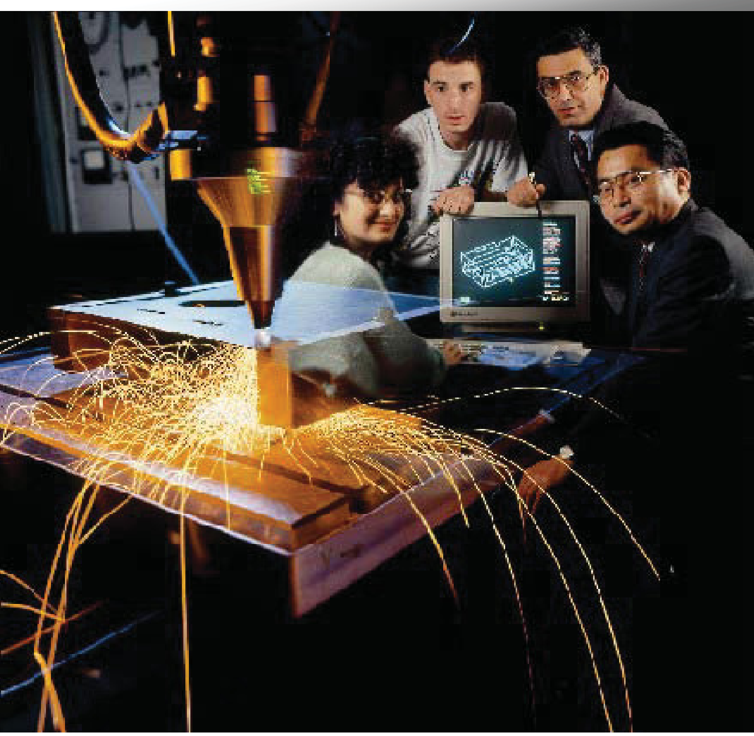


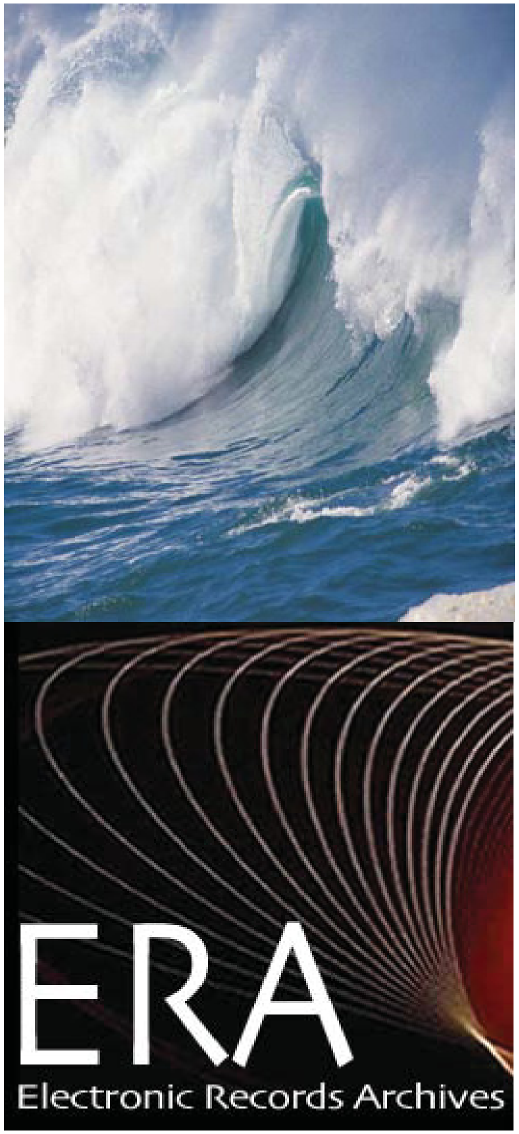
‘Set-Aside’ in the Semantic Web:

Findings and Implications of Other Government Case Studies

L’Associazione Nazionale Archivistica Italia – ANAI
Seminario Internazionale “I risultati di InterPARES2”
Centro Congresso Le Stelline
Milano
12 – 13 dicembre 2006

Kenneth Hawkins, Ph.D
Electronic Records Archives Program
U.S. National Archives and Records Administration
College Park, Maryland

NARA's Mission and the ERA



archives.gov/era

- To preserve and make accessible the permanently valuable records of the United States government

- Scope / Variety / Complexity / Volume / Obsolescence

- *"ERA will authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software, enabling NARA to carry out its mission into the future."*

Focus 3: eGovernment

InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

CS18 – Computerization of Alsace-Moselle's Land Registry

CS19 – Preservation and Authentication of Electronic Engineering and Manufacturing Records

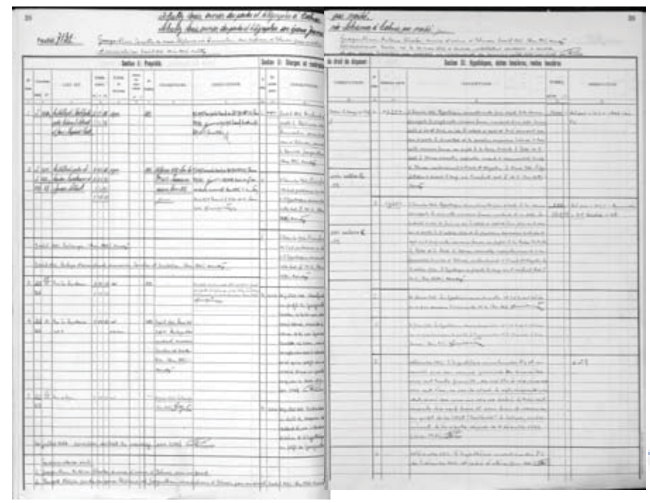
- documentary entities throughout systems
- specifying how d/e's relate to each other
- fixed or memorial *vs.* enabling or prospective
- static *vs.* actionable metadata

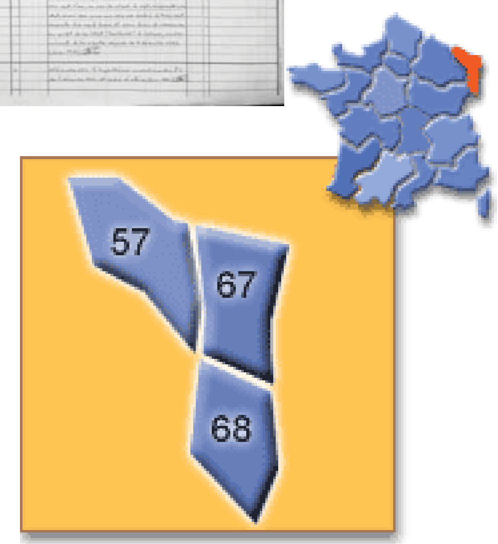
Two indicative case studies

- Case Study 18
  - Computerization of Alsace-Moselle's Land Registry
- Case Study 19
  - Preservation and Authentication of Electronic Engineering and Manufacturing Records

CS18: Business Context

- the French region of Alsace-Moselle's inscription of real estate transactions in a *livre foncier*
- Making public record of property rights, sales, transfers, and legal verification of inscriptions in the registry
- Operated by authority of Ministry of Justice under mandates dating to 1891





- In 1994, the functions transferred to a new agency, GILFAM
(*Groupement pour l'Informatisation du Livre Foncier d'Alsace-Moselle*)

CS18: Case Study Background

- Computerization of all land registry activities, including the judge's certification using digital signatures
- Situated within context of European Union directives
  - equating digital signatures with handwritten signatures
  - promoting transparency and interoperability of govt functions
- Incorporates standards for digital signatures & dates
- CS18 team:
  - Department of Information Studies, UCLA
  - Archives de France
  - InterPARES project, University of British Columbia
- Responds to InterPARES 1 call for further study of *authentication* versus *authenticity* of digital records
- Addresses IP2 objective to examine the record within interactive, dynamic, and experiential environments

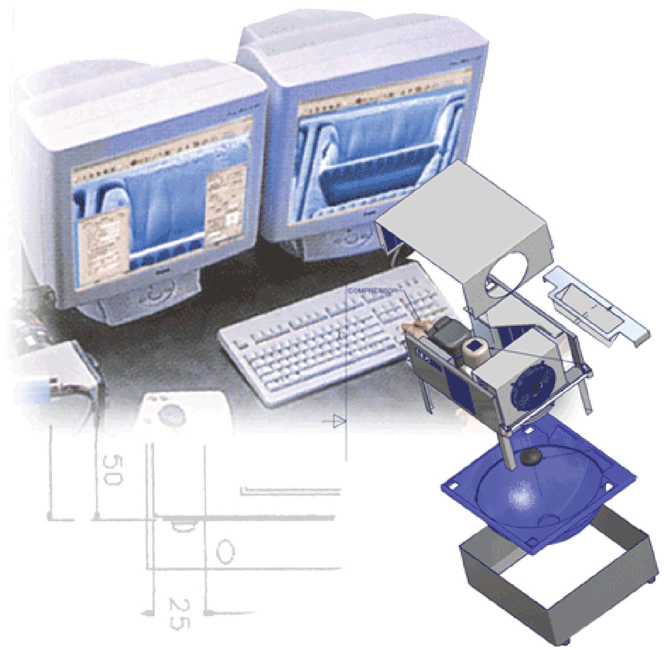
CS18: Salient Aspects of the Case

- Juridical system incorporates technologies with distributed, interactive characteristics
- Fixed data memorializes past transactions *and* enables future ones
- Transactions are *ad hoc* and actuated by human agents
- Activities of the transactions and interactions of documentary entities fit into patterns
- Heterogeneous / compound nature of both system and digital entities it produces, leaving some documentary entities more amenable to preservation than others
- Computerization of the land registry has required GILFAM to plan for the retirement of archival records to the French National Archives while maintaining functionality for current business needs

CS18: Findings of the Case

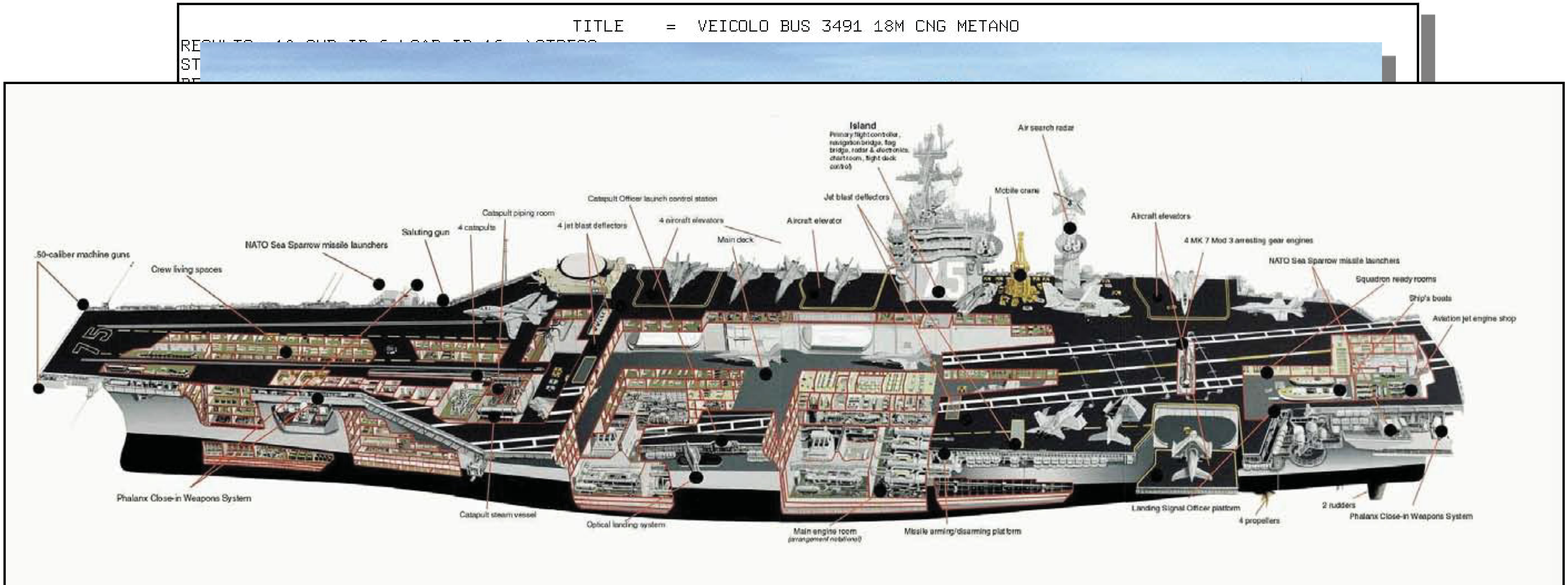
- The records may be unpreservable if the whole database continues to be considered ‘the record’
- In 2004, GILFAM stated “The only important data is the signed data accompanied by a fixed date.” – singling out the most proprietary entity within the system(!)
- Preservation is held hostage by the market
- Moderns systems of record combine fixed/memorial and enabling/prospective documentary entities
- Modern systems combine static and actionable metadata
- Recordkeeping and preservation metadata schemas must consider the full range of documentary entities, their interrelationship(s), dependencies, and collaborating services

CS19: Business Context

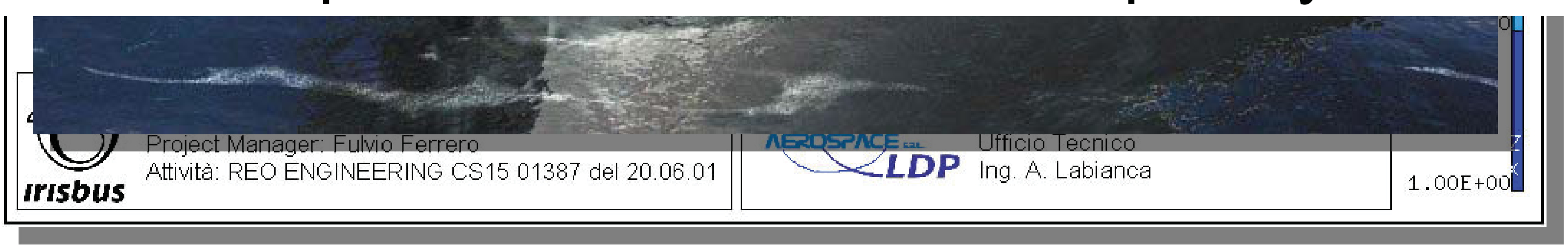


- United States agency involved in the design and manufacture of complex technology objects
- High-assurance, high-confidence manufacturing
- Relies on digital models and records to represent and produce machined piece parts in accordance with manufacturing process knowledge and specifications
- Long-term requirements for the creation, exchange, and preservation of digital models of machined piece parts and manufacturing process knowledge

Exponential Increases in Complexity and Volume



- Over one million individual CAD files of discrete pieces
- Expected lifespan of Roosevelt-class ship: 75 years



CS19: Case Study Background

- Partnership between manufacturing agency, NARA/ERA, and San Diego Supercomputer Center
- Builds on InterPARES 1 to examine methods for trusting the content (reliability, accuracy) and authenticity of records as used by the creator
- Incorporates ISO 10303, Standard for the Exchange of Product Model Data (STEP) and ISO 14721, Open Archival Information System reference (OAIS) model
- Addresses InterPARES 2 objective of extending research to new record types and aggregations in interactive, dynamic, experiential systems

CS19: Salient Aspects of the Case

- Design/manufacturing systems use distributed technologies with interactive and dynamic characteristics
- Examines records used in production environment far exceeding human capacity to discover, authenticate, and operationalize for business needs*
- Explores the potential role of a digital archives as a mechanism for transmitting authentic electronic records over time, independently of the specific technology used to create, reproduce, or use the record
- Actionable metadata and enabled actions fit into patterns
- Attempts to assess authenticity and reliability using metadata characteristics outside of proprietary limits – and whose trustworthiness can be verified absolutely

CS19: Tools Across a Semantic Web

- Using prototype trusted persistent computational environment set up by NARA and SDSC
- Storage Resource Broker (SRB) and metadata cataloging system (MCAT).
- Original part geometry created in CAD tool and converted to STEP (ISO standard format)
- Logic-based program to derive authenticating knowledge about required shapes and relationships into a knowledge representation (KR) language
- Translator from KR language to WC3 Web Ontology Language (OWL) XML format

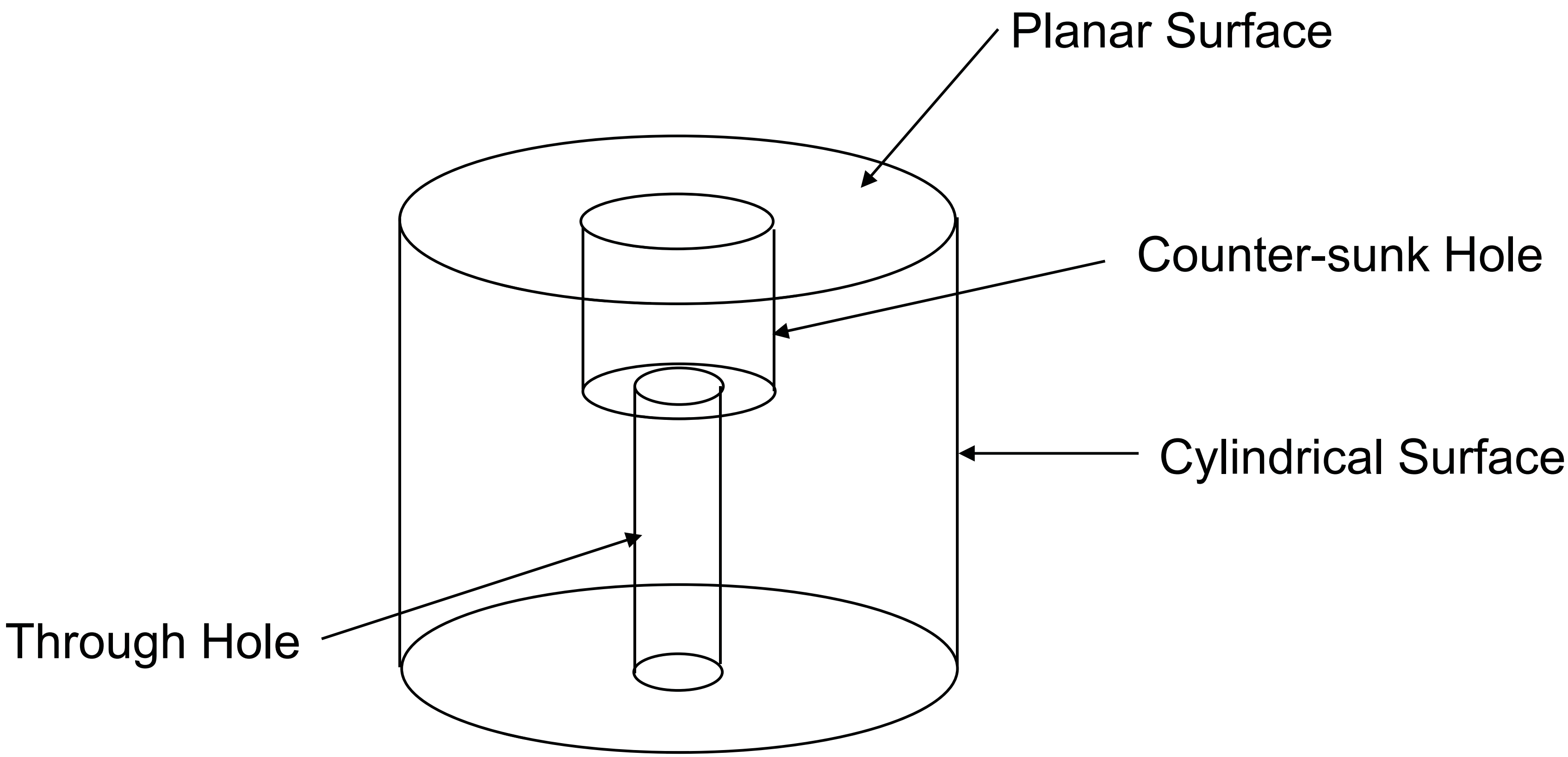
CS19: Tools Across a Semantic Web, cont.

- Approach
  - Digital model of part generated by records creator
  - Converted to archival format that supports reasoning
  - Transferred to prototype archival system at the San Diego Super Computer Center (SDSC)
  - Record set retrieved by creator across secure network
  - Tested for authenticity, reliability, usefulness with logic-based and mathematical tools

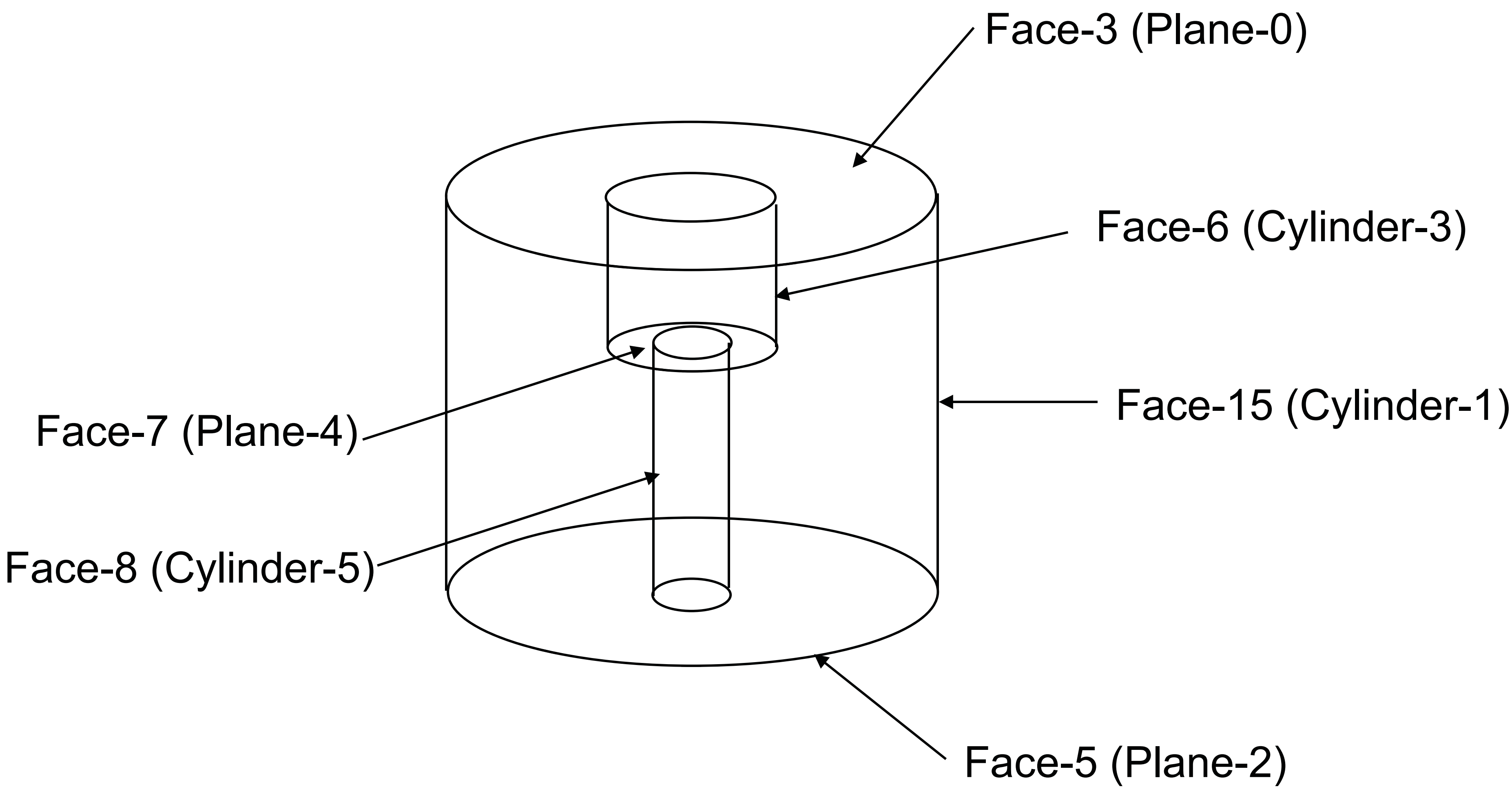
CS19: an experiment to authenticate part shape for digital preservation

- Tools used by team of veteran AI programmers to create a “shape fingerprint” incorporating geometry and geometric relationships of the piece part
- Fingerprint stored as metadata in OWL XML to authenticate part upon storage and retrieval
- OWL file and digital model sent across web to SDSC for storage in ERA prototype
- OWL file and digital model retrieved across web and run through logic-based program to authenticate model as basis for manufacturing piece part

Features of Shape



Part Topology and Geometry

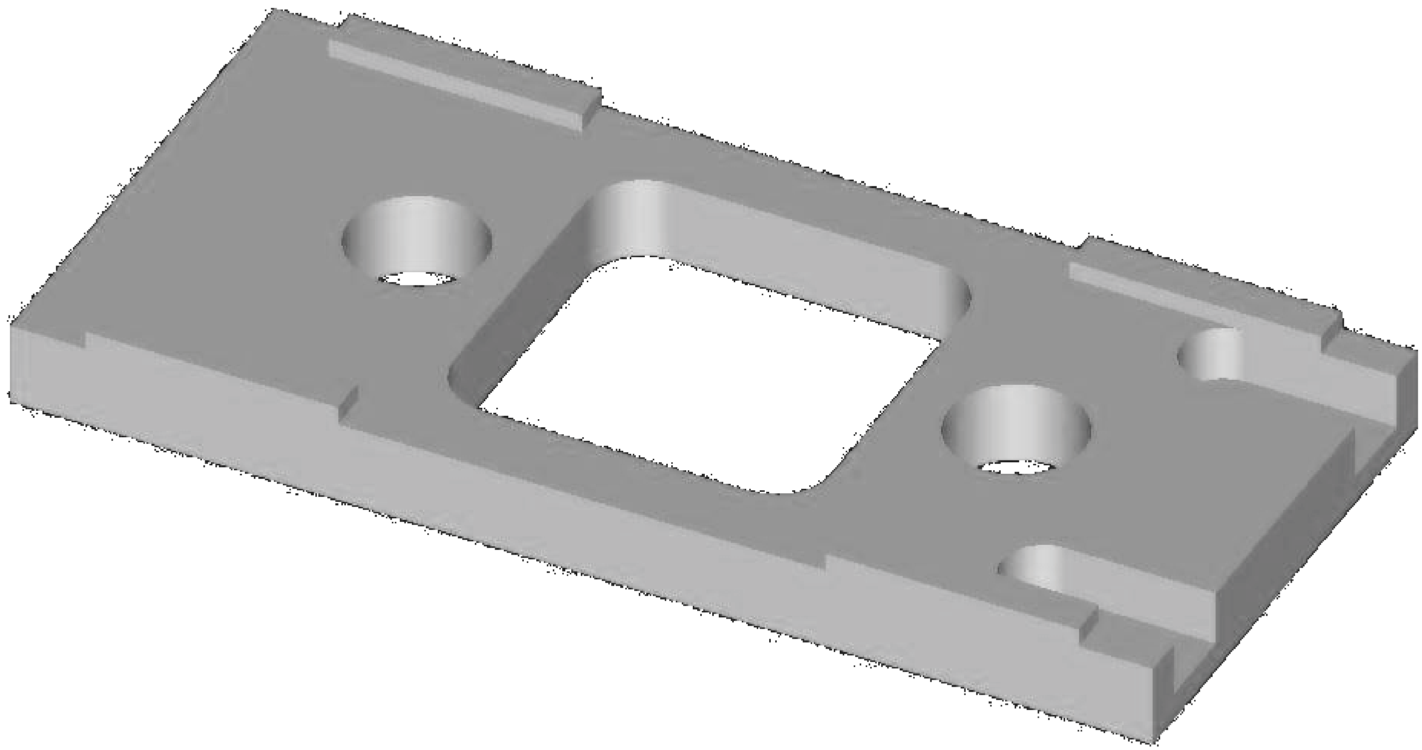


an example

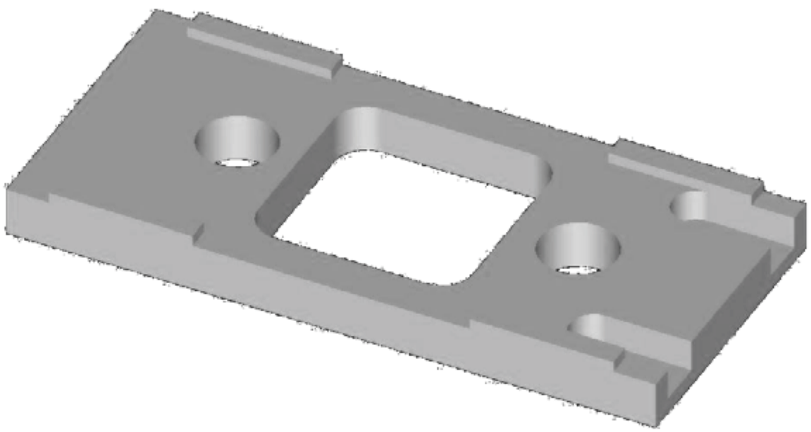
Shape: spherical
+
Diameter: 1.68”
+
Surface: divits
+
Gravity: centered
=

a golf ball!

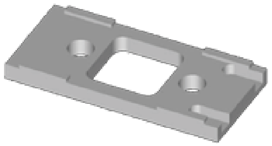
The Mechanical Part that was the Subject of the Experiment



Deduced Features



($cutout 2)
($thru-round-hole-0 4)
($closed-pocket 0)
($blind-round-hole-0 0)
($rectangular-cutout 0)
($closed-rectangular-pocket 0)
($boss 4)
($rectangular-boss 4)
($open-pocket 2))



Same Meaning for Topology and Geometry but Many Formats

ISO 10303 STEP

```
#190=CYLINDRICAL_SURFACE
#191=EDGE
#192=EDGE
#194=EDGE
#195=EDGE
#196=EDGE_LOOP(#191,#192,#194,#195)
#197=FACE_OUTER_BOUND(“,#196,.F.)
#198=ADVANCED_FACE(“,(#197),#190,.F.)
```

Data Format

Logistica

```
(Face Face-3
  :area 112.916
  :convexity Smooth
  :surface Plane-0
  :loops (unordered
    (Loop (Edge-18)
    (Loop (Edge-16)))
```

Reasoner Format

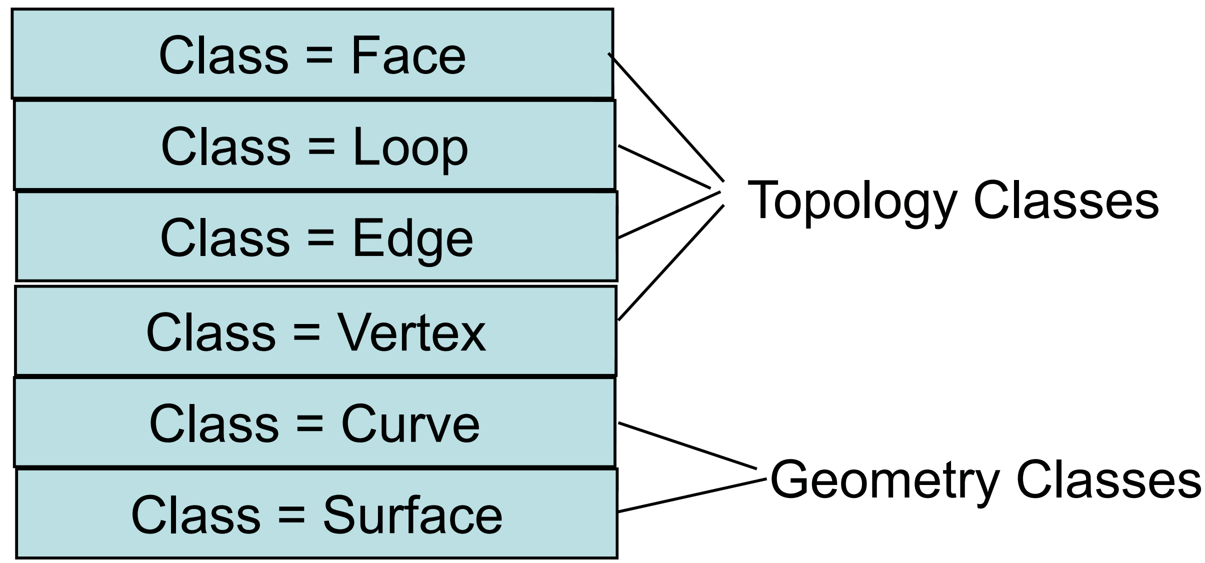
OWL

```
<owl:class rdf:ID=“face”>
<owl:equivalentClass>
<owl:unionOf rdf:parseType=“Collection”
<owl:intersectionOf rdf:parseType=“Collection”
<owl:restriction>
<owl:onProperty rdf:resource=“name”/>
<owl:hasValue rdf:resource=“face-8”/>
</owl:restriction>
<owl:onProperty rdf:resource=“convexity”/>
<owl:hasValue rdf:resource=“smooth”/>
</owl:restriction>
```

Knowledge Format

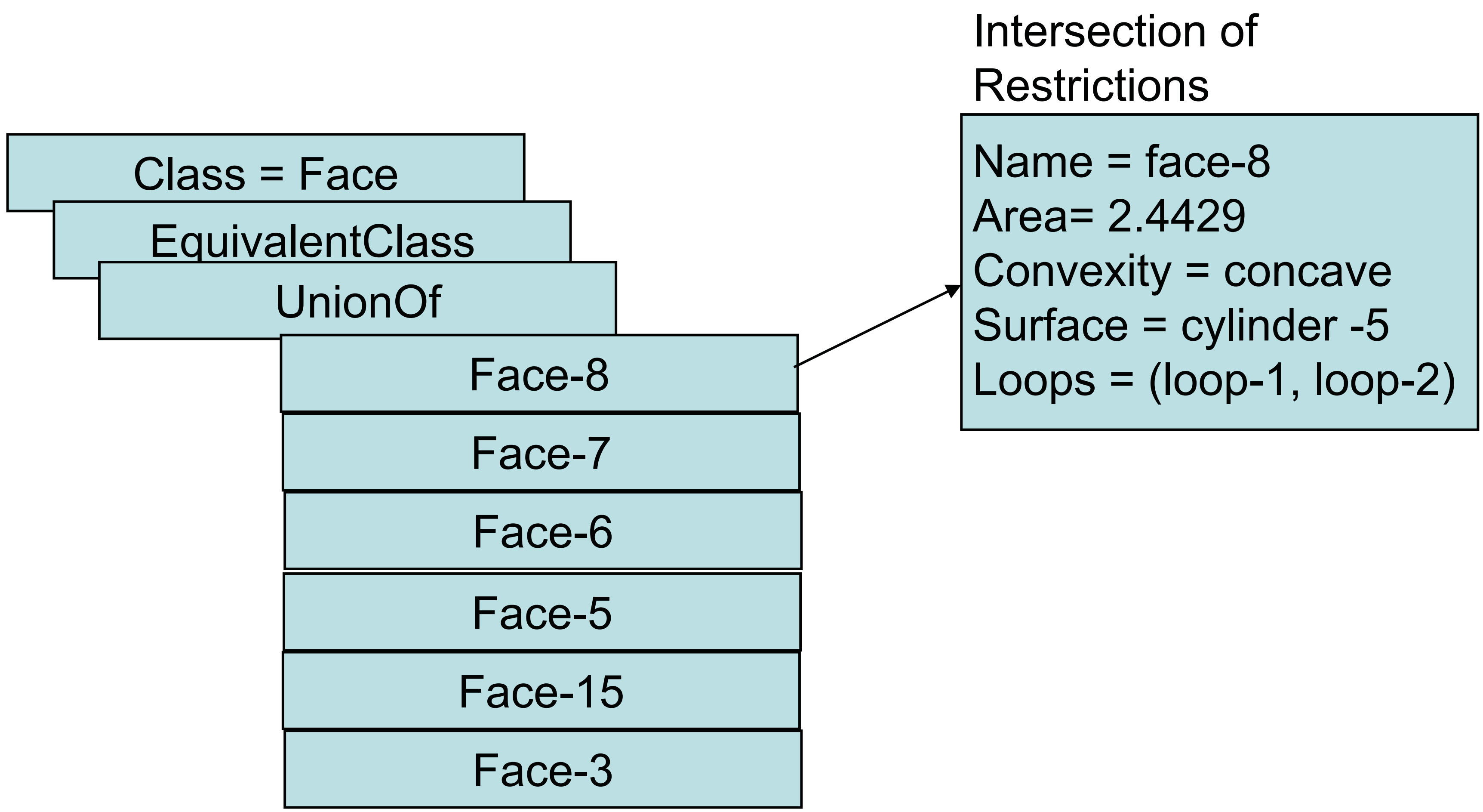
Common Semantics

OWL Mapping: A Collection of Six Classes



- expressed in Unified Modeling Language (UML)

OWL Mapping: Class Defined by Extension



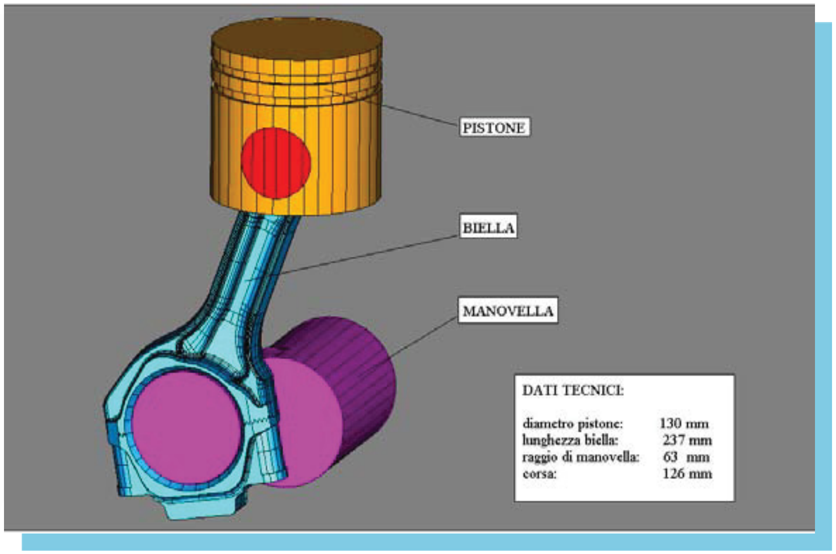
CS19: Experiment Findings

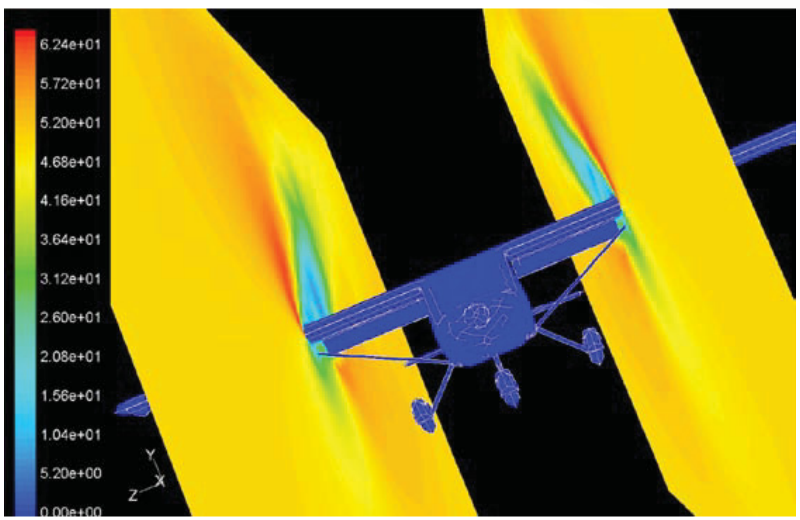
- OWL XML could not take in full range of process-based knowledge representation
- These elements were left in the logic-based program
  - (Preservation is held hostage by the market)
- Precise specifications of part shapes and relationships were successfully transformed and authenticated
- Long-term archiving to meet business requirements was declared incomplete because *some* KR elements were not stored in OWL
- However, expression of metadata to include relationships and cardinality of data elements for authentication was a success

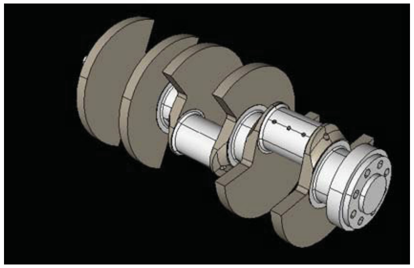
Looking to the Future

- Partnerships are important – no single partner has knowledge or skills sufficient to address range of issues
- Recordkeeping and archival metadata schemas alone are not sufficient
- Authentication is better predicated on logical predicates specific to business domains and rules, not proprietary technologies
- Logic-based programs need to be optimized to accommodate “if-then” scenarios common to object production
- Complexity of authenticating complex engineering and manufactured objects raises issues pertinent to the authentication of other objects within creative processes

broader domains new logical preservation format applies







Implications for records like CS18

- Specifiying metadata constraints based on documentary forms of the land records
  - specific rules and constraints condition the expressed form of land registry inscriptions
  - these rules and constraints can be formalized using first order logic into propositions testable by reasoning programs
  - the accumulation of characteristic static identity and integrity elements, in the patterns associated with the rules governing the business activity, can be evaluated by automated means
  - for example, in the land law governing Alsace-Moselle, the presence or absence of certain elements renders a land sale invalid –logical predicates that provide a litmus for authenticity
  - required because the characteristics traditionally used to authenticate are now latent

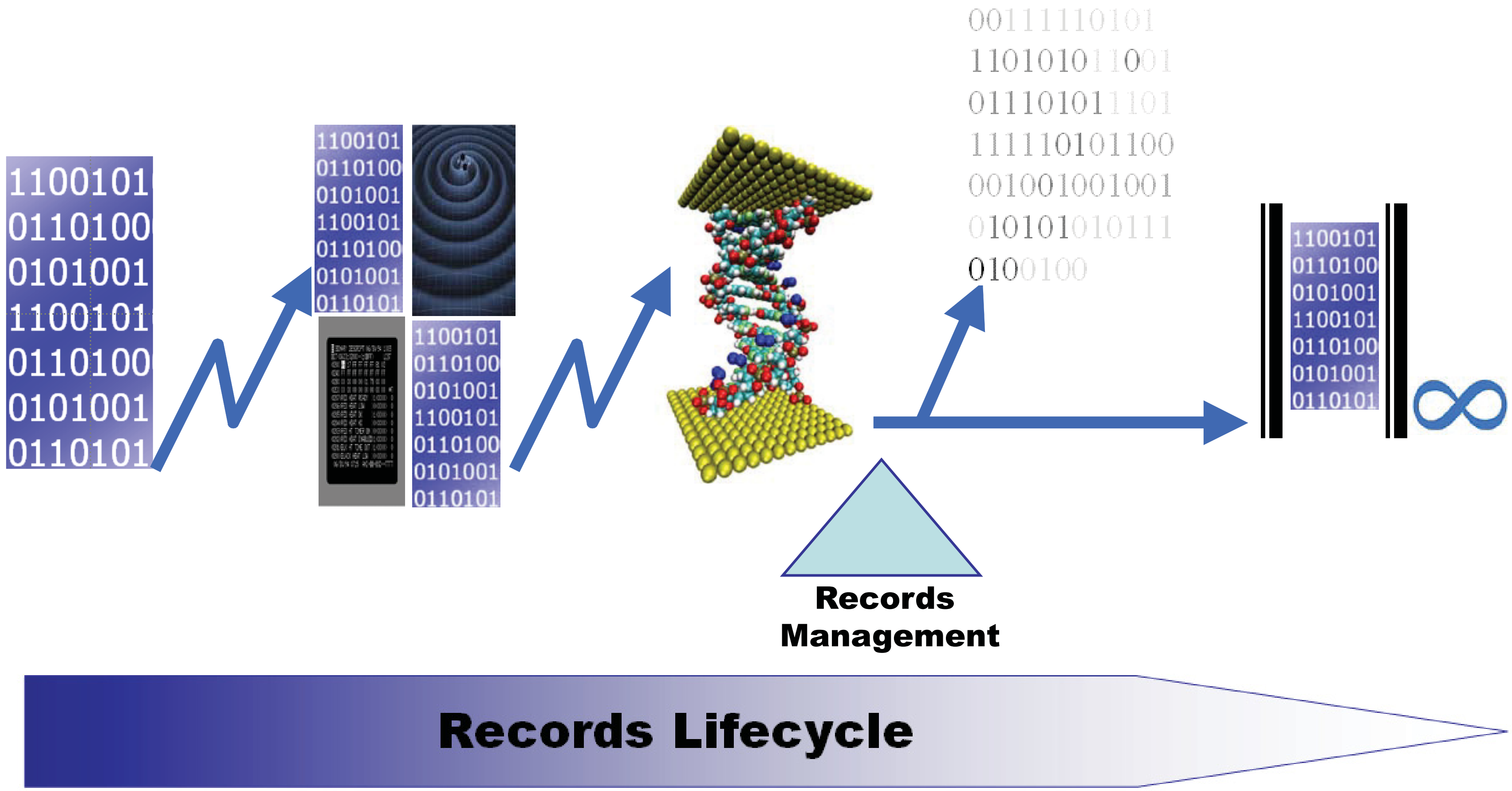
Allied efforts

- Industry or domain-specific standards governing metadata, services, and software architecture
- New legal requirements for compliance
- Dissemination of recordkeeping and archival principles into business domains (EA, KM, ECM, etc.)
- United States: the ERA Program at NARA funded a program soliciting requirements for Records Management Services from the 18 largest Federal agencies
- Record Capture / Provenance / Category / Authenticity Disposition / Case File / Reference

OBJECT MANAGEMENT GROUP

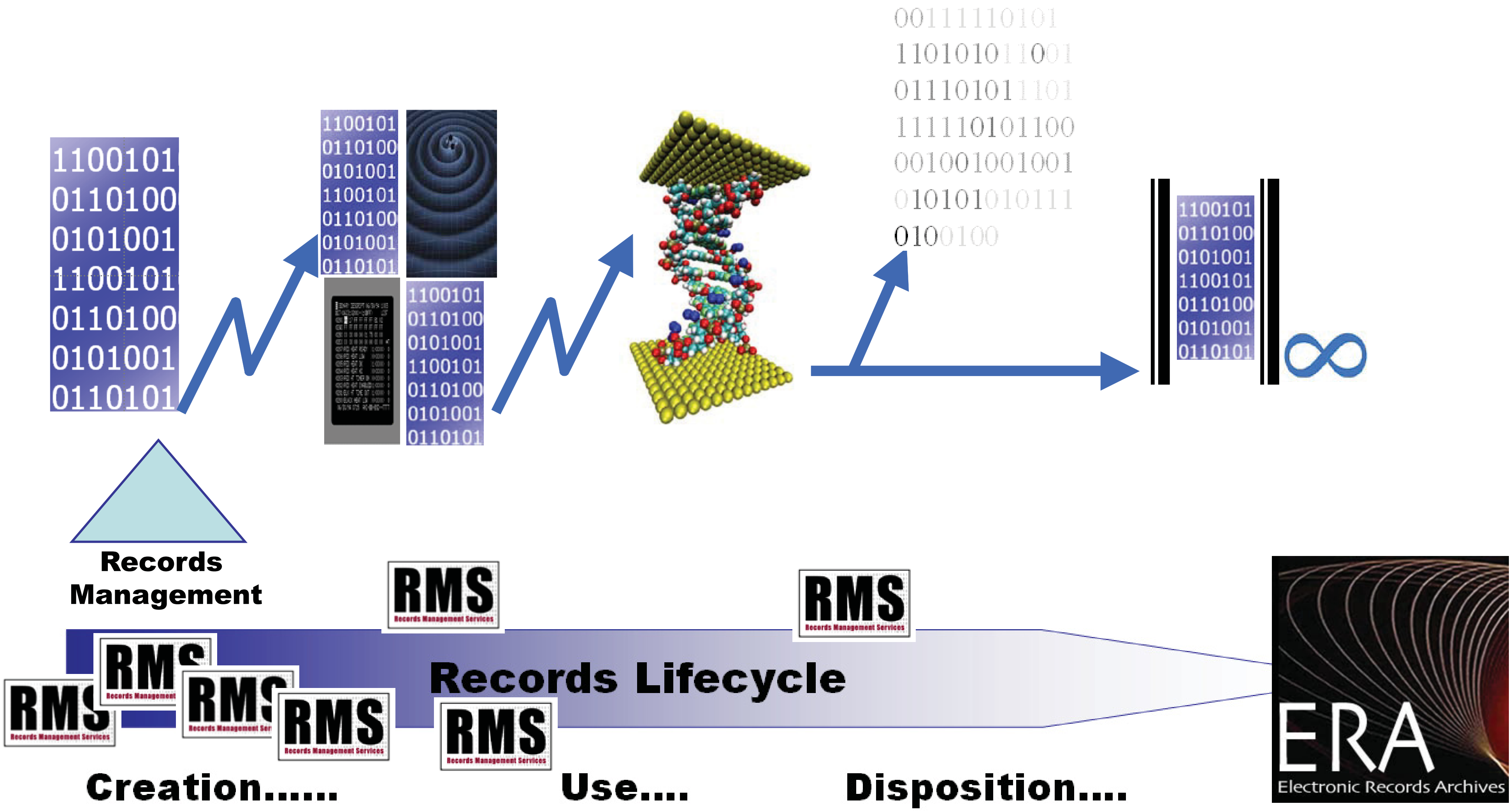
RMS
Records Management Services

Where Records Management (usually) comes in



Creation...... **Use....** **Disposition....**

Where we want Records Management to come in



questions?

ERA
Electronic Records Archives

RMS
Records Management Services

archives.gov/era

archives.gov/era/rms

gov.omg.org

kenneth.hawkins@nara.gov