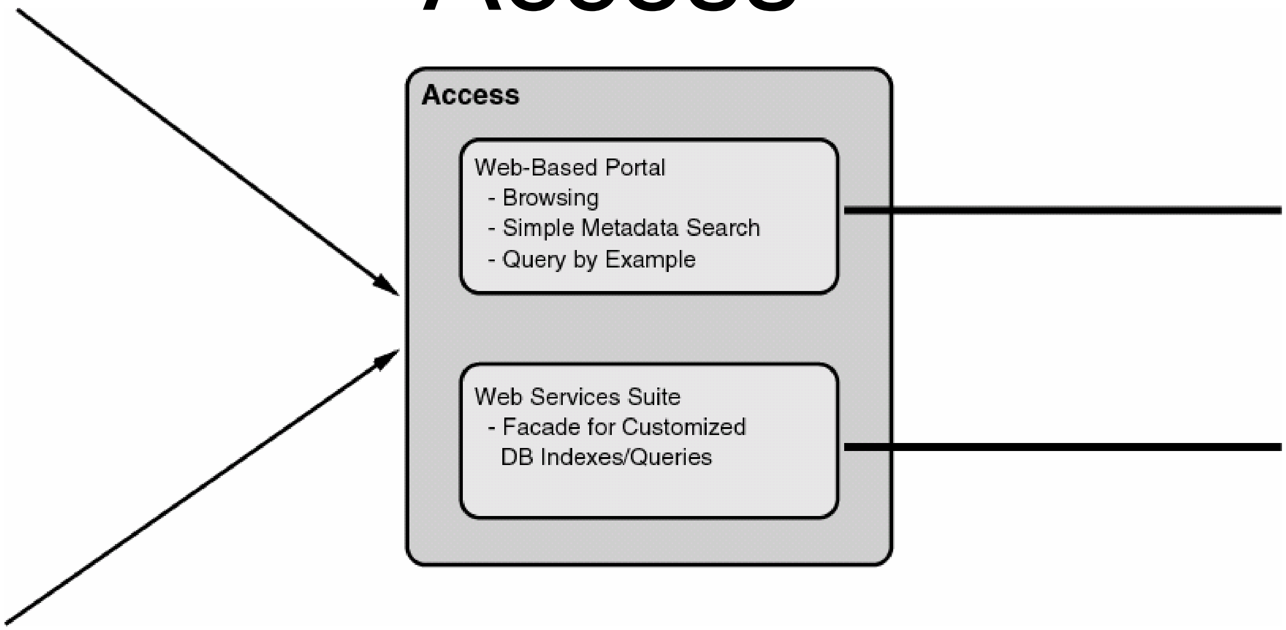


Ingest



- Provide several methods of model upload
  - Web, FTP, etc
- Backend translation of native formats into a "preservation vector"
  - E.g. XMT -> STEP, VRML, STL, etc
- Pre-processing for 3D and CAD Search

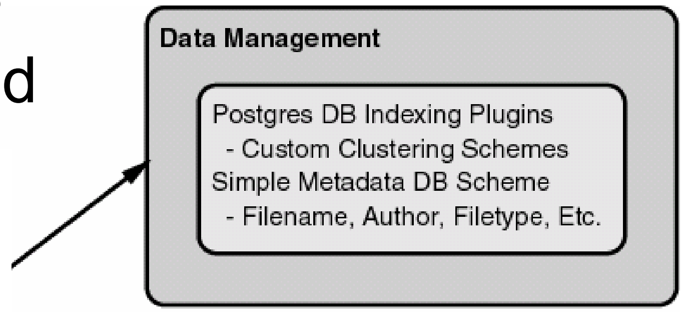
Access



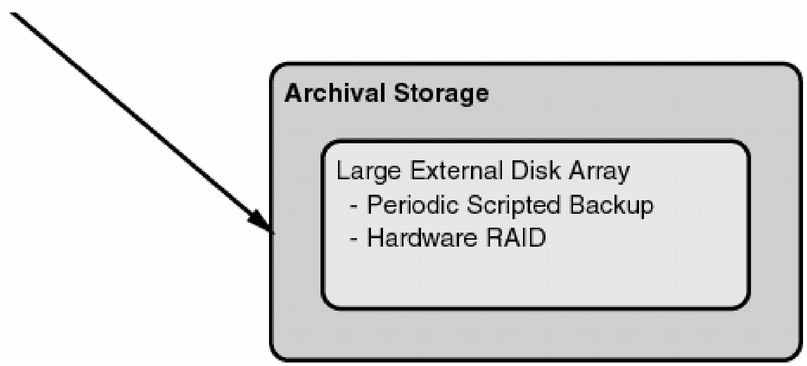
- Several techniques under development
  - Web portal, Wiki, version control (SVN), query/navigate
- Techniques need evaluation/testing to determine +/-

Backend Data Management

- Management issues
  - Metadata tagging and migration
  - Clustering, indexing and organization

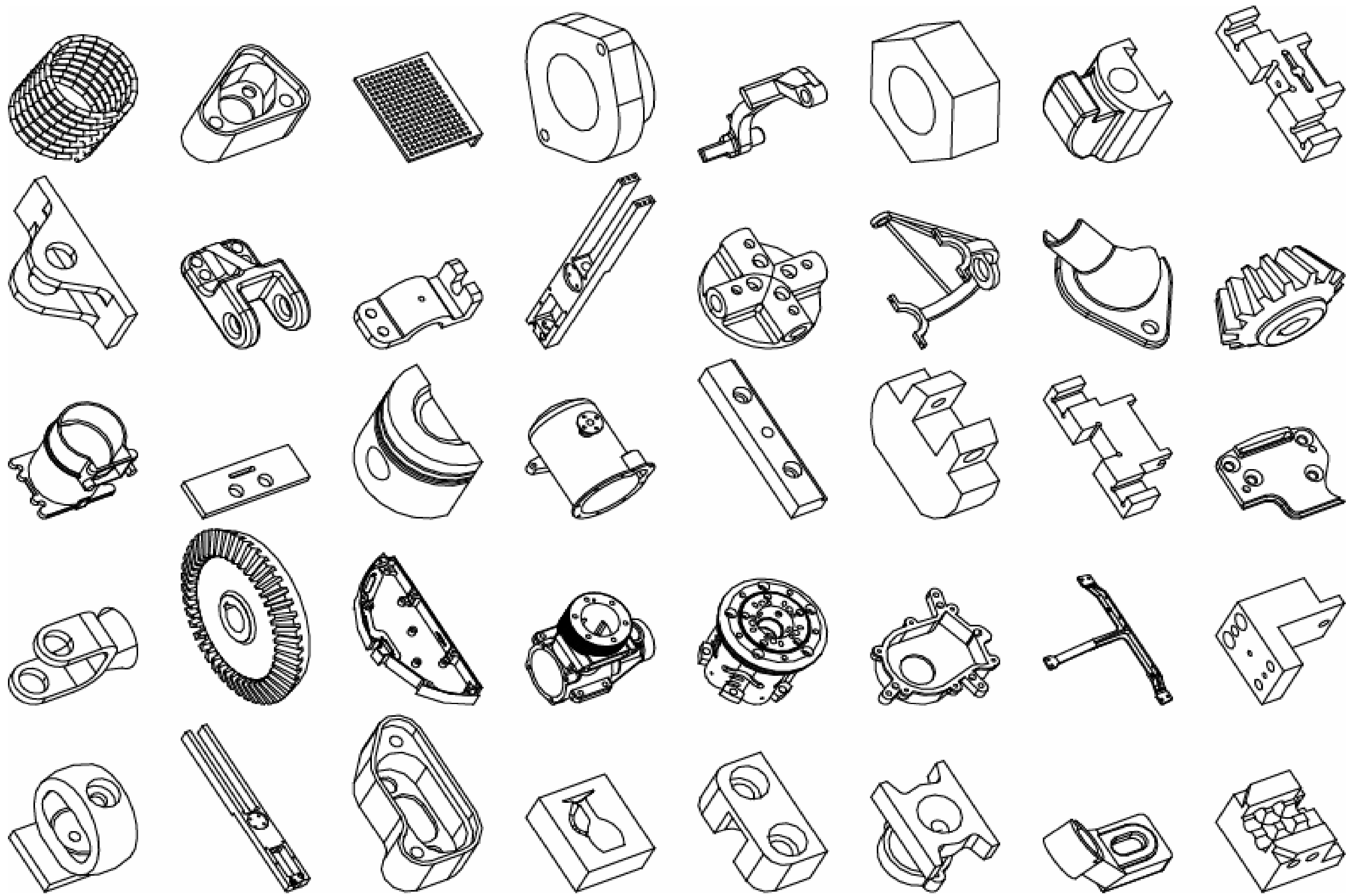


- Storage approaches
  - Basic: DB, files
  - More Advanced: SDSC SRB or Grid



Testbeds

National Design Repository Dataset



CI-TEAM: Creation and Use of Multi-Disciplinary Engineering Models

NSF CISE/SCI-0537370

- Lead Institution: Drexel University
  - PI William Regli, co-PI Michael Piasecki
- University of Maryland @ College Park
  - SK Gupta
- University of North Carolina @ Chapel Hill
  - Ming Lin and Dinesh Manocha
- University of Wisconsin @ Madison
  - Nicola Ferrier, Vadim Shapiro, Krishnan Suresh

Project Objectives

- Develop a multi-disciplinary approach to Engineering Informatics education
- Design curricular materials around the theme of bio-inspired and snake-inspired robots
  - Make materials accessible to undergraduates and high-school students
- Create cyber-tools for design/analysis of bio-inspired robots
- Populate the cyber-infrastructure with information on bio-inspired robot design

Megan's Course Project

From GICLWiki

Final Paper

Contents

Documentation

This section includes a list of files that can be used to reconstruct

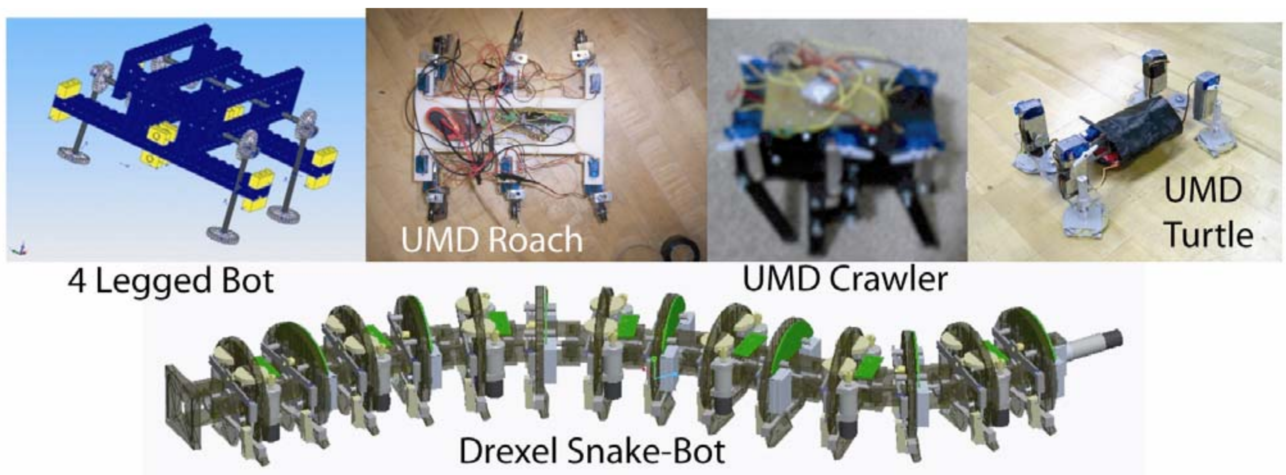
Bill of Materials

Media:ComponentList.jpg

Final Videos

Physical
Media:Physcial.zip

Virtual
Media:Simulation.avi



Expected Outcomes

- Runs thru 2009
- Production of dozens of models of bio-inspired robots
  - Complete robots, robot sub-mechanisms
- Simulation models, physics models
  - Academic tools, commercial (ADAMS), etc
- All documented on the CI Wiki

CI-TEAM: Cyber-Infrastructure-Based Engineering Repositories for Undergraduates (CIBER-U)

NSF CISE/SCI-0537125

- Lead Institution: Penn State University
  - PI Timothy Simpson (IE)
- State University of New York @ Buffalo
  - Kemper Lewis (ME)
- University of Missouri @ Rolla
  - Rob Stone (ME)
- Drexel University
  - William Regli

NSF Grant Number: SCI-0537220 (Lead) Start Date: January 1, 2006

PI: Timothy W. Simpson Lead Institution: Penn State University

Collaborators: Kemper Lewis (U. Buffalo), William Regli (Drexel), Robert B. Stone (U. Missouri-Rolla)

Title: Collaborative Research: Cyber-Infrastructure-Based Engineering Repositories for Undergraduates (CIBER-U)

• Research Objectives:

1. Demonstrate how digital design repositories can be used to enhance instruction and learning in undergraduate curricula
2. Implement CIBER-U in seven undergraduate courses at Penn State, University at Buffalo, and University of Missouri-Rolla
3. Assess the impact on the educational experience, learning, and practical knowledge of the undergraduate students who interact with the National Design Repository via CIBER-U

• Research Questions:

Phase 1: Issues in Repository Entry, Control, and Security

1. How are data entry, access, and storage effectively controlled for educational design repositories within CIBER-U?
2. What are effective search, sharing, and security protocols for educational design repositories within CIBER-U?

Phase 2: Issues in Repository and CAD Model Reuse

1. What are the best reuse strategies for the repository data depending upon the course and education level?
2. Given the distributed yet collaborative nature of the undergraduate engineering students, how can these strategies be used to enhance their educational experience?

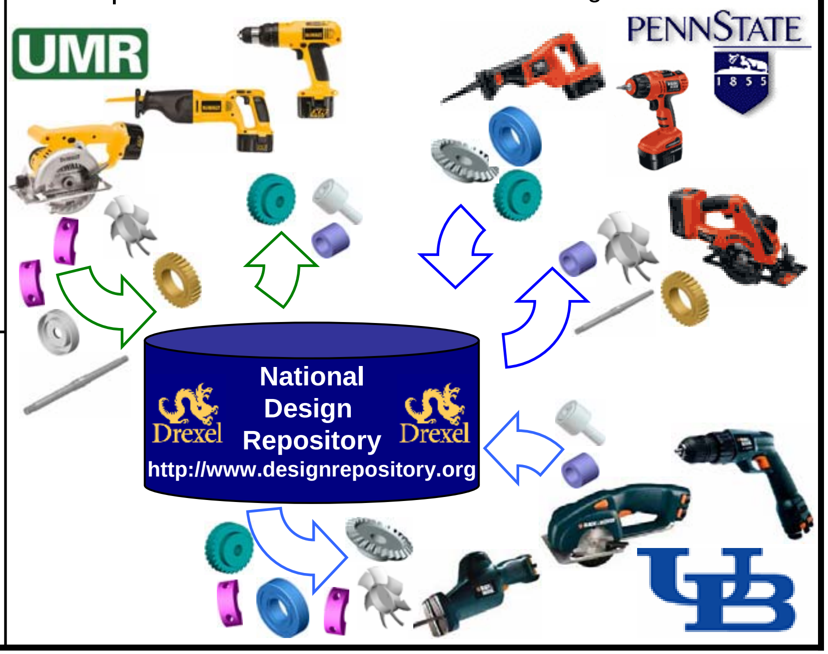
•Broader Impact:

- In two years, 1700+ undergraduate students will participate in CIBER-U, helping to prepare them to enter the workforce with a more effective understanding of working in a distributed, technology-mediated environment
- We will engage 200+ high school students in CIBER-U through ongoing summer high school outreach activities and integrate it within *Project Lead the Way* to expose potentially thousands of additional high school students to cyber-infrastructure

• Intellectual Merit:

1. CIBER-U will realize a high-impact application of advanced cyber-infrastructure for engineering education
2. CIBER-U will foster novel approaches for teaching engineering design through the use of collaborative design repositories
3. The National Design Repository will be frequently populated, updated, and utilized while simultaneously helping advance systematic and coordinated methods to access, store, search, and reuse CAD models and data

• Graphic: CAD Model Creation and Sharing with CIBER-U



Example Product Dissection

Kodak Waterproof One-Time-Use Camera

From GICLWiki

Contents

Figure 1: Fully assembled product

Resulting Dissection

Table 3.1: Disposable Camera Bill of Materials

Part #	Part Name	# Req'd	Mat'l	Manufacturing Process	Image
1	Back Interior	1	ABS Plastic	Injection Molding	
2	Back Plastic Cover	1	ABS Plastic	Injection Molding	
3	Camshaft	1	ABS Plastic	Injection Molding	
4	Eyehole for Shutter	1	ABS Plastic	Injection Molding	
5	Film Advance Gear	1	ABS Plastic	Injection Molding	
6	Film Advance Lock	1	ABS Plastic	Injection Molding	

Table 4.1: Disassembly of a Disposable Camera

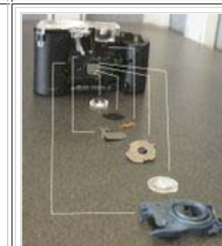
Step	Procedure	Image
1	Pop film advance wheel up and out from body; separate covers	 Figure 5.1: Exterior Disassembly
2	Remove film canister and film spindle	 Figure 5.2: Large Component Removal
	Remove lens holder, outer lens, shutter cap, shutter spring, shutter, eyehole, and inner lens	 Figure 5.3: Lens Assembly Dissection

Table 5.1: List of Models for Download
(Zipped ACIS, IGS, STL, STP, AD_PRT Files)

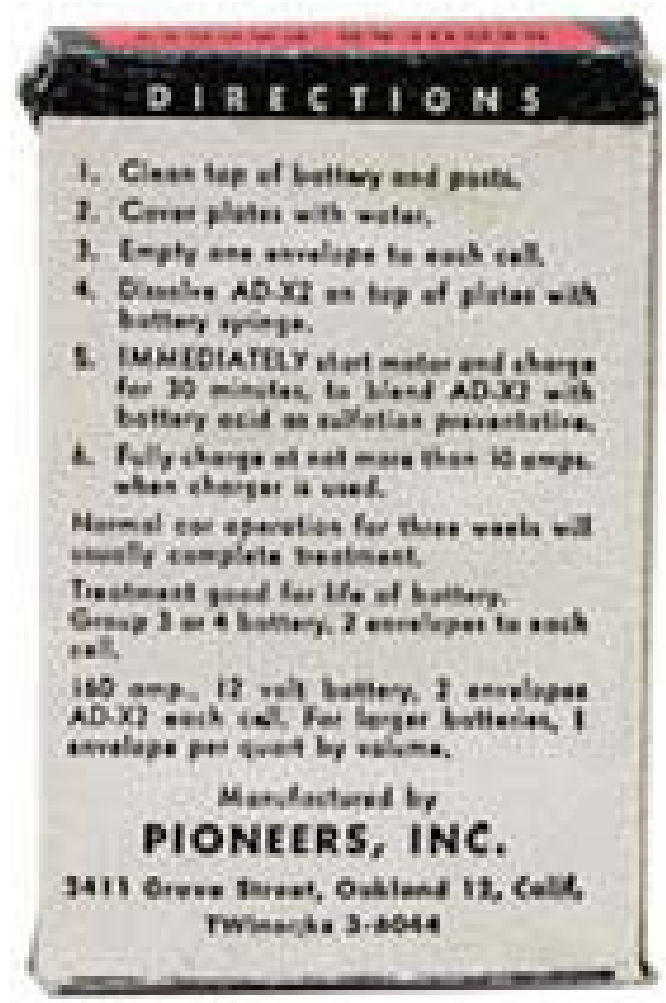
Part #	Part Name	Image	3D Models	2D Drawings
1	Back Interior		3D Package	2D Package
2	Back Plastic Cover		3D Package	2D Package
3	Camshaft		3D Package	2D Package
4	Eyehole for Shutter		3D Package	2D Package

Expected Outcomes

- Runs thru 2009
- 2000+ students per year at 9 institutions will be engaged in product dissection activities
- Production of (potentially) hundreds of product dissection models
  - CAD models, laser scans, photos, functional models, materials, etc.
- All documented on the CI Wiki

Processes and Best Practices

- The Story of Battery Additive AD-X2



Processes and Best Practices

- Good policies and practices may be more important than representations and search tools
- What is needed
  - Guidelines or heuristics
  - Validation procedures
  - Implementation examples

Going forward

- Whose problem is this?
  - What are the best roles for gov't, university and industry?
- Several obvious pieces of fruit
  - Format Registries, Best Practices
- Testbeds becoming available
- Representation and capture of information is fundamental and difficult
  - Going beyond STEP

Q&A

National Science Foundation (NSF)

Digital Archiving and Long-Term Preservation (DIGARCH)
Award NSF CISE/IIS-0456001

Cyber-Infrastructure TEAMs Demonstration Program
Grant SCI-0537125, CIBER-U
Grant SCI-0537370, Multi-Disciplinary Engineering Models